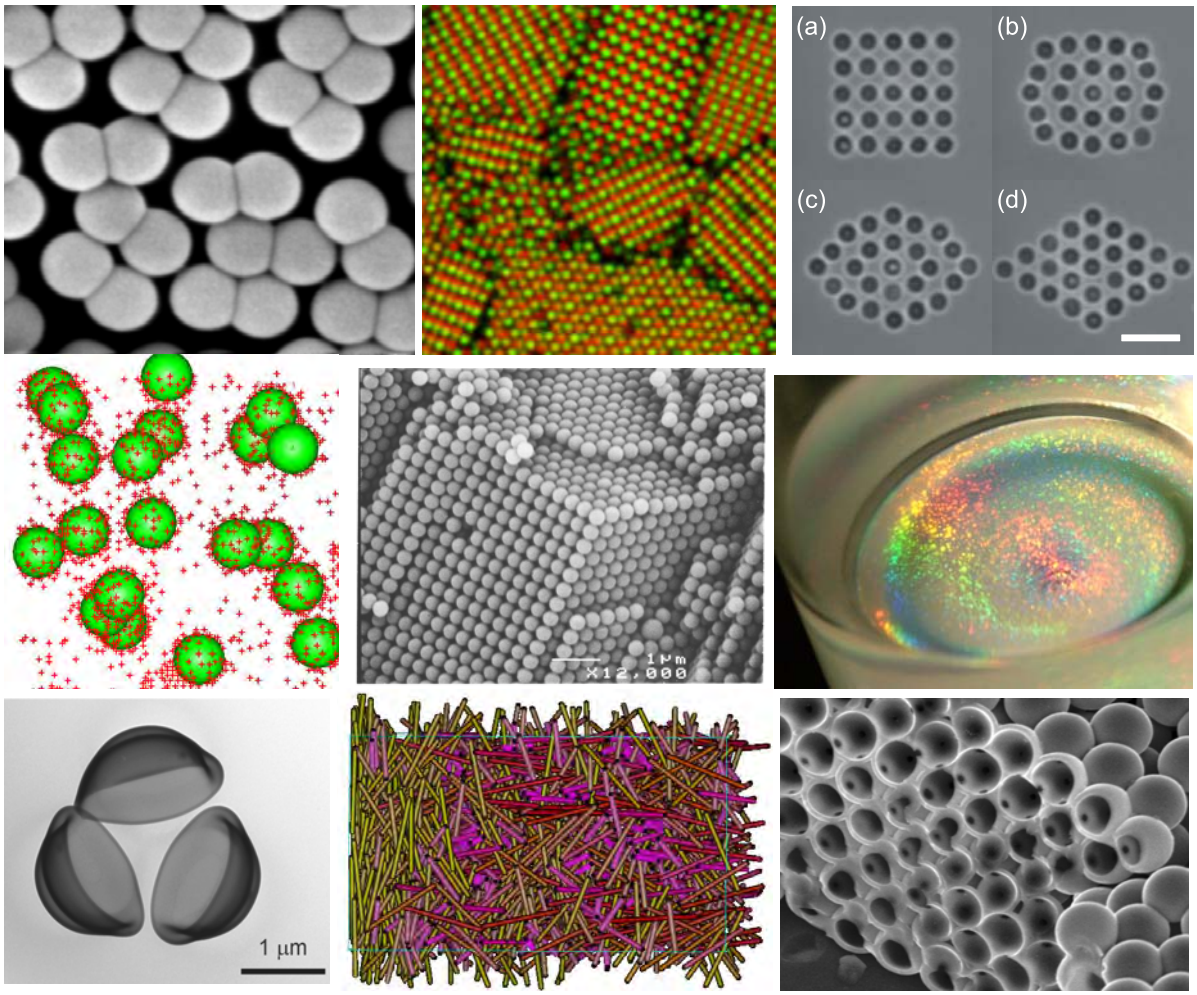
Universiteit Utrecht

# Soft Condensed Matter




Utrecht University

September 2015

# Contents

# 1. Introduction to soft condensed matter

Alfons van Blaaderen
*Soft Condensed Matter, Debye Institute*
*Utrecht University*

**Preface**

In these lectures we want to give an introduction to the sub-field of condensed matter that is generally referred to as '*soft matter*' or '*complex fluids*'. This part (marked with an *) serves as a Masters course at the Utrecht University. In addition, we will treat colloids, already studied in Utrecht for 100 (!) years at a high level, in a PhD course in more depth. We will introduce the main systems that are studied in this fields, colloids, polymers, surfactants and liquid crystals. In soft matter physics $\hbar \cong 0$ (however, see Section 1.5), $kT \cong 1$ and at least one characteristic length scale, $l$, falls in the range: *several nm* $< l <$ *several* $\mu$*m*. As we will explain this relatively large length scale implies that we can integrate out degrees of freedom that are on faster and smaller length scales. In order to understand the procedures that we will apply in (approximate) derivations of the interactions between colloids, we will introduce and recall the (statistical) thermodynamics necessary. This will include an introduction to the treatment of liquids. Existing knowledge of statistical and classical thermodynamics is strictly required, but is highly recommended. The level of mathematical background necessary to follow the lectures is similar to what is needed to follow introductory level lectures on mechanics and Maxwell theory. We will see that some of the complex fluid systems presented are actually some of the most versatile experimental systems by which conventional condensed matter can be studied on a fundamental level and by which new advanced materials can be designed. Of great help in both understanding the properties of soft matter systems in general and of the implications of a theoretical coarse-grained description, are computer simulations, which are introduced as well. Scattering techniques and the strongly advancing quantitative real space microscopy techniques will be treated next to experimental methods used to measure forces in this domain of physics. Applied soft matter science has strong links to the emerging field of nano-science and is also delivering more and more materials of which the properties can be changed dynamically, so-called *advanced* or *smart materials*, some examples of which will be given. The effects of a fluid flow or shear and the dynamics of colloids together with their phase behavior are treated in the advanced part of the course.

Also, soft matter science is truly *multidisciplinary*. For instance, the increased level of control achieved in the synthesis of many of these systems is what recently is driving rapid progress. As an illustration, an introduction to the synthesis of colloids is given in this course. Subjects that overlap with (micro) biology are not treated, despite the fact that many subjects studied in the field of biology inspired physics nowadays draw heavily on knowledge and techniques from the field of soft matter. A reverse influence is starting to appear as well. An example is research in which the complementarity of DNA strands is used to create specific colloidal interactions. The influence of soft matter science on materials science and more specifically nano-science and the design of 'smart' materials is touched upon. The same is true for the increasing role of soft matter systems as model system for investigating fundamental condensed matter questions such as melting, freezing and the glass transition.

The Problems after each chapter are intended to deepen understanding on the subject treated and to bring forward the general principles behind the phenomena discussed and are part of problem classes accompanying the lectures.

# 1. Introduction

## 1.1 Soft Condensed Matter vs. Complex Fluids

An important distinction between conventional or '*simple*' liquids and solids is that the former quickly take the shape of the container in which they are kept, while the latter maintain their shape indefinitely. Almost all '*complex fluids*' are intermediate between a solid and a liquid: while they maintain their shape for some time, they eventually flow. They are solids on a short time scale and liquids at long times: they are *viscoelastic*. Clearly, this rough definition is connected to a human time scale through the designation 'for some time'. Glaciers do 'flow' on geological time scales and solids such as metals creep under large loads by defect motion. Nevertheless, these time scales are so far separated compared to the human lifespan that it does make sense to define a class of materials based upon their ability to flow or not during a (human) experiment, or said in another way, of having mechanical properties in between that of simple liquids and solids. Examples of such 'complex liquids' that are also treated in these lectures are: *polymers*, *colloids*, (*micro*-) *emulsions*, *foams* and *surfactant solutions*.

There are also complex fluids that change from solid-like to liquid-like, or vice versa, when subjected to a small deformation. Examples of these are different kind of gels that can consist of many of the examples mentioned above. Some fluids change to solids after application of an external electric or magnetic field; these are called electro- or magneto-rheological fluids. Classical solids or liquids do not in general change state in response to a weak field.

Another important distinction between classical solids and liquids is that the former have properties that depend on the orientation of the crystallographic axes of the material, these properties, like elastic constants, are *anisotropic*, while the latter are the same in all directions, *isotropic*. Liquid Crystals (LC's) take the shape of the container they are in immediately, because they flow like liquids, but their mechanical properties are anisotropic like that of crystals (see Fig. 1.1). Just as there are many types of crystalline symmetries, there are also many types of liquid crystal phases depending on the number of degrees of freedom that are solid- or liquid- like. These macroscopic anisotropic properties are the manifestation of some kind of microscopic anisotropy. This can be the shape of the molecules forming the

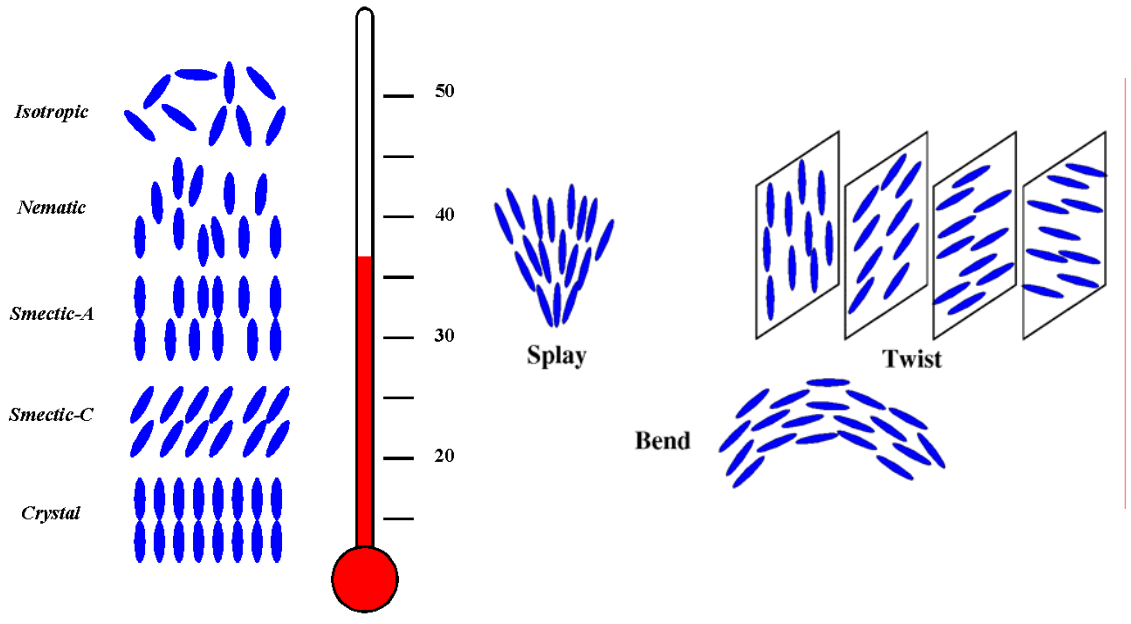



**Figure 1.1 Rod-like molecules can display a gradual decrease of positional and orientational order. Left shows some of the possible liquid crystalline phases that are possible, right shows the three fundamental anisotropic modes of deformation of a nematic liquid crystal that has only long-range order in the orientation of the rods.**

liquid crystal, but can also be on larger length scales involving self-organized subunits of surfactant molecules or colloidal particles as will be explained below.

A distinguishing feature of complex fluids compared to ordinary condensed matter is, that there always is a length scale involved that is large with respect to the size of individual atoms or small molecules. This separation of length scales makes it possible to integrate out many degrees of freedom and give a more simplified description of the problem. It basically means that the ordinary approach of statistical mechanics to treat a complicated many-body system, such as simple liquids, is taken a few steps further. Usually simplifying the description. Essential to our definition of complex fluids or soft matter is that a final description can be based on statistical mechanics. This sets a limit on the larger end of the length scale that can be included as will become clear below and from the rest of the course. Crossing this upper length scale gets us into the domain of granular matter. In this very interesting and, active field a simple connection with a statistical thermodynamic description cannot be made (at present). It should therefore be mentioned that there are researchers that also include slurries, like cement or wet sand, and foams with mm size air pockets to the realm of complex liquids. Through the general definition given above one can argue that they do belong to this class. However, as mentioned we would like to treat these granular matter systems as separate from the systems that can be dealt with using a thermodynamic description. In a strict sense liquid crystals consisting of small anisotropic molecules do not have a large length scale associated with them, nevertheless the defect structures that are part of these phases determine in a lot of cases their properties (and thus most of the times also their behavior in applications). The length scales of the defects include many molecular sizes

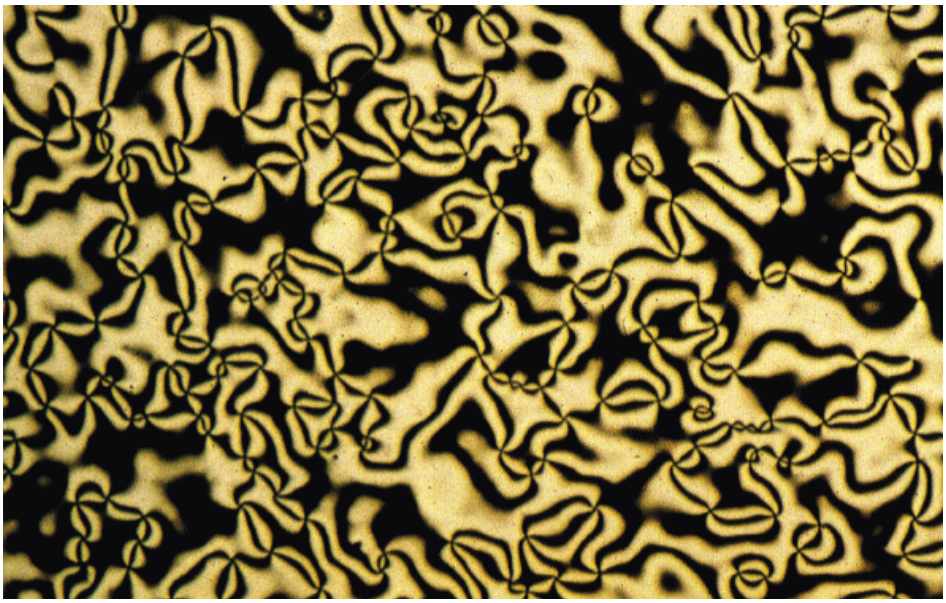


**Figure 1.2 Defect structures made visible through crossed polarizers. Different intensities are caused by different local orientation of the molecules, image several mm$^2$.**

(Fig. 1.2).

Colloidal particles dispersed in a medium encompass a large class of complex fluids. They are ‘solutions’ of one phase of matter (solid, liquid, gas), the colloidal particle, into another that acts as the continuous phase (liquid, gas). Roughly, the sizes of the particles are between several nm and several μm (see the section below on time and length scales). The name for solutions of particles in this size range is a *dispersion*; in the case that the

continuous phase is a gas such a dispersions is also called an *aerosol*. In case the particle is a liquid the colloidal system is called an (*micro*)-*emulsion*. The distinction between micro-emulsion and 'ordinary' emulsion is made on its thermodynamic stability. Micro-emulsions are thermodynamically stable and form spontaneously; emulsions need input of external free energy (e.g., in the form of violent stirring) to be formed and are metastable (although they can be very long lived). Dispersions of a gas in a liquid are called *foams*, while one can also disperse gases in gases through (*soap*) *films*. Again, for completion we mention that, based on the colloidal size range, some researchers also consider porous matter (liquid dispersed in a solid), solid suspensions (solid in a solid, e.g. wood) and solid foams (gas in a solid) as part of colloidal systems. We do not, because also in this case the description of these systems can generally not be made by a (coarse grained) thermodynamical approach.

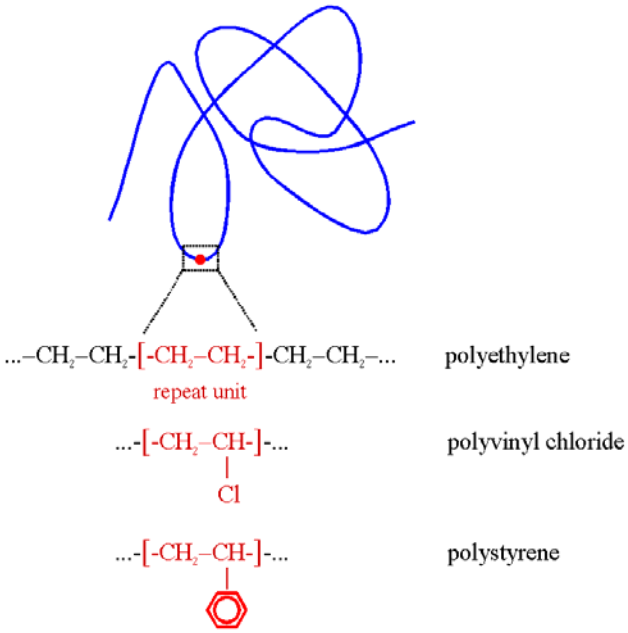



**Figure 1.3 The simplest polymer molecules are macromolecules consisting of covalently linked monomers.**

Polymers are macromolecules that consist of many subunits connected to each other through chemical bonds (Fig. 1.3). More and more complex polymers are manmade, but still by far the most advanced types are found in nature. In the cell polymers are not only the carriers of the genetic code (DNA, RNA) they also catalyze or, literally, do all the work in the form of proteins. When polymer chains are in a collapsed state they form so-called polymer colloids or latex particles. However, the distinction between a polymer in a solution is not so clear. Generally, the synthetic pathways man has developed to make polymers renders them with a relatively broad length distribution, i.e., they are *polydisperse* in length. This is in stark contrast with many biopolymers which are exact copies of one another, a property referred to as *monodisperse*. However, recent new synthetic approaches making use of self-similar structures called dendrimers or architectures where polymer arms are attached to a central core-unit (so called star polymers) can also lead to monodisperse polymer colloids.

Surfactants are molecules with a 'schizophrenic' character in the sense that part of the molecule is happy in oil, usually an alkane-like tail, while another part of the molecule, usually with dissociable or polar groups, likes water (Fig. 1.4). The distinction between 'oil' and a polar solvent, usually water, is made because generally liquids that have such a large difference in polarity, pay a large enthalpic price if they would mix and that is why they will not, despite the favorable entropic contribution to mixing. This explains the word *amphiphile* (from the greek 'loving both') that is often also used to describe these molecules. Therefore, if an oil and immiscible polar liquid like water are into contact the amphiphilic molecules go to the interface with their polar part in the water and apolar part in the oil (chapter 3). In the process they lower the free energy of the interface between the two phases significantly, i.e.,

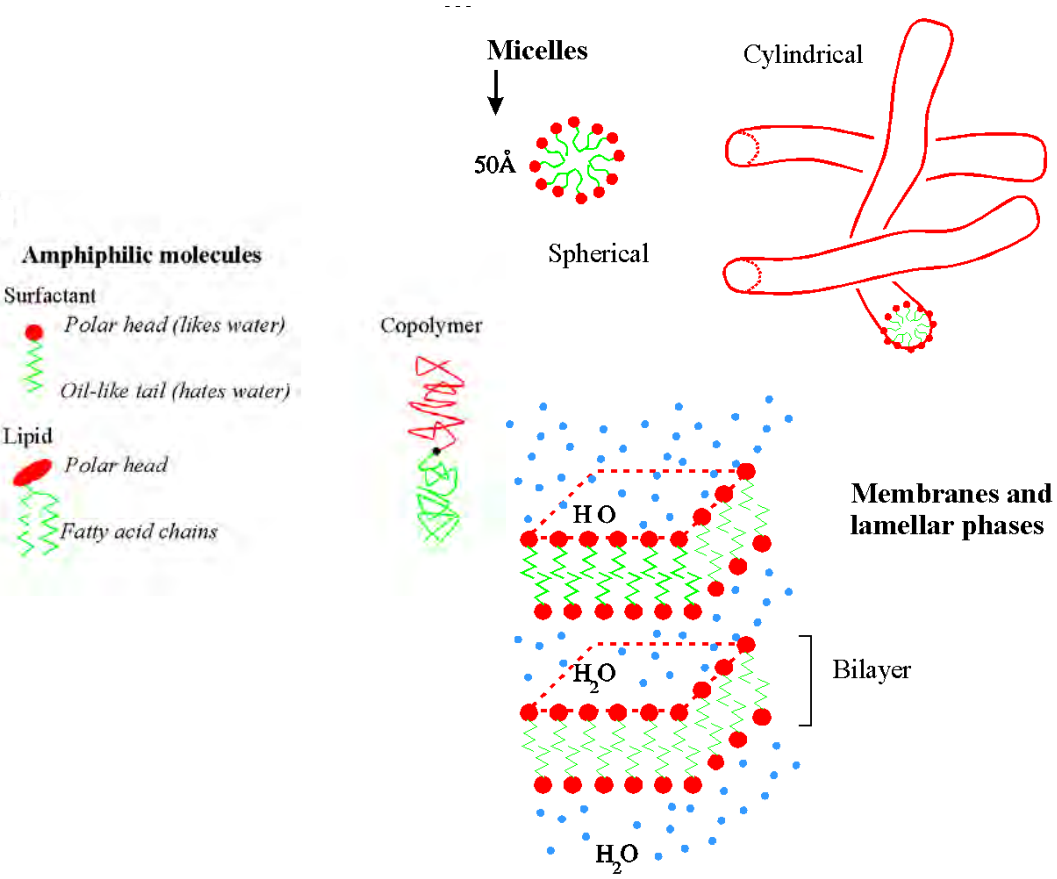



**Figure 1.4 Amphiphiles are molecules with a 'water loving' part (dark color) and 'oil loving' part (light); these are found in nature or are man-made and can even be polymeric. At higher concentrations they form complex shapes some of which are shown on the right. Not shown are the droplets of water in oil or vice verse that these 'schizophrenic' molecules can also stabilize: (micro)emulsions.**

they lower the surface tension. In the case where the surface tension gets really low and almost vanishes, the entropy of mixing can become large enough that even a thermodynamically stable mixture of water droplets in oil or oil droplets in water (water should be read here as 'immiscible polar solvent') can result: a micro-emulsion. For this to occur, it turns out that the droplets need to be very small (~nm's) to give enough mixing entropy and the surface free energy very small as the created interface surface is large. Under some conditions the amount of the phase of oil or water inside the droplets can be extremely small or even absent. In the case of solutions of surfactant in pure liquids; the colloidal entities that than form are called *micelles* (Fig. 1.4). The more general term referring to these

kind of systems, which do not necessarily need to be spheres is: *association colloids*. The non-droplet phases that can form, are characterized by the symmetry of the way the one phase is dispersed in the other or, in the case of only one solvent, according to the shape of the association colloid. Thus bi-continuous or even liquid crystalline arrangements of surfactant arranged matter is known. To distinguish the liquid crystals made up from self-organized amphiphiles from those consisting of anisotropic molecules, the former are referred to as *lyotropic* ('placement of liquid') and the second as *thermotropic*. The term thermotropic comes from the principle way to change the phase behavior: a change in temperature. And as mentioned, there are also liquid crystals where the smallest anisotropic units giving rise to the liquid crystalline behavior are colloidal particles; consequently these LC's are called colloidal liquid crystals (chapter 12).

Droplets for which the surface tension is not low enough that they form thermodynamically stable phases can still be dispersed in another liquid by the action of surface-active molecules. In this case mechanic energy is needed to overcome the energy barrier necessary to form the surface between the two phases. As the amount of forces that can be applied on a liquid droplet in a violently flowing liquid determines the size of the droplets, they can generally not be smaller in size than ~100 nm and have usually a broad distribution in sizes (are polydisperse). Because these emulsion droplets would lower their free energy by merging together and reduce the total surface area, a mechanism needs to be in place preventing two droplets from coming into contact and coalesce. This is achieved either by charge repulsion or a so-called steric repulsion between the surfactant molecules, or a combination of both these two mechanisms.

All the examples of complex fluids mentioned above will be introduced in these lectures. The trend nowadays is however, mostly because of increased level of our ability to synthesize and control these systems on an ever-increasing level, to 'mix' these basic soft matter constituents in all kinds of new ways. Some of which will be mentioned briefly in chapter 15 of these lecture notes that deals with new materials made from soft matter and the use of colloids as condensed matter model system. Examples of these more complex complex fluids are: emulsions of thermotropic liquid crystals, amphiphilic and self-associating polymers, polymers with liquid crystalline (side) groups, dispersions of colloidal particles inside thermotropic liquid crystals, emulsions stabilized by colloidal particles etc. etc.

The use of the term *complex fluids* to describe the field as explained above can also be seen in a more negative light. With some exaggeration: physicists are used to describing the hydrogen atom and consider anything larger as 'complex'. This is a like the author that just finished his new textbook and proudly names it 'Modern Mathematics' to set it aside from everything written previously. Because of this possible negative connotation, we prefer the term *soft matter* to describe this research field. Our preference may also indicate that we are from Europe, as there is a strong preference on this side of the Atlantic for use of the term *soft matter*, while Americans usually prefer *complex fluids*. In any case, also the term soft matter needs an explanation, which we will give by focusing on the high-density phases of colloidal particles. The explanation has to do with the large length scale that is, as argued above, also a characteristic of soft matter. Colloidal particles at sufficiently high osmotic pressure, the equivalent in the colloidal domain of mechanical pressure (see Section 1.2), will crystallize forming 3D regular structures in ways that are completely analogous to how molecules freeze as will be explained in more detail in chapter 12 of these notes. Next to having lattice constants that are now in the range of the wavelength of visible light and time scales of crystallization that are much closer to the human time scale, there is another important consequence of the very large size of colloidal particles compared to atomic dimensions. Compared to molecular crystals colloidal crystals are tremendously soft. Nothing will happen to a collection of salt crystals if you put them in a jar and shake it. A colloidal crystal will not

survive such a treatment and will be completely destroyed. This dramatic difference is caused by the ~$10^{12}$ difference in the characteristic quantities, the elastic moduli, that determine the strength of a crystal. A solid is characterized by the fact that it keeps its shape if a small force is applied to it. The proportionality constant between the deformation or *stress* and the applied force per unit area or *strain* are called elastic moduli. Depending on the direction of the forces with respect to the deformation, one can distinguish different kind of constants describing the relation between stress and strain. These are the 3D analogs of the spring constant that according to Hook's law describes the proportionality between the force and displacement from its equilibrium position for a harmonic spring. In case the force per area (stress) is given as in Fig. 1.5, the resulting relative deformation (strain) is a twisting of the body that does not change its volume. This kind of stress-strain situation is called *shear*. The elastic constant describing it is called the *shear modulus*, $\mu$, and is defined by:

$$\frac{F_p}{L^2} = \mu \frac{\Delta L}{L} \tag{1.1}$$

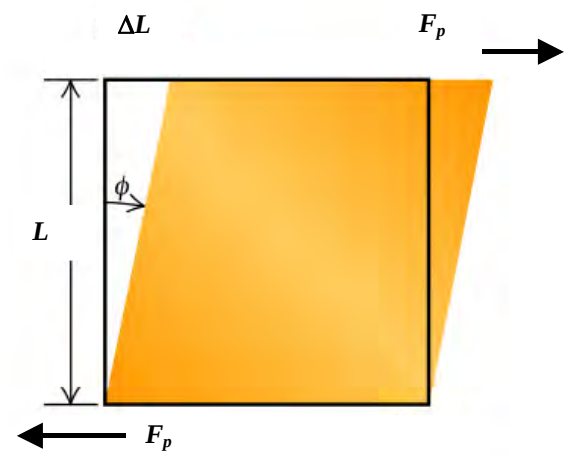



**Figure 1.5 Crystal with dimension $L^3$ deformed by shear forces parallel to the two opposing faces.**

Here $F_p$ is the shear force of a crystal with linear size $L$ and $\Delta L$ is the crystal deformation (Fig. 1.5). Thus, the shear modulus has the dimension of a pressure or energy/(length)$^3$. Intuitively, it is clear that the strength of a crystal is originating from the forces that bind the particles together in their 3D arrangement. This energy density is proportional to the number of these bonds in the crystal per unit of volume. As colloidal particles are about $10^3$-$10^4$ as large as atoms, the elastic modulus is a factor $10^9$-$10^{12}$ less; soft matter indeed!

The association of small energy densities with large length scales is similar in other manifestations of soft matter and is accompanied by a dramatically increased sensitivity to external fields such as the already mentioned flow fields (shear), or electric and gravitational fields.

## 1.2 Historical notes

It goes much too far to give here a detailed historic account of the main sub fields of soft matter. It is however, important to know something about when certain phenomena were first studied, why certain names were given to certain fields and what the impact was of understanding gained in one field to other disciplines. Again, because of space we can only

touch upon these issues and have to refer to text books for a more elaborate description. However, really historical accounts dealing specifically with soft matter are rare. We also arbitrarily do not mention those developments here that took place less than 50 years ago.

Therefore, although particle systems were important for human civilization much through all of history, be it in the form of making ceramics, paints, inks or later to make steel, it is appropriate to start by the researcher who coined the word colloid. This was Thomas Graham who in 1861 studied solutions and classified that what could pass through, what we would now call a semi-permeable, membrane and what did not. What did not pass he called colloids after the Greek κολλα meaning glue. This reflected the fact that many of the substances that Graham dissolved and that did not pass his membrane were polymeric in nature and often displayed a sticky behavior when dried.

As a small aside, we mention here some other experiments performed with semi-permeable membranes, in this case on dilute molecular solutions, because they lead in the mid-1880's van 't Hoff in Amsterdam to a law that now bears his name:

$$\Pi V = nRT \tag{1.2}$$

This equation, which bears great resemblance to the ideal gas law, describes the relation, for a dilute amount of $n$ moles of dissolved molecules, between the osmotic pressure, $\Pi$, the volume, $V$, and the temperature, $T$. So what *is* the osmotic pressure? It is the excess pressure that is needed to achieve equilibrium between a solution containing $n$ moles of molecules that cannot pass a (stiff) membrane that separates this solution with a compartment containing pure solvent. When brought into contact with each other through the membrane, solvent will start to stream from the compartment with the pure liquid into the compartment containing the dissolved species, thereby increasing the pressure, until equilibrium is reached. This excess pressure is called the osmotic pressure and assures that the flows of solvent going both ways through the membrane become equal again. It is no coincidence that van 't Hoff's law states that this excess pressure exactly equals the pressure an amount of $n$ moles of ideal gas would have exerted were it to be placed in a compartment of the same volume and temperature.

However, before the term 'colloid' was coined by Graham, important experiments with colloidal particles had already been performed. In 1827 the botanist Robert Brown studied the thermal motion of pollen grains he observed through a microscope. Contrary to others who had tried to explain this erratic motion before him, he correctly concluded by studying a range of finely divided substances that this motion had nothing to do with life or a 'life force'. It took till Albert Einstein derived in 1905 the relationship between the diffusive Brownian motion and the thermal energy of the solvent molecules that causes it (see Eq. (1.8) in the next Section) this correct conclusion was given a theoretical basis. It is much less known that in the same year, and independently, W. Sutherland from Australia derived the same relation. Despite this fact Eq. (1.8) is generally referred to as the Stokes-Einstein relation. In 1910 Perrin used it to experimentally determine Avogadro's number by analyzing the diffusive motion observed through a microscope of a model dispersion of colloidal spheres he had painstakingly had made monodisperse by repeated centrifugation. These and other experiments he performed to determine this fundamental quantity earned him in 1926 the Nobel prize in Physics for putting '*a definite end to the long struggle regarding the real existence of molecules'* (committee report).

Important experiments before Graham were also performed by Faraday (1791-1867) on gold sols which he flocculated by adding salt. Without salt however, they can be quite stable; several of Faraday's gold sols are still on display in the British museum. Faraday also discovered that small particles could be detected by focusing light into a conical region. This lead tot the development of the utramicroscope by Zsigmondy & Siedentopf in 1903, later

used by Perrin. The theory for the scattering of particles small compared to the wavelength was developed by Lord Rayleigh (1881) and finally as a general solution to Maxwell's equations by G. Mie (1908). Because of the increased length scale characterizing soft matter (light) scattering is an important technique that will also be treated in these notes in chapter 5.

It took until after the development of Quantum Mechanics in the 1940's before the first general description of the interaction forces between two colloidal particles were given. This theory was developed independently of each other by Derjaguin and Landau in the former Soviet Union and Verwey and Overbeek in the Netherlands. This so-called DLVO theory describes the interactions resulting from the (between identical particles) attractive Van der Waals forces and the (between identical particles) repulsive forces resulting from charges residing on the particles surfaces (Fig. 1.6). This potential is still a cornerstone of colloid science and experimental ways to measure it will be presented in chapter 9, together with its derivation.

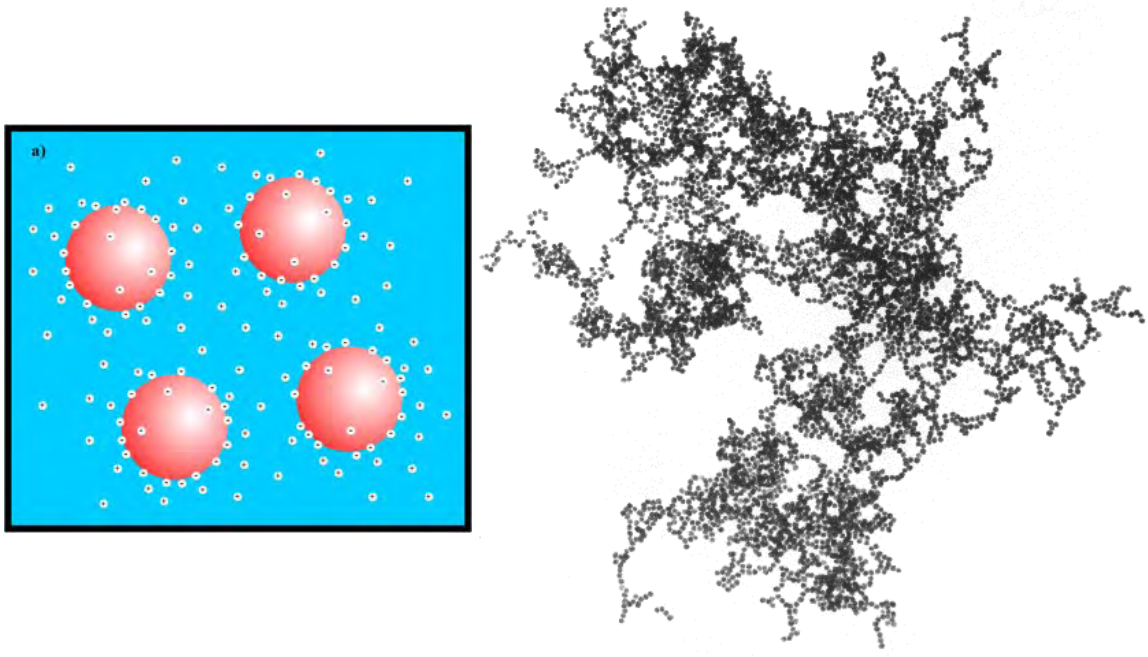



**Figure 1.6 Many colloidal particles are stabilized against the attractive Van der Waals forces by repulsion of charges on the particles surface (left). After addition of salt this repulsion cannot prevent aggregation. On the right a fractal-like aggregate of gold colloids (20 nm) after addition of salt (. The potential describing these interactions is the DLVO potential.**

Significant contributions on the study of surfactants (chapter 3) can be traced back to Benjamin Franklin's observations in 1757 of pouring oil on turbulent waters. He noticed that the wakes behind ships were calmed after the cook dumped greasy material in the water. Although he did obtain patents on this effect, they did not turn out to be too practical. It was Agnes Pockels who at the end of the 19$^{th}$ century studied and build the apparatus to measure the pressure versus area curves for monolayers of surfactants. This set the stage for Langmuir's work on the same subject in the beginning of the 20$^{th}$ century and our ability to characterize and understand the very complex phase behavior of amphiphiles.

In 1920 H. Staudinger showed that polymers (chapter 6) are not micellar aggregates, but real macromolecules, in which the monomers from which they formed are held together by covalent bonds. In 1931 Carothers produces the first nylon polymers. Soon thereafter W. Kuhn derives in 1934 the probability distribution for the average size and shape of a random coil.

Early studies on liquid crystals include those of another botanist L. Reinitzer who in 1888 observed two separate melting temperatures in cholesterol nonanoate. The term liquid crystal was coined by the physicist Otto Lehmann, who demonstrated that Reinitzer's phase changes were thermodynamic transitions. In the early 1920's work on the identification of many of the new LC phases took place, while also a description of the unusual defects and textures present in these phases was started. Most notably among the scientists studying LC's at that time was G. Friedel.

### 1.3 Separation of Time- and Length-Scales: Coarse Graining and Characteristic Forces

In this section we will focus on colloidal spheres with a radius *R*, but a lot of the reasoning is valid for most of the other soft matter systems as well. The fact that in the description of the properties of a dispersion one can take a so-called *coarse grained* approach is actually what defines the *colloidal domain,* as Evans and Wennerström call the size range between a few nm and a few μm. It also sets it apart from the molecular world on the one hand and granular matter on the other. If there is still a well-defined set of thermodynamical variables, the following approach in the statistical mechanical description can be taken. Coarse graining means averaging over a set of variables. If there is a large separation of length and time scales between the variables describing the constituents of the dispersed phase (the liquid or gas) and that what is dispersed in it (the colloid) one can average over, in other words trace-out, the fast variables that change on small length scales. This means we can, for instance, use continuum descriptions of the mechanical response of the dispersion medium (hydrodynamics). This approach simplifies both theories describing structure and dynamics of colloidal matter as it does computer simulations of these systems. On the other hand it is equally important for describing these properties and for that matter of what we define to be soft matter that not *all* thermodynamic variables are integrated out. In a physical description it means that there is a connection between the thermodynamics of the continuous phase with that of the dispersed particles. For instance, by linking the same fluctuations that cause Brownian motion to the dissipation that is characterized by the viscosity (first example of a fluctuation-dissipation relationship). It is exactly this loss of connection that makes 'sand in water' not a colloidal dispersion and instead a granular matter system; contrary to a solution of glass beads the size of 1 μm. It also means that the description of sand piles or slurries of sand in water is much harder and actually at this moment in time even lacks a general accepted theoretical framework to tackle this problem. The description of the properties of the ~1000 times smaller glass colloids, on the other hand, can be dealt with through the usual statistical mechanical approach. This is also why the first chapters (2-3) of these lecture notes on soft matter will start with both a recapitulation of statistical mechanics and an introduction to the description of liquids. These theoretical foundations are necessary to understand the methods for deriving, for instance, potentials between colloidal particles that do not contain the details of the liquid in which they are dispersed. In the case of charged colloids interacting through Van der Waals forces and (screened) Coulombic forces from charges residing on the particle surface, we will perform, or more accurately stated, sketch, such a derivation resulting in the DLVO potential after those that first derived it (Derjaguin, Landau, Verwey and Overbeek, chapter 9).

Interestingly, the size range that is defined as *colloidal* is dependant on those who do the observations, not through a collapse of a wave function, but by defining the time scale of experimental observations. Statistical mechanics shows that if an object has a well-defined thermodynamic temperature it has 0.5 $kT$ of kinetic energy per degree of freedom (equi-partition theorem) where $k$ stands for Boltzmann's constant ($1{,}381 \times 10^{-23}$ J/K) and $T$ the absolute temperature. This thermal energy leads on the scale of molecules to kinetic chaos

that on the particle level, characterized by a radius $R$, can be interpreted as a *Brownian force* with a magnitude of $O(kT/R)$. Before we can continue our analysis of relevant time and length scales and the role of different interactions that are of importance in a colloidal dispersion, we first have to take a short digression to that part of continuum physics that describes the laws of Newton in a continuous medium like a fluid or gas: hydrodynamics.

First, let us consider what happens to a fluid of which the density does not change, a so-called incompressible fluid ($\nabla \bullet U = 0$, usually a very good approximation), when we apply the same shear stress $F_p/L^2$ as in the shear deformation of the crystal mentioned above in Section 1.1. In the deformation of the solid described by Eq. (1.1) in that section we saw that the shear strain was, for small values, proportional to the shear strain. In a fluid the shear strain increases continuously and without limit as long as the force is applied: a fluid flows (Fig. 1.7). Therefore, the stress does not depend on the shear strain, but on its rate of change also called the strain rate or the shear rate (usually depicted by $\dot{\gamma}$). The *viscosity* of a fluid, $\eta$, is defined as the ratio of the shear stress, to the strain rate:

$$\eta = \frac{F_p / L^2}{v / L} = \frac{\text{Shear stress}}{\text{Strain rate}} \tag{1.3}$$

For a *Newtonian fluid* the viscosity does not depend on the speed $v$ (or the shear rate) making the speed directly proportional to the applied force (stress). The viscosity, or as it is also called the 'internal friction', determines for a fluid how easy a liquid will flow by an applied force and how much energy is dissipated as heat. Viscous forces oppose the motion of one portion of a liquid relative to another. The equations that describe an incompressible

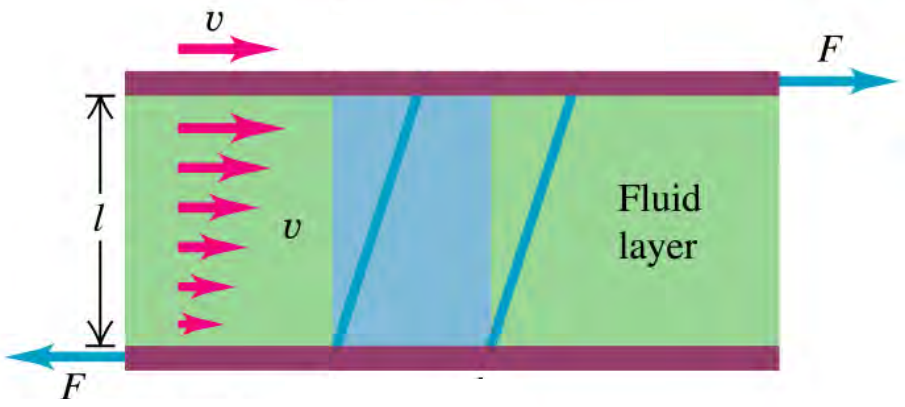



**Figure 1.7 Shear forces applied onto a liquid lead to a shear rate characterized by a viscosity.**

Newtonian liquid are called the *Navier-Stokes* equations. These describe the momentum conservation principles of elementary physics, for a system that includes friction, as applied to a stationary control volume through which fluid may enter or leave. Similarly as in Eq. (1.3), a sphere with a stick boundary condition on which a constant force is applied will obtain a constant speed. A stick boundary condition means that the fluid just adjacent to the particles surface does not move with respect to the particle. Again the proportionality constant is given by the viscosity and by the particle size. This friction factor, $f$, has Stokes' name associated with it, because it follows as a solution to the *Stokes* hydrodynamic *equations*.

$$f = 6\pi\eta R \tag{1.4}$$

The Stokes equations are a limit of the Navier-Stokes equations and are obtained by neglecting effects of inertia. Inertia and viscous forces affect a mechanical mass-point system in similar ways as in a hydrodynamical description. The *Brownian time*, $\tau_B$, is defined as the relaxation time of momentum of a particle. It is given by the ratio of the mass and the friction factor and describes the time it takes a particle with mass *m* on which a force works to obtain its steady state velocity as dictated by the Stokes friction factor:

$$\tau_B = \frac{m}{f} = \frac{m}{6\pi\eta \mathrm{R}} \tag{1.5}$$

The importance of inertia, as given by $O(R^2\rho U^2)$, with *U* a typical velocity, relative to viscous forces, $O(\eta RU)$, is given by the dimensionless group called the *Reynolds* number:

$$R_e \equiv \frac{\rho UR}{\eta} \tag{1.6}$$

If we fill in some typical numbers (see Table 1.1 for values chosen) we find 2 x $10^{-10}$ s for the Brownian time and $10^{-6}$ for the Reynolds number. Because of its small mass, momentum transferred to a Brownian particle is very quickly lost and inertial effects can be neglected for all relevant velocities. As our day-to-day experience with hydrodynamics, for instance while swimming or paddling a canoe, is in the limit where inertia effects cannot be neglected, one has to be careful when applying human intuition to the colloidal domain. For instance, if we shrink a person swimming to the size of a μm, his mass is so small that it becomes impossible to swim. Swimming relies heavily on pushing oneself off against the water to gain forward momentum, something that is not possible in the colloidal domain.

Filling in Eq. (1.5) for a solvent molecule tells us that on the molecular scale momentum is relaxed in $10^{-15}$ s. This is the time scale on which the solvent exerts forces on the colloidal particle. Before Perrin scientists tried to infer from the Brownian motion observed through a microscope the mean velocities of the particle. In principle the mean kinetic energy of a colloidal particle $<0.5mv^2>=1.5kT$, $0.5kT$ per degree of translational freedom as stated earlier. However, through our estimates of the relevant time and length scales we can now see that this method of analyzing Brownian motion is doomed to fail. It would mean accessing the particle displacements on time scales much shorter than $\tau_B$, and length scales much shorter than can be resolved through a light microscope. What one observes through a microscope is already the result of many different uncorrelated forces on the particle. A process that is characterized by a great many realizations of uncorrelated events is a diffusion process characterized by a diffusion coefficient $D_o$. We will revisit diffusion not only in chapter 13 on the dynamics of colloids, but also find that it describes the basic shape of a polymer molecule (chapter 6). The subscript '0' here designates that we are dealing with single particle diffusion, not influenced by other particles. In a diffusion process particle displacements are not proportional to time as in Newtonian free flight, but instead scale with the square root of time:

$$\left\langle x^2(t)\right\rangle = 2D_0 t \tag{1.7}$$

Eq. (1.7) describes the mean square displacement, $<x^2(t)>$, as a function of time, *t*, for a 1D process. For each extra dimension a factor 2 needs to be added. It was Einstein who was the

first to derive a value for the diffusion coefficient by connecting the random diffusion process with the average kinetic energy of the colloids:

$$D_0 = \frac{kT}{f} = \frac{kT}{6\pi\eta \mathrm{R}} \tag{1.8}$$

In his derivation Einstein used the Stokes friction factor, Eq. (1.4), van 't Hoff's law, Eq. (1.2) and the fact that Brownian motion is described by a diffusion process, Eq. (1.7).

The time it takes a Brownian particle to diffuse a distance $2R$, $\tau_I$, is thus given by:

$$\tau_I = \frac{2R^2 f}{kT} = \frac{12\pi R^3 \eta}{kT} \tag{1.9}$$

This time is sometimes called the interaction time and we can use it to calculate the mean particle velocity resulting from the Brownian force as manifested through the irregular bombardment of solvent molecules, we find: $kT/(Rf)$ and thus indeed that the Brownian force is of $O(kT/R)$.

The interaction time, $\tau_I$, also has physical meaning in a concentrated dispersion; it is the time it takes for a collectoin of colloidal particles that are close together to significantly change their configuration. The colloids experience direct interactions with their neighbors in addition to hydrodynamic friction in the process. This is not to say that the presence of the neighbors is not felt at shorter times. On the contrary, because hydrodynamic interactions are very long-range and very fast, as we will see shortly, the Brownian particle feels already the presence of the other particles on the so-called hydrodynamic time scale, $\tau_H$:

$$\tau_H = \frac{\rho R^2}{\eta} \tag{1.10}$$

This hydrodynamic time, which is on the order of the time it takes a hydrodynamic shear wave to traverse a distance $R$, comes naturally out of the (Navier-Stokes) equations describing the hydrodynamics in which temporal accelerations are taken into account. After some thought it is not surprising that this hydrodynamic time, after some rearrangements turns out to be of the same order as the Brownian time, $\tau_B$, defined earlier. Similarly as we did before we can calculate that the distance a Brownian particle diffuses in this time, $l_B$, equals:

$$l_B = \frac{\sqrt{2mkT}}{f} \tag{1.11}$$

Filling in our usual assumptions gives, $l_B \approx 0.1$ nm, which is as was already stated, very short. Thus, for times (much) longer than $\tau_B$ and distances (much) further than $l_B$ we can forget about the transients and consider velocities as determined by the friction factor. Said in another way: for distances longer than $l_B$ colloidal motions are overdamped. These facts make the calculation of hydrodynamic effects, which is still a formidable, many body problem, a lot easier and is also at the basis of a computer simulation technique called Brownian dynamics (chapters 7 & 12).

As we will see in chapter 9 it is actually almost always the case that particles dispersed in a liquid acquire a net charge. Although we will also derive in that chapter that the interaction between two charged spheres is mediated in important ways by the ions in solution

that reside around the particle because of its electric potential, an order of magnitude estimate from just a Coulombic repulsion of two spheres with a surface potential of $\zeta$, is given by $\varepsilon\varepsilon_0\zeta^2$, here $\varepsilon_0$ is the dielectric permittivity of free space (8.85 x $10^{-12}$ C/Vm). As seen in Table 1.1 the electrical forces are usually larger than the Brownian forces thus explaining why these repulsions can protect two colloids from the attractive Van der Waals forces. As we will also see in chapter 9, additivity of the Van der Waals forces between molecules will lead to interactions between colloids that are always attrictive between identical particles and that are characterized by the Hamaker constant, $A_{eff}$. This constant, which has a unit of [energy], depends on the dielectric properties of both the particles and the intervening dispersion medium and gives rise to a Van der Waals force of $O(A_{eff}/R)$.

Lastly, as was already mentioned, soft matter is much more amendable to external fields than conventional condensed matter. Here we will look at one field, gravity, that on earth is always present and is an important factor limiting experiments on particles that are larger than several μm, next of course to the time restraint a human lifetime puts on the time an experiment can run. The result of gravity is sedimentation: and the velocity a single colloidal sphere in a dilute dispersion will attain either moving down or up under the influence of this external field is given by the sedimentation velocity, $U_s$. If $\Delta\rho$ is the density difference between the particle and the solvent, the gravitational force is $O(R^3\Delta\rho g)$ with $g$ the gravitational acceleration (9.89 m/s$^2$), than balancing frictional and gravitational forces gives:

$$U_s = \frac{2R^2\Delta\rho g}{9\eta} \tag{1.12}$$

With Eq. (1.6) we can check that indeed for typical values the Reynolds number is still small for all sedimentation processes. Another dimensionless number the *Peclet* number, here for sedimentation, gives the relative importance of diffusion as compared to sedimentation:

$$P_e = \frac{2RU_s}{D_0} \tag{1.13}$$

Even gas molecules experience the gravitational pull of the earth resulting in a barometric height distribution in which the gas density is characterized by an exponential. The decay length of this exponential distribution is called the gravitational length, $l_g$, and gives the height one has to lift a particle of (buoyant) mass $\Delta m$ to raise its potential energy in the gravitational field by $kT$:

$$l_g = \frac{kT}{\Delta mg} \tag{1.14}$$

For an ideal gas this equation is easily derived by assuming local mechanical equilibrium: the gas pressure at a certain height should be the same as the weight of the gas above that point, i.e., hydrostatic pressure balance. Similarly, but now taking the osmotic pressure as given by Van 't Hoff's law instead of the ideal gas law, we can derive an exponential density distribution for colloids if they behave ideally as well. However, while for an ideal gas the gravitational height is several km, for colloids it can be even on the order of a particle size or smaller. Measuring the gravitational height of a colloidal dispersion was another way in which Perrin determined Avogadro's number.

Table 1.1 summarizes the importance of several forces for our chosen example parameters. Firstly, for all colloids inertia effects can generally be neglected and for almost all time and length scales one can assume diffusive motions and velocities governed by a friction factor. As mentioned, electrical forces keep it suspended and stable against aggregation by attractive Van der Waals forces, while stirring or a modest speed results in viscous forces that are of the same order as those exerted on the particle by thermal motion of the solvent molecules. Gravity effects are not dominant, but can in the long run not be neglected. However useful these quick order-of-magnitude estimates are, it should also not be forgotten that all the forces mentioned have very different dependencies on distance, decaying with different power laws or even exponentially. This is why the derivation of some of these laws will be dealt with in coming chapters as well as the experimental methods to measure them.

Table 1.1 *Order of Magnitude of Characteristic Forces:* $R = 1\ \mu m$, $\eta = 10^{-3}$ kg/ms, $U = 1\ \mu m/s$, $\rho = 10^3$ kg/m$^3$, $\Delta\rho/\rho = 10^{-2}$, $g = 10$ m/s$^2$, $A_{eff} = 10^{-20}$ J, $\zeta = 50$ mV, $\varepsilon = 10^2$

| | | |
|---|---|---|
| $\frac{\text{electrical force}}{\text{Brownian force}}$ | $\frac{R\varepsilon\varepsilon_0\zeta^2}{kT}$ | $\sim 10^2$ |
| $\frac{\text{Van der Waals force}}{\text{Brownian force}}$ | $\frac{A_{eff}}{kT}$ | $\sim 1$ |
| $\frac{\text{viscous force}}{\text{Brownian force}}$ | $\frac{\eta U R^2}{kT}$ | $\sim 1$ |
| $\frac{\text{gravitational force}}{\text{Brownian force}}$ | $\frac{\Delta\rho R^3 g}{\eta U R}$ | $\sim 10^{-1}$ |
| $\frac{\text{inertial force}}{\text{viscous force}}$ | $\frac{\rho R^2 U^2}{\eta U R}$ | $\sim 10^{-6}$ |

## 1.4 Crossroad of Disciplines and Fields

The *Colloidal Domain*: where *Physics*, *Chemistry*, *Biology* and *Technology* meet. This is the appropriate title of a recent textbook on colloid science. Even in recent years this statement is gaining more and more momentum, as it is driven by a rapid increase in synthetic methods, increased power of computer simulations and an increasing theoretical understanding combined with powerful methods, including quantitative 3D microscopy, to study and manipulate soft matter. As this course is intended for those who in principle have never heard of complex fluids, our focus is on the physics of both the theoretical basis behind the approach to describe soft matter and on the physics of the experimental and computer simulation tools that are used to study this field. We have limited our focus further to an equilibrium treatment for the Masters subjects (marked with an *), and leave phenomena that have to do with soft matter under flow conditions (shear) and interacting diffusing systems for the advanced part of the course. We can only touch upon the very interesting physical chemical or chemical physical processes that underlie many of the synthetic approaches used to design soft matter systems. The reason is that without a basic understanding of the forces

and phase behavior of the systems in question it is not possible to describe a synthetic pathway in any detail. For instance, the route used most in industry to produce polymers, emulsion polymerization, illustrates this point nicely. The reaction heat that comes free during the chemical routes used to make polymers is large, but more problematically it comes free in a short amount of time. Without going into too much detail, this is mostly due to the fact that the reaction pathways are autocatalytic and therefore have a tendency to 'run away'. One way to prevent the reactions to get out of hand is to perform them in small droplets of monomers in a sea of an inert (heat absorbing) liquid, such as water. To keep the emulsion droplets from creaming, they have to be stabilized by a surfactant. Subsequently, an initiator has to start the reaction by diffusing from the water phase into the droplets of monomers. A complex set of conditions depending on the diffusion constants of the species involved, the phase behavior of the growing polymer in the monomer liquid and the details of the chemical reactions, etc., now determine the final size of the polymers and the distribution of their length. In short, many synthetic pathways for soft matter components are complicated soft matter research issues where by far not all problems have been solved and where a lot of active research is going on. It is however important to understand how the increased length scale that is present in soft matter systems gives chemical control over the properties. To take colloids as an example: quantum mechanics dictates the interactions between molecules, while for a colloidal spheres with a radius $R$ the interaction potential can be tailored to a very large

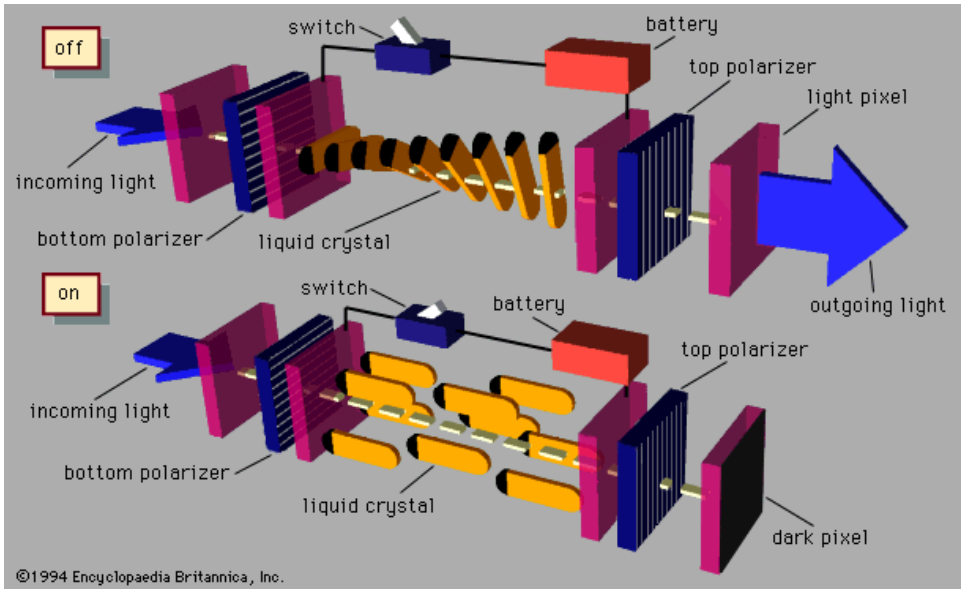



**Figure 1.8 Twisted nematic liquid crystal display. Two surfaces that are treated to align the rod-like molecules in such a way that they make a 90$^{o}$ twist in the average direction in which they are pointing. This twist is followed by the polarization vector of the light that passes thus allowing light to transmit the device even for crossed polarizers. After switching on a moderate electric field, all molecules align reversibly perpendicular to the polarizer planes and the pixel becomes dark.**

degree both by changing the properties of the particle surface (and/or bulk) and by changing the properties of the dispersion medium. Using particles with core-shell structures even increases the possibilities.

In many industrial processes soft matter systems play a role, this is clear if they are the main components such as in paints, polymers, cosmetics or foods, but is often also less obvious such as in oil recovery, or in the making of IC's. Almost all these systems are very complex involving many soft matter components that all interact with each other. This is why

most of these systems are not treated in this introductory course. It is certainly true however, that progress in the field of soft matter is being made at such a pace that the 'model' systems being studied are starting to become more and more complex as well, bridging the gap between the academic and industrial labs. Moreover, the progress mentioned is also creating a whole new class of materials that are directly based on the model systems with their well-controlled properties. These are advanced materials of which the properties can be adjusted dynamically by external stimuli such as: liquid crystal displays (Fig. 1.8), sensors based on photonic crystals or electro-rheological fluids. These materials are also called 'smart' materials (chapter 15).

As already mentioned, almost all the important content of every cell falls in the colloidal domain. This includes on the small end proteins and on the large end red blood cells. Despite a lot of recent research that uses almost exclusively theoretical and experimental techniques that were developed by soft matter researches, there is still a lot not known. For instance, even a simple question like: how does the cell or its components exert forces, is largely unknown

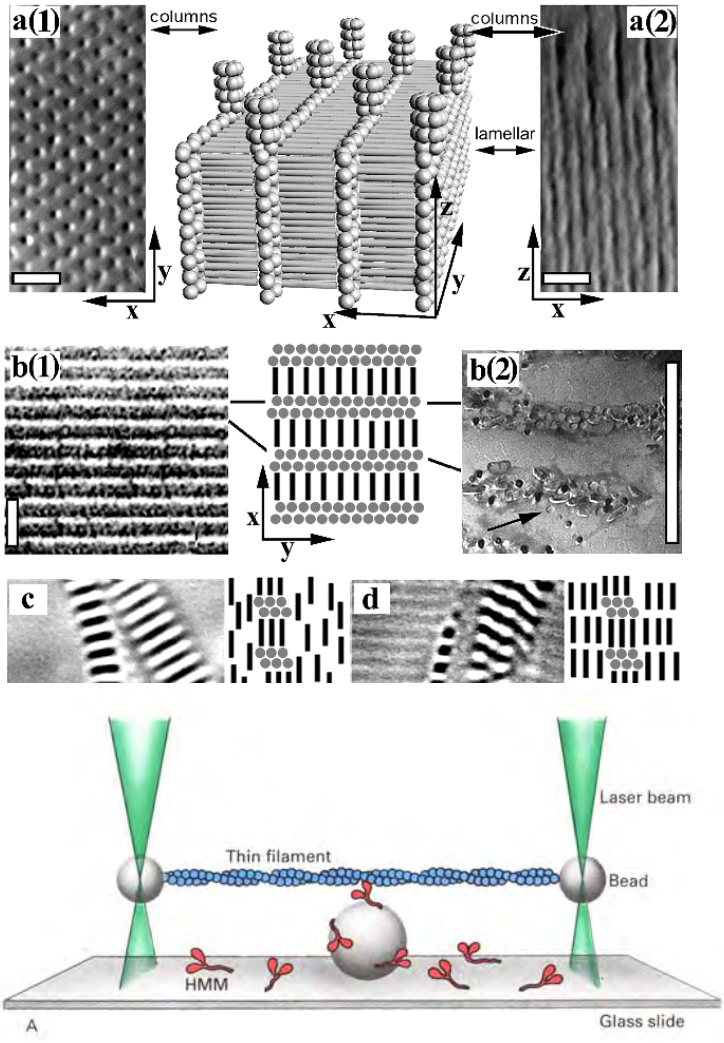



**Figure 1.9 Cross-fertilization between the (micro)biology inspired physics and soft condensed matter fields. Top: Mixture of monodisperse colloidal rods (virus particles) and spheres show an intriguing and complex phase behavior, next to schematic pictures the phases are demonstrated with (polarization)microscopy images (M. Adams, et al., *Nature*, 393, 349-352 (1998)). Bottom: Optical tweezers grabbing colloidal latex spheres with an biological actin filament attched (explained in Chapter 9) are used to measure forces of molecular motors attached to another colloidal particle.**

and an active field of research. It is fair to say that because of the large number of researchers in this field of biology inspired physics, there is also a flow of knowledge starting to go the other way (Fig. 1.9). One nice example is the possibility to induce specific, temperature reversible, interactions between colloids by coating them with single strands of DNA that will only interact with other particles that are coated with the complementary sequence. Similar cross-fertilization is also starting in the field of biology inspired (soft) materials science.

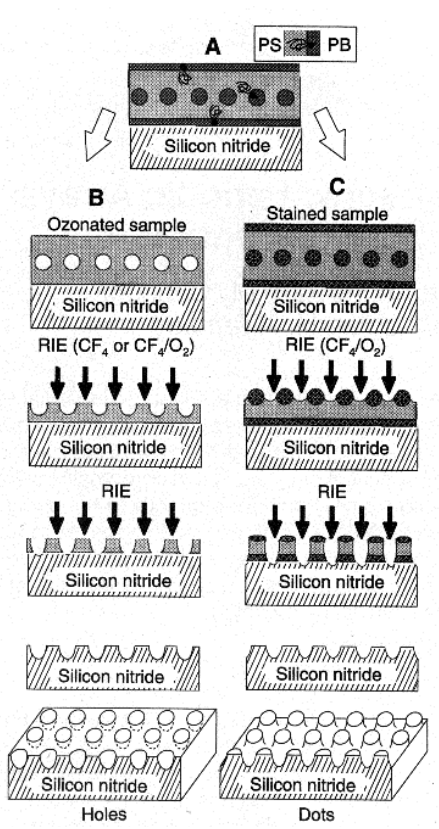



**Figure 1.10 Self-assembled array of spherical di-block domains are used in lithographic procedures to make a pattern of holes or pillars in silicon nitride with dimensions of only a few nm. (M. Park, et al., *Science*, 276, 1401 (1997)).**

## 1.5 Connection with NanoScience and NanoTechnology

The (materials) science 'buzzword' at present is certainly anything that has 'nano' in it. Not only have the promises of 'nanoscience' and 'nanotechnology' inspired governments to increase funding in this area, the general public is starting to become aware of this 'new field' of science as well. Already, there are serious warnings in editorials in the journal *Nature* to the scientists working in this field that 'objections and worries about the dangers of this emerging field need to be taken seriously'. By definition anything that has at least one dimension with feature sizes under 100 nm is determined to be part of nano-science. This length scale is chosen completely arbitrarily. Part of the motivation to study systems below this size is that this is roughly the size at which commercial IC's are produced at the moment. And although it has already been predicted for over 20 years, there will indeed be an end to the doubling of the number of transistors on a silicon chip every (now) two years. Also without the 'hype' it is fair to say that the ability to make and design structures below 100 nm

is increasing rapidly. As may be clear from the above, soft matter scientists are playing an important role in this field.

Another reason for increased expectations is that new effects arise when one enters dimension truly close to a nm. For particles of this size one can no longer state that $\hbar = 0$ and quantum mechanical effects manifest itself. This can be understood qualitatively by considering the QM description of an electron in a box. Confinement of the wave function leads to a discrete spectrum of energy levels where the spacing is determined by the size of the box. This explains why metal particles or semiconductor crystals with sizes of several nm have strong size dependant properties and why, for instance their absorption spectrum shows discrete absorption bands. In similar ways as we can tune the interparticle interactions for larger colloids, one can now chemically tune the wave function of these '*quantum dot*s' (chapter 8).

What we have hoped to have achieved after following these SCM lectures is the ability to understand why in nano-science is not is not possible to just scale down a steam engine to a micron size and expect it to still work. It is amazing how many 'serious' scientists do not realize this.

**(General)Textbooks**


1) *Colloidal Dispersions*, W. B. Russel, D. A. Saville, W. R. Schowalter, Cambridge Un. Press, Cambridge, 1995.
2) *An Introduction to Dynamics of Colloids*, J. K. G. Dhont, in Studies in Interface Science, series Eds. D. Möbius, R. Miller, Elsevier, Amsterdam, 1996.
3) *The Colloidal Domain, where physics, chemistry, biology and technology meet*, D. F. Evans, H. Wennerström, in Advances in interfacial engineering series, VCH, New York, 1994.
4) *Principles of condensed matter physics*, P. M. Chaikin, T. C. Lubensky, Cambridge Un. Press, Cambridge, 1997.
5) *The Structure and Rheology of Complex Fluids*, R. G. Larson, in Topics in Chemical Engineering, Oxford Un. Press, Oxford, 1999.
6) *Intermolecular and Surface Forces*, J. Israelachvili, 2nd ed., Academic Press, London, 1992.
7) *Foundations of Colloid Science*, *Volumes I and II*, R. J. Hunter, Clarendon Press, Oxford, 1991.
8) *Structured Fluids, polymers, colloids, surfactants*, T.A. Witten, Ph. A. Pincus, Oxford Un. Press., 2004.
9) *Soft Condensed Matter*, A. L. Jones, Oxford Un. Press, 2002.


**Origin Figures**

All figures in this chapter have been obtained from the Timeline of soft condensed matter that was presented at the 100 year APS celebration, a version of this timeline can be found at: http://www.nat.vu.nl/~fcm/ComplexFluids/ComplexFluids.html

# 2. Statistical mechanics of fluids

Arnout Imhof
*Soft Condensed Matter*
*Utrecht University*

## 2.1. Introduction

Many properties of soft condensed matter can be described accurately by considering the dispersing liquid, or solvent, as a continuum, which is characterized by bulk macroscopic properties such as the viscosity and the permittivity. In this approach the colloidal particles play the role of molecules. In such a 'supramolecular fluid' the colloids interact via effective forces that are mediated by collections of ions and molecules. This is possible thanks to the large separation in length and time scales between the colloidal particles of interest and the 'true' molecules of which the system exists. For example, the interaction between two electrically charged colloidal particles in a solvent is really an interaction between collections of ions distributed over the particles and in their immediate surroundings. But because the latter move much faster they are (almost) always in equilibrium with the pair. Thus, at every instant the so-called 'mean force' between the colloids arises from the statistical average over the equilibrium ensemble of possible configurations of ions around the fixed colloids.

In many systems of interest the number of colloidal particles is still much too large, so that we will need to turn to statistical mechanics to obtain a description of their macroscopic behavior. The statistical mechanics of fluids was developed for molecular liquids and gases. But the results can be directly transplanted to soft condensed matter systems by considering them as supramolecular fluids. This has been a very successful approach in the study of concentrated dispersions. In this chapter we will present the main results from this theory. When we speak of 'particles' in this chapter, therefore, they can be either the 'ordinary' atoms or molecules for which the theory was originally intended, or the larger particles that make up the supramolecular fluid. It should be kept in mind, however, that the analogy is valid only for a system's equilibrium properties. Dynamical properties, such as diffusion and flow of particles, must be derived from a different set of starting equations, which reflect the fact that they move not through a vacuum but through a low molecular, viscous liquid. This lies mostly outside the scope of this course.

## 2.2. Van der Waals theory

One of the oldest and best known theories of the liquid state is that of van der Waals (1873). We start with it because it shows the connection between molecular properties on the one hand and the macroscopic phase behavior on the other. It will set the stage for the more modern treatments of fluids in the following sections.

The classical ideal gas laws[1] $p = NkT/V$ and $U = 3NkT/2$ do not hold for real gases at finite density $\rho = N/V$. Van der Waals realized that this is caused by interactions between the molecules. It is remarkable that he sought to describe both gas and liquid phases in one equation of state. He distinguished two intermolecular forces: a short ranged steep ("hard") repulsion due to the finite size of the molecules, and a longer ranged ("soft") attraction that is responsible for the cohesion of liquids. The repulsion reduces the space available to the molecules by an amount $bN$. The attraction reduces the pressure exerted on the walls of a container by an amount proportional to the number of pairs of molecules, or $a\rho^2$. The *van der Waals equation of state* is therefore

$$p = \frac{kT}{v-b} - \frac{a}{v^2} = \frac{\rho kT}{1-b\rho} - a\rho^2, \tag{2.1}$$

where $v = V/N = 1/\rho$ is the volume per molecule. The van der Waals constants $a$ and $b$ are dependent on the type of gas. Figure 2.1 shows the dependence of the pressure on density. At

[1] Here, $p$ is the pressure, $N$ the number of molecules, $V$ the volume, $k$ Boltzmann's constant, $U$ the internal energy, and $T$ the temperature.

high enough temperatures the pressure increases monotonically and there is no difference between the gas and liquid phase. But for sufficiently low temperatures there is a "van der Waals loop": along an isotherm the pressure decreases when the density is raised. This is where the system is unstable and phase separates into a dilute (gas) and a dense (liquid) phase.

The critical isotherm separates the two regimes. It has a point where the first and second derivatives both vanish, known as the critical point:

$$\begin{cases} \left(\frac{\partial p}{\partial v}\right)_{T_c} = 0 \\ \left(\frac{\partial^2 p}{\partial v^2}\right)_{T_c} = 0 \end{cases} \Rightarrow \begin{cases} v_c = 3b \\ kT_c = \frac{8a}{27b} \\ p_c = \frac{a}{27b^2} \end{cases}. \tag{2.2}$$

Using Eq. (2.2) we can introduce the reduced variables $\tilde{p} = p/p_c$, $\tilde{v} = v/v_c$, $\tilde{T} = T/T_c$. Subsitution into Eq. (2.1) produces the result

$$\tilde{p} = \frac{8\tilde{T}}{3\tilde{v} - 1} - \frac{3}{\tilde{v}^2}. \tag{2.3}$$

This is known as the law of corresponding states. Its importance is in the fact that van der Waals was able to classify a large number of systems with just a single equation of state: different gases behave the same if the reduced parameters are the same.

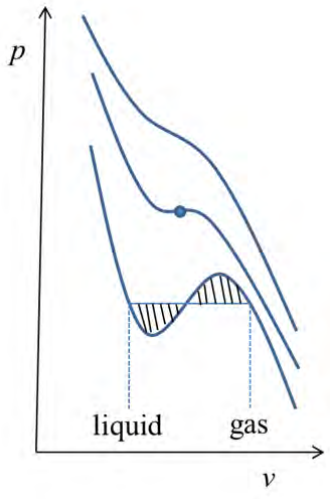



**Figure 2.1. Van der Waals equation of state for a number of different temperatures.**

The densities of the coexisting phases can be found (ad hoc) by the so-called *Maxwell equal area construction*: the segment of the isotherm where the loop occurs must be replaced by a horizontal line that is chosen such that the two shaded areas are equal. This can be seen as follows: In equilibrium, the isothermal transfer of an infinitesimal amount of material from the coexisting liquid (L) to the coexisting gas (G) phase should result in a zero change in Gibbs free energy: $(dG)_{T,N} = Vdp = 0$. Thus,

$$\int_L^G Vdp = [pV]_L^G - \int_L^G pdV = 0, \tag{2.4}$$

from which the Maxwell construction follows immediately.

Although the van der Waals equation is only qualitatively valid, it is useful to consider how it manages to capture molecular interactions. The parameter *b* is the *excluded volume* per

molecule. A spherical particle of diameter $d$ excludes a spherical volume of radius $d$ to another particle, so per particle

$$b = \tfrac{2}{3}\pi d^3 . \tag{2.5}$$

The interpretation of $a$ can be found by first deriving the equation of state for the internal energy of the van der Waals gas. By differentiating $dU = TdS - pdV$ with respect to $V$ at constant $T$ we find

$$\begin{aligned}\left(\frac{\partial U}{\partial V}\right)_T &= T\left(\frac{\partial S}{\partial V}\right)_T - p \\ &= T\left(\frac{\partial p}{\partial T}\right)_V - p\end{aligned} \tag{2.6}$$

The last equality follows from the equation for the Helmholtz free energy: $dA = -SdT - pdV$. Making use of (2.1) we find

$$\left(\frac{\partial U}{\partial V}\right)_T = \frac{a}{v^2} . \tag{2.7}$$

Realizing that $v = V/N$, we obtain

$$U = -\frac{N^2 a}{V} + f(T) . \tag{2.8}$$

The second term in this equation is an unknown function of temperature. The first term depends only on the density of the fluid. It is the average potential energy of the molecules, which changes only if the distance between them changes. This "configurational energy" can be estimated with a molecular picture: Suppose that the potential energy of two molecules at a separation $r$ is $\phi(r)$ and that the local density of surrounding molecules is constant $\rho$ for $r > d$. Then the average potential energy of a molecule in the field its neighbors is

$$\rho \int_{2R}^{\infty} \phi(r) 4\pi r^2 dr . \tag{2.9}$$

Then the configurational energy is

$$\Phi_{\text{conf}} = \frac{1}{2} N \rho \int_{d}^{\infty} \phi(r) 4\pi r^2 dr , \tag{2.10}$$

where the factor ½ avoids double counting. Comparing this expression with (2.8) we find

$$a = -\frac{1}{2}\int_{d}^{\infty} \phi(r) 4\pi r^2 dr . \tag{2.11}$$

The magnitude of this constant can therefore be calculated if an expression is known for the pair interaction between molecules. We now call this the van der Waals interaction.

## 2.3. Virial expansion of the pressure

For temperatures above the critical point and at low densities Kamerling Onnes (1901) proposed a general series expansion of the pressure:

$$p = \rho kT\left(1 + B_2(T)\rho + B_3(T)\rho^2 + \ldots\right) \tag{2.12}$$

The temperature dependent coefficient $B_i$ is called the *i-th virial coefficient*. It is possible, for example, to expand the van der Waals equation in a Taylor series. This gives

$$\begin{aligned} B_2 &= b - \frac{a}{kT} . \\ B_3 &= b^2 \end{aligned} \tag{2.13}$$

The second virial coefficient has a repulsive and an attractive component. With decreasing temperature $B_2$ can become zero (at the Boyle temperature), or even negative. With the help of statistical thermodynamics it is possible to derive general expressions for the virial coefficients. We will show how this is done for $B_2$.

Using the canonical ensemble the Helmholtz free energy *A* is given by

$$A = -kT \ln Z_N , \tag{2.14}$$

in which

$$Z_N = \frac{1}{N!}\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}N} q_{\text{int}}^N Q_N , \tag{2.15}$$

$$Q_N = \int_V \exp\left[-\Phi\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right)/kT\right] d\mathbf{r}_1 \ldots d\mathbf{r}_N . \tag{2.16}$$

Here, *h* is Planck's constant and *m* the mass of a particle. $q_{\text{int}}$ is the partition function of a single particle and accounts for vibrations and rotations. The *configurational integral* $Q_N$ contains the potential energy $\Phi$ of the collection of molecules as a function of their positions $\mathbf{r}_i$; each of the *N* integrals is over the system volume *V*. Clearly, for an ideal gas $\Phi \equiv 0$ so that we find $Q_N = V^N$. The pressure is then given by

$$p = -\left(\frac{\partial A}{\partial V}\right)_{N,T} = kT\left(\frac{\partial \ln Q_N}{\partial V}\right) = kT\frac{N}{V} = \rho kT , \tag{2.17}$$

as expected.

With increasing density intermolecular forces will start to give ever more important contributions to the pressure. First, interactions between the $\frac{1}{2}N\left(N-1\right)$ *pairs* of molecules become important. We let $\phi\left(r_{ij}\right)$ again be the interaction between the pair of molecules *i* and *j* separated by the distance $r_{ij} = \left|\mathbf{r}_j - \mathbf{r}_i\right|$. Then we can write

$$\exp\left[-\Phi/kT\right] = \exp\left[-\sum_{\text{pairs}} \phi\left(r_{ij}\right)/kT\right] = \prod_{\text{pairs}} \exp\left[-\phi\left(r_{ij}\right)/kT\right]. \tag{2.18}$$

If we now define the *Mayer function*

$$f_{ij} = \exp\left[-\phi\left(r_{ij}\right)/kT\right] - 1, \tag{2.19}$$

then it is clear that $f_{ij}$ is small compared to unity, unless the particles approach each other closely (at high densities). Continuing with (2.18),

$$\exp\left[-\Phi/kT\right] = \prod_{\text{pairs}}\left(1+f_{ij}\right) = 1 + \sum_{\text{pairs}} f_{ij} + \text{terms like } f_{ij}\, f_{kl} \tag{2.20}$$

Ignoring the second and higher order terms we find after substitution in Eq. (2.16):

$$\begin{aligned} Q_N &= \int_V d\mathbf{r}_1 \ldots d\mathbf{r}_N \left(1 + \sum_{\text{pairs}} f_{ij}\right) \\ &= V^N + \tfrac{1}{2}N\left(N-1\right)V^{N-2}\int_V\int_V f_{12}\left(r_{12}\right) d\mathbf{r}_1 d\mathbf{r}_2 + \ldots \end{aligned} \tag{2.21}$$

Since $f_{12}$ only depends on the distance $\left|\mathbf{r}_2 - \mathbf{r}_1\right| \equiv r$, but not on the two positions separately, we can perform one more integration over *V*. We end up with

$$Q_N = V^N\left(1 + \frac{N\left(N-1\right)}{2V}\int_V f_{12}\left(r\right) d\mathbf{r} + \ldots\right). \tag{2.22}$$

If we differentiate $\ln Q_N$ the same way as in Eq. (2.17) we finally find

$$p = kT\left(\rho + B_2\rho^2 + \ldots\right) \tag{2.23}$$

with

$$\begin{aligned} B_2 &= -\tfrac{1}{2}\int_V f_{12}(r)\,d\mathbf{r} \\ &= \tfrac{1}{2}\int_V \left(1-\exp\left[-\phi(r)/kT\right]\right)d\mathbf{r}. \end{aligned} \tag{2.24}$$

This is the general expression we were looking for. A derivation of the higher order virial coefficients starts from the grand canonical ensemble, but we will not pursue that here.

It is straightforward to show that for hard spherical particles

$$B_2 = \tfrac{2}{3}\pi d^3. \tag{2.25}$$

Compare this with the van der Waals results Eqs. (2.5) and (2.13).

Notice that if the hard spheres also have a long range interaction that is weak, such that $\phi(r) \ll kT$ for $r > d$, the exponential in Eq. (2.24) can be linearized and we recover the van der Waals result of Eqs. (2.13) and (2.11). Hence, if we neglect interactions between three and more particles, *and* we assume that the attractive interactions are weak, then the van der Waals equation is obtained. In practice, this equation is an improvement of the ideal gas law (above $T_c$) and it qualitatively describes liquid-gas phase separation (below $T_c$). In the following we will consider methods to derive equations of a more general validity starting from statistical mechanics. But first we have to make a few remarks on the application of these methods to supramolecular fluids.

## 2.4. Osmotic pressure

The statistical mechanical formalism introduced above, Eq. (2.15), can of course also be applied to a much wider class of systems such as solutions of colloids or macromolecules. These solutions consist of a low molecular liquid (or a solution of low molecular components) in which a component of much larger size is dissolved (or dispersed). Their size is typically between a nanometer and a micrometer. The lower limit should be larger than the size of the low molecular components so that these form a continuous background. The upper limit should not be so large that the particles sediment before they have had the time to reach a thermodynamic equilibrium. This way such solutions resemble a gas or a liquid of giant atoms floating in a continuous background. This view reduces the number of particles enormously by removing those in which we are not interested. Nevertheless, the number of supramolecular particles is still so large that statistical mechanical methods are needed.

In such a treatment the usual thermodynamic parameters, such as the temperature and the system volume, still apply. Only the pressure has to be replaced by the *osmotic pressure*. A natural way to introduce it is by considering a *membrane equilibrium*, as shown schematically in Figure 2.2. The colloids are separated from a reservoir containing the solvent by means of a semipermeable membrane. The membrane must be permeable to the low molecular components and impermeable to the colloids or macromolecules. When the system is in equilibrium the chemical potentials of the low molecular components on both sides of the membrane are equal and there is a pressure difference across the membrane: the osmotic pressure $\Pi$. It can be measured from the liquid rise in the colloid phase and the mass density of the suspension:

$$\Pi = p_{\text{colloid}} - p_{\text{reservoir}} \tag{2.26}$$

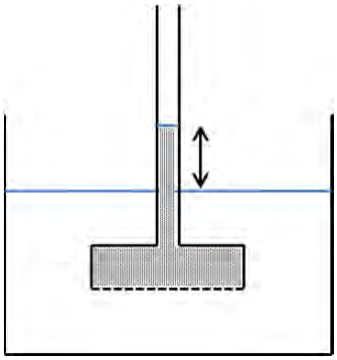


**Figure 2.2. Membrane equilibrium between a macromolecular solution (white) and its solvent (grey) separated by a semipermeable membrane (dashed line).**

For the statistical mechanical formulation of the problem the size of the particles enclosed by the membrane is of no consequence for the osmotic pressure (as Einstein argued for the first time). So, for dilute solutions as well as dilute suspensions van 't Hoff's law is valid:

$$\Pi = \rho kT \ . \tag{2.27}$$

Osmotic pressures for colloidal systems will therefore tend to be significantly smaller than for molecular solutions. For higher concentrations we can use a virial expansion analogous to (2.12):

$$\Pi = \rho kT\left(1 + B_2(T)\rho + B_3(T)\rho^2 + \ldots\right). \tag{2.28}$$

We have seen that the second virial coefficient depends on the interactions between a *pair* of particles, see Eq. (2.24). We only need to describe the meaning of the potential energy $\phi(r)$ in that expression. It is the reversible work (at constant *T*) needed to move a particle from an infinite distance to a distance *r* of another particle. During the move the molecular components of the suspension (solvent, ions, etc.) always remain in equilibrium with the pair of particles. Put differently, $-d\phi/dr$ is the statistically averaged force that the particles exert on each other at the separation *r*. Therefore, $\phi$ is called the *potential of mean force*. As we will see in a later chapter this force has its origin in the forces between electric double layers, London-van der Waals forces, and steric forces that may arise from the presence of polymers on the particle surface.

Colloidal dispersions often contain several low molecular components that can pass the membrane, such as electrolytes. This is called the *Donnan equilibrium*. Apart from an osmotic pressure difference there is now also an electric potential difference, the *Donnan potential* $\psi$. An elementary derivation that assumes ideal solution behavior of the monovalent ions shows that

$$\Pi = kT\left[\rho + 2\rho_e\left\{\sqrt{1+\left(Z\rho/2\rho_e\right)^2} - 1\right\}\right], \tag{2.29}$$

$$\rho_i^{+/-} = +/-\tfrac{1}{2}Z\rho + \rho_e\sqrt{1+\left(Z\rho/2\rho_e\right)^2}\ , \tag{2.30}$$

$$\psi = -\frac{kT}{e}\ln\left[\rho_i^{+}/\rho_e\right]. \tag{2.31}$$

In these equations, the colloid carries *Z* negative charges, *e* is the elementary charge unit, $\rho_e$ the electrolyte concentration (number density) in the reservoir, and $\rho_i^{+/-}$ the concentration of the cations/anions inside the membrane.

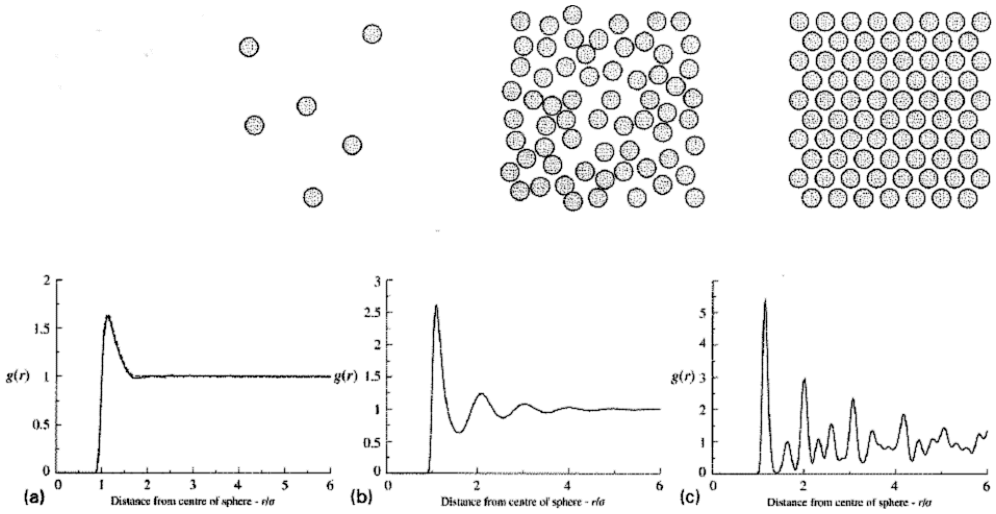



**Figure 2.3. Typical radial distribution functions for (a) a gas, (b) a liquid, and (c) a solid. They have been calculated using a molecular dynamics simulation for atoms interaction through a Lennard-Jones potential. The state parameters are:** $T = 2\varepsilon/k$, $\varepsilon/k$, $0.2\varepsilon/k$ **and** $\rho d^3 =$ **0.05, 0.8, 0.9, respectively. From: Barrat and Hansen, *Basic concepts for simple and complex liquids*, 2003.**

## 2.5. Distribution functions in dense fluids

We now ask ourselves the question how particles in a concentrated fluid are distributed in space. It will qualitatively be clear that this will be different in gases, liquids, and solids (see Figure 2.3). In gases (and dilute colloids) the particles rarely interact with each other, and when they do, no more than two particles will interact with each other. The positional ordering will be extremely small. In liquids (and concentrated colloids) the particles will interact much more often and with many particles at the same time. The particles will be ordered over a distance of a few particle diameters, because the neighbors of a given particle have to stay at least one diameter away, and the next-neighbors two diameters. The order will only be short-lived because the positions of particles are not fixed due to thermal motion. In solids the order becomes long ranged. The particles even assume positions that are periodic in space. The order is also long-lived.

In this section we will introduce a number of distribution functions that are used to quantify the order just described. We will see that some of them can be measured, for example with light scattering, and that they are directly related to thermodynamic quantities.

We again consider a system of $N$ particles in a volume $V$ at a constant temperature $T$ (canonical ensemble). The probability of finding a specified particle, say particle 1, in a volume element $d\mathbf{r}_1$, and another specified particle, say particle 2, in volume element $d\mathbf{r}_2$, etc., is equal to (compare Eq. (2.16))

$$\frac{\exp\left[-\Phi\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right)/kT\right]d\mathbf{r}_1\ldots d\mathbf{r}_N}{Q_N} \equiv \rho_{\text{spec}}^{(N)}\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right)d\mathbf{r}_1\ldots d\mathbf{r}_N\,. \tag{2.32}$$

The function $\rho_{\text{spec}}^{(N)}$ is an example of a probability density function. But it is much too detailed. More useful is a function describing a configuration of fewer particles, say of the first $n$. This is the probability density function

$$\rho_{\text{spec}}^{(n)}\left(\mathbf{r}_1,\ldots,\mathbf{r}_n\right) = \frac{1}{Q_N}\int_V \exp\left[-\Phi\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right)/kT\right]d\mathbf{r}_{n+1}\ldots d\mathbf{r}_N \tag{2.33}$$

It is clear from the normalization that

$$\int_V \rho_{\text{spec}}^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n)d\mathbf{r}_1\ldots d\mathbf{r}_n = 1. \tag{2.34}$$

Real particles cannot be labeled, so it is more realistic to ask for the probability to find *any* particle in $d\mathbf{r}_1$, any other particle in $d\mathbf{r}_2$, etc. … This probability is

$$\rho^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n) = \frac{N!}{(N-n)!}\rho_{\text{spec}}^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n). \tag{2.35}$$

From (2.34) the normalization follows:

$$\int_V \rho^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n)d\mathbf{r}_1\ldots d\mathbf{r}_n = \frac{N!}{(N-n)!}. \tag{2.36}$$

Case $n = 1$:

The function $\rho^{(1)}(\mathbf{r})$ is called the *singlet distribution*. For a crystal it has sharp peaks, in an inhomogeneous fluid it is a smooth function of $\mathbf{r}$, and in a homogeneous fluid it is a constant. In the latter case it follows from (2.36) that

$$\rho^{(1)} = \frac{N}{V} = \rho. \tag{2.37}$$

Case $n = 2$:

The function $\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2)$ in this very important case is called the *pair distribution function*. It is the probability to find any particle in $d\mathbf{r}_1$ and at the same time any other particle in $d\mathbf{r}_2$. We often want to compare this probability with the case of non-interacting particles at the same density. This defines the (dimensionless) *pair correlation function*

$$g^{(2)}(\mathbf{r}_1,\mathbf{r}_2) = \frac{\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2)}{\rho^{(1)}(\mathbf{r}_1)\rho^{(1)}(\mathbf{r}_2)}. \tag{2.38}$$

In homogeneous fluids $g^{(2)}(\mathbf{r}_1,\mathbf{r}_2)$ must be translationally invariant, and so it can only depend on the distance $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$ between the particles:

$$g^{(2)}(r_{12}) = g^{(2)}(\mathbf{r}_1,\mathbf{r}_2) \tag{2.39}$$

This function is called the radial distribution function. Often the subscripts are dropped. If $g(r)$ has, for example, the value 2 at some distance then the probability to find two particles at that distance from each other is twice as high as it would have been for an ideal gas at the same density $\rho$. From (2.36), (2.37), and (2.38) we see that the normalization is

$$\int_V\int_V g^{(2)}(r_{12})\,d\mathbf{r}_1 d\mathbf{r}_2 = \frac{N(N-1)}{\rho^2}. \tag{2.40}$$

General case:

In general we can define triplet, quadruplet, etc. … correlation functions as

$$g^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n) = \frac{n^{(n)}(\mathbf{r}_1,\ldots,\mathbf{r}_n)}{n^{(1)}(\mathbf{r}_1)n^{(1)}(\mathbf{r}_2)\cdots n^{(1)}(\mathbf{r}_n)}. \tag{2.41}$$

These are mainly used in finding approximative equations for the pair correlation function (so-called closures) and we will not consider them in the following.

Some examples:

1. The most important of the distribution functions is the pair distribution function. Calculation of this function (let alone that of the higher ones) is not an easy task. The most

accurate method is to solve the problem numerically, using computer simulations. Some examples are shown in Figure 2.3 for a Lennard-Jones (LJ) fluid, the particles of which interact through

$$\phi_{\mathrm{LJ}}(r) = 4\varepsilon\left(\left(\frac{d}{r}\right)^{12} - \left(\frac{d}{r}\right)^{6}\right). \tag{2.42}$$

This potential has a repulsive part and an attractive part. The parameter $\varepsilon$ characterizes the strength of the interactions. The LJ interaction describes the behavior of small quasi-spherical atoms such as argon quite well.

2. Another example is shown in Figure 2.4 for the case of hard spheres (HS), which form an important reference system. Their pair interaction potential is

$$\phi_{\mathrm{HS}}(r) = \begin{cases} \infty & (r \le d) \\ 0 & (r > d) \end{cases}. \tag{2.43}$$

The density was chosen such that the volume fraction of particles was $\eta = \frac{\pi}{6}\rho d^3 = 0.49$. This is just below the volume fraction of 0.494, where the hard sphere fluid is known to begin forming a crystalline solid.

3. Although the distribution functions are hard to calculate, they are (relatively) easy to measure experimentally. Scattering methods have been used the most, but more recently 3D confocal microscopy is also used to measure $g^{(2)}(r_{12})$. In Figure 2.5 an example is shown of colloids interacting through an electric double layer. The interactions were modeled with the hard-core Yukawa interaction, which correctly describes weak double layer overlap:

$$\phi_{\mathrm{Yuk}} = \begin{cases} \infty & (r \le d) \\ \varepsilon \dfrac{\exp\left[-\kappa(r-d)\right]}{r/d} & (r > d) \end{cases} \tag{2.44}$$

It consists of a hard sphere interaction dressed with a repulsive region characterized by a width $\kappa$ and strength $\varepsilon$. Monte Carlo simulations were used to find the parameter values that best fitted all radial distribution functions simultaneously. This was the case with $\varepsilon/kT = 140$ and $\kappa d = 5$.

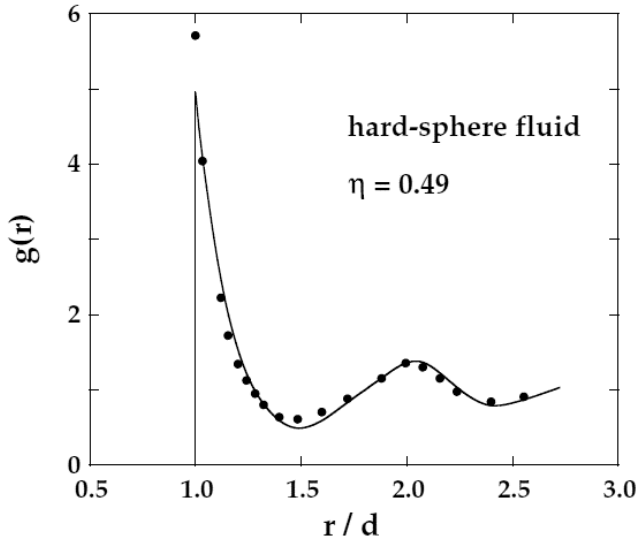



**Figure 2.4. The radial distribution function of a hard-sphere fluid close to the liquid-solid transition. The curve is a calculation using the Percus-Yevick approximation and the points are the results of a Monte Carlo simulation. From: Hansen and McDonald, *Theory of simple liquids*, 3rd ed., 2005.**

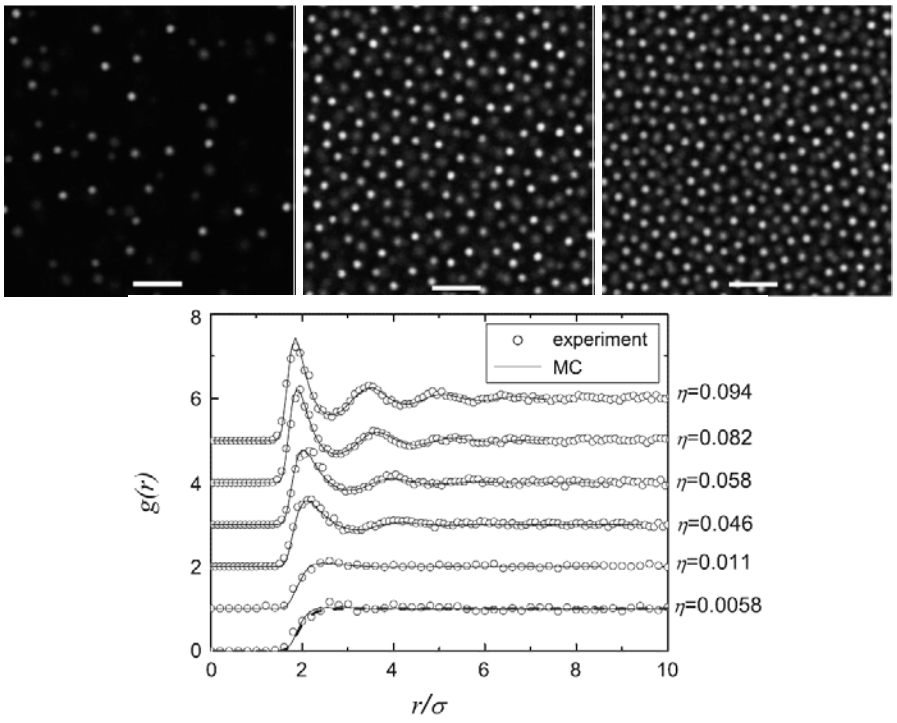



**Figure 2.5. Snapshots taken from a 3D confocal microscopy dataset of charged colloids at volume fractions of 0.011, 0.058, and 0.094. Below are the radial distribution functions calculated from these data (points) and Monte Carlo simulations using the Yukawa potential with $\varepsilon = 140kT$ and $\kappa d = 5$. Scale bars are 10 μm. From: Royall et al., *J. Phys.: Condens. Matt.* 15, S3581 (2003).**

## 2.6. Relation between $g^{(2)}(r_{12})$ and thermodynamic quantities

Apart from direct measurements of $g^{(2)}(r_{12})$ it is also possible to measure thermodynamic quantities, for example the pressure as a function of density. There are three independent routes that relate the pair correlation function to thermodynamic state functions. The first two routes assume pairwise additive interactions, the third does not. Pairwise additive means that the potential energy of the system can be written as

$$\Phi(\mathbf{r}_1,\ldots,\mathbf{r}_N)=\sum_{i>j}\phi(r_{ij}). \tag{2.45}$$

### 2.6.1. The caloric route

The first route is derived by calculating the internal energy using the canonical ensemble, Eq. (2.15). Ignoring the internal degrees of freedom of the particles,

$$\begin{aligned} U &= -\left(\frac{\partial \ln Z_N}{\partial(1/kT)}\right)_{N,V} \\ &= \frac{3}{2}NkT+\frac{1}{Q_N}\int_V \Phi(\mathbf{r}_1,\ldots,\mathbf{r}_N)\exp\left[-\Phi(\mathbf{r}_1,\ldots,\mathbf{r}_N)/kT\right]d\mathbf{r}_1\ldots d\mathbf{r}_N. \end{aligned} \tag{2.46}$$

The first term is the average kinetic energy (it is the equipartition theorem). The second term is the average potential energy in the system. We assume pairwise additivity, and since all particles are identical the second term becomes

$$\begin{aligned} &\frac{1}{2}N(N-1)\int_V \phi(r_{12})\frac{1}{Q_N}\int_V \exp\left[-\Phi/kT\right]d\mathbf{r}_3\ldots d\mathbf{r}_N\, d\mathbf{r}_1 d\mathbf{r}_2 \\ &=\frac{1}{2}\int_V \phi(r_{12})\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \\ &=\frac{1}{2}\rho^2\int_V \phi(r_{12})g^{(2)}(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \end{aligned}$$

where we made use of (2.33), (2.35), and (2.38). Assuming a homogeneous fluid, (2.39), we can perform one more integration to obtain

$$U=\frac{3}{2}NkT+\frac{1}{2}N\rho\int_V \phi(r)\,g(r)\,d\mathbf{r}. \tag{2.47}$$

This derivation was rather formal, but the result can be derived more simply: The probability of finding a particle in a volume element $d\mathbf{r}$ at a distance $r$ from a given particle is per our definition $\rho g(r)\,d\mathbf{r}$. The average interaction energy is therefore $\rho\phi(r)\,g(r)\,d\mathbf{r}$. Integrating this over all space, summing over all particles, and correcting for double counting we find (2.47). Compare the second term with (2.10): we now see that for the van der Waals gas it is assumed that $g(r)\equiv 1$.

### 2.6.2. The virial theorem

Another route is found by calculating the pressure from $p=kT\left(\partial \ln Z_N/\partial V\right)_{T,N}$. Assuming pairwise interactions this leads to (see problems)

$$p=\rho kT-\frac{\rho^2}{6}\int_V r\frac{d\phi}{dr}g(r)\,d\mathbf{r}\ . \tag{2.48}$$

### 2.6.3. The compressibility equation

The third route relates to the compressibility of the system. To calculate this quantity we must allow the number of molecules in the system to fluctuate. We imagine the system to be a fixed subvolume in a large reservoir that can exchange particles at constant temperature and chemical potential $\mu$. This is the situation in, for example, a scattering experiment where a part of the sample is illuminated by a beam of radiation. There is a direct relationship between the isothermal compressibility $\kappa_T$ and the fluctuations in the number of particles in a subvolume $V$:

$$\kappa_T \equiv \frac{1}{\rho}\left(\frac{\partial \rho}{\partial p}\right)_T = \frac{V}{kT}\frac{\left\langle N^2\right\rangle - \left\langle N\right\rangle^2}{\left\langle N\right\rangle^2}. \tag{2.49}$$

On the right hand side we recognize the variance in particle number. It is derived in the appendix. This relation shows that a large compressibility is directly related to large density fluctuations. This is understandable, because in order to get a more than average number of particles in $V$ pressure work has to be done. This work is large if the compressibility is low. Conversely, spontaneous fluctuations will be smaller if more work is needed to produce them. Increased fluctuations are a sign of an approaching phase separation and even diverge at the critical point. Recall Figure 2.1; the critical isotherm has a zero slope at the critical point, so here the compressibility diverges.

Using this relation and the normalization of the pair distribution function, Eq. (2.40), Ornstein and Zernike derived the important *compressibility theorem*:

$$kT\left(\frac{\partial \rho}{\partial p}\right)_T = 1 + \rho\int_V \left(g\left(r\right)-1\right)d\mathbf{r}\,. \tag{2.50}$$

It is also derived in the appendix and remains valid even if the interactions cannot be written as pairwise contributions. It should be remembered that in a colloidal system the pressure $p$ is to be interpreted as the osmotic pressure $\Pi$. It is then the *isothermal osmotic compressibility* that is directly related to the radial distribution function. In the next chapter we will see that it is also related to the forward scattering of radiation and can therefore be measured. The increased scattering near the critical point is called *critical opalescence.*

## 2.7. Potential of mean force

The formalism developed in the previous section can of course be used for colloidal systems if we use the potential of mean force for the pair interactions, and replace the pressure $p$ with the osmotic pressure $\Pi$. One can ask whether the concept of potential of mean force could also be extended to a concentrated system, rather than to a pair of particles. Consider two particles, either molecules or colloids, among a large number of others. What is the average force that these particles, say 1 and 2, exert on each other when they are kept at fixed positions? Denoting $\nabla_1 = \left(\partial/\partial x_1, \partial/\partial y_1, \partial/\partial z_1\right)$, the force on particle 1 is

$$\begin{aligned}\langle \mathbf{F}_{12} \rangle &= \langle -\nabla_1 \Phi \rangle \\ &= \frac{\int d\mathbf{r}_3 \ldots d\mathbf{r}_N \left(-\nabla_1 \Phi\right) \exp\left(-\Phi/kT\right)}{\int d\mathbf{r}_3 \ldots d\mathbf{r}_N \exp\left(-\Phi/kT\right)} \\ &= \frac{kT\nabla_1 \int d\mathbf{r}_3 \ldots d\mathbf{r}_N \exp\left(-\Phi/kT\right)}{\int d\mathbf{r}_3 \ldots d\mathbf{r}_N \exp\left(-\Phi/kT\right)} \\ &= kT \frac{\nabla_1 g\left(r_{12}\right)}{g\left(r_{12}\right)} \\ &= kT \nabla_1 \ln g\left(r_{12}\right).\end{aligned}$$

Therefore, the *potential of mean force* is

$$V\left(r_{12};\rho,T\right) = -kT \ln g\left(r_{12};\rho,T\right), \tag{2.51}$$

where we made explicit that it depends on the density and temperature. It is the reversible work we have to do when bringing particle 1 to a distance $r_{12}$ from infinity, while keeping particle 2 fixed and allowing all the other particles to remain at equilibrium. Note that for low density the contribution of the surrounding particles will become negligible, so that the potential of mean force approaches the pair potential:

$$\phi\left(r_{12}\right) = -kT \ln g\left(r_{12};\rho \to 0\right). \tag{2.52}$$

## 2.8. Weak external fields: linear response theory

We now consider a system that is perturbed by the action of a (weak) external field. The presence of this field causes the system to be (slightly) inhomogeneous; the singlet distribution $\rho^{(1)}\left(r\right)$ is no longer exactly equal to $N/V$. Let $\delta u\left(\mathbf{r}\right)$ describe the potential energy of a particle at position **r** due to interactions with the field. We will assume it to be small compared to the potential energy due to interactions between particles:

$$\delta\Phi = \sum_i \delta u\left(\mathbf{r}_i\right) \ll \Phi . \tag{2.53}$$

The density of the system can be calculated from the definition (2.35), after which we linearize with respect to $\delta u$:

$$\begin{aligned}\rho^{(1)}\left(\mathbf{r}\right) &= N \frac{\int \exp\left[-\left(\Phi+\delta\Phi\right)/kT\right] d\mathbf{r}_2 \ldots d\mathbf{r}_N}{\int \exp\left[-\left(\Phi+\delta\Phi\right)/kT\right] d\mathbf{r}_1 \ldots d\mathbf{r}_N} \\ &= \frac{\frac{N}{Q_N}\int \exp\left[-\Phi/kT\right] d\mathbf{r}_2 \ldots d\mathbf{r}_N - \frac{N}{Q_N}\int \left(\delta\Phi/kT\right) \exp\left[-\Phi/kT\right] d\mathbf{r}_2 \ldots d\mathbf{r}_N}{1 - \frac{1}{Q_N}\int \left(\delta\Phi/kT\right) \exp\left[-\Phi/kT\right] d\mathbf{r}_1 \ldots d\mathbf{r}_N} + O\left(\delta\Phi^2\right) \\ &= \rho_0 - \frac{N}{Q_N}\int \left(\delta\Phi/kT\right) \exp\left[-\Phi/kT\right] d\mathbf{r}_2 \ldots d\mathbf{r}_N + \frac{\rho_0}{Q_N}\int \left(\delta\Phi/kT\right) \exp\left[-\Phi/kT\right] d\mathbf{r}_1 \ldots d\mathbf{r}_N\end{aligned} \tag{2.54}$$

In the last step we recognized the singlet distribution of the unperturbed system: $\rho_0 = N/V$.

Let's consider the third term first. Substituting (2.53) gives a sum of *N* terms. But since all particles are identical the result is just *N* times the term for *i* = 1:

$$\text{(third term)} = \frac{\rho_0}{Q_N} N \int \left(\delta u(\mathbf{r}_1)/kT\right) \exp\left[-\Phi/kT\right] d\mathbf{r}_1 \ldots d\mathbf{r}_N$$

$$= \rho_0 \int \left(\delta u(\mathbf{r}_1)/kT\right) \frac{N \int \exp\left[-\Phi/kT\right] d\mathbf{r}_2 \ldots d\mathbf{r}_N}{Q_N} d\mathbf{r}_1$$

$$= \rho_0^2 \int \left(\delta u(\mathbf{r}_1)/kT\right) d\mathbf{r}_1$$

In the last step we again recognized the singlet distribution in the unperturbed system.

The second term can be handled similarly, but we have to distinguish the term $i = 1$ from the $N - 1$ other terms:

$$\text{(second term)} = -\rho_0 \, \delta u(\mathbf{r}_1)/kT - N(N-1) \int \left(\delta u(\mathbf{r}_2)/kT\right) \frac{\int \exp\left[-\Phi/kT\right] d\mathbf{r}_3 \ldots d\mathbf{r}_N}{Q_N} d\mathbf{r}_2$$

$$= -\rho_0 \, \delta u(\mathbf{r}_1)/kT - \rho_0^2 \int \left(\delta u(\mathbf{r}_2)/kT\right) g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_2$$

Now we are ready to take all the terms together, and seeing that the integration variables are just dummy variables we write:

$$\rho^{(1)}(\mathbf{r}_1) = \rho_0 - \rho_0 \, \delta u(\mathbf{r}_1)/kT - \rho_0^2 \int \left(\delta u(\mathbf{r}_2)/kT\right) \left(g^{(2)}(r_{12}) - 1\right) d\mathbf{r}_2 . \qquad (2.55)$$

This is called the Yvon equation. It has the following interpretation: the external field *directly* changes the density at a point by an amount $-\rho_0 \, \delta u(\mathbf{r}_1)/kT$. This is expected from the Boltzmann distribution (after linearization). But in a nearby volume element $d\mathbf{r}_2$ the number of particles has also changed, namely by $-\rho_0 d\mathbf{r}_2 \, \delta u(\mathbf{r}_2)/kT$. Each extra particle there *indirectly* changes the density at $\mathbf{r}_1$ by $\rho_0 \left(g^{(2)}(r_{12}) - 1\right)$. After integrating over $d\mathbf{r}_2$ we find (2.55).

An interesting case to consider is that of a periodic field, which we shall write as $\delta u(\mathbf{r}) = \delta u_0 \exp(i\mathbf{q} \cdot \mathbf{r})$. From (2.55) we see that the resulting density change is

$$\delta\rho(\mathbf{r}_1) = \rho^{(1)}(\mathbf{r}_1) - \rho_0$$

$$= -\rho_0 \delta u_0 \exp(i\mathbf{q} \cdot \mathbf{r}_1)/kT - \rho_0^2 \, \delta u_0/kT \int \exp(i\mathbf{q} \cdot \mathbf{r}_2) \left(g^{(2)}(r_{12}) - 1\right) d\mathbf{r}_2 \qquad (2.56)$$

$$\equiv -\rho_0 \frac{\delta u(\mathbf{r}_1)}{kT} S(\mathbf{q})$$

where *structure factor* is defined as

$$S(\mathbf{q}) = 1 + \rho_0 \int \exp(i\mathbf{q} \cdot \mathbf{r}) \left(g^{(2)}(r) - 1\right) d\mathbf{r} . \qquad (2.57)$$

The linear response to a periodic field therefore consists of a periodic change in density with the same wavelength and a magnitude that is proportional to the field strength and the structure factor. In the next chapter we will see that this quantity can be measured in a scattering experiment. It is essentially the Fourier transform of the radial distribution function (minus one) and so it is the spectrum of density oscillations around a particle. The response to a field is large when its wavelength coincides with a wavelength that is naturally present in the system. In that case the *susceptibility* of the system is said to be large[2].

One may wonder what would happen if we were to apply an external field that is identical to the pair interaction potential $\phi(r_{12})$ of a particle. This would be like inserting an

[2] The susceptibility is often defined as $\chi(\mathbf{q}) = \rho_0 S(\mathbf{q})/kT$ .

extra particle into the system, say at $\mathbf{r}_1$. But the change in the density at a nearby position $\mathbf{r}_2$ should by definition be $\rho_0\left(g^{(2)}\left(r_{12}\right)-1\right)$. According to (2.55), then, we would have found an integral equation that relates $g^{(2)}$ to itself, and we could hope to solve it! The problem, of course, is that particle insertion does not constitute a weak field. But the interpretation of Eq. (2.55) in terms of a direct and an indirect effect on the density can be used to define a new function, the *direct correlation function* $c(r)$ as

$$h\left(r_{12}\right)=c\left(r_{12}\right)+\rho_0\int c\left(r_{13}\right)h\left(r_{23}\right)d\mathbf{r}_3 \tag{2.58}$$

Here, the *total correlation function* $h(r)$ is defined as

$$h(r)=g(r)-1 \tag{2.59}$$

Eq. (2.58) is known as the *Ornstein-Zernike* (OZ) equation. It can be seen as a definition of $c(r)$, which makes it an exact equation. Its importance is, however, that it is often possible to find suitable approximations to it. These approximations are based on the observation from the Yvon equation that $c(r)$ should resemble the (negative of the) interaction potential. Or, if we want to "undo" the linearization that we used, it should resemble $\exp\left(-\phi\left(r_{12}\right)/kT\right)-1$. One of the approximations, due to Percus and Yevick, works very well for hard spheres. We will not examine these approximations further here.

## 2.9. *Appendix*: Derivation of the compressibility theorem

We imagine the system to be a fixed subvolume in a large reservoir that can exchange particles at constant temperature and chemical potential $\mu$. The average number of particles in the system can be calculated from the grand canonical ensemble, of which the partition function is

$$\Xi\left(\mu,V,T\right)=\sum_{N'=0}^{\infty}e^{\frac{\mu N}{kT}}Z_{N'}\,. \tag{2.60}$$

Here $Z_{N'}$ is the canonical partition function of a system containing $N'$ particles. The average number of particles in $V$ is then

$$\left\langle N\right\rangle=\frac{\sum_{N'=0}^{\infty}Ne^{\frac{\mu N}{kT}}Z_{N'}}{\sum_{N'=0}^{\infty}e^{\frac{\mu N}{kT}}Z_{N'}}\,. \tag{2.61}$$

By differentiation with respect to $\mu$ we find

$$kT\left(\frac{\partial\left\langle N\right\rangle}{\partial\mu}\right)_{T,V}=\left\langle N^2\right\rangle-\left\langle N\right\rangle^2. \tag{2.62}$$

The right hand side is precisely the variance in particle number. We can relate this to the compressibility by making use of

$$\left(\frac{\partial N}{\partial \mu}\right)_{V,T} = \left(\frac{\partial N}{\partial p}\right)_T \left(\frac{\partial p}{\partial \mu}\right)_{T,N}$$
$$= \left(\frac{\partial N}{\partial p}\right)_T \left(\frac{\partial N}{\partial V}\right)_{T,p} \tag{2.63}$$
$$= N\left(\frac{\partial \rho}{\partial p}\right)_T$$

We therefore find

$$kT\left(\frac{\partial \rho}{\partial p}\right)_T = \frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle}, \tag{2.64}$$

which is Eq. (2.49).

To derive the compressibility theorem we start with the pair distribution function defined in the grand canonical ensemble:

$$\rho^{(2)}_{\mathrm{GCE}}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{N=2}^{\infty} p(N)\,\rho^{(2)}_{\mathrm{CE}}(\mathbf{r}_1,\mathbf{r}_2;N), \tag{2.65}$$

where $p(N)$ is the probability to find $N$ particles in $V$. We now use $\rho^{(2)}_{\mathrm{CE}}(\mathbf{r}_1,\mathbf{r}_2;N)$ to denote the pair distribution function of an $N$-particle system in the canonical ensemble, as defined previously by (2.35). Using that equation we find

$$\int_V\int_V \rho^{(2)}_{\mathrm{GCE}}(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 = \sum_{N=2}^{\infty} p(N)\frac{N!}{(N-2)!} = \langle N(N-1)\rangle. \tag{2.66}$$

Defining the pair correlation function $g^{(2)}$ as before and assuming a homogeneous system we obtain

$$\frac{\langle N\rangle^2}{V^2}\int_V\int_V g^{(2)}(r_{12})\,d\mathbf{r}_1 d\mathbf{r}_2 = \langle N(N-1)\rangle. \tag{2.67}$$

This is analogous to (2.40) with $\rho = \langle N\rangle/V$. We perform one more integration and find

$$\int_V g(r)\,d\mathbf{r} = V\frac{\langle N^2\rangle - \langle N\rangle}{\langle N\rangle^2}. \tag{2.68}$$

When this is combined with (2.64), for example by eliminating $\langle N^2\rangle$, we find Eq. (2.50), the compressibility theorem.

# 3. Scattering techniques in soft condensed matter

Arnout Imhof
*Soft Condensed Matter*
*Utrecht University*

## 3.1. Introduction

Scattering techniques form an important class of tools for the study of soft condensed matter. Most laboratories have equipment for doing light scattering measurements, and many commercial instruments (in electrophoresis or particle sizing equipment) use light scattering internally. A light source is used to illuminate the sample and the scattered light is measured as a function of the angle between the incident beam and the detector. Light scattering is a suitable technique for soft condensed matter because the typical length scales (the particle size or the average distance between particles) is on the order of the wavelength of visible light. A few other types of radiation are also used, namely neutrons and X-rays. These scattering techniques are analogous to light scattering, but the wavelength is typically much shorter (0.1 – 10 nm). This makes these techniques useful for systems containing small particles. Even when the particles are not small, however, X-rays and neutrons are still useful when the samples scatter light too strongly (think of milk for example). As we shall see later, the information in this case is contained at small scattering angles (up to a few degrees), and the techniques are called small-angle X-ray or neutron scattering, or SAXS and SANS in short.

Wave scattering is a general phenomenon that occurs whenever the medium through which the wave propagates is inhomogeneous. Sound waves scatter if the medium contains objects with a different sound velocity, X-rays scatter from inhomogeneities in electron density, and light scatters from inhomogeneities in the refractive index. In this chapter we will concentrate on scattering of light, but scattering of other types of waves is almost completely analogous so that the theory we will develop applies equally well to those other types of radiation.

Suppose that an electromagnetic wave is incident on an object. Matter is composed of discrete electric charges, electrons and protons. They are set into oscillatory motion by the electric field of the incident electromagnetic wave. From electrodynamics it is known that the accelerated electric charges must radiate waves in all directions. This secondary radiation is called the radiation *scattered* by the object. Apart from scattering, part of the incident wave may also be converted to other forms of energy, such as heat. This process is called *absorption*. Both scattering and absorption remove light energy from the primary beam, which is thereby attenuated. Together, they are called *extinction*.

The light scattered by a particle in an otherwise homogeneous medium consists of the sum of wavelets scattered by all the subvolumes making up the particle. It is the *interference* between all these wavelets that leads to a characteristic angular dependence of the scattered light. This is the reason that light scattering can be used to measure the properties (size and shape) of colloidal particles. If there are other particles near the first particle then the waves scattered by different particles will also interfere with each other. Then the angular dependence also contains information on the average interparticle distances. As we shall see later, this angular dependence is directly proportional to the structure factor of the dispersion. Because of thermal motion the positions of the scattering particles relative to each other fluctuates continuously. By measuring the time dependence of the scattered light (*Dynamic Light Scattering*) we can obtain information on the dynamics of the scattering particles.

Why doesn't light scatter when it propagates through a homogeneous medium like, say, glass or water? Surely, the charges in these materials also start oscillating and must radiate in all directions! In fact they do, but it should be remembered that the scattered wave seen by an observer is the sum of wavelets originating from every little subvolume illuminated by the light beam. In a homogeneous medium there always exists a second subvolume, about half a wavelength away, which scatters exactly out of phase with a given subvolume. This leads to complete cancellation in all directions but the forward direction. Only when the medium is

inhomogeneous this cancellation does not take place, because half a wavelength away there is a different material. This means that some light is now scattered away from the forward direction. Of course, no medium except vacuum is truly homogeneous. Glass, water and air all consist of atoms and molecules, and therefore even these media scatter light, albeit only very weakly.

We now know that light scattered by a particle depends on the incident wave. In practice, scattering media contain many particles. Thus, the light incident on a given particle is the sum of the primary wave and the waves scattered by all the other particles. This seems like a hopeless problem, but we can simplify it if the total amount of scattered light is small enough so that it does not contribute significantly to the wave seen by the scattering particles. This is called *single scattering*, because a light wave scattered by a single particle does not scatter again from any of the other particles. The assumption is valid if the distance between the particles is large enough or if the amplitude of the scattered wave is very small compared to that of the incident wave. *Multiple scattering* takes place in systems in which this condition is not met. Examples are milk and clouds.

In the cases we shall consider the scattered light has a frequency equal to that of the incident wave. This is called *elastic scattering*. The angle-dependence gives information on the spatial structure of the sample. In *inelastic scattering* the frequencies of the scattered waves are measured in addition to the angle dependence. This gives information on dynamic processes, such as diffusion or relaxation. Many forms of inelastic scattering exist, such as Brillouin scattering and Raman scattering in which the scattered light contains both higher and lower frequencies. And in X-ray scattering the so-called Compton modified scattering is lowered in frequency. In dynamic light scattering the frequency changes are only very small (but not strictly zero), so it is sometimes also called quasi-elastic light scattering.

## 3.2. Measures of scattered light

In light scattering experiments the sample, containing particles suspended in a medium, is irradiated with a beam of light and light scattered at different angles is measured. We will call the direction of the incident beam the *z*-direction. Scattered light is measured by simply placing the detector on a rotation stage. In general, there are two angles to be varied, $\theta$ and $\phi$, but in most instruments only $\theta$ is varied (see Figure 3.1). The *yz*-plane is defined to be the plane containing the incident beam and the detector and is called the *scattering plane*. A special scattering angle is zero degrees, where one measures the sum of the incident wave and the forward scattered wave. This is what one would measure in a normal transmission spectrophotometer. But what are the measured quantities? The *intensity* of the light, of course. (Or, to be more precise: the *irradiance*, the units are $\mathrm{J.m^{-2}.s^{-1}}$.) But to completely characterize the scattered light the polarization state and the phase are also needed. Phases cannot be directly measured, but the polarization state must be carefully characterized by placing polarizers in the incident beam and/or in front of the detector[1].

First, let us consider a sample with only one particle. If we describe the angle dependence of the scattered intensity $I_s$ with the function $F(\theta,\varphi)$ then the intensity measured by a detector placed at a large distance *r* from the sample can be written as

$$I_s(\theta,\varphi) = \frac{I_0 F(\theta,\varphi)}{k^2 r^2} \tag{3.1}$$

[1] A complete discussion of polarization dependent scattering would complicate the discussion considerably. However, in most (but not all) cases relevant for soft condensed matter experiments the two polarization components shown in Figure 3.1 are independent. This is the case when the particles scatter light only weakly. Since this is already a requirement for single scattering it does not impose extra restrictions.

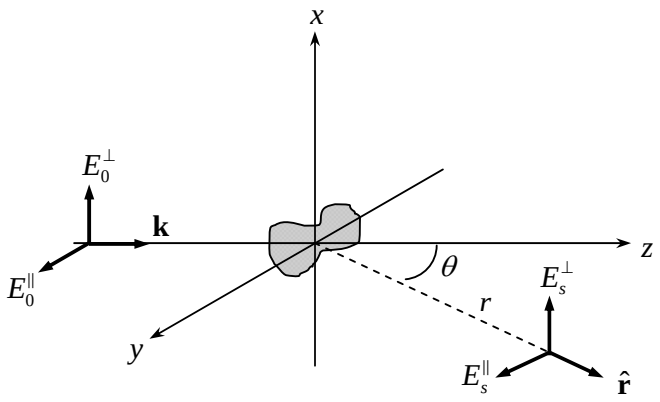



**Figure 3.1. The scattering geometry.**

$I_0$ is the intensity of the incident beam. The factor $1/r^2$ arises because the total scattered intensity over the surface of a sphere must be independent of its radius[2]. Further, $k = 2\pi/\lambda$ is the length of the *wave vector* of the light (measured in the medium). The factor $1/k^2$ is included to make $F$ a dimensionless function. If we integrate (3.1) over the surface of a sphere with radius $r$ we get the total *power* of scattered light $P_s$. The scattered power per unit incident intensity has units of area and is called the *scattering cross section* of the particle:

$$C_{sca} = \frac{P_s}{I_0} = \frac{1}{k^2}\int_{4\pi} F(\theta,\varphi)\, d\Omega \tag{3.2}$$

where $d\Omega = \sin\theta d\theta d\varphi$ is shorthand for the infinitesimal element of solid angle. What is the physical significance of $C_{sca}$? Conservation of energy dictates that the power removed from the incident beam must equal the total scattered power (if there is no absorption). Therefore the power in the light beam is lowered by an amount $C_{sca}I_0$. This is just as if the particle "casts a shadow" of area $C_{sca}$ on a detector placed in the transmitted beam. If there is also absorption then the power received by the transmission detector is reduced by a further $C_{abs}I_0$. The total reduction of the light power must be the sum of scattering and absorption, and is given by the *extinction cross section:*

$$C_{ext} = C_{sca} + C_{abs} \tag{3.3}$$

Now consider a sample containing many identical particles at number density $\rho$. Every particle in the beam reduces the power in the beam further. Assuming only single scattering as usual, it is clear that every particle reduces the power by the amount $C_{ext}I(z)$, and that a thin slab of thickness $dz$ reduces it by $C_{ext}I(z)\rho A dz$, with $A$ the cross sectional area of the beam. Therefore,

$$dI = -\rho C_{ext} I(z) dz\,.$$

Integrating from 0 to $L$, the thickness of the sample, we see that the transmitted intensity is given by

$$I_t = I_0 e^{-\rho C_{ext} L}\,. \tag{3.4}$$

If there is no absorption and the number density of particles is known then their scattering cross section can be measured with a transmission measurement, using (3.4).

[2] This assumes that the medium is nonabsorbing.

A small detector placed not in the transmitted beam, but at an angular position $(\theta,\varphi)$ sees an intensity per particle given by (3.1). If the detector intercepts light coming from a sample volume $V_s$ (the scattering volume) then the quantity that is often used to measure the scattered light is the *Rayleigh ratio*, defined as

$$R(\theta,\varphi) \equiv \frac{r^2 I_s(\theta,\varphi)}{V_s I_0} = \frac{\rho}{k^2} F(\theta,\varphi) \tag{3.5}$$

Another quantity that is often encountered in the literature is the *differential scattering cross section* $dC_{sca}/d\Omega$:

$$\frac{dC_{sca}}{d\Omega} \equiv \frac{r^2 I_s(\theta,\varphi)}{N I_0} = \frac{1}{k^2} F(\theta,\varphi), \tag{3.6}$$

where $N = \rho V_s$ is the number of scattering particles. It is called this way because (compare with (3.2)) it is the power received by the detector per unit incident intensity and per unit solid angle, but the notation should *not* be interpreted as the derivative of a function $C_{sca}$.

## 3.3. Light scattering by single particles

We will now consider scattering by a single particle placed at the origin and embedded in a homogeneous medium. It is illuminated with a monochromatic plane wave with angular frequency $\omega$. The situation is again shown in Figure 3.1. The angle $\theta$ is the angle between the incident beam and the direction of the detector. We start with a particle that has a size much smaller than the wavelength in the medium $\lambda$. This means that the particle behaves approximately like a point dipole. The incident wave is denoted in complex notation[3] by its electric field vector

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E_0} e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t}. \tag{3.7}$$

The wave vector **k** points in the direction of propagation of the wave and has length $|\mathbf{k}| = k = 2\pi/\lambda$. The velocity of the wave in the medium is $c = \omega/k$. We will take the particle to be in the origin. The electric field of the incident beam will give rise to an induced dipole moment $\mathbf{p} = \alpha\mathbf{E}$ with $\alpha$ the polarizability of the particle. Far from the particle, at a distance *r*, the electric field of the wave radiated by this oscillating dipole is (see for example Griffiths, page 457):

$$\mathbf{E}_s = \frac{1}{4\pi\varepsilon_m c^2 r}\left[\hat{\mathbf{r}}\times\left(\hat{\mathbf{r}}\times\ddot{\mathbf{p}}(t - r/c)\right)\right].$$

Here $\hat{\mathbf{r}}$ is a unit vector in the direction of the detector. The two dots on **p** mean the second time derivative, which should be evaluated at time $t - r/c$. Thus, we obtain

$$\mathbf{E}_s = -\alpha \frac{k^2}{4\pi\varepsilon_m r}\left[\hat{\mathbf{r}}\times\left(\hat{\mathbf{r}}\times\mathbf{E_0}\right)\right] e^{ikr - i\omega t}, \tag{3.8}$$

where we have used $\omega = ck$. We decompose the electric field in components perpendicular and parallel to the scattering plane (see Figure 3.1). The result is then

$$\begin{pmatrix} E_s^{\parallel} \\ E_s^{\perp} \end{pmatrix} = \alpha \frac{k^2}{4\pi\varepsilon_m r} e^{ikr - i\omega t} \begin{pmatrix} E_0^{\parallel}\cos\theta \\ E_0^{\perp} \end{pmatrix} \tag{3.9}$$

[3] In complex notation the real electric field is the real part of the complex electric field, $\mathbf{E}_r = \mathrm{Re}(\mathbf{E}) = \mathbf{E_0}\cos(\mathbf{k}\cdot\mathbf{r} - \omega t)$. The intensity of the wave is proportional to the square of the modulus of the complex electric field: $|\mathbf{E}|^2 = \mathbf{E}\cdot\mathbf{E}^*$.

This describes a spherical wave: in every direction it is looks like a plane wave with an amplitude that falls like 1/*r*. Also note that at 90 degrees the scattered light has complete perpendicular polarization. At this angle the dipole has a vanishing component in the $E_s^{\parallel}$ direction.

The polarizability for a particle with volume $V_p$ and permittivity $\varepsilon_p$ in a medium with permittivity $\varepsilon_m$ (both linear and isotropic) is given by the Clausius-Mosotti (or Lorentz-Lorenz) relation

$$\begin{aligned}\alpha &= 3\varepsilon_m \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m} V_p \\ &= 3\varepsilon_m \frac{m^2 - 1}{m^2 + 2} V_p \quad \text{with } m = \frac{n_p}{n_m}\end{aligned} \qquad (3.10)$$

*m* is the ratio of the refractive indices of the particle and the medium. In general, $\alpha$ depends on the frequency; at frequencies where the particle absorbs it becomes a complex number. The scattered intensity is $I_s \sim |\mathbf{E_s}|^2$ so that for unpolarized incident light

$$\begin{aligned}I_s(\theta) &= I_0 \frac{k^4}{32\pi^2 \varepsilon_m^2 r^2} |\alpha|^2 \left(1 + \cos^2\theta\right) \\ &= I_0 \frac{9\pi^2}{2\lambda^4 r^2} \left|\frac{m^2 - 1}{m^2 + 2}\right|^2 V_p^2 \left(1 + \cos^2\theta\right)\end{aligned} \qquad (3.11)$$

This result was derived by Rayleigh, and scattering by particles much smaller than the wavelength is therefore called Rayleigh scattering. Two things about (3.11) are important to notice. Firstly, the scattered intensity depends on the inverse *fourth* power of the wavelength. This means that short wavelengths are scattered much more strongly than long wavelengths. Molecules in the air scatter only slightly, but they scatter blue much more strongly than red. This is why the sky looks blue. The setting sun looks red because a good deal of the blue light has scattered out of the rays by the time they reach your eyes. (Why doesn't the sun look red during most of the day?) Another important thing about (3.11) is that the scattered intensity increases with the square of the volume, i.e. the *sixth* power of the diameter of a particle! Finally, if a particle and its environment have the same refractive index (*m*=1) then the scattered intensity vanishes. In this case the particle is said to be *index matched*.

Now we move on to consider light scattering by particles that are not small compared to the wavelength. When the plane wave (3.7) is incident upon such a particle every volume element becomes an oscillating dipole with the same frequency $\omega$. But since the dipoles are at different positions they each carry a different phase. The scattered wave seen by a detector is the superposition of the dipole fields radiated by all the dipoles. At this point we have to make an approximation. Each dipole responds to the field incident upon it according to (3.8). Since the field incident on a volume element is the superposition of the primary beam with the fields coming from all the other dipoles, the problem becomes extremely complicated. Only for a few types of particles exactly solving the complete Maxwell equations can solve the problem. (For spherical particles this is called Mie theory, after the person who first solved this in 1908. The solution is in the form of infinite series that must be evaluated using a computer.) The approximation that we will make is that the field incident on a volume element can be approximated by the field of the primary beam. The fields from the other dipoles are neglected. In quantum mechanical scattering problems this is called the 1st order Born approximation; in light scattering the Rayleigh-Gans-Debye theory. The conditions for validity are the following:

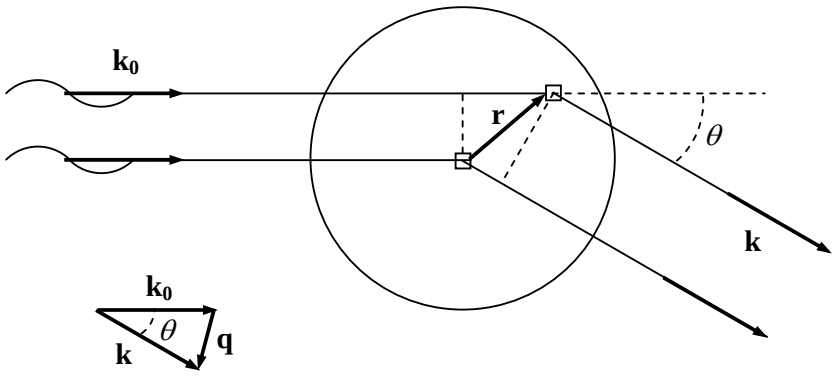



**Figure 3.2. Waves scattered by volume elements separated by a distance r differ in phase by an amount of $(\mathbf{k}_0 - \mathbf{k})\cdot\mathbf{r}$.**

$$\begin{aligned} |m-1| &\ll 1 \\ kd\,|m-1| &\ll 1 \end{aligned} \tag{3.12}$$

where *d* is a characteristic linear dimension of the particles. The first condition means that the scattering is weak so that the intensity of the incident beam is hardly attenuated inside the particles. This way, all dipoles experience the same incident field and radiate only a little. The second condition means that the phase fronts do not become distorted on passage through the particle.

With this approximation we can sum the waves of equation (3.9) over all volume elements in the particle. Figure 3.2 shows how to take the phase differences into account. The incident wave is characterized by a wave vector $\mathbf{k_0}$. Two volume elements $d\mathbf{r}$ are shown. One sits in the origin, the other at position $\mathbf{r}$. Each scatters a wave with wave vector $\mathbf{k}$ towards the detector, which is at a large distance *R*. The phase difference between these two waves is

$$\begin{aligned} \Delta\phi &= -\mathbf{k}\cdot\mathbf{r} + \mathbf{k_0}\cdot\mathbf{r} \\ &= -(\mathbf{k}-\mathbf{k_0})\cdot\mathbf{r} \\ &\equiv -\mathbf{q}\cdot\mathbf{r} \end{aligned} \tag{3.13}$$

We call $\mathbf{q}=\mathbf{k}-\mathbf{k_0}$ the *scattering vector*. It is the difference between the scattered and incident wave vectors. Since we are considering elastic scattering these both have length $2\pi/\lambda$. It is then easy to show that

$$q = |\mathbf{q}| = \frac{4\pi}{\lambda}\sin(\theta/2) \tag{3.14}$$

Remember that $\lambda$ is the wavelength in the medium, not in vacuum!

The field scattered by the element at $\mathbf{r}$, as seen by the detector, is then

$$dE_s^{\perp} = \frac{k^2}{2\pi R}\left(m(\mathbf{r})-1\right)E_0^{\perp}e^{ikR-i\mathbf{q}\cdot\mathbf{r}-i\omega t}d\mathbf{r}\,. \tag{3.15}$$

Since *m* is close to unity we have simplified $(m^2-1)/(m^2+2) \approx \frac{2}{3}(m-1)$. We also allow that the particle is inhomogeneous, so *m* can depend on position. The parallel component of the field has an extra factor $\cos\theta$. This must be integrated over the volume of the particle $V_p$ to get the total field scattered in the direction of $\mathbf{k}$:

$$E_s^{\perp} = \frac{k^2}{2\pi R}E_0^{\perp}e^{ikR-i\omega t}f(\mathbf{q}) \tag{3.16}$$

with

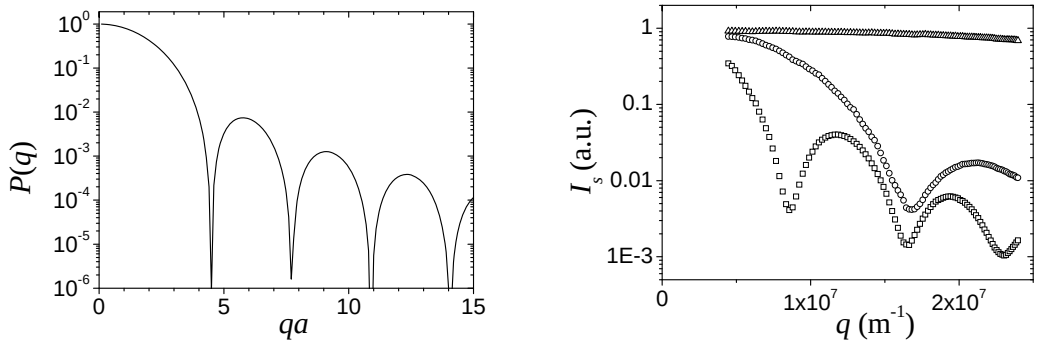



**Figure 3.3. (Left) Form factor of a spherical particle. (Right) Intensity scattered by silica spheres in ethanol with radii of 40 nm (triangles), 264 nm (circles), and 435 nm (squares). The wavelength was 633 nm.**

$$f(\mathbf{q}) = \int_{V_p} \left(m(\mathbf{r}) - 1\right) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \,. \tag{3.17}$$

We have separated the interesting part $f(\mathbf{q})$ from the boring constants. It can be interpreted as the amplitude of the wave scattered at wave vector $\mathbf{q}$ by the particle as a whole. Finally, the square of the modulus gives us the scattered intensity. After a bit of rearranging we find

$$I_s = I_0 \frac{k^4 V_p^2 \left|\bar{m} - 1\right|^2}{8\pi^2 R^2} P(\mathbf{q}) \left(1 + \cos^2\theta\right) \tag{3.18}$$

with

$$P(\mathbf{q}) = \left|\frac{f(\mathbf{q})}{f(\mathbf{0})}\right|^2 = \left|\frac{\int_{V_p} \left(m(\mathbf{r}) - 1\right) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}}{\int_{V_p} \left(m(\mathbf{r}) - 1\right) d\mathbf{r}}\right|^2 \tag{3.19}$$

Here $\bar{m}$ is the value of $m$ averaged over the volume of the particle. The prefactor in the formula gives the absolute scattered intensity and has just the same properties as in the Rayleigh formula. The part that interests us more is the *form factor* $P(\mathbf{q})$ of the particle, which describes the angular dependence of the scattering[4]. The form factor is determined only by the size and shape of the particle and is normalized in the forward direction (where $\mathbf{q}=\mathbf{0}$). By measuring the angular dependence of the scattered light we can immediately obtain the form factor and learn about the particle properties. We must then compare this with a model.

The simplest model is of course the sphere. For a sphere with radius $a$ the form factor can be calculated by choosing the $z$-axis along $\mathbf{q}$ and first integrating over the angles. The result is

$$P(q) = \left[3\frac{\sin(qa) - qa\cos(qa)}{(qa)^3}\right]^2 . \tag{3.20}$$

This function is shown in Figure 3.3. Note that the form factor depends only on the product $qa$, so there is in principle no limit to the number of oscillations. Of course, $q$ can never be larger than $4\pi/\lambda$, so the number of oscillations in the range of scattering angles between 0 and 180° is a sensitive function of the particle radius. This can be used to measure the size of

[4] In the crystallography literature it is often the function $f(\mathbf{q})$ which is called the form factor.

spheres. If the particles are monodisperse the minima and maxima can be observed in the scattered intensity, as is also shown in Figure 3.3. Also note the strong forward scattering for large particles.

Equations (3.18) and (3.19) as they stand describe scattering by a single particle with a fixed and given orientation in space. In practice we almost always measure the light scattered by a large number of particles which each have a different (and constantly fluctuating) orientation. In a dilute suspension all the particles scatter incoherently so that we can add the intensities scattered by each particle. The scattered intensity thus becomes *N* times larger and the form factor must be averaged over the orientations of the particles. Results are known for a number of particle shapes, such as ellipsoids, rods and disks. These formulas can be found in the scattering literature.

We will not study all the consequences of particle anisotropy, but just consider the case of small particles. This is a useful case because many particles are much smaller than the wavelength of the light used to study them, for example most polymers, proteins, nanocrystals and micelles. In this case $\mathbf{q}\cdot\mathbf{r}$ is small so that we can only measure the initial part of the form factor. What we get in return is that the particles are allowed to be nonspherical and polydisperse. First, we must decide on an origin for an irregularly shaped particle. (Since only differences in phase are important this choice should not influence the result.) The easiest choice is the center of mass:

$$\mathbf{R}_{cm} = \frac{1}{V_p}\int_{V_p} \mathbf{r}d\mathbf{r} = \mathbf{0}$$

We can now expand the exponent in the form factor (3.19) in a Taylor series. The first order term is zero due to our choice of the origin. This leads to

$$P(\mathbf{q}) = \left|1 - \frac{1}{2}\left\langle \frac{\int_{V_p}(m-1)(\mathbf{q}\cdot\mathbf{r})^2\,d\mathbf{r}}{\int_{V_p}(m-1)\,d\mathbf{r}}\right\rangle + \ldots\right|^2 \approx 1 - \frac{\int_{V_p}(m-1)q^2r^2\left\langle\cos^2\theta\right\rangle d\mathbf{r}}{\int_{V_p}(m-1)\,d\mathbf{r}}$$

The angled brackets $\langle\cdots\rangle$ indicate an average over the ensemble of particles. Since the particles have random orientations $\left\langle\cos^2\theta\right\rangle = 1/3$. The result is the *Guinier approximation*:

$$P(q) = 1 - \frac{1}{3}q^2R_g^2 + \ldots \tag{3.21}$$

with the *radius of gyration* defined as

$$R_g^2 = \frac{\int_{V_p}(m-1)r^2d\mathbf{r}}{\int_{V_p}(m-1)\,d\mathbf{r}}\,. \tag{3.22}$$

Note that this is still a volume integral and that *r* is the distance of a volume element to the particle's center of mass. For spheres it is easy to derive that $R_g = \sqrt{3/5}\,a$, but the result is most useful for irregularly shaped particles, because it provides a uniquely defined particle size that is directly measurable and which can be compared to a model. For example, a rodlike particle with length *L* has $R_g = L/\sqrt{12}$ and a Gaussian polymer coil of *n* random chain (or Kuhn) segments of length *l* has $R_g = l\sqrt{n/6}$.

## 3.4. Light scattering by an ensemble of particles

Next, we consider scattering by a more concentrated suspension. This means that many particles in the scattering volume $V_s$ contribute to the wave scattered towards the detector. To do this we take exactly the same approach as in the case of the not-so-small particle. We again assume that every volume element scatters a dipole field, and that the field incident on a given particle can be approximated by the external field. The waves scattered by the other particles are neglected. This is the same as saying that there is only *single* scattering. If a concentrated suspension is measured the scattered intensity per particle must be very small. This is the case for very small particles or if $m \approx 1$. The latter condition can sometimes be satisfied if the refractive index of the medium is adjusted so that it almost equals that of the particles (*index-matching*).

Assuming that the single scattering assumption is satisfied we again refer to Figure 3.2. This time, imagine that the large circle is the entire scattering volume $V_s$, which contains a large number of particles $N$. The volume elements $d\mathbf{r}$ can lie either in a particle or in the medium. The waves scattered towards the distant detector are once more summed, taking their phases into account:

$$E_s^{\perp} = \frac{k^2}{2\pi R} E_0^{\perp} e^{ikR - i\omega t} \int_{V_s} \left( m(\mathbf{r}) - 1 \right) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{3.23}$$

Only volume elements lying in the particles have a nonzero $m-1$ and contribute to the integral. We can thus separate the integral into contributions coming from each particle:

$$E_s^{\perp} = \frac{k^2}{2\pi R} E_0^{\perp} e^{ikR - i\omega t} \sum_{j=1}^{N} \int_{V_j} \left( m(\mathbf{r}) - 1 \right) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{3.24}$$

where $V_j$ is the volume occupied by particle $j$. Let $\mathbf{r}_j$ denote the position of the center of particle $j$. The integration variable $\mathbf{r}$ can then be written as $\mathbf{r} = \mathbf{r}_j + \mathbf{r}'$. The new integration variable $\mathbf{r}'$ ranges over the volume of the particle with its center translated to the origin. Equation (3.24) now becomes

$$E_s^{\perp} = \frac{k^2}{2\pi R} E_0^{\perp} e^{ikR - i\omega t} \sum_{j=1}^{N} f_j(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{r}_j} \ . \tag{3.25}$$

We have recognized the integral appearing in (3.17). It describes the interference of waves scattered by one particle. The exponential functions containing the position coordinates $\mathbf{r}_j$ describe interference of light scattered by different particles. Because the particles move continuously the latter kind of interference leads to rapid fluctuations in the scattered light (typically to μs to ms). What is measured in a *static light scattering* experiment is the light scattered from a large sample volume containing very many particles, and averaged over times of the order of 1 s. In other words the measured quantity is the ensemble averaged intensity $I_s \sim \left\langle \mathbf{E}_s \cdot \mathbf{E}_s^{*} \right\rangle$. We will assume that all $N$ particles are identical[5]. Remembering that the parallel field component has a factor $\cos\theta$ we obtain

$$I_s(\mathbf{q}) = I_0 \frac{N k^4 V_p^2 \left| \bar{m} - 1 \right|^2}{8\pi^2 R^2} P(\mathbf{q}) S(\mathbf{q}) \left( 1 + \cos^2\theta \right), \tag{3.26}$$

with

[5] The expression can be generalized for a mixture of nonidentical particles, but the result can no longer be separated into the product of a form factor and a structure factor. This also the case for nonspherical particles. In that case, identical particles with different orientations also have different $f_j$ 's.

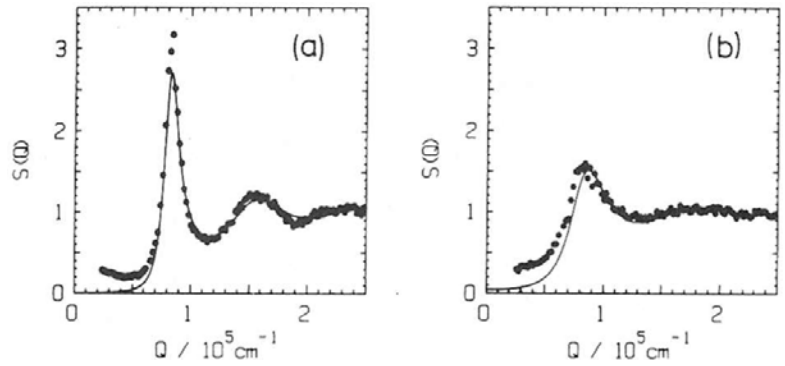



**Figure 3.4. Structure factors of 80 nm diameter polystyrene spheres in water measured with light scattering. The particles surfaces carry ionizable $-SO_3H$ groups. The concentration was $1.8\times10^{18}$ particles/m$^3$. (a) 3.1 μmol/L NaOH added; (b) 3.1 μmol/L NaCl added. The drawn lines are fits using the rescaled mean spherical approximation (RMSA). From: Härtl and Versmold, *Langmuir* 8, 2885 (1992).**

$$S(\mathbf{q}) = \frac{1}{N}\left\langle \sum_{j=1}^{N}\sum_{k=1}^{N} e^{i\mathbf{q}\cdot(\mathbf{r}_k - \mathbf{r}_j)} \right\rangle . \tag{3.27}$$

This is the *structure factor* from Chapter 2.

This result means that the structure factor can be measured in a scattering experiment as follows. The scattered intensity of the sample under study, which has particle number density $\rho$, is measured as a function of angle. Then a small amount is diluted to $\rho_{\text{dil}}$ by a large factor, say 100 times, so that its structure factor becomes equal to unity. The scattered intensity is again measured. Call this $I_{\text{dil}}$. The structure factor of the original sample is then found from

$$S(\mathbf{q}) = \frac{\rho_{\text{dil}}}{\rho}\frac{I_s(\mathbf{q})}{I_{\text{dil}}(\mathbf{q})} . \tag{3.28}$$

Some examples of measured structure factors are given in Figure 3.4. The large oscillations imply strong liquid-like ordering. Remember that a peak in the structure factor at wave vector $q$ implies that there exists a (sinusoidal) density fluctuation with wavelength

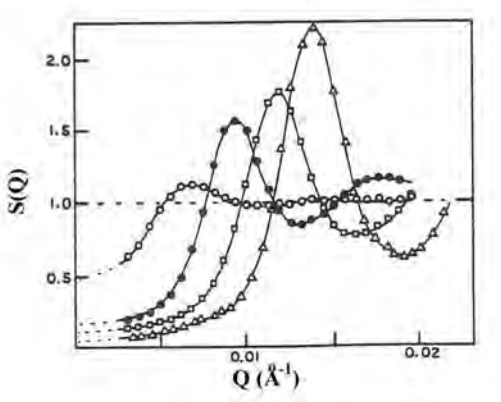



**Figure 3.5. Structure factors measured with SANS of 16 nm radius polystyrene spheres in water containing $10^{-4}$ mol/L NaCl. Particle volume fractions were 0.01, 0.04, 0.08, and 0.13. From: Goodwin et al. *Makromol. Chem. Suppl.* 10/11, 499 (1985).**

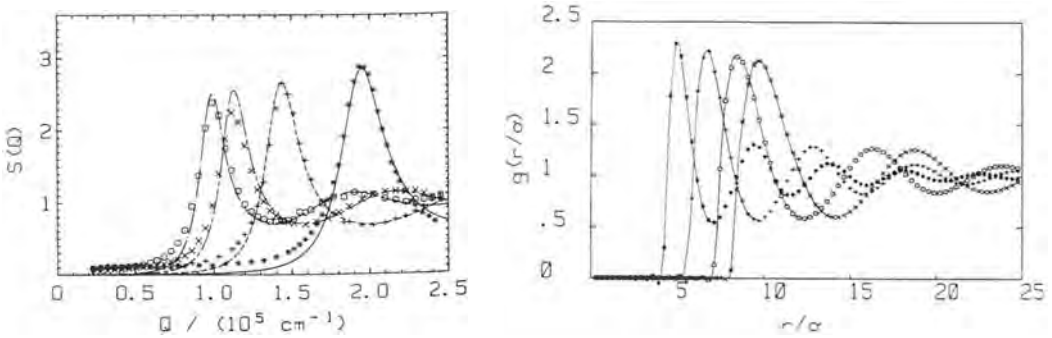



**Figure 3.6. The left panel shows structure factors of 75 nm diameter, negatively charged polystyrene spheres in deionized water measured with light scattering. Particle concentrations were (from left to right curve) $2.9\times10^{18}$, $4.4\times10^{18}$, $9.0\times10^{18}$, and $22.6\times10^{18}$ $m^{-3}$. The curves are fits with the RMSA theory. The right panel shows the corresponding radial distribution functions, which were obtained by inverting the fitted theoretical curves. From: Härtl and Versmold, *J. Chem. Phys.* 88, 7157 (1988).**

$2\pi/q$. The particles in this experiment were negatively charged. Addition of NaOH ionizes all the sulfonate groups on the surfaces, but does not reduce the Debye length because the $H^+$ counterions are merely replaced by $Na^+$. Addition of an equal concentration of NaCl, on the other hand, reduces the Debye length making the liquid-like structure less pronounced. The dependence of the structure factor on the particle concentration is demonstrated in Figure 3.5 for charged polystyrene spheres with more salt added. It is clear that as the concentration is increased the liquid-like structure becomes much more pronounced. Also the first peak shifts to larger $q$, corresponding to smaller average interparticle distances.

It can be shown that expression (3.27) is identical to the structure factor defined in Chapter 2 (see the appendix for a derivation):

$$S(\mathbf{q}) = 1 + \rho \int \exp(i\mathbf{q}\cdot\mathbf{r})\big(g(\mathbf{r}) - 1\big)\, d\mathbf{r}\,. \tag{3.29}$$

This equation holds that the structure factor is the *Fourier transform* of the radial distribution function (minus one). In principle, it should then be possible to Fourier invert the structure factor to obtain the radial distribution function. In practice, however, this turns out to be very difficult. The reason is that $S(q)$ can only be measured over a limited range of $q$, and the data become more noisy at larger $q$. Instead, one often finds a suitable fit of the structure factor with an appropriate theory and one then inverts the theoretical curve. This is done in Figure 3.6. Again, the increase in particle concentration leads to a shift of the structure factor peak to larger $q$, while the primary peak in $g(r)$ moves to smaller $r$. In this case the height of the peaks does not increase very much with concentration. This is because in a deionized dispersion most ions in the system are the ions dissociated from the particle surfaces. As the particle concentration is increased the ionic strength of the solvent is also increased leading to more effective screening.

By taking the limit $q \to 0$ in Eq. (3.29) and comparing the result with the well-known compressibility equation (Chapter 2) we find that the structure factor at $q$=0 is proportional to the osmotic compressibility of the system:

$$\lim_{q\to 0} S(q) = kT\left(\frac{\partial\rho}{\partial\Pi}\right)_T. \tag{3.30}$$

As an example of this relation between scattering and thermodynamics Figure 3.7 shows measurements of the value of $S(q$=0) for a colloidal dispersion of particles that behave like

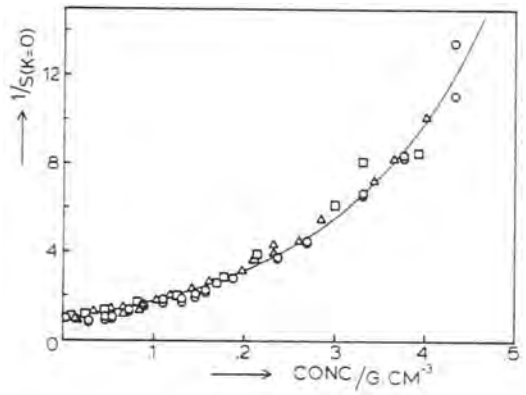



**Figure 3.7. Scattering measurements of the structure factor extrapolated to zero wave vector on a system of stearyl alcohol coated silica spheres dispersed in cyclohexane. These particles behave like hard spheres. The data agree with the Percus-Yevick expression for the compressibility (drawn line). Triangles are light scattering data on particles with radius 35 nm, squares are SANS measurements on the same particles. Circles are light scattering data on spheres with a radius of 22 nm. From: De Kruif, Jansen, and Vrij, In: "Physics of Complex and Supramolecular Fluids", Wiley, New York, 1987.**

hard spheres. The data are described correctly by the Percus-Yevick equation. This has been an important test of this theory, and demonstrated at the same time that silica spheres coated with stearyl alcohol and dispersed in cyclohexane behave like true hard spheres. The values of $S(q=0)$ are extrapolated from a plot of $S(q)$ versus $q^2$. Great care should be taken to ensure that the measurements are not affected by the presence of dust or clusters of particles, which contribute mainly to the scattering at low $q$.

In many cases, for example molecular weight determinations in polymers or the measurement of particle sizes in association colloids, one is not so much interested in the form of the structure factor but instead treats it as a correction of the scattering for the finite concentration of the sample. In such cases a range of weight concentrations $c$ is prepared and each sample measured over a range of $q$. The scattered intensity at low $q$ and low $\rho$ can be approximated as:

$$\frac{K\rho}{I(q)} \approx \left(1+\frac{1}{3}R_g^2 q^2\right)\left(1+2B_2\rho\right). \tag{3.31}$$

Zimm has proposed a method to perform the double extrapolation in a so-called Zimm-plot, and thus to obtain the radius of gyration and the second virial coefficient from the slopes. Since $\rho = cN_{av}/M$ the molecular weight $M$ can be determined from the intercept provided all the constants in the factor $K$ are correctly accounted for.

## 3.5. Scattering by crystals

As a special case we consider the structure factor of a crystalline array of particles. This is usually called diffraction. You may recognize the equations that we shall derive from the theory of X-rays diffraction by molecular crystals. In the previous paragraph we have seen that the peaks in the structure factor become sharper when the particles are forced to occupy a smaller volume. This reflects the tendency of the particles to order in order to use space more efficiently. It is therefore not hard to imagine that when the system starts to develop periodic order the structure factor peaks become very high and narrow, while it is zero in between the peaks. Of course, this transition cannot be truly continuous because a first order phase

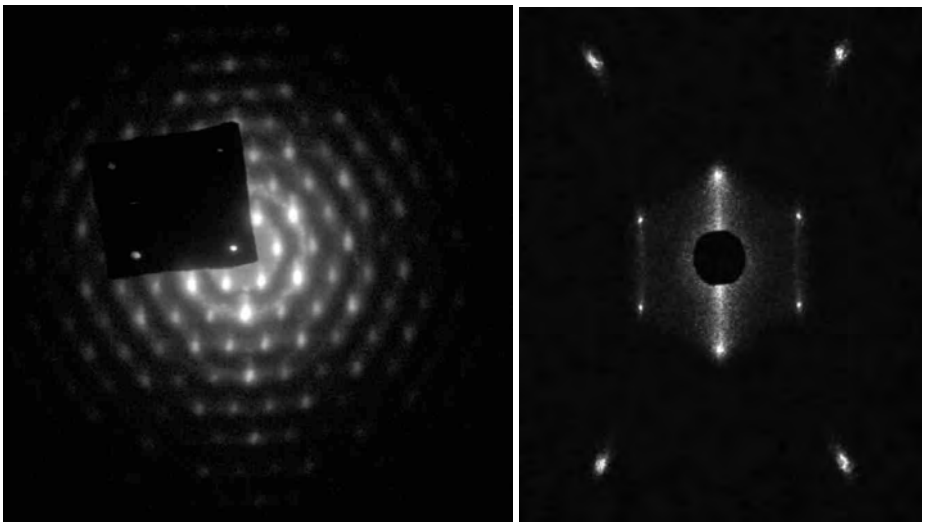


**Figure 3.8. Diffraction patterns from colloidal crystals. *(Left)* Small-angle x-ray diffraction on an fcc crystal of silica spheres in water, *(Right)* Laser light diffraction on a body-centered tetragonal crystal of silica spheres in a water/DMSO mixture. Thee black square and circle are beam stops that block the transmitted beam.**

transition takes place during the process. The peaks in the structure factor are the Bragg spots that appear in the diffraction pattern. Two examples are shown in Figure 3.8: one made with light scattering, the other with small-angle x-ray diffraction.

By definition a crystal has translational symmetry along three independent axes. The basis vectors describing these translations are denoted $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. Together they define the unit cell, which when translated makes up the whole crystal. We will assume that the crystal is finite, extending a number of $M_1$, $M_2$, and $M_3$ unit cells along the three basis vectors, respectively. It therefore has the shape of a parallelepiped. Since each unit cell is identical we can write the location of a particle $i$ sitting at position $\mathbf{R}_p$ relative to the origin of a unit cell as $\mathbf{r}_i = \mathbf{R}_p + m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3$, with the $m$'s integer numbers. For simplicity, we will again assume that all particles are identical, but it is easy to generalize to unit cells with more than one particle. Expression (3.27) tells us to take the sum of $\exp(-i\mathbf{q}\cdot\mathbf{r}_i)$ over all particles, and then multiply the result by its complex conjugate and divide it by the total number of particles. So, first we write down the sum, as follows:

$$\sum_{i=1}^{N} e^{-i\mathbf{q}\cdot\mathbf{r}_i} = \sum_{p=1}^{n} e^{-i\mathbf{q}\cdot\mathbf{R}_p} \sum_{m_1=0}^{M_1-1} e^{-im_1\mathbf{q}\cdot\mathbf{a}_1} \sum_{m_2=0}^{M_2-1} e^{-im_2\mathbf{q}\cdot\mathbf{a}_2} \sum_{m_3=0}^{M_3-1} e^{-im_3\mathbf{q}\cdot\mathbf{a}_3} \tag{3.32}$$

The sum over $p$ runs over all the $n$ particles in a single unit cell. The other sums run over all the unit cells. So the total number of particles is $nM_1M_2M_3$. Notice that each of the sums over the $m$'s have the form of a geometric progression

$$\sum_{m=0}^{M-1} x^m = \frac{1-x^M}{1-x}.$$

where $x$ is of the form $\exp(-i\mathbf{q}\cdot\mathbf{a})$. We might as well multiply this by its complex conjugate at once. The result is

$$\frac{1-\exp(-iM\mathbf{q}\cdot\mathbf{a})}{1-\exp(-i\mathbf{q}\cdot\mathbf{a})}\cdot\frac{1-\exp(iM\mathbf{q}\cdot\mathbf{a})}{1-\exp(i\mathbf{q}\cdot\mathbf{a})} = \frac{2-2\cos(M\mathbf{q}\cdot\mathbf{a})}{2-2\cos(\mathbf{q}\cdot\mathbf{a})} = \frac{\sin^2(\frac{1}{2}M\mathbf{q}\cdot\mathbf{a})}{\sin^2(\frac{1}{2}\mathbf{q}\cdot\mathbf{a})}$$

We do this for all the factors so that

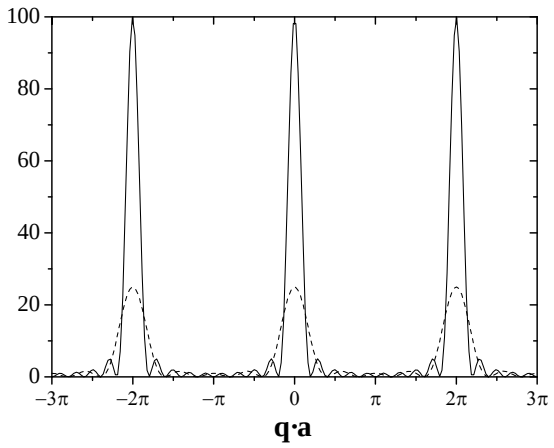



**Figure 3.9. The function** $\sin^2(\frac{1}{2}M\mathbf{q}\cdot\mathbf{a})/\sin^2(\frac{1}{2}\mathbf{q}\cdot\mathbf{a})$ **for** *M=5* **(dashed) and** *M=10* **(solid).**

$$S(\mathbf{q}) = \frac{1}{M_1 M_2 M_3} F(\mathbf{q}) \frac{\sin^2\left(\frac{1}{2}M_1\mathbf{q}\cdot\mathbf{a}_1\right)}{\sin^2\left(\frac{1}{2}\mathbf{q}\cdot\mathbf{a}_1\right)} \frac{\sin^2\left(\frac{1}{2}M_2\mathbf{q}\cdot\mathbf{a}_2\right)}{\sin^2\left(\frac{1}{2}\mathbf{q}\cdot\mathbf{a}_2\right)} \frac{\sin^2\left(\frac{1}{2}M_3\mathbf{q}\cdot\mathbf{a}_3\right)}{\sin^2\left(\frac{1}{2}\mathbf{q}\cdot\mathbf{a}_3\right)} \tag{3.33}$$

with

$$F(\mathbf{q}) = \frac{1}{n}\left|\sum_{p=1}^{n} e^{-i\mathbf{q}\cdot\mathbf{R}_p}\right|^2 \tag{3.34}$$

Note that the function $F(\mathbf{q})$ is just the structure factor of the unit cell[6].

Let us interpret this result by first taking a look at the function $\sin^2\left(\frac{1}{2}Mx\right)\big/\sin^2\left(\frac{1}{2}x\right)$. It is plotted in Figure 3.9 for *M*=5 and *M*=10. It is seen that this function is sharply peaked around *x* values of an integer times $2\pi$. The height of the peaks is $M^2$ and their width is $2\pi/M$. This means that for a crystal with more than a few unit cells the scattered intensity is always zero except for those values of **q** which satisfy the following conditions simultaneously:

$$\begin{aligned} \mathbf{q}\cdot\mathbf{a}_1 &= 2\pi h \\ \mathbf{q}\cdot\mathbf{a}_2 &= 2\pi k \\ \mathbf{q}\cdot\mathbf{a}_3 &= 2\pi l \end{aligned} \tag{3.35}$$

Here *h*, *k*, and *l* must be integers. These equations are called the *Laue equations*. They are equivalent to the Bragg law, which we will show later. For a given crystal with basis vectors $\mathbf{a}_i$ they determine the angles where diffraction is observed. The physical basis for this result is that because all unit cells scatter equivalently they must all scatter in phase if there is to be diffraction. Otherwise, even the smallest phase difference will cause complete cancellation of the wave sum, due to the large number of cells. "Fortuitous" constructive interference only occurs at very special angles, which are characteristic for the crystal. This of course forms the basis for structure identification in crystallography. For small crystallites some light also ends up in a small angular range around the peaks, so that the width of the peaks can sometimes be used to estimate the size of the crystals.

[6] In the crystallography literature it is the complex function $\sum f_p \exp(-i\mathbf{q}\cdot\mathbf{R}_p)$ that is called the structure factor. But in expressions for the diffracted intensity only its modulus squared appears.

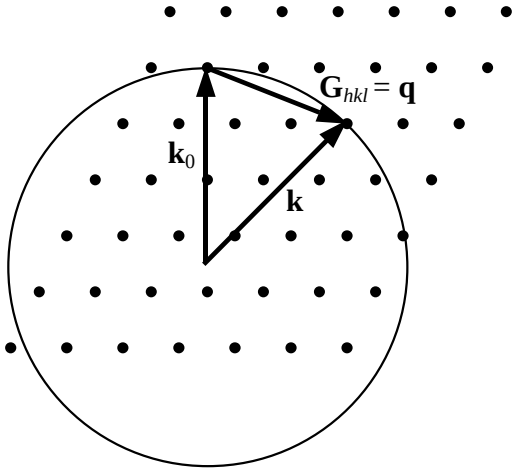



**Figure 3.10. A two-dimensional representation of the Ewald construction. Diffraction maxima appear for reciprocal lattice points that fall on the surface of the Ewald sphere.**

We now understand why only particular diffraction peaks are seen in crystals. But note that equation (3.33) also contains the factor $F(\mathbf{q})$, describing interference between the particles within the unit cell. If this cell contains only a single particle then it is equal to unity. But if it contains two or more particles then $F$ determines the relative strengths of the diffraction peaks. It could even equal zero at a $\mathbf{q}$ value which otherwise satisfies (3.35). In this case a diffraction, which is allowed by the periodicity of the lattice, disappears because it is not allowed by the symmetry of the unit cell. One speaks of *systematic vanishings*. It may seem unlikely that the few $\mathbf{q}$ values allowed by (3.35) would happen to coincide exactly with a zero in $F(\mathbf{q})$, but it is in fact a common occurrence in many lattice types, as we will demonstrate shortly in the example of a face centered cubic lattice. Systematic vanishings are an important aid for crystallographers in recognizing the lattice type.

From crystallography we know that with every crystal lattice we can associate a *reciprocal lattice*. The basis vectors of the reciprocal lattice are cunningly defined as[7]

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3},\ \mathbf{b}_2 = 2\pi \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3},\ \mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}, \tag{3.36}$$

so that we have

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}. \tag{3.37}$$

In other words, the each reciprocal lattice vectors is orthogonal to two of the basis vectors of the direct lattice. Every vector in reciprocal space, such as the $\mathbf{q}$ belonging to a particular diffraction peak, can now be represented by the linear combination

$$\mathbf{q} = p_1\mathbf{b}_1 + p_2\mathbf{b}_2 + p_3\mathbf{b}_3.$$

The $p$'s are the components of $\mathbf{q}$ along the three axes. By virtue of (3.37) they can be determined by taking the dot product with the basis vectors of the real lattice, for example

$$\mathbf{q} \cdot \mathbf{a}_1 = 2\pi p_1.$$

Comparing this with the Laue conditions we see that for diffraction to occur we must have $p_1 = h$, and similarly for $p_2$ and $p_3$. Thus, we find that

$$\mathbf{q} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \equiv \mathbf{G}_{hkl}. \tag{3.38}$$

[7] Contrary to what is common practice in solid state physics most crystallographers define the reciprocal lattice vectors without the factor $2\pi$.

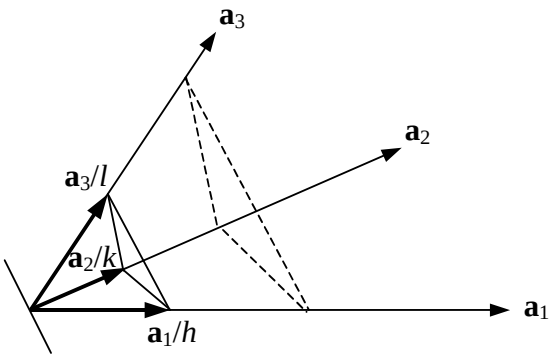



**Figure 3.11. Real space representation of the crystallographic planes *hkl*.**

The integers *hkl* are called the *Miller indices* belonging to the particular diffraction. Every reciprocal lattice point is labeled by a set of Miller indices. Thus, equation (3.38) tells us that the scattering vector **q** must precisely coincide with a reciprocal lattice vector **G**. This is the basis for the *Ewald construction* (see Figure 3.10), which says that if we place the incident wave vector $\mathbf{k}_0$ with its head pointing at the origin of reciprocal space and draw a sphere around it, then diffracted waves **k** occur wherever this sphere intersects with a reciprocal lattice point. By rotating the crystal with respect to the incident beam we can scan (a part of) reciprocal space in search of diffractions.

The real space interpretation of the Miller indices is that of a set of parallel equidistant planes, one of which passes through the origin, and the next nearest intercepts the $\mathbf{a}_1$ axis at $a_1/h$, the $\mathbf{a}_2$ axis at $a_2/k$, and the $\mathbf{a}_3$ axis at $a_3/l$ (see Figure 3.11). These planes are perpendicular to $\mathbf{G}_{hkl}$, because so are two independent vectors in the planes: $\mathbf{a}_1/h - \mathbf{a}_2/k$ and $\mathbf{a}_2/k - \mathbf{a}_3/l$. (To see this, take the dot product of each of these vectors with $\mathbf{G}_{hkl}$ as given by (3.38)). A normal vector to the lattice planes is then $\mathbf{G}_{hkl}/|\mathbf{G}_{hkl}|$. The distance between two planes of the set $d_{hkl}$ can thus be found from

$$d_{hkl} = \frac{\mathbf{a}_1}{h} \cdot \frac{\mathbf{G}_{hkl}}{|\mathbf{G}_{hkl}|} = \frac{2\pi}{|\mathbf{G}_{hkl}|}. \tag{3.39}$$

Using this result in (3.38) and (3.14) we finally obtain the familiar Bragg law:

$$2d_{hkl} \sin(\theta/2) = \lambda. \tag{3.40}$$

The factor ½ in the sine probably appears unfamiliar, but that is only because crystallographers define their scattering angle as $2\theta$ instead of $\theta$.

We conclude this section with an example: diffraction by a face centered cubic (*fcc*) lattice. This type of lattice is common in colloidal crystals. The unit cell contains 4 particles. The basis vectors of this cubic lattice are simply

$$\mathbf{a}_1 = a\hat{\mathbf{x}}, \quad \mathbf{a}_2 = a\hat{\mathbf{y}}, \quad \mathbf{a}_3 = a\hat{\mathbf{z}}. \tag{3.41}$$

Using (3.36) the reciprocal lattice vectors are found to be

$$\mathbf{b}_1 = \frac{2\pi}{a}\hat{\mathbf{x}}, \quad \mathbf{b}_2 = \frac{2\pi}{a}\hat{\mathbf{y}}, \quad \mathbf{b}_3 = \frac{2\pi}{a}\hat{\mathbf{z}}. \tag{3.42}$$

The diffraction condition (3.38) tells us that diffraction peaks may be expected at scattering vectors **q** that are linear combinations of integer multiples of these three vectors. The corresponding diffraction angles can be found by taking the modulus squared on both sides of (3.38), and using (3.14) and (3.42):

$$\left(\frac{2\sin(\theta/2)}{\lambda}\right)^2 = \frac{h^2+k^2+l^2}{a^2}. \tag{3.43}$$

So we may expect a large number of diffraction peaks. But it turns out that many of them are missing due to systematic vanishings. This can be seen by calculating the structure factor using Eq. (3.34). The position of a particle in its unit cell can be written as $\mathbf{R} = x\mathbf{a}_1 + y\mathbf{a}_2 + z\mathbf{a}_3$, and $\mathbf{q}$ is given by the diffraction condition (3.38). Then for a diffraction *hkl* we have

$$\mathbf{q}\cdot\mathbf{R} = 2\pi(hx+ky+lz). \tag{3.44}$$

If one particle is located at (*x*,*y*,*z*)=(0,0,0) then other particles are present at (½,½,0), (½,0,½) and (0,½,½). The structure factor (3.34) is then

$$\begin{aligned} F(\mathbf{q}) &= \frac{1}{4}\left|1+e^{-i\pi(k+l)}+e^{-i\pi(h+l)}+e^{-i\pi(h+k)}\right|^2 \\ &= \frac{1}{4}\left(1+(-1)^{k+l}+(-1)^{h+l}+(-1)^{h+k}\right)^2 \\ &= \begin{cases} 4 & hkl \text{ all even or all odd} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3.45}$$

We see that not all combinations of *hkl* give rise to diffraction. Hence, the *fcc* lattice can be recognized by the fact that all diffractions with one odd or one even Miller index vanish systematically. If we are dealing with an unknown crystal we list $(2\sin(\theta/2)/\lambda)^2$ of the various diffractions and look for a common divisor (1/*a*) that produces small integers. These integers equal $h^2+k^2+l^2$ if the crystal is cubic. The missing *hkl* values tell us whether the structure is really fcc[8].

The reason for the fact that systematic vanishings are not so unlikely at all is seen to lie in the special position of the particles within the unit cell. In fact, only unit cells with special positions of the particles can be repeated periodically to fill all space. These crystal lattices are called the *Bravais lattices*, and there exist only 14 of them.

## 3.6. Dynamic light scattering

In the previous sections it was mentioned in passing that the particles in soft matter systems are in constant thermal motion (*Brownian motion*). In principle this may be expected to lead to fluctuations in the scattered intensity. In *dynamic light scattering* (DLS)[9] these fluctuations are measured. This provides information on the dynamics of particles in solution. In dilute suspensions DLS has become a standard technique for the characterization of particle size and polydispersity. Also, the dynamics of internal degrees of freedom can be studied, such as chain fluctuations in polymers and shape fluctuations of microemulsion droplets. When applied to concentrated suspensions DLS provides information on the length and time scale dependent dynamics of particles.

### 3.6.1. Fluctuations in the scattered intensity

In static light scattering a wide beam is used (typically a few mm) which illuminates very many particles. This leads to the ensemble average in Eqs. (3.26) and (3.27). Since the structure factor describes interference of waves scattered by different particles and these particles are constantly moving it is clear that the there must also be fluctuations in the

[8] In crystallography comparing the relative intensities of the diffraction peaks with their calculated values must then further validate the proposed structure.

[9] Other names are Photon Correlation Spectroscopy (PCS) and Quasi-Elastic Light Scattering (QELS).

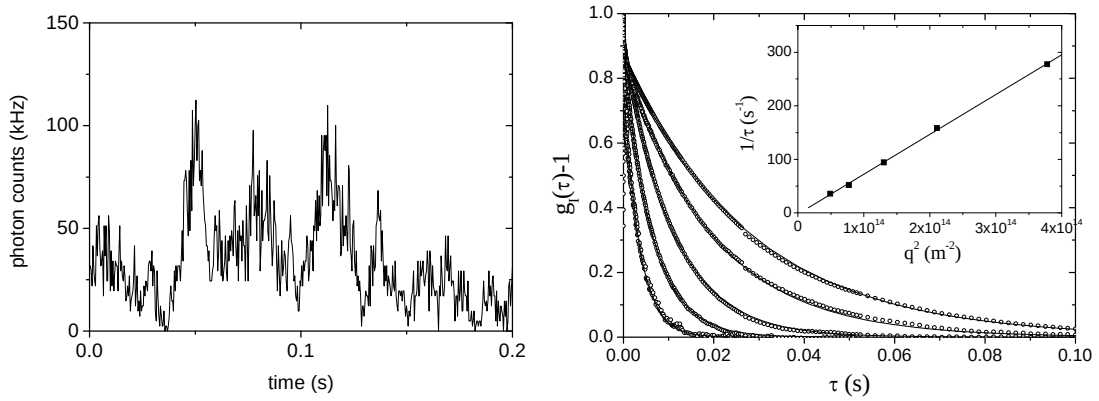



**Figure 3.12. Left: fluctuating intensity of a speckle measured in DLS. Right: Normalized IACF measured at various angles. The curve with the slowest decay corresponds to the data in the left graph. The inset shows the *q*-dependence of the decay rate. The sample was a dilute suspension of *R*=435 nm silica spheres in ethanol.**

scattered intensity. In dynamic light scattering a much smaller beam of coherent radiation is used: about 0.1 mm, but still containing many particles. The fluctuations are then readily visible to the naked eye: If we use a white screen as the detector we see a field containing many bright spots in a dark background. This is known as *(laser) speckle*. The origin of the speckle pattern can be understood as follows: Each particle scatters a spherical wave toward the screen with a phase that depends sensitively on its position, shape, and orientation. At each point on the screen the intensity is determined by the sum of all these scattered waves:

$$E_s(t) = \sum_{j=1}^{N} f_j(\mathbf{q},t)\exp\left(-i\mathbf{q}\cdot\mathbf{r}_j(t)\right) \tag{3.46}$$

At some points the waves happen to cancel each other, while at others they reinforce each other. Particles only need to move over about one wavelength to let individual speckles blink on and off. The idea of dynamic light scattering is now to measure the (average) rate at which the speckles blink, and relate this to the rate with which particles change their relative positions. This is done by making the detector also very small, by placing a pinhole in front of it, such that it collects light in only a single speckle.

The angular size (in radians) of a speckle is determined only by the ratio of the wavelength and the apparent diameter $d\sin\theta$ of the scattering volume as seen by the detector:

$$\Delta\theta_{speckle} \approx \frac{\lambda}{d_{app}} = \frac{\lambda}{d\sin\theta} \tag{3.47}$$

This is the same as the width of the primary diffraction maximum of a slit. For wavelengths in the visible and a scattering volume of 0.5 mm this is ~$10^{-3}$ rad, or 0.05°. With a detector placed at a distance of 20 cm the pinhole size should be about 200 μm. Incidentally, we now see that the fact that fluctuations are small in static light scattering lies not in the larger number of illuminated particles, but in the larger number of (very small) speckles collected by the detector. In SLS we can therefore also get away with using an incoherent light source (though monochromatic). In fact, this even improves the speckle averaging. In DLS the use of a coherent light source is required.

Mathematical functions that are used for studying fluctuating quantities are so-called correlation functions. In the case of DLS we measure the *intensity autocorrelation function*, or IACF:

$$g_I(\mathbf{q},\tau) = \langle I(\mathbf{q},t)\, I(\mathbf{q},t+\tau)\rangle \tag{3.48}$$

This means that the intensity measured at a fixed time $t$ is multiplied with the intensity measured a time $\tau$ later. This is repeated for a large number of speckles and averaged to form the ensemble average, as indicated by the angled brackets. In practice, of course, instead of moving the detector to a different speckle a great number of times (while keeping **q** almost constant!), we prefer to keep the detector fixed and to take a time average[10]. The measurement in Figure 3.12 illustrates how this works. The intensity looks completely random, but there is a characteristic time scale associated with the fluctuations: The intensity never jumps from a maximum to a minimum in an instant, but this takes a finite amount of time. At very short times $\tau$, when the particles have not had enough time to move much, the intensity is unchanged, so that $g_I(0) = \langle I^2 \rangle$. After a somewhat longer time the particles have moved enough to let the intensity change and the IACF begins to decay. After a sufficiently long time a speckle's intensity has changed so much that it has become completely uncorrelated with its initial value. The average of the product in (3.48) is then equal to the product of the averages, making $g(\tau \to \infty) = \langle I \rangle^2$. It is clear that $\langle I^2 \rangle \geq \langle I \rangle^2$. In fact, if only a single speckle is measured it can be shown that $\langle I^2 \rangle = 2\langle I \rangle^2$. In other words, after one correlation time the speckle has become a different speckle. The ensemble average is obtained by measuring during a time interval equal to many times (>$10^4$) this correlation time.

### 3.6.2. Brownian diffusion

Fluctuations in the scattered light are primarily caused by random displacements of the particles resulting from the constant bombardment by solvent molecules, known as Brownian motion (but in general also by rotations or shape fluctuations). We will now make a short detour and derive a few results to describe Brownian displacements, which we need to calculate the IACF. We will follow the ingenious and simple argument used by Einstein (1905).

Consider an equilibrium system of non-interacting particles suspended in a liquid. A steady external force **K** derivable from a potential $\Phi(\mathbf{r})$ acts on the particles and drives them to an impermeable boundary. (Think of the force of gravity driving the particle to the bottom of the container.) The velocity imparted to the particles by this force will be $\mathbf{K}/\gamma$ where $\gamma$ is called the friction factor. Random movements due to thermal agitation drive the particles away from the boundary. In this state of thermal equilibrium the probability density for the position of the particles is given by the Boltzmann distribution as

$$P(\mathbf{r}) = P_0 \exp(-\Phi / kT), \tag{3.49}$$

where $P_0$ is a normalization constant. In equilibrium the particle flux caused by the external force, $P\mathbf{K}/\gamma$, must be balanced by the flux due to Brownian diffusion at every position in the system:

$$-(P/\gamma)\nabla\Phi - D_0 \nabla P = 0\,. \tag{3.50}$$

Substitution of (3.49) then shows that the diffusion coefficient must have the value

[10] This assumes that the system is ergodic, so that time averages and ensemble averages are equal. For nonergodic systems we have no choice but to measure different speckles, although tricks exist to make this process more efficient.

$$D_0 = \frac{kT}{\gamma}, \tag{3.51}$$

which is known as the Einstein relation. For spherical objects the hydrodynamic friction factor $\gamma$ was calculated by Stokes:

$$D_0 = \frac{kT}{6\pi\eta a}. \tag{3.52}$$

Here $\eta$ is the shear viscosity of the solvent.

It is interesting to note that the particle flux caused by Brownian motion is the same as if a steady force acted on the particles opposite to the external force. In view of (3.50) this so-called Brownian, or thermodynamic, force is equal to

$$\mathbf{F}_{\mathrm{Br}} = -kT\nabla \ln P \tag{3.53}$$

This does not mean, of course, that the collisions of solvent molecules really cause the exertion of a steady force on the particle. It just means that when the probability density is non-uniform the mean Brownian displacement is such as to cause a diffusive flux equal to that caused by a steady external force $\mathbf{F}_{\mathrm{Br}}$.

The above arguments are equally valid for particles on which no external force is acting. If a suspension is non-uniform initially then the only particle flux is the diffusive flux $\mathbf{J} = -D\nabla P$. Particle conservation demands that its divergence must be equal to the (negative) rate of change of the local density, as expressed by the continuity equation

$$\frac{\partial P}{\partial t} = -\nabla \cdot \mathbf{J}. \tag{3.54}$$

Combining this with the diffusive flux results in the diffusion equation

$$\frac{\partial P}{\partial t} = D_0 \nabla^2 P. \tag{3.55}$$

In one of the problems you will be asked to solve this equation subject to the initial condition that the particles all start diffusing from the origin. (It would be wise, however, not to attempt this until after reading the next paragraph.) An important and characteristic property of the solution is that the mean-square displacement of the particles increases linearly with time:

$$\left\langle \Delta r^2 \right\rangle = 6D_0 t. \tag{3.56}$$

### 3.6.3. Dilute suspensions

We will now make the discussion a little more precise and derive the form of the IACF for a suspension of independent, but identical particles. Particles are independent if they do not interact with each other. This is the case if a suspension is sufficiently dilute. We will also assume that the particles have orientational degrees of freedom. If (3.46) is then substituted in (3.48) we get

$$g_I(\tau) = \left\langle f^4 \sum_{j,k,l,m=1}^{N} \exp\left(-i\mathbf{q}\cdot\left[\mathbf{r}_j(0) - \mathbf{r}_k(0) - \mathbf{r}_l(\tau) + \mathbf{r}_m(\tau)\right]\right)\right\rangle. \tag{3.57}$$

When particles are independent the average of a product is equal to the product of the averages. So, if only one of the numbers *jklm* is different from any of the others the IACF will contain a factor of the form $\left\langle \exp\left(-i\mathbf{q}\cdot\mathbf{r}_j(\tau)\right)\right\rangle$. This average is zero, since the particles are distributed randomly so that the phase is a random quantity (except in the uninteresting case $\mathbf{q}=\mathbf{0}$). As a result, only three kinds of terms survive: (i) $N^2$ terms for which $j=k$, $l=m$ , (ii) $N^2-N$ terms for which $j=m$, $k=l$, $j\neq k$, and (iii) $N^2-N$ terms for which $j=l$, $k=m$, $j\neq k$. In case (i) the terms equal unity. Terms of type (ii) give

$$\left\langle \exp\left(i\mathbf{q}\cdot\left[\mathbf{r}_j(\tau)+\mathbf{r}_j(0)\right]\right)\right\rangle\left\langle \exp\left(-i\mathbf{q}\cdot\left[\mathbf{r}_k(\tau)+\mathbf{r}_k(0)\right]\right)\right\rangle.$$

Both these averages are zero for the same reason as before: their phase depends on the absolute position of the particle, which is random. Finally, terms of type (iii) give

$$\left\langle \exp\left(i\mathbf{q}\cdot\left[\mathbf{r}_j(\tau)-\mathbf{r}_j(0)\right]\right)\right\rangle\left\langle \exp\left(-i\mathbf{q}\cdot\left[\mathbf{r}_k(\tau)-\mathbf{r}_k(0)\right]\right)\right\rangle.$$

These terms are *not* zero, because they depend on particle *displacements*, which are small for small $\tau$ and large for large $\tau$. Taking these things together we obtain

$$g_I(\tau) = N^2 f^4 + \left(N^2 - N\right) f^4 \left|\left\langle \exp\left(i\mathbf{q}\cdot\Delta\mathbf{r}_j(\tau)\right)\right\rangle\right|^2,$$

where $\Delta\mathbf{r}_j(\tau) = \mathbf{r}_j(\tau) - \mathbf{r}_j(0)$. Because $\langle I\rangle = Nf^2$ and $N^2 \gg N$ we finally get for the IACF:

$$g_I(\tau) = \langle I\rangle^2\left[1 + \left|\left\langle \exp\left(i\mathbf{q}\cdot\Delta\mathbf{r}_j(\tau)\right)\right\rangle\right|^2\right]. \tag{3.58}$$

We now need to evaluate the ensemble average in (3.58) for particles in Brownian motion. Let $P(\Delta\mathbf{r},t)$ be the probability of finding that a given particle has undergone a displacement $\Delta\mathbf{r}$ in a time *t*. The equation of motion for this process is the diffusion equation (3.55) with the initial condition that the particle has not moved at time $t = 0$:

$$\begin{cases} \dfrac{\partial P(\Delta\mathbf{r},t)}{\partial t} = D_0\nabla^2 P(\Delta\mathbf{r},t) \\ P(\Delta\mathbf{r},t=0) = \delta(\Delta\mathbf{r}) \end{cases} \tag{3.59}$$

We first recognize that the average we want to calculate is

$$\left\langle \exp\left(-i\mathbf{q}\cdot\Delta\mathbf{r}\right)\right\rangle = \int P(\Delta\mathbf{r},t)\exp\left(-i\mathbf{q}\cdot\Delta\mathbf{r}\right)d(\Delta\mathbf{r}).$$

But this is just the Fourier transform of *P* with respect to $\Delta\mathbf{r}$. Thus, we Fourier transform Eq. (3.59), giving

$$\begin{cases} \dfrac{\partial P(\mathbf{q},t)}{\partial t} = -q^2 D_0 P(\mathbf{q},t) \\ P(\mathbf{q},t=0) = 1 \end{cases},$$

with the simple solution

$$P(\mathbf{q},t) = \exp\left(-q^2 D_0 t\right).$$

Substitution into Eq. (3.58) then yields the result

$$g_I(\tau) = \langle I\rangle^2\left[1 + \exp\left(-2q^2 D_0\tau\right)\right]. \tag{3.60}$$

This result tells us that the IACF of particles in (independent) Brownian motion is an exponential with a characteristic time $1/2q^2D_0$. This is the behavior seen in the experiment of Figure 3.12. The IACF will decay faster if the particles diffuse faster, as expected. There is also a strong *q*-dependence (slow decay at small angles). This is because $q^{-1}$ can be seen as the distance over which particles need to diffuse in order to cause the IACF to decay. Since in a diffusion process mean-square displacements are proportional to time this leads to a $q^{-2}$ dependence.

Equation (3.60) is often used to measure the diffusion coefficient of particles in a dilute solution. This value is then related to an (average) particle radius, using (3.52). The radius obtained from this equation is usually called the *hydrodynamic radius*, because it is the hydrodynamic friction on the sphere that determines its diffusion coefficient. If a layer of solvent close to the particle is entrained the hydrodynamic radius may be somewhat larger than the actual radius. This can be significant if the particles are rough, or if they contain a

stabilizing coating of long polymer molecules. Non-spherical particles have of course a more complicated relation between $D_0$ and $R$. Results have been derived for different geometric shapes.

An important application of DLS is the measurement of the particle size distribution of polydisperse suspensions. In essence, the decay of the IACF becomes multi-exponential due to the presence of particle with different sizes. It can be inverted using numerical methods to obtain the size distribution, which is used routinely on all commercial instruments.

### 3.6.4. Concentrated suspensions

In concentrated suspensions the particles are no longer independent and the average in Eq. (3.57) cannot be factorized so easily. Nevertheless, it is possible to derive some general results. We will state the most important results and refer to the literature cited at the end of this chapter for details and derivations. It can be shown that under certain general assumptions the *Siegert relation* is valid:

$$g_I(\tau) = \langle I \rangle^2 + \left| g_E(\tau) \right|^2, \tag{3.61}$$

where the electric field autocorrelation function (EACF) $g_E$ is defined by

$$g_E(\tau) = \left\langle E_s(t) E_s^*(t+\tau) \right\rangle. \tag{3.62}$$

The assumptions are that (i) the scattering volume contains a large number of particles, (ii) the scattering volume is much larger than the range over which the particles are correlated, and (iii) the system is ergodic. For most systems these assumptions are easily satisfied[11]. By inserting Eq. (3.46) it is seen that

$$g_E(\tau) = \langle I \rangle \left\langle \frac{1}{N} \sum_{j,k=1}^{N} \exp\left( i\mathbf{q} \cdot \left[ \mathbf{r}_j(0) - \mathbf{r}_k(\tau) \right] \right) \right\rangle. \tag{3.63}$$

The factor between angled brackets is called the *dynamic structure factor* (or intermediate scattering function). Notice that it equals the (static) structure factor, Eq. (3.27), at $\tau = 0$. Also notice that for independent particles the only nonzero terms are those with $j=k$, so that we obtain (3.58) again.

The analysis of the measured IACF is much more complicated than in the dilute case. Particle correlations affect both the decay rate and the $q$-dependence of the IACF. This is described phenomenologically with an equation similar to (3.60) but with a $q$ and $t$ dependent *collective* diffusion coefficient $D_c$:

$$g_I(\tau) = \langle I \rangle^2 \left[ 1 + \exp\left( -2q^2 D_c(q,\tau)\tau \right) \right] \tag{3.64}$$

Note that this does *not* mean that the IACF still decays according to a single exponential. Also, there is no longer a pure $q^{-2}$ dependence. How then should the collective diffusion coefficient be interpreted? Remember that thermal motion of particles constantly creates and dissipates small fluctuations in the particle number density. These fluctuations have a (Fourier) spectrum of wavelengths. The collective diffusion coefficient can be interpreted as describing the relaxation of a sinusoidal density wave with a wavelength $2\pi/q$. $D_c$ is also time dependent because at short times particles move only a small distance relative to each other (much less than a particle diameter). But on longer time scales particles must increasingly 'pass each other' in order to make progress. Thus, the diffusion coefficient relevant to mean-square displacements at longer times is decreased.

[11] The electric field seen by the detector is then a Gaussian random variable.

As an illustration of these ideas we will conclude this section with a few examples, namely that of concentrated suspensions of spherical colloids. Figure 3.13 shows a number of dynamic structure factors (curves 1-3) of charged spheres measured at different $q$. It is clear that the decay is not single-exponential (curve 2), contrary to what is measured at low volume fractions (curve 4). However, in the limit of short times a single exponential can still describe the initial decay. The inverse of the 'short-time' diffusion coefficients assigned to the data in this way are also shown in Figure 3.13 together with the static structure factor. It is seen that the shape of the $1/D_c(q)$-curve looks similar to that of the structure factor. Apparently, density fluctuations with a length scale similar to the mean interparticle spacing relax slowly (at short times). Diffusion on shorter and longer length scales is more rapid. Although DLS results from various kinds of soft matter can be quite different this same behavior is generally found: diffusion on length scales with strong particle ordering is slow and vice versa.

The collective diffusion coefficient tells us how particles move collectively to dissipate spontaneous density fluctuations. It can be much more insightful to consider the motion of individual particles in the suspension. This is described by the self-diffusion coefficient. It can be measured with DLS by adding a small number of strongly scattering 'tracer' particles to a suspension containing a large number of index-matched 'host' particles. Because the tracers are the only particles that scatter light the only nonzero terms in (3.63) are those for which $j$ and $k$ both refer to a tracer particle. Since the tracers are very small in number they behave independently (i.e. they almost never interact with each other). Thus, the averages of terms with $j \neq k$ separate into products, producing zero as before. Only terms with $j=k$ survive, and these lead to

$$g_E = \langle I \rangle \left\langle \exp\left(i\mathbf{q}\cdot\Delta\mathbf{r}(\tau)\right)\right\rangle. \tag{3.65}$$

Accordingly, we can write

$$g_I(\tau) = \langle I \rangle^2 \left[1+\exp\left(-2q^2 D_s(\tau)\right)\right], \tag{3.66}$$

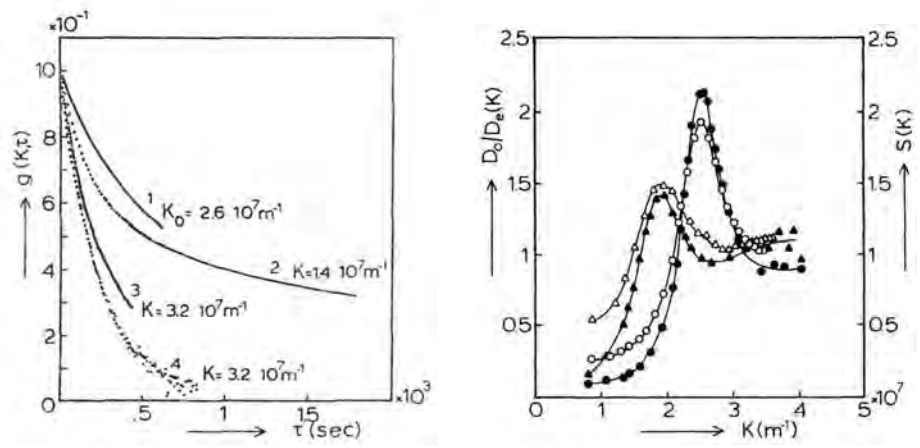



**Figure 3.13. DLS measurements on a suspension of charged silica spheres ($q$=K). Left graph: Decay of the dynamic structure factor at different wave vectors at a volume fraction of 0.10 (curves 1,2,3), and 0.001 (curve 4). Right graph: $D_0/D_c$ (open symbols) and $S(q)$ (closed symbols) measured at volume fractions of 0.10 (circles) and 0.043 (triangles). From: A. P. Philipse and A. Vrij, *J. Chem. Phys.* 88, 6459 (1988).**

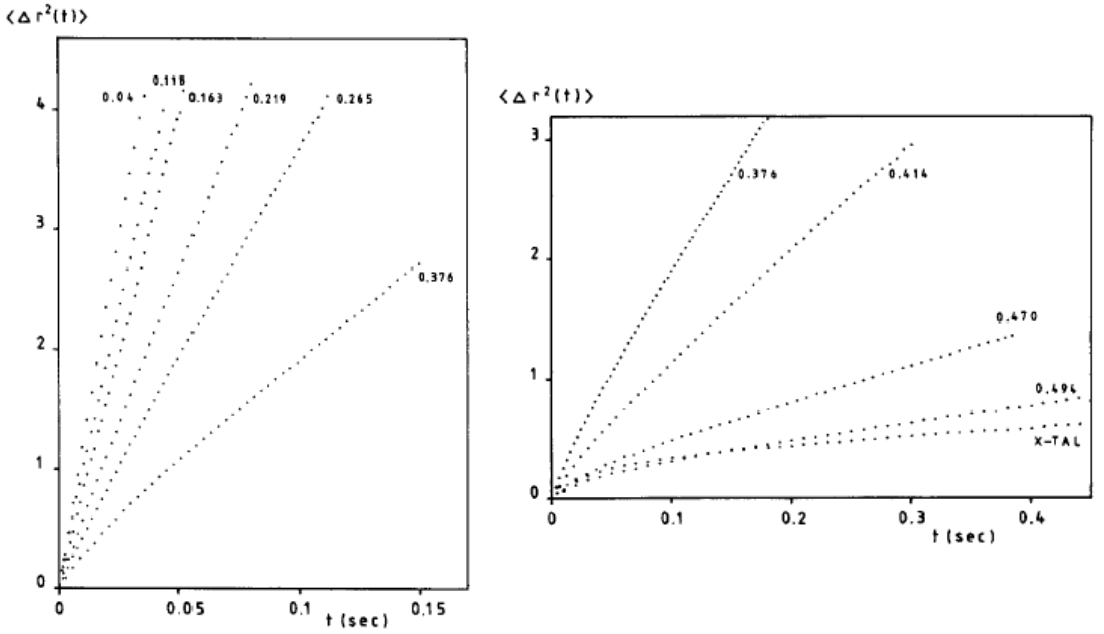



**Figure 3.14. Mean square displacements of tracer spheres (silica) in a suspension of hard sphere colloids (pmma), measured with DLS. The particles had a radius of 330 nm and were suspended in a mixture of $CS_2$ and decalin, which index-matched the pmma spheres. The particle volume fractions are shown on the graphs. From: W. van Megen and S. M. Underwood, *J. Chem. Phys.* 91, 552 (1989).**

where $D_s$ is the *self-diffusion coefficient*. It does not depend on $q$ because distances between the independent tracers do not play a role. $D_s$ is time dependent, however, because the tracer particles do still feel their nonscattering neighbors. Some experimental results are shown in Figure 3.14 for concentrated suspensions of hard sphere colloids. They are presented in terms of mean square displacements (msd) as a function of time. It can be seen that at low volume fractions the msd increases linearly with time, as expected for normal diffusion. But at higher volume fractions there is a crossover from a short-time to a slower long-time regime. In the crossover regime the tracer particles start feeling the presence of their neighbors, which slow down their progress. On very long time scales these particle interactions have averaged out and the msd increases linearly with time again, but much more slowly. One therefore speaks of short-time and long-time self-diffusion coefficients, equal to one-sixth the slope of the msd.

In the experiments shown effort was taken to make the interactions between the tracer and host particles identical, so that it may be assumed that the motion of the tracers is identical to that of the host particles. This is not necessary, however. Similar experiments are performed in a wide variety of systems. For example, tracers can be added to polymer solutions, or attached to cell membranes. The dynamics of the tracers then provides information about the properties of the host material, such as the local viscosity or elasticity.

## 3.7. Scattering of other types of radiation

As mentioned before not only light can be used to obtain information on the structure of matter, but in principle type of information is obtained with other types of radiation. Of these, X-rays and neutrons are routinely used in the study of soft condensed matter and, indeed, in

the study of all kinds of condensed matter. We discuss some of the differences and advantages compared to light scattering.

### 3.7.1. X-rays

The wavelength of X-rays is typically on the order of 0.1 nm, which places us in the large $q$ limit. Information on the structure of larger particles therefore appears at small angles, on the order of 1º. This application of X-rays is therefore called small-angle X-ray scattering, or SAXS. To reach such small angles the detector is placed at large distance from the sample, typically ~10 m. Smaller length scales such as polymer segments and intramicellar structure are studied at wider angles (WAXS). Length scales that can therefore be probed with X-ray scattering range from 0.01 to $10^3$ nm. Since the refractive index of most materials at these short wavelengths is very close to unity (to within $10^{-6}$ to $10^{-4}$) multiple scattering is absent even in samples that strongly scatter light. Since X-rays are electromagnetic radiation they are scattered by charged particles, usually electrons. Because of there short wavelength the scattering strength is related not to the polarizability of bulk matter but to the scattering strength of an electron. For an individual electron which oscillates in the e.m. field the scattered intensity is again a dipole field (*Thomson scattering*):

$$I_s = I_0 \left( \frac{e^2}{4\pi\varepsilon_0 m_e c^2 R} \right)^2 \frac{1+\cos^2\theta}{2}. \tag{3.67}$$

Here $m_e$ and $e$ are the electron mass and charge, and $c$ the speed of light. To calculate scattering by larger objects such as atoms, polymer molecules or colloidal particles we sum the fields scattered by the electrons taking into account the proper phase differences, as explained before. It is clear that the resulting formulas will be completely analogous to those for visible light. Because heavy elements contain many more electrons than light elements scattering by the former dominates the measured intensity. This may make it difficult to measure light elements in a sample.

Classically, X-rays are produced in X-ray tubes, in which a metal foil is bombarded with electrons. The resulting X-rays have a well-defined wavelength and can be used to irradiate samples. Since X-rays are hard to focus and reflect from mirrors a collimated beam must be produced with a narrow slit. This seriously limits the available power. Stronger X-ray beams are available from modern synchrotrons in which magnetic bending or undulation of a relativistic electron beam produces an almost collimated beam containing a “white” spectrum of wavelengths (typically 0.05 to 0.2 nm). The extremely high irradiance makes it possible to study weakly scattering samples, even after monochromation and collimation. Worldwide a growing number of synchrotrons is available at national or international facilities. European synchrotrons with international access are in Grenoble (“European Synchrotron Radiation Facility”), Daresbury (“Daresbury Synchrotron Radiation Source”), Hamburg (“Hamburger Synchrotronstrahlungslabor”), Villigen (“Swiss Light Source”), and Aarhus (“Institute for Storage Ring Facilities”).

Although the X-rays produced in synchrotrons (like those from tubes) are incoherent the beams are so intense that a small pinhole can be placed in the beam to produce a nearly coherent source. In this way Dynamic X-ray Scattering (DXS) has become possible in recent years.

### 3.7.2. Neutrons

Neutrons interact with matter mainly through two types of interaction. The magnetic dipole of the neutron interacts with the magnetic field of unpaired electrons. Although this is very important in the study of magnetic materials it does not usually play a role in soft condensed matter systems. The other interaction is the strong nuclear interaction between the

neutron and the atomic nuclei. Despite the word "strong" the scattering cross sections of atoms (*i.e.* of their nuclei) are very small, so neutrons have a high penetrating power and multiple scattering is of no concern. Neutrons are produced in nuclear reactors. Some of the neutrons are needed to continue the nuclear fission process, while others need to be transported out. In reactors used for scientific research these neutrons are channeled into relatively well-collimated beams. Neutrons are described as a matter wave with De Broglie wave vector

$$k = \frac{1}{\hbar}\sqrt{2m_n E}\ . \tag{3.68}$$

$E$ is the kinetic energy of the neutrons and $\hbar$ Planck's constant divided by $2\pi$. This energy depends on the reactor and is often given in terms of the neutron temperature, as $E = \frac{3}{2}kT$. In many reactors the neutrons have undergone many collisions with the nuclei present in the moderator and will have "thermalized" upon leaving the reactor, so that their temperature is about 300 K. These neutrons have a wavelength of 0.2 nm. (By using a cooled moderator wavelengths of up to 2 nm are accessible.) This is the same order of magnitude as that of X-rays, so that neutron scattering provides information on the same length scales. Correspondingly, small angle neutron scattering is called SANS.

Since the strong interaction is very short range the atomic nuclei can be considered point scatterers for neutrons at room temperature or below. Thus the neutron waves scattered by each atomic nucleus has a form similar to the oscillating dipole in light and X-ray scattering. Scattering by larger objects follows by summing over all nuclei including the correct phase, just like before. The strength of scattering is expressed as a *scattering length*, which depends on the number of protons and neutrons in the nucleus, but not in a systematic way. This has the advantage that light elements may scatter as much as heavy elements. Even different isotopes of the same element may have completely different scattering lengths. Since the isotopes will be distributed randomly through the sample, and their nuclear spins normally have a random orientation, the scattered intensity contains an often-large background of so-called incoherent scattering. An important difference with electromagnetic scattering is that neutron scattering lengths for some elements, most notably $^1$H, are *negative*. The analogous thing in light scattering would be to have a negative $f_j$ in equation (3.25) for certain components of the sample. By mixing normal solvents with deuterated solvents ($^2$H has a positive scattering length) the scattering contribution from selected parts of the sample can be made to vanish while minimally affecting the chemical composition. This process is called *contrast variation* and is an important technique in neutron scattering. In light scattering contrast variation is only possible by replacing the solvent by one with a refractive index equal to that of the selected part of the sample. (In equation (3.25) the $f_j$ of this part would become zero.) But this always comes at the price of changing the chemical makeup of the system under study, which is almost certain to lead to other, unwanted, changes.

Several nuclear reactors provide beamlines for scientific research. In the Netherlands test reactors offering neutron beamlines to researchers are located in Delft ("Interfacultair Reactor Instituut") and in Petten ("Energieonderzoek Centrum Nederland"). More powerful neutron sources open to European researchers are in Jülich ("Forschungszentrum Jülich") and in Grenoble ("Institut Laue-Langevin").

## 3.8. *Appendix*: Relation between *S*(q) and g(r).

To derive Eq. (3.29) we start with Eq. (3.27) and notice that the quantity to be averaged is a sum of terms that each depend on only two variables. Each of these terms is identical and can be related to the pair distribution function $\rho^{(2)}$:

$$
\begin{aligned}
S(\mathbf{q}) &= \frac{1}{N}\left\langle \sum_{j=1}^{N}\sum_{k=1}^{N} e^{i\mathbf{q}\cdot(\mathbf{r}_k-\mathbf{r}_j)} \right\rangle \\
&= 1+\frac{1}{N}N(N-1)\left\langle e^{i\mathbf{q}\cdot(\mathbf{r}_2-\mathbf{r}_1)} \right\rangle \\
&= 1+\frac{1}{N}\int_V\int_V e^{i\mathbf{q}\cdot(\mathbf{r}_2-\mathbf{r}_1)}\rho^{(2)}(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2 \\
&= 1+\frac{\rho^2}{N}\int_V d\mathbf{r}_1\int_V d\mathbf{r}_2 e^{i\mathbf{q}\cdot(\mathbf{r}_2-\mathbf{r}_1)} g^{(2)}(\mathbf{r}_2-\mathbf{r}_1) \\
&= 1+\rho\int_V d\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} g(\mathbf{r})
\end{aligned}
$$

Since $g(r)$ approaches unity for large $r$ it is convenient to rewrite this as

$$
\begin{aligned}
S(\mathbf{q}) &= 1+\rho\int_V d\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\left(g(\mathbf{r})-1\right)+\rho\int_V d\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} \\
&= 1+\rho\int_V d\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\left(g(\mathbf{r})-1\right)+(2\pi)^3\rho\delta(\mathbf{q})
\end{aligned}
$$

The last term is irrelevant as long as the scattering vector does not vanish:

$$
S(\mathbf{q}) = 1+\rho\int_V d\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}}\left(g(\mathbf{r})-1\right).
$$

This is Eq. (3.29). Note that the limit $q \to 0$ is still defined properly because $\lim_{q\to 0}\delta(q) = 0$.

## 3.9. More reading

The derivation of the field radiated by an oscillating dipole can be found in:

- Griffiths, D. J., *Introduction to Electrodynamics* (3rd ed., Prentice Hall, Upper Saddle River, 1999).

or in most other textbooks on electrodynamics.

Famous books on light scattering and absorption by individual particles are:

- H. C. van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957)
- M. Kerker, *The scattering of light and other electromagnetic radiation* (Academic Press, New York, 1969).
- C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles* (John Wiley & Sons, New York, 1983).

Note that these books only treat scattering by single particles. For discussions of structure factors we refer to textbooks on soft condensed matter, which are listed elsewhere in this syllabus.

A classic text on dynamic light scattering is

- B. J. Berne and R. Pecora, *Dynamic Light Scattering – With applications to chemistry, biology, and physics* (Wiley, New York, 1976).

# 4. Self-assembly in soft matter

Willem K. Kegel
Van't Hoff Laboratory for Physical and Colloid Chemistry
Utrecht University
The Netherlands

June 2015

# Contents

# Chapter 1

# Surfactants and Micelles

## 1.1 Introduction: surfactants

An important type of molecules that adsorb positively at (water-air or water - oil) interfaces are so-called surface-active agents or briefly *surfactants*. The common characteristic of these materials is that their molecules consist of a polar 'head' (hydrophilic part) and a hydrocarbon 'tail' (hydrophobic part), as schematically depicted in Fig. 1.1. The amphipolar nature of surfactants makes them adsorb positively at relatively low surfactant concentrations.

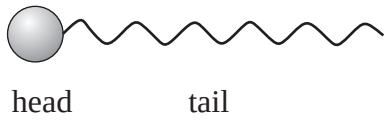



Figure 1.1: The common characteristic of surface active agents, or briefly surfactants, is that they consist of a polar 'head' and a hydrocarbon 'tail'.

Depending on their chemical structure and properties of the head group, a surfactant is called anionic, cationic, amphoteric, or non-ionic.

Classical 'soaps' are a well-known example of anionic surfactants. These soaps are usually sodium or potassium salts of fatty acids (carboxylates) with a chain length of typically 10 to 20 hydrocarbon groups; $RCOO^-M^+$, for instance sodium lauraat, $C_{11}H_{23}COO^-Na^+$. With increasing tail length, the solubility in water decreases and hence the surface activity increases.

In the case of cationic surfactants, the soap is a cation, i.e., the headgroup is positively charged. Examples are tetra- alkyl ammonium salts such as cetyl-trimethyl ammonium bromide (CETAB).

Surfactants containing head groups that have both positive and negative sites are referred to as amphoteric. These surfactants are well-miscible with all other type surfactants and less antagonizing for skin and eyes. Therefore they are frequently encountered in cosmetics. Typical examples are the betaines.

The remaining approximately 21% of the total of surfactants are comprised by non-ionic surfactants. The head group of this type does not contain charges but has a propensity to forming hydrogen bonds. They are well-miscible with all other types of surfactants.

An important class of non-ionic surfactants are the poly(glycol ethers), $R - O(C_2H_4O)_nH$, or $C_mE_n$. The hydrocarbon tail R is usually linear and is abbreviated as $C_m$. The ethylene oxide head-group is frequently denoted as $E_n$. For small head-groups ($n = 1 - 6$) these surfactants are frequently applied as oil/water emulsifiers. Intermediate ethylene oxide head-group ($n = 6 - 15$) are found in detergents, whereas large head-groups ($n > 15$) in special emulsifiers.

Positive adsorption at interfaces decreases the interfacial tension as dictated by the Gibbs adsorption equation. However, beyond a certain surfactant concentration, the critical micelle concentration (cmc), the interfacial tension remains more or less constant. The situation has been shown schematically in Fig. 1.2.

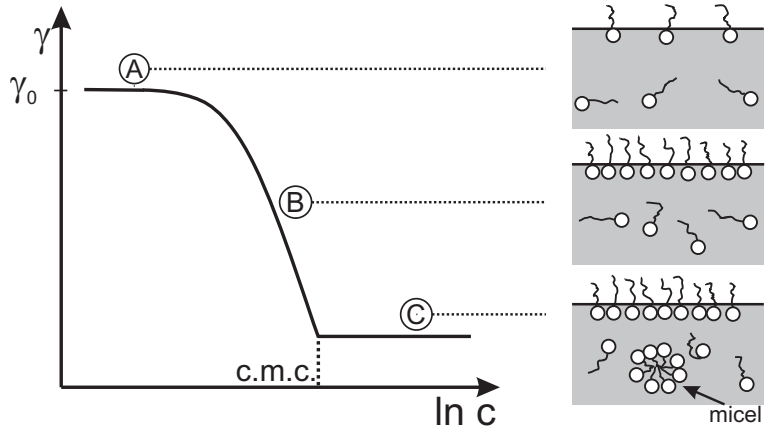



Figure 1.2: Charactistic change of the interfacial tension as a function of surfactant concentration. The dashed vertical line indicates the critical micelle concentration, cmc.

## 1.2 Micelles: The critical micelle concentration (cmc)

In this treatment I will follow Debye, [1]. Forming a micelle containing $n$ surfactant monomers can be described as the equilibrium

$$nA \rightleftharpoons A_n \tag{1.1}$$

In the above equilibrium, $A$ stands for surfactant monomers and $A_n$ for a surfactant aggregate (micelle) containing $n$ surfactant molecules. Typically the value of $n$ is 50-100 for spherical micelles. Now Debye's arguing goes as follows. Define the equilibrium constant as

$$K = \frac{x_n}{x_1^n} \tag{1.2}$$

where $x_1$ is the concentration (molefraction) surfactant monomers, and $x_n$ the concentration (molefraction) micelles. In general the dimension of the equilibrium constant as defined by eq. 1.2 is $concentration^{(1-n)}$. We use this dimensional argument to *define* the concentration

$$x_0 = K^{\frac{1}{1-n}} \tag{1.3}$$

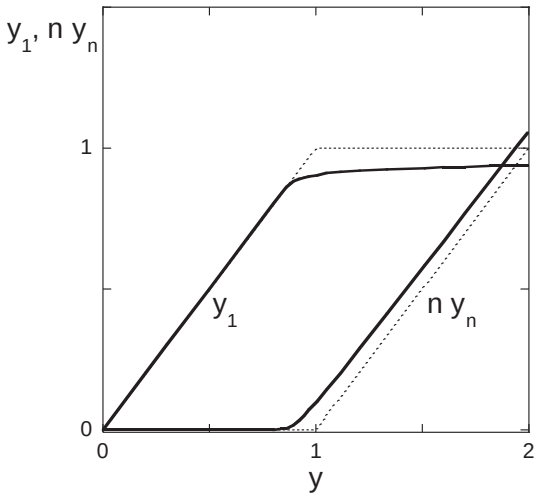



Figure 1.3: Iterative solution of Eq. 1.5. The surfactant concentration $y_1$ and the micelle concentration $ny_n$ with $n = 65$ as a function of the total surfactant concentration, $y$, in the system, in reduced units. Below the c.m.c. ($y = 1$) micelles are barely present, whereas above the c.m.c. the monomer concentration is virtually constant. The dashed lines represent the limiting case for $n \to \infty$, given by Eq. 1.6

Now the total concentration of surfactant molecules in the system is the sum of monomeric surfactant molecules, and surfactant molecules in the form of micelles. In other words, the total surfactant concentration, $x$, follows by *mass conservation*

$$x = x_1 + nx_n \tag{1.4}$$

Writing $x_n = x_1^n K = x_1^n / x_0^{n-1}$, and defining the reduced concentrations $y = x/x_0$, $y_1 = x_1/x_0$, and $y_n = x_n/x_0$ eq. 1.4 becomes

$$y = y_1 + ny_1^n \tag{1.5}$$

with the typical values of $n \sim 50 - 100$, Eq. 1.5 can be solved iteratively for $y_1$, or plotted as $y$ as a function of $y_1$ and subsequently interchanging the $x$ and $y$ axis. It can be verified from eq. 1.5, that with typical values of $n$ stated above, we have

$$y_1 \begin{cases} = y, & \text{if } y < 1; \\ \approx 1, & \text{if } y > 1. \end{cases}, \tag{1.6}$$

The results of a numerical calculation using $n = 65$ has been shown in Fig. 1.3.

The crossover from the behaviors described by eq. 1.6 occurs around y=1, as can be seen in Fig. 1.3. The value of y=1, therefore, corresponds to a *critical point*. $y = 1$ implies that $x = x_0$. Therefore, $x_0$ is defined as the *critical micelle concentration*. As long as $x < x_0$, all surfactant is present in the form of monomers. But if $x > x_0$, micelles are being formed; the monomer ($x_1$) concentration remaining (almost) constant.

## 1.3 Thermodynamics of micelle formation

In section 1.2 we presented a *phenomenological* description: we *presumed* that micelles of size $n$ form, and subsequently studied their properties. We now address the question: what is the thermodynamic condition for the formation of micelles of size $n$? We return to eq. 1.1, but leave the value of $n$ unspecified. We address the question as to what the size distribution of micelles is, on the basis of thermodynamics. The general thermodynamic condition for chemical equilibrium is

$$\sum_i \nu_i \mu_i = 0, \tag{1.7}$$

where $\nu_i$ are the stoichiometric coefficients and $\mu_i$ the chemical potentials. Applying eq. 2.1 to the equilibrium 1.1 we get

$$\mu_n = n\mu_1 \tag{1.8}$$

In this equation, the subscripts 1 and $n$ again refer to surfactant monomers, and micelles consisting of $n$ surfactant molecules, respectively. As long as monomers and micelles are dilute, we may write the chemical potentials as

$$\mu_i = \mu_i^0 + kTlnx_i \tag{1.9}$$

where $\mu_i^0$ is the standard chemical potential ($\mu_i = \mu_i^0$ iff $x_i = 1$), k and T are Boltzmann's constant and absolute temperature, and i in this case can be monomer ($i = 1$) of micelle ($n$mer: $i = n$). Combining eqs. 2.2 and 2.3 leads to the distribution of micelles of size $n$

$$x_n = x_1^n exp\left(\frac{-(\mu_n^0 - n\mu_1^0)}{kT}\right) \tag{1.10}$$

First of all, note that the cmc is related to the above expression by (compare with eqs. 1.2 and 1.3

$$x_0 \equiv x_{cmc} = K^{1/(1-n)} = exp\left(\frac{\mu_n^0 - n\mu_1^0}{(n-1)kT}\right) \tag{1.11}$$

In order for the system to have a cmc of $x_0 < 1$ (note that $x_0 > 1$ is *unphysical*; in that case there will be no cmc at all), it follows from Eq. 1.11 that there must be a value of $n$, or a range of values, such that

$$\mu_n^0 < n\mu_1^0 \tag{1.12}$$

In other words, for micelles of size $n$ to form, it follows from eq. 2.4 that the function $\frac{\mu_n^0 - n\mu_1^0}{kT}$ should have a (deep) minimum at $n$.

## 1.4 Influence of molecular properties of the surfactant on the cmc

First of all we write the argument of the exponent in Eq. 1.11 as

$$\frac{\mu_n^0 - n\mu_1^0}{(n-1)} \equiv \Delta\mu^\theta = \left(\Delta\mu^\theta\right)_{\text{head}} + \left(\Delta\mu^\theta\right)_{\text{tail}}. \tag{1.13}$$

In order to form micelles it is required that $\Delta\mu^\theta < 0$, as has been discussed below Eq. 1.11. The head groups will be closely packed upon micelle formation, which is in general unfavorable compared to the situation where surfactant molecules are monomers. We may therefore expect $\left(\Delta\mu^\theta\right)_{\text{head}} > 0$. The packing will be favorable for the hydrophobic tails and hence $\left(\Delta\mu^\theta\right)_{\text{tail}} < 0$. Since micelle formation requires $\Delta\mu^\theta < 0$, we find $\left|\left(\Delta\mu^\theta\right)_{\text{tail}}\right| > \left|\left(\Delta\mu^\theta\right)_{\text{head}}\right|$.

Despite the vast amount of literature on the calculation of $\Delta\mu^\theta$, we restrict ourselves to observed trends in c.m.c. and the subsequent conclusions for $\left(\Delta\mu^\theta\right)_{\text{head}}$ and $\left(\Delta\mu^\theta\right)_{\text{tail}}$.

### 1.4.1 Influence of chain length on cmc

Upon increasing chain length, it is expected that $\left(\Delta\mu^\theta\right)_{\text{tail}}$ gets more negative and the critical micelle concentration decreases. Experimental values in Table 1.1 confirm this expectation. The cmc is lowered by roughly a factor of 2 for each added $CH_2$-group in anionic surfactants. According Eq. 1.11 this implies that $\Delta\mu^\theta$ decreases by (per mole!) $RT\ln 2 \approx 1.72\,\text{kJ/mole}$ per added $CH_2$-group. The critical micelle concentration decreases stronger for non-ionic surfactants for each additional $CH_2$-group; rougly a factor of 3. This in turn implies that $\Delta\mu^\theta$ decreases approximately 2.72 kJ/mol for each additional $CH_2$ tail-unit .

Table 1.1: The c.m.c. and $\Delta\mu^\theta = RT\ln x_{\text{cmc}}$ of four types of surfactants as a function of the number of C-atoms $n_C$ in the alkyl tail at 298 K.

| | $R$ | $c_{\text{cmc}}$ [mol/l] | $x_{\text{cmc}}$ | $\Delta\mu^\theta$ [kJ/mol] |
|---|---|---|---|---|
| $RCOO^-Na^+$: | $C_{12}$ | $2.3\times10^{-2}$ | $4.2\times10^{-4}$ | -19.3 |
| $\Delta\mu^\theta = 0.65 - 1.66n_C$ kJ/mol | $C_{14}$ | $6.0\times10^{-3}$ | $1.1\times10^{-4}$ | -22.5 |
| | $C_{16}$ | $1.5\times10^{-3}$ | $2.7\times10^{-5}$ | -26.0 |
| | $C_{18}$ | $4.0\times10^{-4}$ | $7.3\times10^{-6}$ | -29.2 |
| $ROSO_3^-Na^+$: | $C_8$ | $1.3\times10^{-1}$ | $2.3\times10^{-3}$ | -15.0 |
| $\Delta\mu^\theta = -1.35 - 1.70n_C$ kJ/mol | $C_{10}$ | $3.3\times10^{-2}$ | $6.0\times10^{-4}$ | -18.3 |
| | $C_{12}$ | $8.3\times10^{-3}$ | $1.5\times10^{-4}$ | -21.7 |
| | $C_{14}$ | $2.1\times10^{-3}$ | $3.8\times10^{-5}$ | -25.2 |
| $R(CH_3)_3N^+Br^-$: | $C_{10}$ | $6.5\times10^{-2}$ | $1.1\times10^{-3}$ | -16.8 |
| $\Delta\mu^\theta = 0.59 - 1.74n_C$ kJ/mol | $C_{12}$ | $1.6\times10^{-2}$ | $2.9\times10^{-4}$ | -20.2 |
| | $C_{16}$ | $9.2\times10^{-4}$ | $1.7\times10^{-5}$ | -27.2 |
| $R.-.O(C_2H_4O)_6.H$ ($C_{n_C}E_6$): | $C_8$ | $7.6\times10^{-3}$ | $1.4\times10^{-4}$ | -22.0 |
| $\Delta\mu^\theta = 0.40 - 2.80n_C$ kJ/mol | $C_{10}$ | $8.0\times10^{-4}$ | $1.4\times10^{-5}$ | -27.6 |
| | $C_{12}$ | $8.3\times10^{-5}$ | $1.5\times10^{-6}$ | -33.2 |
| | $C_{14}$ | $8.7\times10^{-6}$ | $1.6\times10^{-7}$ | -38.8 |

The change in standard chemical potential of a surfactant with tail length $n_c$ can according to the experimental results be described emperically by the Klevens equation [2]

$$\Delta\mu^\theta = RT \ln x_{\rm cmc} = A - Bn_c \tag{1.14}$$

The value of $A$ is dominated by the nature of the head group. Generally one finds $-3.1 < A < +3.7\,\mathrm{kJ/mol}$. $B$ values are typically found in the range $1.5 < B < 2.9\,\mathrm{kJ/mol}$. The Klevens equation for some characteristic examples is given in the left column of Table 1.1.

The Klevens equation, 1.14, only holds for $n_C \lesssim 16$, presumably because for larger tail lengths the alkyl tail will fold to prevent hydrophobic interactions. On the other hand, $n_C \gtrsim 8$ to have enough thermodynamic driving force for micelle formation, i.e., $\left|\left(\Delta\mu^\theta\right)_{\rm tail}\right| > \left|\left(\Delta\mu^\theta\right)_{\rm head}\right|$.

### 1.4.2 Effect of salt on the cmc

We expect $\left(\Delta\mu^\theta\right)_{\rm head} > 0$ as a consequence of the interactions between charged or polar head-groups. The charges of the head groups are screened by the addition of an indifferent salt which decreases the interactions between the head groups. Therefore, we expect $\left(\Delta\mu^\theta\right)_{\rm head}$ to decrease upon increasing elektrolyte concentration. From Table 1.2 it is verified that this is indeed the case.

Table 1.2: The c.m.c. and $\Delta\mu^\theta = RT \ln x_{\rm cmc}$ of sodium octyl sulphate, $\mathrm{H(CH_2)_8OSO_3^-Na^+}$, as a function of salt concentration at 298 K.

| | $c_{\rm cmc}$ [mol/l] | $x_{\rm cmc}$ | $\Delta\mu^\theta$ [kJ/mol] |
|---|---|---|---|
| water | 0.134 | $2.4 \times 10^{-3}$ | -14.9 |
| water + 0.01 mol NaCl/l | 0.121 | $2.2 \times 10^{-3}$ | -15.2 |
| water + 0.03 mol NaCl/l | 0.102 | $1.8 \times 10^{-3}$ | -15.6 |
| water + 0.1 mol NaCl/l | 0.069 | $1.2 \times 10^{-3}$ | -16.6 |
| water + 0.3 mol NaCl/l | 0.035 | $6.3 \times 10^{-4}$ | -18.3 |

The critical micelle concentration as a function of salt concentration generally depends on the valence of the added ions, in particular that of the counter ions. Addition of salt increases the sensitivity of the cmc with increasing tail-length. That is, the cmc decreases steeper with longer carbon tails and, at very high ionic strength, may even resemble the behaviour of non-ionic surfactants. There is almost no influence of salt on the critical micelle concentration of non-ionic surfactants.

## 1.5 Geometry of surfactant molecules and micellar shape.

What is the physical origin of the minimum of the function $\Delta\mu^\theta$ as a function of $n$ in eqs. 1.11, 1.13? Note that the quantity $\Delta\mu^\theta$ equals the difference in Gibbs free energy of a molecule inside a micelle containing $n$ surfactant molecules, and a freely moving surfactant monomer. Thus, the position of the minimum suggests that at that particular value of $n$, the surfactant molecules are most comfortably packed in a micelle. This suggests that the geometry of surfactant molecules the relative size of their headgroups, the lengths of their hydrophobic tails, may have something to do with it. Indeed, packing cones in the form of a sphere will naturally lead to a

number of cones that are close packed in a sphere. Squeezing in more cones than the number corresponding to close packing will lead to repulsion, and the function $\Delta\mu^\theta$ will rise with $n$. Other shapes of the surfactant molecules will lead to different structures of the aggregates. See Fig 1.4, taken from Ref. [3]. Referring to this Figure, we will now derive the geometrical conditions (on surfactant molecules) for aggregate shapes.

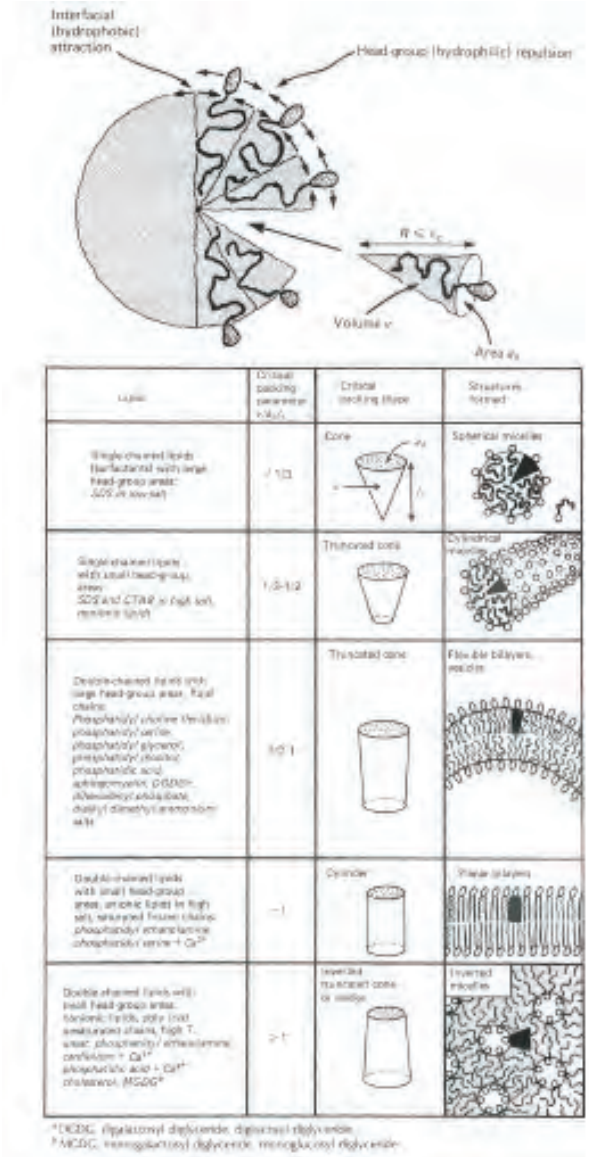



Figure 1.4: Relation between surfactant parameter and micelle shape - from Ref [3]

The three geometrical parameters defined in Fig. 1.4 are the effective molecular headgroup area $a_0$, the effective length of the hydrophobic tail $\ell_0$, and the effective volume of the surfactant molecule, $v_0$. Note that these parameters are not fixed for a certain surfactant molecule: they depend on conditions such as the ionic strength and temperature. That particularly applies to $a_0$: this quantity includes the effect of the electrical double layer around the headgroup (if the headgroup is charged)

We start with the condition for spherical micelles: In that case the volume $V_m$ and surface area $A_m$ of a micelle of radius $R$ are

$$V_m = \frac{4\pi}{3}R^3 = nv_0 \qquad A_m = 4\pi R^2 = na_0 \quad \text{(spherical micelle)} \tag{1.15}$$

From eq. 1.15 we may write $n$ in two ways as

$$n = \frac{4\pi}{3v_0}R^3 \;=\; \frac{4\pi}{a_0}R^2 \qquad \text{(spherical micelle)} \tag{1.16}$$

The above eq. is true iff

$$R = \frac{3v_0}{a_0} \quad \text{(spherical micelle)} \tag{1.17}$$

Since obviously, we also have that $R \leq \ell_0$ , we arrive at the geometrical condition for spherical micelles

$$\frac{v_0}{a_0\ell_0} \;<\; \frac{1}{3} \quad \text{(spherical micelle)} \tag{1.18}$$

The quantity $\frac{v_0}{a_0\ell_0}$ is often referred to as the surfactant parameter. The number of surfactant molecules in a micelle can now be estimated by $n = \frac{36\pi v_0^2}{a_0^3} \;=\; \frac{4\pi\ell_0^3}{3v_0}$. For the surfactant sodium dodecyl sulfate (SDS), the values of $\ell_0$ and $v_0$ are approximately 1.93 $nm$ and 0.4 $nm^3$, respectively [3], leading to $n \;\approx\; 75$ being in fair agreement with experiments.

Now lets investigate the condition for cylindrical micelles. For a cylinder of radius R and length L we have

$$V_m = \pi R^2 L = nv_0 \qquad A_m = 2\pi RL = na_0 \quad \text{(cylindrical micelle)} \tag{1.19}$$

By the same reasoning as in the case of spherical micelles we arrive at the condition $R = 2v_0/a_0$. Using again that $R \leq \ell_0$, we get $\frac{v_0}{a_0\ell_0} \;<\; \frac{1}{2}$. Combination with the condition for spherical micelles, eq. 1.18 finally gives

$$\frac{1}{3} \;<\; \frac{v_0}{a_0\ell_0} \;<\; \frac{1}{2}\text{(cylindrical micelle)} \tag{1.20}$$

Micelles also come in plate shapes; in that case we have, for plates with thickness $d$ and interfacial area $A$

$$V_m = Ad = nv_0 \qquad A_m = 2A = na_0 \quad \text{(plates)} \tag{1.21}$$

So in case of plates we get $d = 2v_0/a_0 < 2\ell_0$ , so that

$$\frac{1}{2} \;<\; \frac{v_0}{a_0\ell_0} \;<\; 1\text{(plates)} \tag{1.22}$$

Finally, it is easy to see that the condition for *inverse micelles* is that $\frac{v_0}{a_0\ell_0} \;>\; 1$.

## 1.6 Problems

1. Caloric experiments show that the cmc is almost independent of temperature (typically $\left|\frac{d\ln x_{\rm cmc}}{dT}\right| \approx 10^{-3}\,{\rm K}^{-1}$). Is micelle formation an enthalpically or entropically dominated process?
Hint: Make use of the relation between $\Delta\mu^\theta$ and $x_{cmc}$ as well as Tables 1.1 and 1.2. Use (and verify if it looks alien to you) the thermodynamic relation

$$\left(\frac{\partial \Delta\mu^\theta/T}{\partial T}\right)_p = -\frac{\Delta h^\theta}{T^2},$$

where $\Delta h^\theta$ is the molar enthalpy of micelle-formation.

2. Consider the formation of dimers from single molecules via

$$2A \rightleftharpoons A_2$$

Let the total mole fraction be $x$, the monomer mole fraction $x_1$, and the dimer mole fraction $x_2$.
a. Show that at equilibrium, $x_2 = x_1^2\ exp(-\Delta\mu^0/kT)$. Here, $\Delta\mu^0 = \mu_2^0 - 2\mu_1^0$, with $\mu_i^0$ the standard chemical potential of $i \in \{1,2\}$.
b. Show that for small $x$ (more specifically: $xe^{-\Delta\mu^0/kT} << 1$ ), $x_1 = x$. Also show that for large $x$ (more specifically: $xe^{-\Delta\mu^0/kT} >> 1$ ), $x_1 \propto \sqrt{x}$. Hint: use mass conservation.
c. How does the behavior under b. compare to the the formation of surfactant micelles?

3. The standard-chemical potential of a surfactant molecule in a micellar aggregate of size $n$ in principle depends on the occupied area per surfactant molecule, $a$. A primitive way to take that dependence into account is by writing the standard chemical potential per surfactant molecule in an aggregate of size $n$ as $\overline{\mu}_n^0 = \frac{\mu_n^0}{n} = \gamma a + \frac{k}{a}$.
a. Provide (a) possible interpretation(s) of the two terms in this equation.
b. Show that the optimal area per molecule is given by $a_0 = \sqrt{k/\gamma}$. Provide an interpretation of this results.
hint: consider the function $\overline{\mu}_n^0(a)$.
c. Show that $\overline{\mu}_n^0$ can be written as $\overline{\mu}_n^0 = 2\gamma a_0 + \frac{\gamma}{a}(a-a_0)^2$.
d. Show that the fraction of molecules with head group area $a$ relative to the optimal area $a_0$ is approximately given by

$$\frac{x_a}{x_{a_0}} = \exp\left(-\frac{\gamma}{a}(a-a_0)^2/kT\right). \tag{1.23}$$

What assumption(s) have you made?

# Chapter 2

# Reversible aggregation and adsorption

## 2.1 Introduction

In this chapter we will start with reversible aggregation of effectively single components in one dimension. One-dimensional aggregation underlies so-called 'living polymers', that is, one-dimensional aggregates of, e.g., surfactant molecules or taper-shaped crown ether derivates. Also, one-dimensional aggregation is relevant in biological stuctures such as rod-like viruses and microtubuli. We will first consider the situation for effectively single-component living polymers, which are shown to be very poly disperse. We will gradually increase complexity and move to the situation where a second component as a 'lattice' is involved. This adsorption problem will be tackled by using the grand ensemble. We will end up with allostery using as an example cooperative binding of oxygen to hemoglobin.

## 2.2 One-dimensional aggregation of single components: living polymers

Our first goal is to calculate the distribution $x_n$ of one-dimensional aggregates of aggregation number $n \geq 1$. Here $x_n$ refers to the mole fraction of aggregates with aggregation number $n$. The situation has schematically been depicted in Fig. 2.1. To calculate the size distribution, consider the equilibrium $nA \rightleftharpoons A_n$. The general

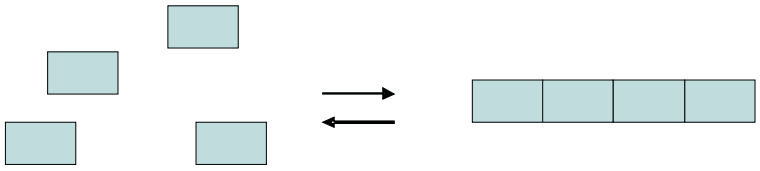


Figure 2.1: A special case of 4 monomers in equilibrium with a linear aggregate of size $n$=4.

thermodynamic condition for chemical equilibrium is

$$\sum_i \nu_i \mu_i = 0, \tag{2.1}$$

where $\nu_i$ are the stoichiometric coefficients and $\mu_i$ the chemical potentials. Applying eq. 2.1 to the equilibrium defined above we get

$$\mu_n = n\mu_1 \tag{2.2}$$

In this equation, the subscripts 1 and $n$ refer to monomers, and aggregates consisting of $n$ of those monomers, respectively. As long as monomers and aggregates are dilute, we may write the chemical potentials as

$$\mu_i = \mu_i^0 + kTlnx_i \tag{2.3}$$

where $\mu_i^0$ is the standard chemical potential ($\mu_i = \mu_i^0$ iff $x_i = 1$), $k$ and $T$ are Boltzmann's constant and absolute temperature, and $i$ in this case can be monomer ($i = 1$) or aggregate ($n$ - mer: $i = n$). Combining eqs. 2.2 and 2.3 leads to the distribution of aggregates of size $n$

$$x_n = x_1^n exp\left(\frac{-(\mu_n^0 - n\mu_1^0)}{kT}\right) \tag{2.4}$$

This result is independent of the dimensionality of the aggregates. In the case of one-dimensional aggregates, we take

$$\mu_n^0 - n\mu_1^0 = (n-1)w \tag{2.5}$$

where $w$ is the reversible work to create a bond, or, equivalently, the average free energy per bond, in an aggregate. The underlying assumption here is that the interaction between the building blocks in the one-dimensional aggregate is short-ranged, making the bond free energy $w$ independent of aggregate size $n$. Note that in general one would indeed have $w = w(n)$. To obtain the size distribution $x_n = x_1^n e^{-(n-1)w/kT}$ explicitly, we need the monomer concentration $x_1$ which can be obtained by imposing mass conservation, at least in principle. The total monomer concentration is given by

$$x = \sum_{n=1}^{\infty} n x_1^n e^{-(n-1)w/kT} = y \sum_{n=1}^{\infty} n z^n = \frac{yz}{(1-z)^2}, \tag{2.6}$$

where we have defined $y = e^{w/kT}$ and $z = x_1 e^{-w/kT}$. In the last step in Eq. (2.6) use has been made of the properties of the binomial series $\sum_{n=0}^{\infty} x^n = 1/(1-x)$ for $x \leq 1$, see the Problems section. Solving Eq. (2.6) for $x_1$ leads to

$$x_1 = \frac{1 + 2xe^{-w/kT} - \sqrt{1 + 4xe^{-w/kT}}}{2xe^{-2w/kT}}. \tag{2.7}$$

As long as $xe^{-w/kT} << 1$, that is, at small total monomer concentration and / or small value of $-w/kT$, we may expand the root in eq. (2.7) up to second

order and find $x_1 = x$. That is, under these conditions there are only monomers and no aggregates. In the other extreme, i.e., when $xe^{-w/kT} >> 1$, we have $x_1 \approx e^{w/kT}$. So in that case, adding monomer to the system will increase the fraction of aggregates and leave the concentration of free monomers constant. Thus, the quantity $x_0 = e^{w/kT}$ can be seen as the 'critical aggregate concentration', similar to the 'critical micelle concentration' for surfactants. The average size of the aggregates is given by

$$< n >= \frac{\sum_{n=1}^{\infty} n x_n}{\sum_{n=1}^{\infty} x_n} = z \frac{\partial \ln \left(\sum_{n=1}^{\infty} x_n\right)}{\partial z} = \frac{1}{1-z}. \tag{2.8}$$

It is left as an exercise to the reader to show that as long as $xe^{-w/kT} >> 1$, we have $< n >\approx \sqrt{x/x_0}$. Hint: write $x_1 = e^{w/kT}\left(1 - \frac{1}{\sqrt{xe^{-w/kT}}}\right)$. We define the polydispersity by

$$\sigma^2 =< n^2 > - < n >^2 = z \frac{\partial < n >}{\partial z} = \frac{z}{(1-z)^2} \tag{2.9}$$

From eqs. (2.8) and (2.9) it follows then that the relative polydispersity is given by

$$\frac{\sigma}{< n >} = \sqrt{z}. \tag{2.10}$$

Beyond the cmc, or when $x > e^{w/kT}$, it can be inferred from eq. (2.10) and the definition of $z$ that the relative polydispersity is always close to 1, independent of the interaction potential between monomers. The size distribution $x_n$ is continuously decreasing with $n$. The fraction of monomers in clusters of size $n$, $nx_n$, is peaked but the distribution is very broad.

## 2.3 Self-assembly on templates

In this section we consider adsorption of monomers on a template. For simplicity we will neglect interactions between monomers and only take into account monomer-template interactions, being effectively adsorption energies. While the treatment is general, i.e., applicable to any dimension, we will limit ourselves to 1-dimensional adsorption in order to make a connection with the previous section. The situation is schematically shown in Fig. (2.2). The template can be a polymer such as DNA

Figure 2.2: One-dimensional adsorption of $n = 6$ monomers on a template with $n_{max} = 12$ adsorption sites. This particular situation, that is, 6 filled sites on a total of 12, can be realized in 924 ways, according to Eq. (2.12).

or RNA, or any other polymer, and the monomers can be proteins or surfactant molecules, for example. Templates, monomers and aggregates are assumed sufficiently dilute so that the mixed solution behaves ideal. Similar to eq. (2.4), we may write down the multi- chemical equilibria which now involves the templates $P$ as $nA + P \rightleftharpoons A_nP$ and do a similar analysis as in the previous section. While that's a plausible route, there is a more straightforward one that we will pursue here: via the

grand ensemble, by good-old Gibbs [**?**]. The grand partition function of a template with $n_{max}$ adsorption sites, see again Fig. 2.2, reads

$$\Xi_1 = \sum_0^{n_{max}} \lambda^n Z(n, n_{max}, T) = \sum_0^{n_{max}} \binom{n_{max}}{n} \lambda^n e^{-\epsilon n/kT} = (1 + \lambda e^{-\epsilon/kT})^{n_{max}}, \quad (2.11)$$

with the adsorption (free) energy $\epsilon$, the fugacity $\lambda = e^{\mu/kT}$, and the canonical partition function for a lattice with uncorrelated adsorption $Z(n, n_{max}, T) = \binom{n_{max}}{n} e^{-\epsilon n/kT}$. The binomial coefficient

$$\binom{n_{max}}{n} = \frac{n_{max}!}{(n_{max} - n)!\, n!}, \quad (2.12)$$

takes into account the number of ways $n$ monomers can be adsorbed onto a template with $n_{max}$ adsorption sites, assuming adsorption is uncorrelated. The subscript '1' in $\Xi_1$ in eq. (2.11) refers to the single lattice that we are dealing with. If the lattices

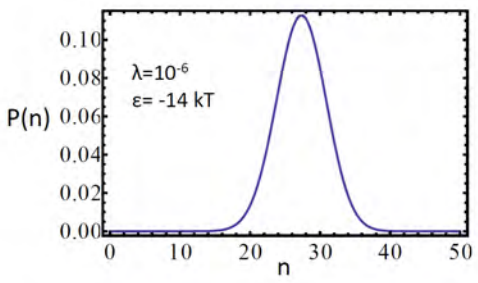



Figure 2.3: Typical size distribution $P(n)$ as defined below eq. (2.13) using the values of $\lambda$ and $\epsilon$ as indicated in the figure, and $n_{max} = 50$.

are independent, that is, if interaction between them can be ignored, we write the grand partition function of $N_p$ of these lattices as

$$\Xi = \Xi_1^{N_p} = (1 + \lambda e^{-\epsilon/kT})^{N_p n_{max}}. \quad (2.13)$$

It is easy to verify that a single lattice with $N_p \times n_{max}$ adsorption sites also leads to eq. (2.13): $\Xi = \sum_0^{N_p n_{max}} \binom{N_p n_{max}}{n} \lambda^n e^{-\epsilon n/kT} = (1 + \lambda e^{-\epsilon/kT})^{N_p n_{max}}$ . Thus, the way the adsorption sites are distributed over the system is not important in this model. The situation would of course be different if adsorbed species or lattices interact. Eq. (2.11) defines the size distribution $P(n) = \Xi_1^{-1} \binom{n_{max}}{n} \lambda^n e^{-\epsilon n/kT}$ which is plotted in Fig. 2.3. The factor $\binom{n_{max}}{n}$ in $P(n)$ has a maximum at $n^* = n_{max}/2$. The peak shifts to $n < n^*$ if $(\lambda e^{-\epsilon/kT}) < 1$ while $n > n^*$ if $(\lambda e^{-\epsilon/kT}) > 1$ Using the definition for the size distribution, the average number of adsorbed species on a lattice of size

$n_{max}$ is given by (**verify that**)

$$< n >= \lambda \frac{\partial \ln \Xi_1}{\partial \lambda} = n_{max} \frac{\lambda e^{-\epsilon/kT}}{1 + \lambda e^{-\epsilon/kT}}, \tag{2.14}$$

Eq. (2.14) is the Langmuir adsorption equation, which may be expressed as $\theta =$

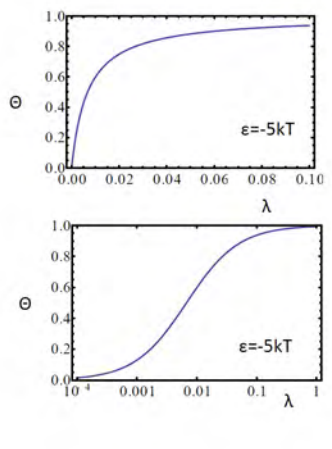



Figure 2.4: Adsorption isotherms $\theta =< n > /n_{max}$ as a function of $\lambda$ with linear and logarithmic scale.

$\frac{<n>}{n_{max}} = \frac{\lambda e^{-\epsilon/kT}}{1+\lambda e^{-\epsilon/kT}}$. As a measure for the polydispersity or number fluctuations we take the variance, given by

$$\sigma^2 =< n^2 > - < n >^2 = \lambda \frac{\partial < n >}{\partial \lambda} = n_{max} \frac{\lambda e^{-\epsilon/kT}}{(1 + \lambda e^{-\epsilon/kT})^2}. \tag{2.15}$$

Now it's time we identify the physical significance of $\lambda$. The chemical potential of the adsorbed species is equal to the chemical potential of the unbound species, $x_{1*}$ at equilibrium and reads $\mu = \mu_1^0 + kT \ln x_{1*}$, so that $\lambda = e^{\mu/kT} = x_{1*} e^{\mu_1^0/kT}$. Now in this case in all relevant thermodynamic properties, the combination $\lambda e^{-\epsilon/kT}$ appears, so we are entitled to 'rescale' the adsorption energy by making an explicit link to the reservoir of unbound species. In other words, there is no harm being done by writing $\lambda e^{-\epsilon/kT} = x_{1*} e^{-(\epsilon - \mu_1^0)/kT}$. Using that, we obtain an expression for $\lambda$ by applying mass conservation - after all, the appropriate role of $\lambda$ in the grand ensemble is a Lagrange multiplier coupled to the constraint that total number of molecules of adsorbing species, N, is conserved, that is, $N = N_{ads} + N_{1*}$, or,

$$N = N_p < n > + \frac{V}{v_s} x_{1*}, \tag{2.16}$$

where the number of unbound species has been written in terms of volume $V$, volume of a solvent molecule $v_s$ and mole fraction $x_{1*}$ as $N_{1*} = \frac{V}{v_s} x_{1*}$. Substitution of $x_{1*} = \lambda e^{-\mu_1^0/kT}$ and eq. (2.14) for $< n >$, leads to a quadratic equation in $\lambda$ that is easily solved to give the physically acceptable solution

$$\lambda = \frac{1}{2e^{-\epsilon/kT}} \left( e^{-\epsilon/kT}(x - x_P n_{max}) - 1 + \sqrt{h(x, x_P, \epsilon)} \right), \tag{2.17}$$

with $h(x, x_P, \epsilon) = 4xe^{-\epsilon/kT} + \left(1 + e^{-\epsilon/kT}(x_P n_{max} - x)\right)^2$, and where we defined the total mole fraction of adsorbing species as $x = Nv_s/V \approx N/N_s$ with $N_s$ the number of solvent molecules and the mole fraction of templates as $x_P = N_p v_s/V \approx N_p/N_s$. Clearly, eq. (2.17) provides an expression for $\lambda$ in terms of total concentrations of the components and adsorption (free) energy $\epsilon$. In view of the rescaling of the adsorption energy, the value of $\lambda$ given by eq. (2.17) is equal to the mole fraction of unbound monomers, $x_{1*}$, if we replace $\epsilon$ by $w = \epsilon - \mu_1^0$, that is,

$$x_{1*} = \frac{1}{2e^{-w/kT}} \left( e^{-w/kT}(x - x_P n_{max}) - 1 + \sqrt{h(x, x_P, w)} \right), \tag{2.18}$$

with $h(x, x_P, w) = 4xe^{-w/kT} + \left(1 + e^{-w/kT}(x_P n_{max} - x)\right)^2$. Eq. (2.18) may also be obtained using the (multi) chemical equilibrium approach, but the route taken here is more straightforward, and is easier to extend to, e.g., multi-component adsorption and to distributions of lattices with different adsorption energies and/ or lattice sites.

Finally, the careful reader may have noticed that the value of the re-scaled interaction energy, see the discussion below eq. (2.15), depends on the chosen concentration unit, here the mole fraction $x$. In fact, the choice of mole fraction is a consequence of the fact that as long as the interactions are short-ranged, the size of a solvent molecule is the relevant length scale. How does this length scale appear? The mole fraction can be written as $x_i = \frac{n_i}{n_i + n_s} \approx \frac{n_i}{n_s}$. As the number density $\rho_i = n_i/V \approx \frac{n_i}{n_s v_s}$, with $v_s$ the molecular volume of a solvent. Thus, under dilute conditions $x_i \approx \rho v_s = \rho\omega^3$. Here, $\omega$ is on the order of the size of a solvent molecule. This is in fact a 'coarse-grained' length scale: in our description the properties of the solvent molecules have been integrated out and do not appear explicitly. In a truly molecular description, it is the De Broglie wavelength $\Lambda$ that acts as the characteristic length scale - for example the expression for the chemical potential of an ideal gas reads $\mu = kT\ln(\Lambda^3\rho)$. The difference between this ideal-gas expression and Eq. (2.3) is an 'integration constant' $\mu^0$ and the length scale in the argument of the logarithm: $\Lambda$ for a molecular description, and $\omega$ for a coarse-grained one.

## 2.4 Allostery: oxygen binding to hemoglobin

In this section allosteric interactions are being addressed. In allostery, a substrate (usually a protein) can be in two or more states, where the states have different affinity to the molecules that bind onto the substrate. Allostery is quite common in biology. Besides oxygen binding to hemoglobin, as used here as an example, it occurs in ion pumps and genetic regulation [4]. Allostery is an example of 'collective' binding or adsorption: under appropriate conditions, binding of several molecules is easier than binding of only one or a few. That leads to steep adsorption isotherms - sometimes close to true phase transitions. In the case of oxygen binding to hemoglobin, several models have been developed with the aim to explain

the experimentally observed steep response to small difference in oxygen pressure. Yet the ultimate proof for an allosteric mechanism came from the observation that hemoglobin can be in two conformational states. Hemoglobin molecules consist of four binding sites for oxygen molecules. It is easy to see that in the simplest case, where each oxygen molecule binds to hemoglobin with energy $\epsilon$, the grand partition function is given by $\Xi = \sum_0^4 \binom{4}{n} \lambda^n e^{-\epsilon n/kT}$ so that the average fraction of occupied sites reads (**verify that**)

$$\Theta = \frac{< n >}{4} = \frac{\lambda e^{-\epsilon/kT}}{1 + \lambda e^{-\epsilon/kT}} \qquad \text{Langmuir adsorption equation.} \tag{2.19}$$

Here $\lambda = e^{\mu/kT}$ is related to the oxygen pressure $p$ via $\mu = \mu^0 + kT\ln(p/p^0)$ with $\mu^0 = \mu(p = p^0)$. Eq. (2.19) is the famous Langmuir adsorption equation. In fact we may use a similar rescaling of $\epsilon$ as in the previous section. Anyway, experimental data of $\Theta$ as a function of oxygen pressure cannot be fitted to the Langmuir adsorption equation, eq. (2.19) as can be seen in Fig. 2.5 : $\Theta$ increases much steeper with oxygen

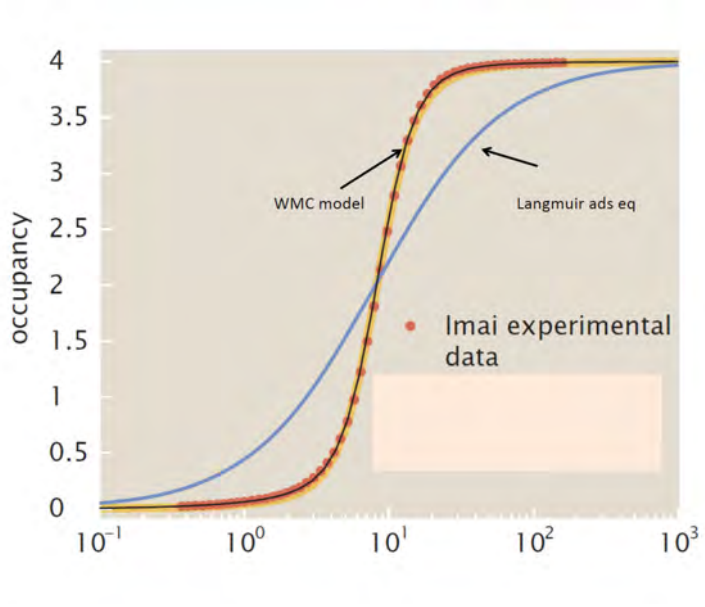



Figure 2.5: Average number of bound oxygen molecules per hemoglobin molecule ($4\Theta$) as a function of partial oxygen pressure. Points are experiments, lines best fits to eqs. (2.19, 2.23) indicated by 'Langmuir ads eq.' and 'MWC model', respectively.

pressure than prescribed by eq. (2.19). Such a much steeper effect than predicted by independent adsorption is generally referred to as a 'collective effect' - the onset of adsorption occurs at higher pressure compared to independent adsorption, but once it starts, it saturates much faster. It is as if oxygen molecules, once adsorbed with some reluctance, drag their friends in to join. While several models are able to quantitatively describe such 'collective' behavior, a model that is consistent with the structure of hemoglobin, pH dependence etc is the so-called MWC model, after Monod, Wyman and Changeux [5, 6], a two-state model for hemoglobin where each state has a different affinity (adsorption energy) for oxygen. The two states are referred to as 'Tense' (T) and 'Relaxed' (R), where T is the ground state of

hemoglobin, yet R has a higher affinity for oxygen. We write the grand partition function of the two state system as

$$\Xi = \Xi_T + \Xi_R, \tag{2.20}$$

where the subscripts refer to the Tense and Relaxed states. The grand partition functions of these states are given by

$$\Xi_T = \sum_0^4 \binom{4}{n} \lambda^n e^{-\epsilon_T n/kT} = \left(1 + \lambda e^{-\epsilon_T/kT}\right)^4, \tag{2.21}$$

and

$$\Xi_R = e^{-\epsilon/kT} \sum_0^4 \binom{4}{n} \lambda^n e^{-\epsilon_R n/kT} = e^{-\epsilon/kT} \left(1 + \lambda e^{-\epsilon_R/kT}\right)^4. \tag{2.22}$$

where $\epsilon$ refers to the self-energy of the R state (not to be confused with $\epsilon$ in eq. (2.19)) assumed unfavorable ($\epsilon > 0$) compared to the T state, at least as long as no oxygen molecules have been adsorbed. The fraction of hemoglobin sites occupied by oxygen is given by

$$\Theta = \frac{< n >}{4} = \frac{1}{4} \frac{\lambda}{\Xi} \frac{\partial \Xi}{\partial \lambda} =$$
$$= \Xi^{-1} \left[ \lambda e^{-\epsilon_T/kT} \left(1 + \lambda e^{-\epsilon_T/kT}\right)^3 + e^{-\epsilon/kT} \lambda e^{-\epsilon_R/kT} \left(1 + \lambda e^{-\epsilon_R/kT}\right)^3 \right]. \quad \text{MWC model} \tag{2.23}$$

This eq has been plotted in Fig. 2.5 along with eq. (2.19) and experimental data. The steep increase of $\Theta$ as a function of $\lambda$ - or as a function of the appropriate experimental quantity, here the partial oxygen pressure, is a consequence of small numbers of adsorbed oxygen molecules being unfavorable. In this case, small $<$ 3. This, in turn, is a consequence of the competition between an unfavorable R state and a favorable adsorption energy for oxygen molecules of that state. The self energy $\epsilon$ is a constant function of the number of adsorbed molecules, while the adsorption energy decreases linearly with the number of adsorbed oxygen molecules. This effective leads to an 'all or nothing' situation where at relatively small oxygen pressures the empty state dominates, followed by a switch to high occupation numbers upon increasing the oxygen pressure. Effectively, increasing adsorption pushes the hemoglobin molecule to an 'excited' state with unfavorable self energy. This scenario can be verified in Fig. (2.6) where we plotted the probabilities that an hemoglobin molecule contains 0, 1, 2, 3, 4 oxygen molecules. These probabilities, for the MWC model, follow from eqs. (2.20 - 2.22) and are given by

$$P(n) = \Xi^{-1} \left[ \binom{4}{n} \lambda^n e^{-\epsilon_T n/kT} + e^{-\epsilon/kT} \binom{4}{n} \lambda^n e^{-\epsilon_R n/kT} \right]. \quad \text{MWC model} \tag{2.24}$$

In case of independent (Langmuir) adsorption, as a comparison, these probabilities (see the grand partition function for Langmuir adsorption defined above eq. (2.19)) are given by

$$P(n) = \Xi^{-1} \binom{4}{n} \lambda^n e^{-\epsilon n/kT}. \quad \text{Langmuir model} \tag{2.25}$$

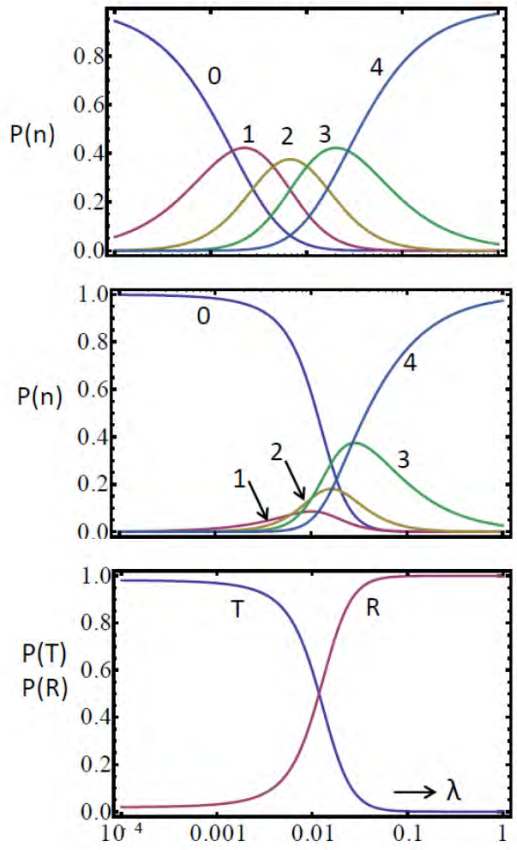



Figure 2.6: Probability distributions that hemoglobin molecules contain 0-4 oxygen molecules as indicated in the Fig., for the Langmuir model eq. (2.25)(Top) using $\epsilon = -5kT$, MWC model eq. (2.24) (middle). In the bottom Fig., the probabilities that the whole MWC object is in the T or the R state, $P(T)$ and $P(R)$, eq. (2.26) are plotted. In the bottom Figs we took $\epsilon_T = 0$, $\epsilon_R = -5kT$, $\epsilon = +4kT$.

The probability that the whole object in the MWC model is in the T or in the R state is plotted on the bottom Fig. 2.6 and is given by

$$P(T) = \Xi_T/\Xi \quad ; \quad P(R) = \Xi_R/\Xi \tag{2.26}$$

A pragmatic means to quantify collective adsorption is by the so-called Hill coefficient. In the language of eqs. (2.19, 2.23), the Hill coefficient $n$ is defined by

$$\Theta = \frac{k\lambda^n}{1 + k\lambda^n} \qquad \text{Hill cooperative binding.} \tag{2.27}$$

In this equation, the value of $n$ is a measure for the steepness or 'collectivity' of the adsorption isotherm: $n = 1$ reduces to the Langmuir adsorption equation, and $n > 1$ results in adsorption isotherms that are qualitatively similar to MWC isotherms.

## 2.5 Problems

1. Verify the last step in Eq. (2.6).

2. The purpose of this Problem is to illustrate that any association equilibrium has a critical concentration or density of the building blocks below which there are only monomers, and no larger structures. The concept 'cmc', as introduced and analyzed in the paragraphs 1.2 and 1.3, is a reflection of that more general property. You are being encouraged to read and digest that part carefully before diving into this problem.
a. Verify that under appropriate conditions Eq. (2.7) reduces to $x_1 = x$ (hint: expand the square root in a Taylor series). Also verify that the limiting value of $x_1$ is given by $x_1 = e^{w/kT}$.
b. The crossover between these limiting behaviors defines the cmc. In paragraph 2.2 we have been dealing with many different sizes $n$ of one-dimensional aggregates compared to the situation in paragraph 1.2 where we only considered two species: monomers and aggregates. Considering the assumptions in paragraph 2.2, argue whether it makes sense to take the *ansatz* for equilibria eq. (1.1) for any discrete $n$ as $K_n = \frac{x_n}{x_1^n} = x_0^{1-n}$.
c. Using the *ansatz* above show that the cmc associated with one-dimensional aggregation and in particular the 'decoupling' eq. (2.5) is given by $x_0 = e^{w/kT}$. Check the order of magnitude of the cmc as defined as such by using other arguments.
d. Show that the same arguments hold for the 'cmc' $x_0 = e^{w/kT}$ for the dimers in the last problem in the previous chapter.
e. Show that also in the situation of adsorption on a template there is a concentration $x$ below which $x_{1*} = x$. To make life easy(er), limit yourself to the situation where $x = x_P n_{max}$.
f. Calculate the 'cmc' of the association equilibrium $2A \rightleftharpoons A_2$ at 298 K where the 'bond energy' is comparable to that of hydrogen in water being 276 $kJ/mole$.

3. Verify Eqs. (2.8), (2.9), (2.14), (2.15), (2.19), (2.23), (2.24), (2.25).

4. The size distribution $x_n = x_1^n e^{-(n-1)w/kT}$ presumes 'ideal' behavior of monomers and aggregates. Argue that non-ideal behavior of monomers leads to larger deviations from the 'ideal' distribution as compared to aggregates. Discuss a 'way out' if strong deviations from ideal behavior occur.

5. The purpose of this exercise is to show that multi-chemical equilibrium and the grand canonical approach are equivalent.
a. Show that the grand potential of the one-dimensional aggregates in section 2.2 reads

$$\Xi = \frac{\lambda}{1 - \lambda e^{-\epsilon/kT}}, \tag{2.28}$$

where we have taken $\epsilon$ as the 'absolute' interaction energy that should be rescaled to a reference state to be provided by $\lambda$.
b. Show that the probability distribution of one - dimensional aggregates of size $n$ is given by

$$P(n) = \Xi^{-1}\lambda^n e^{-(n-1)\epsilon/kT}, \tag{2.29}$$

c. Show that the average aggregation number is given by

$$< n >= \frac{1}{1-\lambda e^{-\epsilon/kT}}. \tag{2.30}$$

Compare this result with eq. (2.8) and with the expression of $z$ in Section 2.2.

6. Derive the variance $\sigma^2 =< n^2 > - < n >^2$ of the number of oxygen molecules adsorbed onto hemoglobin for the MWC model and plot it using the same values as in Fig. 2.6. Compare your results with the variance that follows from a simple Langmuir adsorption model of oxygen binding to hemoglobin, see eqs. (2.19,2.15). You are encouraged to use Mathematica (or something similar) for the MWC model: while the derivation is straightforward, it is pretty tedious. Before embarking on the calculation, have a look at Fig. 2.6 and predict qualitatively (that is, without calculation, just by reasoning) the behavior of $\sigma^2$ as a function of $\lambda$ - would you expect it to be constant, decreasing, peaked, going through a minimum, increasing? Use (and sharpen) your physical intuition!

7. Show that for a MWC model generalized to N sites and P states, the average fraction of occupied sites is given by $\Theta = \lambda\Xi^{-1}\left(\sum_{p=1}^{P} e^{-(\epsilon_p^s+\epsilon_p)/kT}(1+\lambda e^{-\epsilon_p/kT})^{N-1}\right)$. Here $\epsilon_p^s$ is the self-energy of the whole object in state p, and $\epsilon_p$ the adsorption energy of a molecule onto an object in state p.

# Chapter 3

# Self-assembly of complex structures: virus capsids

## 3.1 Introduction

This chapter mainly consists of a copy of a paper (appended) on the self-assembly of viruses and includes a concise reading guide. This paper, reference [7], made it into a standard biophysics textbook: [4].

## 3.2 Reading guide to Equilibrium assembly of empty HBV capsid

Reference [7]. (see appendix)

Here the interactions between the building blocks of spherical virus capsids, the capsomers, are assumed to be dominated by hydrophobic interactions and electrostatic interactions. The total interaction free energy is the sum of those two. This idea is not too remote from DLVO theory, where the competing terms for colloidal aggregation are van der Waals interactions (drives coagulation) and electrostatic interactions (opposing coagulation by providing a barrier). In the situation for viruses, hydrophobic interactions provide the driving force for assembly, while electrostatic interactions weaken the driving force. Assembly - disassembly can in principle be regulated by the factors influencing the magnitudes of the hydrophobic and electrostatic forces. These factors are temperature, that couples to the hydrophobic interactions, pH, influencing the charge density and thereby the strength of the electrostatic interactions, and ionic strength, that decreases the Debye screening length and with that the magnitude of the electrostatic interactions as well. The hydrophobic interactions and the associated temperature dependence is discussed around eqs. (1) and (8) of the paper, as well as in the Problem (1) below. The electrostatics part is discussed around Eqs. (3) and (4) in ref. [7] and in detail in Problem (2) below. As a next step, viruses are assumes to be huge micelles to which Chapter 1 applies. The 'equilibrium constant', Eq. (11, 12) in ref. [7], that describes the equilibrium between capsomer proteins and complete virus capsids, is

related to the cmc as given by Eq. (1.11).
Note that in some cases different symbols have been used in ref. [7] as compared to the previous chapters in the handout. For example in ref. [7] the aggregation number is defined by $q$ rather than $n$, and $G$ defines $(\mu_n^0 - n\mu_1^0)$. Please carefully check the definitions in ref. [7].

## 3.3 Problems

1. A simple way of looking at so-called *hydrophobic interactions* is by considering objects with hydrophobic surfaces that float around in water. The surface tension between the objects and water is $\gamma$.

a. Prove that if two of these objects stick, the gain in free energy will be $w \approx -2\gamma\sigma_{\text{overlap}}$. Here, $\sigma_{\text{overlap}}$ is the contact area between the objects, i.e., the surface area shielded from water.

b. It is often found that the strength of the hydrophobic interactions *increases* with temperature, i.e., $w$ gets more negative upon increasing temperature. Assuming that $\sigma_{\text{overlap}}$ is independent of temperature, identify the thermodynamic property of the surface that is responsible for this remarkable temperature dependence. Hint: expand $\gamma$ around a reference temperature.

2. In this exercise the electrostatic contribution Eq. (4) in ref. [7] is derived stepwise from Eq. (3) in ref. [7]. Fig. 3.1 serves as a possible aid in setting up a coordinate system.

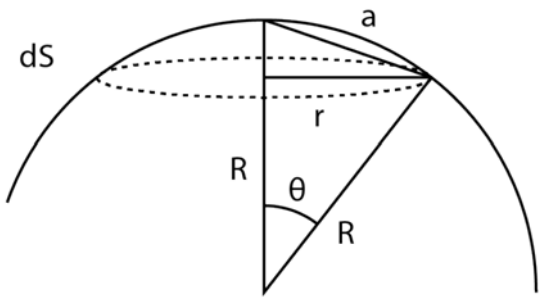



Figure 3.1: Sketch relating the charge separation $a$ and surface element $dS$ to polar coordinates $R$ and $\theta$

a. Write the position of the charges in Eq. (3) in ref. [7] as a scalar quantity and in polar coordinates, in other words show that $a = \mid \vec{r} - \vec{r}' \mid = 2R\sin(\theta/2)$, where $R$ is the radius of the virus capsid.
b. Argue that the total electrostatic interaction energy is given by

$G_C = \frac{kTq}{2} \int_S \rho \frac{z^2 \lambda_B}{a} e^{-\kappa a} dS$. Here $\rho = \frac{q}{4\pi R^2}$ is the surface density of capsomers, $z$ the number of net unit charges on a capsomer and $dS = 2\pi r dr$ the area of an infinitesimal surface element.
c. Write $dS$ in polar coordinates and perform the integration. Make use of the 'standard' integral (which can be verified by partial integration) $\int_0^\pi \frac{e^{-2\kappa R \sin\theta/2}}{\sin\frac{\theta}{2}} \sin\theta \cos\theta \mathrm{d}\theta = \frac{2e^{-2\kappa R}(1+\kappa R)(1+\kappa R+e^{2\kappa R}(\kappa R-1))}{\kappa^3 R^3}$ . Show that, making use of the assumptions discussed in ref. [7] that the result is indeed equal to Eq. (4) in [7].

3. Estimate the cmc of the HBV capsid at the lowest and highest ionic strength at $25^o$ C from Fig. 1 in ref. [7].

# Chapter 4

# Microemulsions

## 4.1 Introduction

Microemulsions are thermodynamically stable mixtures of oil and water. The stability is due to the presence of fairly large amounts (several %) of surfactants. Microemulsions are often transparent, but scattering of light, X-rays, etc. indicate that oil and water are not molecularly dispersed, but are more coarsely mixed. By coarse in this case we mean that oil and water are present in domains of a few to over a hundred nanometers in size. A schematic view of microemulsion droplets is provided by Fig. 4.1.

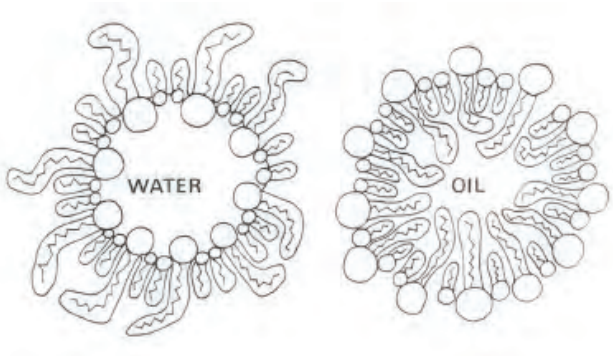



Figure 4.1: Schematic view of water droplets in oil (left) and oil droplets in water (right). In this case two types of surfactant molecules adsorb at the oil-water interface.

Microemulsions contain huge oil-water interfacial areas and to allow stability the interfacial tension must be quite low, usually $<< 1mN/m$. In that case the entropy of mixing, although small on account of the coarseness of the mixture, may be large enough to compensate the positive interfacial free energy and to give the microemulsion a free energy lower than that of the unmixed components. A rough estimate of the value of the interfacial tension, $\gamma$ where spontaneous emulsification occurs is by the condition $4\pi R^2\gamma = kT$, in other words the work to create a drop of radius $R$ should be on the order of the thermal energy $kT$.

Microemulsions can have various textures, such as oil droplets in water, water droplets in oil, (random) bicontinuous mixtures, ordered droplets or lamellar mixtures with a wide range of phase equilibria amongst them and with excess oil and/or

water phases. This great variety is governed by variations in the composition of the whole system and in the structure of the interfacial layers. The situation for relatively small surfactant concentrations and equal volumes of oil and water is shown in Fig. 4.2, showing the transition from Winsor I (oil droplets in water coexisting with excess oil) to Winsor III (bicontinuous structure with excess water and oil phases) to Winsor II (water droplets in oil with excess water).

Qualitatively the thermodynamics of microemulsions is well understood as the interplay between a small interfacial free energy and a small entropy of mixing. However, because of these contributions being small, other small effects, such as the influence of curvature on the interfacial tension, and the influence of fluctuations, become important. In the following we will derive the generalized Laplace equation to illustrate the consequence of curvature contributions to the (interfacial) free energy. Subsequently, we introduce the curvature free energy put forward by Helfrich. The curvature free energy is an important conceptual tool in understanding the physics of (surfactant) monolayers and membranes. It will be used here to explain the structural transitions of microemulsions as a function of temperature and ionic strength. In the last part, we will estimate the value of the interfacial tension of the flat interface between microemulsions and excess phase (see Fig. 4.2) and compare it to the situation without surfactant.

## 4.2 Experimental facts

The situation for relatively small surfactant concentrations and equal volumes of oil and water is shown in Fig. 4.2, showing the transition from Winsor I (oil droplets in water coexisting with excess oil) to Winsor III (bicontinuous structure with excess water and oil phases) to Winsor II (water droplets in oil with excess water).

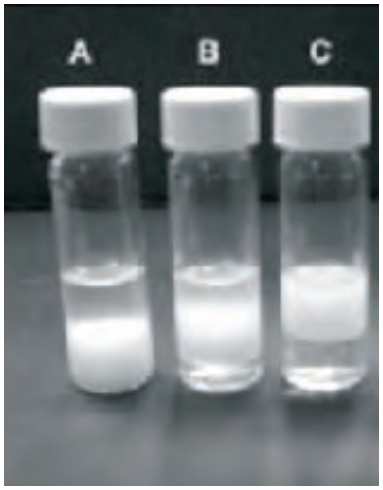



Figure 4.2: A: Oil droplets in water with excess oil (Winsor I); (B) bicontinuous with excess oil and water (Winsor III); (C) water droplets in oil with excess water (Winsor II). From A to C the ionic strength in the system is being increased

At higher surfactant concentrations, the excess water and oil phases are taken up by the microemulsion phase, ultimately leading to single-phase microemulsions.

Upon further increase of surfactant concentrations, excluded volume effects become important, leading to phase transitions to lamellar liquid crystals. The situation for non-ionic surfactants is sketched in Fig. 4.3.

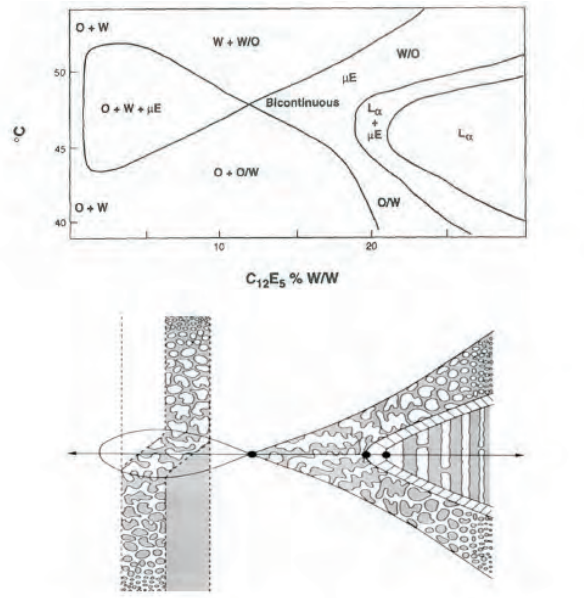



Figure 4.3: Experimental (Top) phase diagram of non-ionic surfactant at equal volumes of oil and water and varying temperature (veritical axis) and surfactant concentration (horizontal axis). An artistic impression of the situation has been shown at the bottom.

## 4.3 Generalized Laplace equation

Consider the formation of a liquid drop of radius $R$. It (hopefully) is well-known that inside the drop the pressure is higher than it is outside. The pressure difference, $\Delta p$, is referred to as the laplace pressure and is given by $\Delta p = 2\gamma/R$, with $\gamma$ the interfacial tension between the drop and its environment (e.g., gas, another liquid).

Now suppose interfacial tension depends on curvature, that is, the droplet radius. Mechanical equilibrium requires that the pressure difference $\Delta p$ times an infinitesimal change in volume of the drop, $dV$, equals the change in interacial free energy $d(\gamma A)$, with $A$ the interfacial area of the drop. In other words, $\Delta p dV = d(\gamma A) = \gamma dA + A d\gamma$. For a spherical drop this leads to

$$\Delta p = \frac{2\gamma}{R} - \frac{2c}{R^2} \tag{4.1}$$

In this equation, the bending moment

$$c = \frac{\partial \gamma}{\partial 2/R} \tag{4.2}$$

Eqs 4.1 is the generalized Laplace equation. Obviously it reduces to the classical Laplace equation if the absolute value of the second term is much smaller than the

first one, being the case if $c/R << \gamma$, implying small bending moment and/or large value of $R$.

## 4.4 Curvature free energy

Helfrich [8] deduced the free energy associated with deformations around a flat surface up to second order in the translational and rotational invariants. The result is

$$F_c = \int_A [\frac{\kappa}{2}(c_1 + c_2 - 2c_0)^2 + \bar{\kappa} c_1 c_2]\, dA. \tag{4.3}$$

In this eq., $c_1$ and $c_2$ are the principle curvatures, $c_0$ the preferred curvature, $\kappa$ the bending elastic modulus, and $\bar{\kappa}$ the modulus associated with Gaussian curvature. The last quantity often is referred to as 'Gaussian bending (elastic) modulus'. For a sphere of radius $R$ we have $c_1 = c_2 = 1/R$.

Eq. 4.3 often is written in terms of mean curvature $H = (c_1 + c_2)/2$ and Gaussian curvature $K = c_1 c_2$, i.e.,

$$F_c = \int_A [2\kappa(H - c_0)^2 + \bar{\kappa} K]\, dA. \tag{4.4}$$

The first term proportional to $\kappa$ in Eqs. 4.3, 4.4 is analogous to the expression of the potential energy of a harmonic spring. The second term is a topological invariant. Without the second term, it can easly be verified that the curvature free energy of, e.g., spheres and cylinders is degenerate. The Gauss-Bonnet theorem states that

$$\int_A K\, dA = 4\pi(1 - g) \tag{4.5}$$

In this eq., $g$ is the genus of a surface being defined by the number of holes. A sphere has $g = 0$, a cylinder and torus have $g = 1$, while torus-like objects with $N$ holes have $g = N$.

The Gaussian modulus $\bar{\kappa}$ can be smaller or larger than zero. Combination of the result Eq. 4.5 with Eq. 4.4 reveals that the sign of $\bar{\kappa}$ reflects the tendency of the surface to form certain topologies: if $\bar{\kappa} < 0$, the second term in Eq. 4.4 is minimal if $g$ is as small as possible, which corresponds to spheres. On the other hand, if $\bar{\kappa} > 0$, large $g$ will minimize the second term. Of course the second term competes with the first one in the curvature energy, and also with the configurational entropy of the objects.

Let's consider the situation where $\kappa, \bar{\kappa} >> kT$, so that the role of configurational entropy is negligible, and $c_0 = 0$. If $\bar{\kappa} = 0$, Eq. 4.4 implies that the curvature free energy of a flat object with $c_1 = c_2 = 0$ equals the curvature free energy of objects where at every point in space $c_1 = -c_2$. The last type of object has been sketched in Fig. 4.4. It is referred to as the 'Schwartz minimal surface' or the 'plumbers nightmare'.

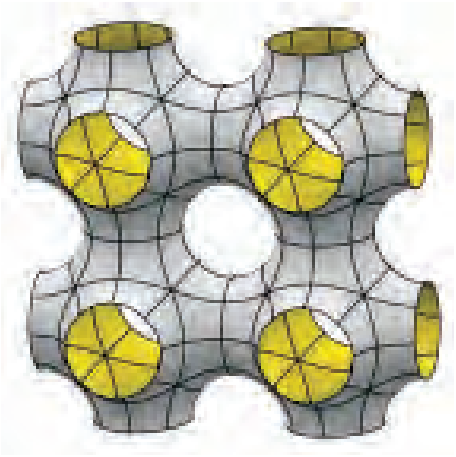


Figure 4.4: Schwartz minimal surface of 'plumbers nightmare'. The surface can be surfactant monolayers separating oil and water, or bilayers with water on both sides of the surface.

Now under the conditions described above, it is the sign of $\bar{\kappa}$ that determines whether plates are stable or the 'plumbers nightmare'. This is relevant for the structure of the 'bicontinuous' microemulsion, i.e., the middle phase in Fig. 4.2B and schematically in the 'head of the fish' in Fig. 4.3.

Away from the 'head of the fish', at small surfactant concentrations, oil droplets in water are stable or water droplets in oil, depending whether temperature is increased or decreased. It is generally accepted that in nonionic surfactants, preferred curvature depends on temperature. The microscopic reason is that temperature influences the level of hydration of the surfactant molecules on the water sides, thereby influencing the sign and absolute value of preferred curvature.

In case of ionic surfactants, the value (and sign) of $c_0$ is determined by the balance between excluded volume interactions between the surfactant chains on the oil side of the surfactant layer, and the electrostatic (screened- Coulomb) interactions between the charged 'heads' of the surfactant molecules at the water side of the surfactant monolayers. By increasing the ionic strength, as in Fig. 4.2, preferred curvature will be more and more towards the water side of the oil-water interface. The reason is that ions screen the electrostatic repulsion between the surfactant headgroups. The values of the bending elastic moduli depend on surfactant chain-length, charge density and ionic strength.

## 4.5 A microscopic model for curvature elasticity: incompressible spring model

In this section a simple microscopic model will be analyzed that will allow some physical insight into the meaning of the curvature elastic moduli. This part is based on chapter 6 in [9]. We model the monolayer of adsorbed surfactant molecules (Fig. 4.1) as springs with spring constant $k_s$ and equilibrium spring length $\ell_s$. The actual stressed or compressed spring length is denoted by $\ell$. The springs are supposed to be incompressible and assume a fixed area per chain at the interface equal

to $\Sigma_0$. In reality, this value is determined by properties of the polar head group. The constant value of $\Sigma_0$ implies that interactions that act on the polar head group are much stronger than the chain stretching energies. The (harmonic) energy per chain is

$$f = \frac{k_s}{2}(\ell - \ell_s)^2, \tag{4.6}$$

and the incompressibility of the chains implies a constant volume of the surfactant layer. In case of a flat layer we have $\Sigma_0 \ell = v_0$, with $v_0$ the molecular volume (volume occupied per chain). In case of a curved layer, it can be shown that

$$v_0 = \Sigma_0 \ell (1 + \ell H + \frac{\ell^2}{3} K), \tag{4.7}$$

We now impose the incompressibility condition: put Eq. 4.7 equal to the volume per chain in a flat layer, i.e., $v_0 = \Sigma_0 \ell_0$, with $\ell_0$ the layer thickness of a flat layer. Solving for $\ell$ and expanding to third order in $\ell_0$ leads to

$$\ell = \ell_0 - \ell_0^2 H + 2\ell_0^3 H^2 - \frac{l_0^3}{3} K, \tag{4.8}$$

Note that in general, in the flat monolayer the chain stretching energy Eq. 4.6 is not minimal as $\ell_0 \neq \ell_s$. Thus, in general, the flat layer will have a preferred curvature related to the imposed $\ell_0$ and the preferred $\ell_s$. Plugging Eq. 4.8 into Eq. 4.6 and keeping the lowest order terms leads to

$$f = \frac{k_s \ell_0^4}{2}[(H - c_0)^2 - \frac{2c_0 \ell_0}{3} K]. \tag{4.9}$$

In this Eq., we defined

$$c_0 = \frac{1}{\ell_0}(1 - \frac{\ell_s}{\ell_0}) = \frac{v_0 - \ell_s \Sigma_0}{\Sigma_0 \ell_0^2}. \tag{4.10}$$

Eq. 4.9 is equivalent to the Helfrich form of the curvature free energy, Eq. 4.4. The bending modulus (the coefficient of $H^2$ in eq. 4.9) and the Gaussian modulus (the coefficient of $K$) both increase as a power of the chain length. Obviously, the spring constant $k_s$ also depends on the equilibrium spring length $\ell_s$. In polymers, and in the limit of small curvatures we have $k_s \sim 1/\ell_s \approx 1/\ell_0$. In that case the bending modulus $\kappa \sim \ell_s^3$. The result that the bending modulus varies with the cube of the thickness also is characteristic for a bent solid elastic plate, see the textbook of Landau and Lifshitz on elasticity theory, ref. [10].

There is a simple physical interpretation of $c_0$ that emerges from the model, eq. 4.10. Any deviation of that quantity from zero arises because of a mismatch of the preferred length $\ell_s$ and the imposed lenght $\ell_0$. The imposed length, in turn, is set by the imposed head area $\Sigma_0$ as compared to the 'optimal' area $v_0/\ell_s$. If $\Sigma_0 > v_0/\ell_s$, preferred curvature is negative and the system prefers to pack with the heads on the 'outside'. The free energy in that case is lower than that of the flat interface: the system accomodates part of the strain induced by the mismatch between the heads and the chains by bending.

## 4.6 Interfacial tension between microemulsion- and excess phase

In this section we calculate the interfacial tension, $\gamma_\infty$, of the macroscopic interface between a droplet-type microemulsion and excess oil or water phase, see Fig. 4.2. Obviously, that quantity is related to the tension of the droplet (of radius R) interface, $\gamma_R$. It should also contain the work related to 'un-bending' the drop into a flat layer. In particular,

$$\gamma_\infty = \gamma_R + \int_{2/R}^{0} \frac{\partial \gamma}{\partial (2/R)}\, d(2/R). \tag{4.11}$$

Combination of Eqs. 4.2 and 4.3 leads to

$$\frac{\partial \gamma}{\partial (2/R)} = c = \frac{\partial^2 F_c}{\partial A \partial (2/R)} = \frac{1}{R}(2\kappa + \bar{\kappa}) - 2\kappa c_0. \tag{4.12}$$

The radius that minimizes the bending energy is

$$R = R_0\ (1 + \frac{\bar{\kappa}}{2\kappa}), \tag{4.13}$$

with $R_0 = 1/c_0$. Here it has been assumed that curvature contributions dominate the total free energy of the system. That only is true if $(2\kappa + \bar{\kappa})\ >>\ kT$. In practice, the values of the bending moduli are on the order of $kT$, and entropy effects are significant. In case of dilute systems, this leads to contibutions to the free energy that are logarithmic in the number density of the microemulsion droplet. For a discussion see Ref. [11] and references therein. However, these contributions only lead to corrections to eq. 4.13 (and the ones that follow) that are logarithmic in the number density of the microemulsion droplets.

Carrying out the intergration eq. 4.11 and using eqs. 4.12, 4.13 leads to

$$\gamma_\infty = \gamma_R + \frac{2\kappa}{R_0 R} = \gamma_R + \frac{2\kappa + \bar{\kappa}}{R^2} \tag{4.14}$$

In general, it is expected that $\gamma_\infty\ >>\ \gamma_R$ so that $\gamma_\infty\ \sim\ R^{-2}$. This is indeed what is to be expected, see the brief discussion on the condition for spontaneous emulsification in the Introduction. In case of liquid-gas interfaces, the interfacial tension $\gamma\ \sim\ 1/d^2$ with $d$ a molecular diameter. Interestingly, for microemulsion the size of the emulsified objects sets the interfacial tension.

## 4.7 Problems

1. Show that for a sphere of radius R, the curvature free energy reads $4\pi(2\kappa + \bar{\kappa}) - 16\pi\kappa c_0 R + 8\pi\kappa c_0^2 R^2$.

2. Prove Eq. 4.13.

3. Verify Eq. 4.5 for a sphere and for a cylinder.

4. What sign of $\bar{\kappa}$ stabilizes the 'plumbers nightmare', Fig. 4.4?

5. Verify Eq. 4.7 by considering the volumes of a spherical and a cylindrical layer.

6. Derive Eq. 4.8 for a sphere of radius R. First verify that the (real) root of the incompressibility condition on Eq. 4.7 for a sphere is given by $\ell \to R\,[(3\ell_0/R{+}1)^{1/3}{-}1]$

# Chapter 5

# Solutions to Problems

**Chapter 1**

1. Using the relation between $x_{\rm cmc}$ and $\Delta\mu^\theta$, Eq. 1.11 with Eq. 1.13, as well as the given thermodynamic relation, we find $\Delta h^\theta = -RT^2\left(\frac{\partial \ln x_{\rm cmc}}{\partial T}\right) \sim 1$ kJ/mol. Moreover, from Table 1.1 and Table 1.2 we find $\Delta\mu^\theta \sim -10$ kJ/mol. Realizing that $\Delta\mu^\theta = \Delta h^\theta - T\Delta s^\theta$, it follows that micelle formation is largely driven by entropy.

2a. See Section 1.3, Eq. 2.4 with $n = 2$.
2b. Use mass conservation $x = x_1 + 2x_2$ and substitute $x_2 = x_1^2\ exp(-\Delta\mu^0/kT) = zx_1^2$ with $z = exp(-\Delta\mu^0/kT)$. We then get $2zx_1^2 + x_1 - x = 0$. The physical root is $x_1 = \frac{-1+(1+8zx)^{1/2}}{4z}$. If $zx << 1$ the root can be expanded as $(1+y)^{1/2} = 1 + \frac{y}{2}$ and we obtain $x_1 = x$. In the limit $zx >> 1$, we have $x_1 = \left(\frac{x}{2z}\right)^{\frac{1}{2}} \sim \sqrt{x}$.
2c. In micelles with $n >> 1$, at the cmc we have a crossover from $x_1 = x$ to $x_1 \approx$ constant. Here we have a crossover from $x_1 = x$ to $x_1 \sim \sqrt{x}$. So, even in the case of the simplest possible association equilibrium, i.e., the formation of dimers, there is a critical concentration below which there are only monomers.

3. a.The first term takes into account hydrophobic interactions between chain parts of the surfactants and water (interfacial tension times exposed area). This term tends to minimize the distance between surfactant molecules. The second term is an approximation for excluded volume interactions (via a second virial approximation), electrostatic interactions, steric repulsion etc. In short: all terms that tend to maximize the average distance between surfactant molecules.
b.Take the first derivative with respect to a and solve, i.e.,solve $\frac{\partial \overline{\mu}_n^0}{\partial a} = 0 = \gamma - k/a_0^2$. Interpretation: interfacial tension tends to minimize $a_0$, repulsion quantified by $k$ tends to maximize $a_0$.
c. Make use of $a_0 = \sqrt{k/\gamma}$ to write the the chemical potential of a surfactant molecule at optimum head group area as $\overline{\mu}_n^0 = 2\gamma a_0$. Substitute $k = \gamma a_0^2$ in $\overline{\mu}_n^0 = \gamma a + \frac{k}{a}$ and rewrite into $\overline{\mu}_n^0 = \gamma a + \frac{\gamma a_0^2}{a}$. QED.
d. There are several ways to do this. A strategy closest to the micelle approach is to put the chemical potentials of surfactants of areas $a$ and $a_0$ equal and solve for the density. We approximate the chemical potentials by writing

$$\mu(a) = \overline{\mu}^0(a) + kT\ln x_a$$

$$\mu(a_0) = \overline{\mu}^0(a_0) + kT\ln x_{a_0}$$

Where we dropped the subscript $n$. The main approximation is the assumption of ideal solution behavior of surfactant molecules in an aggregate. However, by imposing equilibrium and putting the chemical potentials equal, non-ideal terms will compensate and thus, putting $\mu(a) = \mu(a_0)$ and substitution of $\overline{\mu}_n^0 = 2\gamma a_0 + \frac{\gamma}{a}(a - a_0)^2$ as well as $\overline{\mu}^0(a_0) = 2\gamma a_0$, we indeed obtain Eq. (1.23).
An alternative is to calculate the fluctuations using the second derivative which should lead to the same result.

**Chapter 2**
1. Make use of the binomial series $\sum_{n=0}^{\infty} x^n = 1/(1-x)$ for $x \leq 1$. Realize that $\sum_{n=0}^{\infty} x^n = 1 + \sum_{n=1}^{\infty} x^n$, and $\sum_{n=0}^{\infty} nx^n = x\frac{d}{dx}\sum_{n=0}^{\infty} x^n$.
2a. Take $xe^{-w/kT} << 1$ and expand the root via $\sqrt{1+x} \approx 1 + \frac{x}{2} - \frac{x^2}{8}$. Only retaining the linear term will not do here.
2b. It makes sense as long as $w(n)$ is strictly pairwise additive, which in 1 dimension is analogous to writing $w(n) = (n-1)w$. Because of that, and only because of that, we have $K_n = \frac{x_n}{x_1^n} = (e^{-w/kT})^{(n-1)} = (e^{w/kT})^{(1-n)}$ and we identify the 'constant' $x_0 = e^{w/kT}$.
2c. The argument goes in the same way as in paragraph 1.2: write mass conservation Eq. (2.6) in reduced variables so that $y = \sum_1^{\infty} n y_1^n$ with $y = x/x_0$ and $y_1 = x_1/x_0$. The crossover from $y_1 = y$ to $y_1 \approx 1$ occurs around $y = 1$ and thus $x = x_0$. Another, less general, argument that should at least provide an order of magnitude estimate of $x_0$ is the following: take the situation that the monomer concentration is equal to the dimer concentration: $x_1 = x_2 = x_1^2 e^{-w/kT}$ which immediately leads to $x_1 = e^{w/kT}$. This indeed is an order of magnitude estimate for the value of the cmc, as we could as well have taken $x_1 = 2x_2$ as a measure for the crossover. That will lead to a factor two difference. Still the right order of magnitude, though!
2d. That's an easy one now, isn't it?
2e. The situation simplifies considerably with $x - x_P n_{max} = 0$. Expand the root in Eq. (2.18) to linear order and one immediately finds $x_{1*} = x$.
2f. The cmc is $e^{-276/2.48} = 4.17 \times 10^{-49}$ mole / mole $\equiv 2.315 \times 10^{-47}$ M which corresponds to roughly 1 monomer per $7 \times 10^{10}$ km$^3$. That's a low concentration, very low indeed.

3. Hints are in the text.

4. Non-ideal behavior can be taken into account by replacing the mole fractions $x_n$ by the activity (or fugacity) $\lambda_n = f_n x_n$. The coefficients $f_n$ are the 'activity coefficients' which depend on $n$, the composition of the system and temperature, at least in principle. Clearly the influence of non-ideality is strongest for monomers,

as we now have $\lambda_n = \lambda_1^n e^{-(n-1)w/kT}$ and the deviation from ideal solution behavior appears as $f_1^n$ in the size distribution, i.e., small deviations of $f_1$ from unity lead to large deviations in the size distribution due to $f_1^n$, for example if $f_1 = 0.99$ and $n = 200$ then $f_1^n \approx 0.13$. A 'way out' is that the activity can be seen as a Lagrange multiplier coupled to mass conservation. So Eq. (2.6) can still be used, with $x_i \to \lambda_i$, but the connection with a physical meaning of $\lambda_1$ being monomer concentration is lost. Interpretation of that value requires a model for the interactions. Note the analogy with the activity $e^{\mu/kT}$ in the grand ensemble.

5.
a. The grand partition function of one-dimensional aggregates reads
$\Xi = \sum_1^\infty \lambda^n e^{-(n-1)\epsilon/kT} = \frac{\lambda}{1-\lambda e^{-\epsilon/kT}}$.
b. The properly normalized probability distribution is $P(n) = \frac{\lambda^n e^{-(n-1)\epsilon/kT}}{\sum_1^\infty \lambda^n e^{-(n-1)\epsilon/kT}}$
$= \Xi^{-1}\lambda^n e^{-(n-1)\epsilon/kT}$.
c. Use similar properties of binomial series as in Section 2.2. The average size follows from $< n >= \lambda\frac{\partial \ln \Xi}{\partial \lambda}$. Taking $\lambda = e^{\mu/kT}$ and $\mu = \mu_1^0 + kT \ln x_1$ leads to $\lambda = x_1 e^{-\mu_1^0/kT}$. Clearly here $z = x_1 e^{-w/kT} = \lambda e^{-\epsilon\ kT} = x_1 e^{-(\epsilon-\mu_1^0)/kT}$

6. The variance for the Langmuir model for oxygen binding follows from Eq. (2.19) : $\sigma^2 = \lambda\frac{\partial <n>}{\partial \lambda} = \frac{\lambda e^{-\epsilon/kT}}{(1+\lambda e^{-\epsilon/kT})^2}$. Which is just eq. (2.15) as it should. The variance for the MWC model follows in a similar way, that is, via $\sigma^2 = \lambda\frac{\partial <n>}{\partial \lambda}$, but now $< n >$ is given via eq. (2.23) and it's quite a bit of work to differentiate this function. The differentiation is easily done in Mathematica - see the Mathematica notebook *MWC variance.nb* in Blackboard. You are encouraged to play with the numbers and extend to compute more properties.
*A-priori* in both (Langmuir and MWC) situations one expects a peak in $\sigma^2$ as a function of $\lambda$, (see Fig. 2.6): at small values of $\lambda$, no sites are occupied and thus there are no fluctuations. In the limit of large $\lambda$, all sites are occupied and there are again no fluctuations. At intermediate values of $\lambda$, sites are partly occupied so that $\sigma^2$ should be larger compared to both small and large values of $\lambda$. The peak is expected where the spread in occupation number is largest, which is roughly around $\lambda \approx 0.01$ in the upper two Fig. (2.6). Also the peak is expected to be smaller for the MWC scenario since there, the range of $\lambda$ values where several occupation numbers occur is significantly smaller than in the Langmuir scenario. This is indeed the case as can be verified in the Mathematica notebook.

7. In the case of P states and N sites, the generalization of eqs. (2.20 - 2.22) reads $\Xi = \sum_{p=1}^P \left[e^{-\epsilon_p^s/kT} \sum_{n=0}^N \binom{N}{n} \lambda^n e^{-\epsilon_p n/kT}\right]$. Carrying out the summation over the occupation numbers leads to $\Xi = \sum_{p=1}^P e^{-\epsilon_p^s/kT}\left(1 + \lambda e^{-\epsilon_p n/kT}\right)^N$. From there we get the average fraction of occupied sites:

$\Theta = \frac{<n>}{N} = \frac{1}{N}\frac{\lambda}{\Xi}\frac{\partial \Xi}{\partial \lambda} = \lambda \Xi^{-1}\left(\sum_{p=1}^{P} e^{-(\epsilon_p^s+\epsilon_p)/kT}(1+\lambda e^{-\epsilon_p/kT})^{N-1}\right)$.

**Chapter 3** 1a. If the two objects are separated, their surface free energy reads $A_I^\sigma = 2\gamma\sigma$ with $\sigma$ the surface area of one of the objects. If the objects are merged, we have $A_{II}^\sigma = 2\gamma(\sigma - \sigma_{overlap})$. By merging, the gain in free energy is the difference in surface free energy between the situation where the objects are merged and where they are separated, that is $w \equiv A_{II}^\sigma - A_I^\sigma = -2\gamma\sigma_{overlap}$.
b. Write $\gamma$ as a function of temperature as

$$\gamma(T) = \gamma(T_0) + \left(\frac{\partial \gamma}{\partial T}\right)_{T=T_0}(T - T_0). \tag{5.1}$$

Here, $T_0$ is a reference temperature. The temperature coefficient is the negative surface excess entropy per unit area, i.e., $s^\sigma = -\left(\frac{\partial \gamma}{\partial T}\right)_{T=T_0}$. This is the quantity responsible for the temperature dependence on the strength of the hydrophobic interactions, at least to first order. Clearly, $w(T)$ becomes more negative with temperature if $\gamma(T)$ rises with temperature, being the case iff $s^\sigma < 0$. Thus, the remarkable temperature dependence is caused by negative surface excess entropy.

2a. This is straightforward geometry based on Fig. 3.1.
2b. The interactions between a single unit charge with all other charges at the capsid surface is $kT\int_S \rho\frac{z\lambda_B}{a}e^{-\kappa a}dS$. This interaction is multiplied by $qz$ to get the total interaction energy, where we also have to divide by 2 to compensate for double counting. Thus we get $G_\mathrm{c} = \frac{kTq}{2}\int_S \rho\frac{z^2\lambda_\mathrm{B}}{a}e^{-\kappa a}\mathrm{d}S$. The interaction energy is equivalent to summing the interactions over all separations $a$ on the spherical surface of radius $R$.
2c. Substituting $a = 2R\sin\frac{\theta}{2}$ and $\mathrm{d}S = 2\pi r\mathrm{d}r = 2\pi R^2\sin\theta\cos\theta\mathrm{d}\theta$ leads to

$$G_\mathrm{c} = \frac{kTq}{2}\frac{\rho z^2\lambda_\mathrm{B}}{2R}2\pi R^2\int_0^\pi \frac{e^{-2\kappa R\sin\theta/2}}{\sin\frac{\theta}{2}}\sin\theta\cos\theta\mathrm{d}\theta, \tag{5.2}$$

$$= \frac{kTq}{2}\frac{\rho z^2\lambda_\mathrm{B}}{2R}2\pi R^2 \quad \frac{2e^{-2\kappa R}(1+\kappa R)(1+\kappa R+e^{2\kappa R}(\kappa R-1))}{\kappa^3R^3}, \tag{5.3}$$

where in the second step we integrated over the entire sphere, and use has been made of the standard integral provided. Expanding the fraction on the right side we get

$$\frac{2e^{-2\kappa R}(1+\kappa R)(1+\kappa R+e^{2\kappa R}(\kappa R-1))}{\kappa^3R^3} = \frac{-2}{\kappa^3R^3} + \frac{2e^{-2\kappa R}}{\kappa^3R^3} + \frac{4e^{-2\kappa R}}{\kappa^2R^2} + \frac{2}{\kappa R} + \frac{2e^{-2\kappa R}}{\kappa R}. \tag{5.4}$$

For $\kappa R > 1$ only the final terms will matter. We then get

$$G_\mathrm{c} = \frac{kTq^2z^2\lambda_\mathrm{B}}{8R}\left[\frac{-2}{\kappa^3R^3} + \frac{2e^{-2\kappa R}}{\kappa^3R^3} + \frac{4e^{-2\kappa R}}{\kappa^2R^2} + \frac{2}{\kappa R} + \frac{2e^{-2\kappa R}}{\kappa R}\right]. \tag{5.5}$$

$$\approx \frac{kTq^2z^2\lambda_\mathrm{B}}{4R^2}\frac{1+e^{-2\kappa R}}{\kappa} \tag{5.6}$$

which for $\kappa R \gg 1$ will reduce to

$$\approx \frac{kTq^2z^2\lambda_B}{4R^2}\kappa^{-1} \tag{5.7}$$

And so, finally, we end up with the same expression as in the original paper.

3. From chapter (1) it follows that the cmc is given by $\ln x_0 = \frac{1}{1-q}\ln K$. Estimating from Fig. 1 $K \approx 1800$ at an ionic strength of 0.15 M and $K \approx 2150$ at 0.70 M, we have $x_0 \approx 2.7 \times 10^{-7}$ mole /mole and $x_0 \approx 1.4 \times 10^{-8}$ mole / mole at the lowest and highest ionic strength. This translates into $1.5 \times 10^{-5}$ M and $7.9 \times 10^{-7}$ M, respectively. So in virus capsids there is typically a coexisting monomer density on the order of micro molar.

**Chapter 4**

1. For a uniform sphere of radius $R$, Eq. 4.3 reads $F_c = 4\pi R^2[2\kappa((1/R)-c_0)^2+\bar{\kappa}/R^2]$. Subsequently rearrange.

2. Minimize Eq. 4.3 or 4.4 PER UNIT AREA, i.e., find the extremum of $2\kappa((1/R)-c_0)^2 + \bar{\kappa}/R^2$. This reflects the condition that in the system, total interface is conserved, but volume may vary.

3. Sphere: $K = 1/R^2$ and $\int_A K\, dA = 4\pi$; cylinder $K = 0$ and $\int_A K\, dA = 0$.

4. The sign that leads to negative values of $\bar{\kappa}(1-g)$ - compare Eq. 4.4. The plumbers nightmare has $g >> 1$, so that $\bar{\kappa} > 0$ stabilizes that topology.

5. The volume of a spherical shell is

$$V = \frac{4\pi}{3}(R+\ell)^3 - \frac{4\pi}{3}R^3 = 4\pi R^2\ell\left(1 + \frac{\ell}{R} + \frac{\ell^2}{3R^2}\right), \tag{5.8}$$

so that

$$v_0 = \frac{V}{N} = \frac{4\pi R^2\ell}{N}\left(1 + \frac{\ell}{R} + \frac{\ell^2}{3R^2}\right) = \Sigma_0\ell\left(1 + \frac{\ell}{R} + \frac{\ell^2}{3R^2}\right) = \Sigma_0\ell\left(1 + \ell H + \frac{\ell^2 K}{3}\right). \tag{5.9}$$

For a cylinder we have $V = \pi(R+\ell)^2L - \pi R^2L$ leading to $v_0 = \Sigma_0\ell(1 + \frac{\ell}{2R})$. Note that for a cylinder we indeed have $K = 0$.

6. Plugging eq. 5.9 into the incompressibility condition $v_0 = \Sigma_0\ell_0$ leads to $\frac{\ell}{\ell_0}\left(1 + \frac{\ell}{R} + \frac{\ell^2}{3R^2}\right) - 1 = 0$. The physical (that is, real) root is $\ell \to R\,[(3\ell_0/R+1)^{1/3} - 1]$. Expansion in $x = \frac{3\ell_0}{R}$ as the small parameter, i.e., $(1+x)^{1/3} = 1 + \frac{x}{3} - \frac{x^2}{9} + \frac{5x^3}{81}$ leads to Eq. 4.8 for a sphere, i.e., $\ell = \ell_0 - \frac{\ell_0}{R} + \frac{5\ell_0}{3R^2}$.

# Chapter 6

# Appendix: copy of virus capsid paper

# Bibliography


[1] P.J.W Debye. *Ann. NY Acad. Sci.*, 51:575, 1949.

[2] H. B. Klevens. *J. Am. Oil Chem. Soc*, 30:74, 1953.

[3] J. Israelachvili. *Intermolecular and surface forces.* Academic Press, San Diego, 2 edition, 1992.

[4] Rob Phillips, Jané Kondev, Julie A. Theriot, and Hernan G Garcia. *Physical Biology of the Cell.* Garland Science, NY, 2 edition, 2013.

[5] J Monod, J Wyman, and J P Changeux. on the Nature of Allosteric Transitions: a Plausible Model. *Journal of molecular biology*, 12(December):88–118, may 1965.

[6] Jean-Pierre Changeux. Allostery and the Monod-Wyman-Changeux model after 50 years. *Annual review of biophysics*, 41:103–33, jan 2012.

[7] Willem K Kegel and Paul Van Der Schoot Pv. Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly. *Biophysical journal*, 86(6):3905–13, jun 2004.

[8] W. Helfrich. *Z. Naturforsch.*, 28c:693, 1973.

[9] S.A. Safran. *Statistical thermodynamics of surfaces, interfaces, and membranes.* Perseus, Cambridge MS, 1994.

[10] L.D. Landau and E.M. Lifshitz. *Theory of Elasticity.* Pergamon, NY, 2 edition, 1970.

[11] W.K. Kegel, J. Th. G. Overbeek, and H N W Lekkerkerker. Thermodynamics of Microemulsions. In *Microemulsions, Fundamental and Applied Aspects*, page 13. Marces Dekker, NY, 1999.

# Competing Hydrophobic and Screened-Coulomb Interactions in Hepatitis B Virus Capsid Assembly

Willem K. Kegel and Paul van der Schoot
Van't Hoff Laboratory for Physical and Colloid Chemistry, Debye Research Institute, Utrecht University, Utrecht, The Netherlands, and Eindhoven Polymer Laboratories, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT Recent experiments show that, in the range from ~15 to 45°C, an increase in the temperature promotes the spontaneous assembly into capsids of the *Escherichia coli*-expressed coat proteins of hepatitis B virus. Within that temperature interval, an increase in ionic strength up to five times that of standard physiological conditions also acts to promote capsid assembly. To explain both observations we propose an interaction of mean force between the protein subunits that is the sum of an attractive hydrophobic interaction, driving the self-assembly, and a repulsive electrostatic interaction, opposing the self-assembly. We find that the binding strength of the capsid subunits increases with temperature virtually independently of the ionic strength, and that, at fixed temperature, the binding strength increases with the square root of ionic strength. Both predictions are in quantitative agreement with experiment. We point out the similarities of capsid assembly in general and the micellization of surfactants. Finally we make plausible that electrostatic repulsion between the native core subunits of a large class of virus suppresses the formation in vivo of empty virus capsids, that is, without the presence of the charge-neutralizing nucleic acid.

## INTRODUCTION

For many kinds of virus, the self-assembly of the empty virus core shells (or capsids) from their protein subunits resembles that of surfactant micelles (Debye, 1949), in the sense that one either finds only the protein subunits present in the aqueous solution, or the protein subunits in equilibrium with the fully assembled capsids (Tanford, 1980). For both capsids and micelles, the transition between the assembled and disassembled states is a fairly sharp function of the concentration of dispersed material (see e.g., Bruinsma et al., 2003; Caspar, 1963, 1980; Zlotnick, 1994).

Building on the idea that capsid assembly is subject to the laws of equilibrium thermodynamics, Ceres and Zlotnick (2002) recently calculated the free energy of interaction between *Escherichia coli*-expressed hepatitis B virus coat proteins assembled into predominantly $T = 4$ type capsids by fitting an association equilibrium model to their experimental data. These coat proteins, denoted $Cp149_2$, form capsids indistinguishable from the native HBcAg core protein subunits (Zlotnick et al., 1996). At near neutral pH, they found the strength of the attractive interaction between the homodimeric $Cp149_2$ proteins to be quite modest, of the order of a few times the thermal energy per subunit contact. The precise value of the free energy of interaction was found to increase with increasing temperature and with increasing ionic strength, at least within the ranges probed (Ceres and Zlotnick, 2002). The strengthening of the attractive interaction between the subunits with increasing temperature indicates that the in vitro assembly of hepatitis B virus capsids is driven by hydrophobic interactions, which are entropic in nature (Tanford, 1980). To rationalize the effect of the ionic strength, Ceres and Zlotnick (2002) conjecture that the subunits more easily adopt an assembly active conformation at elevated ionic strengths (see also Caspar, 1980).

In this work, we offer an alternative interpretation of the observations of Ceres and Zlotnick, amenable to experimental verification. The basic tenet of our proposal is that the attractive hydrophobic interactions that seem to promote the self-assembly of hepatitis B virus capsids must be counteracted by repulsive electrostatic interactions between the subunits. The latter can be weakened by the addition of inert salt. If we translate this presumption into a coarse-grained potential of mean force acting between the subunits, modeled as structureless, quasimacroscopic objects, we find that it becomes possible to quantitatively describe the scaling of the equilibrium constant of the hepatitis B virus capsid assembly with the temperature, as well as that with the ionic strength. The treatment we adopt allows us to extract information on the properties of the hydrophobic surfaces buried upon assembly of the capsid as well as on the net charge of the water-exposed surface from the available experimental data.

The concept of a potential of mean force that is the sum of attractive and repulsive contributions has proven quite useful in the context of the stability of charged colloids (Israelachvili, 1992; Verwey and Overbeek, 1999), including that of proteins (Broide et al., 1996). For the case of virus coat protein subunits, however, the dominant attractive interactions seem to arise from hydrophobic interactions between parts of the proteins, rather than from the van der Waals interactions that destabilize colloidal dispersions (Curtis et al., 2002; Verwey and Overbeek, 1999). In fact, as already advertised, there seems to be a closer analogy between the assembly of virus capsids and the micellization of surfactant

*Submitted January 13, 2004, and accepted for publication March 8, 2004.*

Address reprint requests to Willem K. Kegel, Utrecht University, Van't Hoff Laboratory, Padualaan 8, Utrecht 3584 CH, The Netherlands. Tel.: 31-30-253-2873; E-mail: w.k.kegel@chem.uu.nl.



0006-3495/04/06/3905/09 $2.00

doi: 10.1529/biophysj.104.040055

molecules in aqueous solutions than that with the demixing of colloidal particles. Surfactants and capsid subunits both unite polar and less polar moieties within a single molecule, and the phenomena of micellization and capsid assembly both presumably involve the microphase separation of the apolar parts of the molecules due to hydrophobic interactions (Caspar, 1980; Ceres and Zlotnick, 2002; Lauffer, 1966). It might be useful to mention in this context that the self-assembly of ionic amphiphiles can be promoted by the addition of inert ions that screen the electrostatic repulsion between the charged headgroups (Gunnarsson et al., 1980; Tanford, 1980).

The remainder of this article is set up as follows. First, we briefly review the kinds of interaction that could be involved in the hepatitis B virus capsid assembly, and eliminate those that are in all likelihood either too weak or that produce a temperature or ionic strength dependence at odds with the observations of Ceres and Zlotnick. For the driving force we presume that these are well described by a phenomenological hydrophobic interaction potential valid for pairs of macroscopic hydrophobic surfaces (Israelachvili, 1992). We make plausible that if there is a net charge present on the water-exposed portions of the capsid proteins, this produces a repulsive interaction inversely proportional to the root of the ionic strength irrespective of the precise details of the local geometry of the capsid, provided its radius is much larger than the subunit size. By analyzing the linear temperature dependence of the resulting net interaction strength, we find that the leading-order term can be associated with the excess surface enthalpy of the hydrophobic interfaces between subunits buried in the capsid.

Next, we apply our findings to the thermodynamics of the association equilibrium between the free dimer subunits and the fully formed capsids, and formulate the connection between the equilibrium constant and the free energy of subunit assembly. We fit our expression for the latter to the experimental data taken from Ceres and Zlotnick (2002) and find, within the bounds of experimental uncertainty, quantitative agreement. The values of the fitting parameters extracted are the effective surface charge density of the water-exposed surface area of the capsid, and the excess thermodynamic potentials associated with the hydrophobic faces of the capsid proteins that are shielded from the aqueous solvent upon aggregation. In analogy with the critical micelle concentration, we introduce the concept of a critical capsid concentration and show that hepatitis B capsid assembly seems to obey a universal law of mass action. Finally, we end the article with an outline of our main conclusions and with a discussion where we make a testable prediction as to the pH dependence of the equilibrium constant.

## FREE ENERGY OF ASSOCIATION

There are many types of interaction that might be involved in the stabilization of hepatitis B virus capsid assemblies, as in fact there are in any type of supramolecular polymer (Ciferri, 2000). Obvious ones are Coulomb interactions between oppositely charged moieties, hydrogen bonds, van der Waals interactions (including specific ion effects, see e.g., Ninham and Yaminsky, 1997), and hydrophobic interactions between apolar patches on the surfaces of the proteins. Of these, only the hydrophobic interactions become potentially stronger with increasing temperature. We furthermore note that all the others are weakened by the addition of inert 1:1 salt (Tanford, 1980). For hepatitis B capsids, the buried contact area between the assembled capsid proteins appears to be largely hydrophobic (Ceres and Zlotnick, 2002). Any putative salt bridges or other types of electrostatic interaction (even at high salt concentration) are unlikely to contribute significantly to the protein binding energy, because the hydrophobic character of the contact area leads to a significant loss of solvation of the groups involved if removed from the aqueous environment in the aggregation process (vanVlijmen et al., 1998). Further support for the importance of hydrophobic interactions in capsid assembly comes from the observation that capsid formation is strongly inhibited by bis-ANS (5, 5′-bis[8-(phenylamino)-1-naphtalenesulfonate]) (Zlotnick et al., 2002). This molecule is known to bind specifically to hydrophobic surface patches.

We are led to conclude that hydrophobic interactions must indeed be the dominant driving force toward the assembly of hepatitis B virus capsids, as was in fact also concluded by Ceres and Zlotnick (2002). A phenomenological potential describing hydrophobic interactions between two identical apolar surfaces in water is (Israelachvili, 1992),

$$V_{\mathrm{H}} \approx -2\gamma A \exp(-D/\lambda), \tag{1}$$

where $\gamma$ denotes the surface tension between hydrophobic material and aqueous phase, $A$ the contact area, $D$ the separation of the surfaces, and $\lambda$ a decay length of the order of a few nanometers provided $D < \lambda$ (at larger separations a longer decay length with a different amplitude takes over; Israelachvili, 1992). Both the surface tension $\gamma$ and the length $\lambda$ depend, if only in principle, upon the salt concentration (Toikka et al., 1996). At contact, $D$ should be on the atomic scale, so $D/\lambda \ll 1$, and the issue of $\lambda$ being ionic strength-dependent or not becomes irrelevant.

Not quite so irrelevant is the ionic strength-dependence of the surface tension $\gamma$. As is well known, the surface tension between water and an apolar medium of low dielectric constant increases with increasing salt concentration. This is caused by a surface depletion effect, induced by the interaction of the ions with their own image charges (Onsager and Samaras, 1934). From the recent theory of Levin (2000) we deduce that the surface tension increment due to the presence of a 1:1 electrolyte of molar concentration $c_{\mathrm{s}} > 0.1$ M must at room temperature obey $\gamma$–$\gamma(0) \approx 0.06 + 1.2\, c_{\mathrm{S}}$ in units of mN/m. Here, $\gamma(0)$ denotes the bare surface

tension in the absence of salt. Measurements of the surface tension of water against air do indeed give a linear dependence on the ionic strength, albeit not quite in agreement with the theoretical prediction, with $\gamma$–$\gamma(0) \approx 1.6\, c_S$ again in units of mN/m for salt concentrations up to one molar (Matubayasi et al., 1999). This suggests that for $c_S < 1$ M, the surface tension increment due to the presence of salt is at most 1.6 mN/m and probably less than that because the effective dielectric constant of the interior of the proteins is likely to be higher than that of air. Perhaps more importantly, corrections due to the finite size of the proteins should fairly strongly diminish the depletion effect and therefore also the surface-tension increment (Messina, 2002). For this reason, we presume that we can ignore the salting-out effect that the electrolyte may have on the dimer subunits through the surface-tension increment it induces. This assumption may fail for salt concentration in excess of ~1 M. We finally note in this context that it was shown by Ceres and Zlotnick (2002) that the electrolyte is unlikely to impact upon the self-assembly by merely altering the activity of bulk water. If this were the case, the addition of nonionic solutes such as sucrose should have a comparable effect. However, adding sucrose at concentrations in excess of 1 M was incapable of inducing capsid formation (Ceres and Zlotnick, 2002).

In the light of these considerations, we write for the free energy of binding upon assembly of a complete aggregate

$$G_H \approx -nq\gamma A, \tag{2}$$

with $n$ the mean number of quasiequivalent hydrophobic contacts of a subunit in the capsid, and $q \gg 1$ the number of subunits per capsid. Note that of the quantities in Eq. 2, both $A$ and $\gamma \approx \gamma(0)$ may be a function of the temperature. In the following, we ignore any temperature dependence of the contact area.

If the salting-out of the hydrophobic surfaces cannot account for the observed influence of the ionic strength on the stability of the hepatitis B capsids, the question is what does? The answer must, in our view, be that the protein subunits carry a net charge at near neutral pH. Note that the presence of charges stabilizes the proteins, and ipso facto also the capsids, against macroscopic phase separation (i.e., demixing). By the same token, however, the charges on the subunits also destabilize the micro-phase-separated state that the capsids represent, because it makes the subunits repel each other. Consequently, hepatitis B capsid assembly must be enhanced with increasing ionic strength because the presence of salt screens the Coulomb interactions between charges that are separated by more than roughly a Debye screening length. So, we put forward *not* that the attractive interactions between the subunits are enhanced by the presence of salt, as was suggested in Ceres and Zlotnick (2002), but that the repulsive interactions are diminished.

The problem now is of course how to calculate the effective electrostatic repulsion between the subunits without actually having to explicitly deal with the complexities of the $T$-shaped structure of the subunits (Conway et al., 1997), as well as that of the fully formed capsid, which is covered with four types of perforating hole ranging in width from ~1 to 3 nm (Zlotnick et al., 1996). Fortunately, we do not need to because the mean capsid diameter of ~30 nm is significantly greater than the various length scales of the structure on it. For example, the maximum shell thickness is ~5 nm at the location of the spikes on the capsids, but it is much smaller than that on most of the shell's surface. This allows us to coarse-grain the charge distribution on the surface of the capsid. We note that the inside and outside surface of the capsid are connected through open passages so the actual distribution of the net charge may in fact be three-dimensional.

Within a Debye-Hückel approximation, the potential of mean force $V_C$ acting between two charges located at positions $\vec{r}$ and $\vec{r}'$ on the water-exposed surface of the virus capsid can be written (Verwey and Overbeek, 1999) as

$$V_C \approx k_B T \frac{\lambda_B}{|\vec{r} - \vec{r}'|} \exp(-\kappa|\vec{r} - \vec{r}'|), \tag{3}$$

where $k_B$ denotes Boltzmann's constant, $T$ the absolute temperature, $\lambda_B = e^2/4\pi\varepsilon_0\varepsilon_r k_B T$ the Bjerrum length, and $\kappa^{-1} = 1/\sqrt{8\pi\lambda_B\rho_S}$ the Debye screening length, with $e$ the unit charge, $\varepsilon_0$ the permittivity of vacuum, $\varepsilon_r$ the relative permittivity of the medium (water), and $\rho_S$ the number density of the 1:1 electrolyte (in units of $m^{-3}$). Obviously, $c_S = 10^{-3}\, \rho_S/N_A$ with $N_A$ Avogadro's number. At room temperature, $\lambda_B \approx 0.7$ nm and $\kappa^{-1} \approx 0.3/\sqrt{c_s}$ nm. To explicitly calculate the (nonlocal) free energy $G_C$ associated with the Coulomb interaction between the subunits that make up the capsid, we need to perform a summation of Eq. 3 over all pairs of charges on the capsid but exclude those present on the same protein subunit. The calculation is greatly facilitated if we presume that the (net) charge is uniformly distributed over a spherical shell of mean radius $R$ and thickness $d \ll R$. Let there be $z$ (net) unit charges on each of the $q$-dimer subunits. Straightforward algebra then gives

$$G_C \approx +\frac{1}{4} k_B T q^2 z^2 R^{-2} \lambda_B \kappa^{-1} \tag{4}$$

to leading order in $d/R \ll 1$, and provided $\kappa R > 1$. On account of the large size of the hepatitis B virus capsids, this latter condition is met when $c_S > 4$ mM. Notice the strong impact of the ionic strength on $G_C$ through the $1/\sqrt{c_S}$-dependence of the Debye length $\kappa^{-1}$. It is in our view precisely this term that makes the self-assembly of hepatitis B virus capsids so sensitive to the ionic strength. Quite similar expressions have been derived in the context of the

self-assembly of charged micelles (Gunnarsson et al., 1980; Tanford, 1980), and in that of end-charged dendrimers (Lyulin et al., 2003).

It is of interest to note that the above expression approximately applies even if $\kappa d > 1$, and one is probing the detailed structure of the capsid ignored in our calculation. That this must be so can be made plausible as follows. At high salt concentration, one would only expect charged patches on neighboring subunits to interact significantly. Presuming that such patches can be represented by more or less flat, parallel surfaces, we have for the potential of mean force (Israelachvili, 1992)

$$V_C \approx +8\pi k_B T A' \sigma^2 \lambda_B \kappa^{-1} \exp(-\kappa D), \tag{5}$$

with $A'$ the surface area of interaction, $\sigma$ the net surface density of charges, and $D < \sqrt{A'}$ again the separation of the surfaces. In analogy to our analysis of the hydrophobic interaction, we thus write

$$G_C \approx +4\pi k_B T n' q A' \sigma^2 \lambda_B \kappa^{-1}, \tag{6}$$

at least for $\kappa D \ll 1$. Because the patches involved in the electrostatic interactions are unlikely to be the same ones as those that interact through hydrophobic interactions, the areas $A$ and $A'$ as well as the (effective) number of contact neighbors $n$ and $n'$ need not be identical. It is easy to verify that, apart for a (geometrical) constant, Eqs. 6 and 4 are indeed identical, for $\sigma = z/n'A'$ and the total surface area involved in Coulomb interactions may be defined as $n'qA'$.

On the basis of the above considerations, we propose the following Ansatz for the free energy of interaction associated with the assembly of a complete hepatitis B virus capsid,

$$G \approx -A_H \gamma + A_C k_B T \sigma^2 \lambda_B \kappa^{-1}, \tag{7}$$

which we expect to apply for concentrations of salt below, say, 1 M. In Eq. 7, we recognize in the first term the hydrophobic attraction that drives the self-assembly of capsid, proportional to the total hydrophobic area $A_H \approx 2nqA$ buried in the formation of the capsid and shielded from contact with the solvent. Counteracting the effects of the hydrophobic attraction is an electrostatic repulsion proportional to the exposed charged surface area, so $A_C \approx n'qA'$ if we absorb an uninteresting dimensionless geometrical constant in the effective area $A_C$. Our approach is not identical to but is similar in spirit to that of Odijk (1994), who studied the impact of hydrophobic interactions on the second and third virial coefficient of charged, rod-like biopolymers.

If we perform a Taylor expansion of $\gamma$ around an arbitrary reference temperature, $T_0$, we obtain

$$\gamma(T) = \gamma(T_0) + \left(\frac{\partial \gamma}{\partial T}\right)_{T=T_0} \times (T - T_0) + \ldots \approx \gamma(T_0) - s(T_0)(T - T_0), \tag{8}$$

for absolute temperatures $T$ that are not too far removed from $T_0$, i.e., for $|T - T_0| \ll T_0$. The quantity $s(T_0) \equiv -(\partial\gamma/\partial T)T_0$ is the (bare) surface excess entropy of the hydrophobic surface in contact with water at the temperature $T = T_0$. It turns out that the quantity $s$ is less than 0 for a whole range of hydrophobic interfaces over a fairly large temperature domain, implying that the hydrophobic attraction between the subunits should indeed increase with increasing temperature, as expected (Claesson et al., 1986; Israelachvili, 1992; Villers and Platten, 1988). The repulsive, Coulomb part of Eq. 7 increases much more slowly with temperature, as we shall see in the next section. As a consequence, the dependence of the strength of the interactions between subunits on temperature is fully determined by the value of $s(T_0)$.

## THE EQUILIBRIUM CONSTANT

The hepatitis B virus capsid of interest consists of 120 dimer subunits making up a spherical capsid shell of icosahedral symmetry and a triangulation number $T = 4$. (Only a small fraction of the capsids have a triangulation number $T = 3$ (Ceres and Zlotnick, 2002).) Let the dimers be denoted by $A_1$ and the fully formed capsid by $A_q$, where $q = 120$ for the hepatitis B capsid. To an excellent approximation, the self-assembly of the capsids can be described by an equilibrium reaction $qA_1 \leftrightarrow A_q$, because partially formed capsids are only present in exceedingly low relative amounts (Ceres and Zlotnick, 2002; Zlotnick, 1994). Thermodynamic equilibrium demands that $q\mu_1 = \mu_q$, with $\mu_1$ chemical potential of the dimers and $\mu_q$ that of the complete capsids. Assuming the dispersion to be dilute, the chemical potential of each component $i$ can be written as $\mu_i = \mu_i^0 + k_B T \ln X_i$, where $\mu_i^0$ denotes a standard chemical potential and $X_i$ the mole fraction of component $i$. In the dilute limit, we may approximate $X_i \approx c_i/55.6$ with $c_i$ the concentration of component $i$ in M. (The molarity of pure water at room temperature is close to 55.6 M.) For the equilibrium concentration of capsids, we thus obtain

$$X_q = g_q X_1^q \exp(-\Delta\mu/k_B T), \tag{9}$$

with $g_q$ the number of distinguishable ways in which the capsid can be realized (the degeneracy), and

$$\Delta\mu = \mu_q^0 - q\mu_1^0 \approx G. \tag{10}$$

Due to the highly ordered, crystal-like structure of the capsid one would sensibly expect $g_q$ to be equal to unity (Zlotnick, 1994). The second, approximate equality in Eq. 10,

involving the binding free energy $G$, holds if it greatly exceeds a thermal energy per subunit; hence, if $G \gg q\, k_B T$. In that case, the effective free energy change due to the bonding into the capsid structure predominates, and we need not consider potential contributions from large-scale elastic deformation or breathing modes (Morse and Milner, 1995), nor from a possible conformational change of the protein dimers upon assembly. There appears to be little evidence for the latter (Wingfield et al., 1995), albeit that it cannot be ruled out altogether (Ceres and Zlotnick, 2002).

We stress that there should be no quantum-mechanical contributions in Eqs. 9 and 10 arising from rotational or any other degrees of freedom, including those associated with the so-called *cratic entropy*, for in the coarse-grained description the solvent degrees of freedom have effectively been integrated out, as have those of the counterions. This means that the smallest length is set by the solvent size, not by the thermal wavelength. (For a discussion of this point, see e.g., Reiss et al., 1996.)

The dimensionless equilibrium constant, $K$, is related to the above quantities by

$$\ln K = \ln \frac{X_q}{X_1^q} \approx -G/k_B T, \tag{11}$$

and to the dimension-bearing equilibrium constant $K_{capsid}$, defined in Ceres and Zlotnick (2002), according to $\ln K \approx \ln K_{capsid} + (q-1)\ln 55.6 = \ln K_{capsid} + 478.2$. Inserting Eq. 7 into Eq. 11, and expanding it to linear order in the temperature relative to a reference temperature $T_0$, we obtain

$$\ln K \approx \ln K(T_0) + \left(\frac{\partial \ln K}{\partial \ln T}\right)_{T=T_0} \times \left(\frac{T}{T_0} - 1\right), \tag{12}$$

with

$$\ln K(T_0) = \frac{A_H \gamma}{k_B T_0} - A_C \sigma^2 \lambda_B \kappa^{-1} \tag{13}$$

and

$$\left(\frac{\partial \ln K}{\partial \ln T}\right)_{T=T_0} = -\frac{A_H h}{k_B T_0} + \alpha A_C \sigma^2 \lambda_B \kappa^{-1} \tag{14}$$

both evaluated at $T = T_0$. Here, we have for simplicity presumed that the interactions areas, $A_C$ and $A_H$, as well as the surface charge density, $\sigma$, are invariants of the temperature. (This may be valid over a limited range of conditions only.) The quantity $\alpha \equiv \frac{1}{2} + \frac{1}{2}(\partial \ln \varepsilon_r / \partial \ln T)_{T=T_0}$ is a measure for the sensitivity of the Coulomb interactions between the subunits and temperature changes, and $h(T_0) = \gamma(T_0) + T_0 s(T_0)$ denotes the (bare) excess surface enthalpy of the hydrophobic surface at the reference temperature $T = T_0$. For bulk water, $\alpha \approx -0.14$ (Weast et al., 1984), suggesting that the reference value of the association constant at $T = T_0$, Eq. 13, must be a much stronger function of the ionic strength than its differential increment with temperature, Eq. 14. As we shall see, this is borne out by the measurements of Ceres and Zlotnick (2002).

The equilibrium constants of the hepatitis B capsid assembly obtained by Ceres and Zlotnick (2002) for a range of temperatures and salt concentrations are plotted in Fig. 1. The model predicts that if we plot $\ln K$ versus temperature $T$ we should get a straight line with a slope $-A_H h(T_0)/k_B T_0^2$, at least if we ignore the expected weak dependence on the ionic strength and set $\alpha = 0$ in Eq. 14. (Below we show that this is justifiable.) Applying a least-squares linear fitting procedure to the data of Fig. 1 did not produce any indication of a systematic dependence of the slopes on the salt concentration within the experimental uncertainty, confirming our expectation. For this reason, we average the values for the fitted slopes. Our theoretical fits are also shown in the Fig. 1, where we see a single slope with a tangent of $5.7 \pm 2.4\ K^{-1}$ fits all the data reasonably well. This implies that for the hepatitis B capsid, $A_H h(T_0)/k_B T_0^2 \approx -5.7 \pm 2.4\ K^{-1}$. If we arbitrarily set the reference temperature at $T_0 = 273.15$ K, and if we make use of the estimated buried hydrophobic surface area of $A_H \approx 120 \times 1.3 \times 10^{-17} = 1.56 \times 10^{-15}\ m^2$ (Ceres and Zlotnick, 2002), we get for the bare excess surface enthalpy a value of $h(T_0) \approx -3.8$ mN/m, equivalent to about $-13\ k_B T$ per subunit. (Note that our binding enthalpy and that defined in Ceres and Zlotnick, 2002, differ in sign by construction.)

At the reference temperature $T_0$, we have $\ln K(T_0) \approx A_H \gamma(T_0)/k_B T_0 - A_C \sigma^2\ \lambda_B \kappa^{-1}(T_0)$. Hence, if we plot the

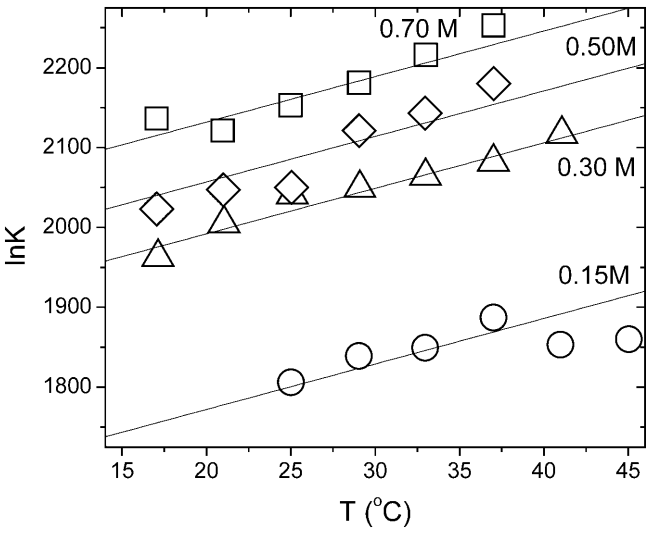



FIGURE 1 The logarithm of the equilibrium constant, $\ln K$, as a function of temperature, $T$, at various salt concentrations as indicated in the figure. Symbols are data from Ceres and Zlotnick (2002). Lines are linear fits with a fixed slope of $+(5.7 \pm 2.4)\ K^{-1}$. The extrapolated values of $\ln K$ at $T = T_0 = 273$ K are plotted in Fig. 2.

experimentally obtained values for $\ln K(T_0)$ as a function of $1/\sqrt{c_S}$, we should get a straight curve with negative slope. Fig. 2 confirms that this is the case. The slope of the curve is a measure for the surface charge density, whereas $\gamma(T_0)$ follows from the intercept at $1/\sqrt{c_S} = 0$ for which the electrostatic interactions are completely screened by the presence of an infinite amount of salt. From the intercept of the linear fit to the data of Fig. 2, we extract $A_H\gamma(T_0)/k_BT_0 \approx 2317 \pm 35$. Note that this value is to be compared (and indeed comparable) to the values of the ''enthalpy per contact'' as reported by Ceres and Zlotnick (2002). Insertion of the estimate $A_H \approx 1.56 \times 10^{-15}$ m$^2$ (Ceres and Zlotnick, 2002) produces a bare surface tension of $\gamma(T_0) \approx +5.5$ mN/m, and an excess surface entropy of $s(T_0) = (h(T_0) - \gamma(T_0))/T_0 \approx -0.03$ mN/mK. The latter finding is consistent with the hydrophobic interaction data of Claesson et al. (1986), which point at typical values for $s$ of the order of $-0.01$ mN/mK for $T \approx 303$ K. The value we find for the surface tension $\gamma$ is roughly an order-of-magnitude smaller than that of ~50 mN/m between water and oil, but comparable to that between water and hexanol (Villers and Platten, 1988). Apparently, the hydrophobic surfaces on the subunits are only weakly hydrophobic, a conclusion that may in fact also be drawn from the slightly negative value of the bare surface excess enthalpy that we found by fitting our theory to the data.

From the fitted line to the data of Fig. 2, we find for the quantity $A_C\sigma^2 \approx (253 \pm 19)/2.1 \times 10^{-19} \approx (1.2 \pm 0.1) \times 10^{21}$ m$^{-2}$. If we naively replace $A_C$ by the estimated surface area of the capsids, equal to $\sim 2.8 \times 10^{-15}$ m$^2$, we get for the effective surface charge density at near-neutral pH a value of $\sigma \approx 6.7 \times 10^{17}$ m$^{-2}$, or ~0.7 charge per nm$^2$. This is close to the limit where one would expect charge renormalization to take place (Trizac et al., 2003), and corresponds to ~15 net charges per dimer subunit or to 7–8 net charges per protein molecule. Whether this is realistic or not is difficult to ascertain at this point, considering that (as far as we are aware) no acid-base titration data are available in the literature. Also, theoretical acid-base titration curves, such as available for foot-and-mouth disease virus capsids (vanVlijmen et al., 1998), seem to be lacking for hepatitis B virus capsids. Note that, ideally, one would determine the net charge of the capsid, e.g., in an electrophoresis or conductivity experiment, and use that to calculate the total area exposed to the water, $A_C$. It is at least reassuring that the number of charges that we find is significantly smaller than the total number of chargeable groups on $Cp149_2$, being 15 negative plus 11 positive charges as an upper bound (Reddy et al., 2001).

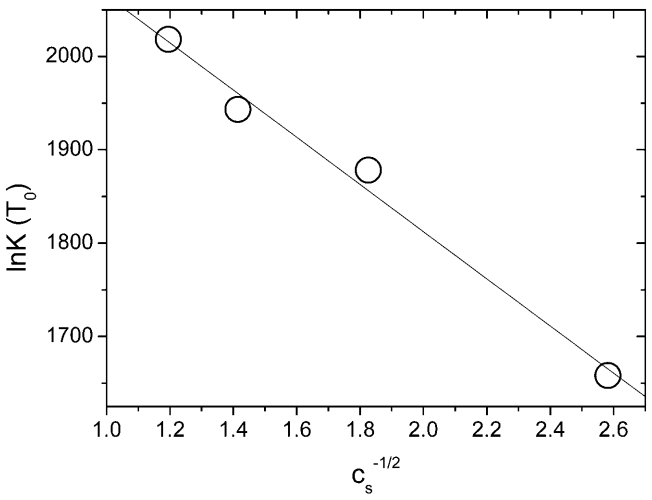



FIGURE 2 The logarithm of the equilibrium constant, ln $K$, at the temperature $T = T_0 = 273$ K as a function of the inverse square-root of the salt concentration $1/\sqrt{c_S}$. The slope of the fitted line is $-(253 \pm 19)$ and its intercept is $+(2317 \pm 35)$. See the main text for an interpretation of these values.

With all the model parameters fixed, we are now in the position to assess the internal consistency of our treatment by evaluating the relative contribution of the Coulomb interaction to the quantity $\partial \ln K/\partial \ln T$ evaluated at $T = T_0$ given in Eq. 14. Inserting the found values for the various model parameters, we obtain for the ratio $\left|\alpha A_C\sigma^2\lambda_B\kappa^{-1}/A_H h k_B^{-1}T_0^{-1}\right| \approx 0.025/\sqrt{c_S}$. This shows that the electrostatic contribution to the slope of Fig. 1 is <6% for $c_S \geq 0.15$ M. This is consistent with the observations, and justifies our neglect of this contribution a posteriori.

We now make connection with the theory of micellization (Israelachvili, 1992; Tanford, 1980), allowing us to predict the equivalent of the so-called *critical micelle concentration*, the critical capsid concentration that we denote by $c_*$. We (arbitrarily) define $c_*$ as that overall concentration of protein subunits at which half of them is absorbed into assemblies. Combining Eq. 11 with the law of conservation of mass, $c = c_1 + qc_q$, where $c$ denotes the molar concentration of subunits in the solution, gives

$$\ln c_* \approx 4.71 - \frac{1}{q-1}\ln qK. \tag{15}$$

Here, $c_*$ is the critical capsid concentration in M, and $\ln K$ is given by Eq. 12. Since according to our theory the equilibrium constant is a sensitive function of the ionic strength, so should be the critical capsid concentration $c_*$. It shifts to lower values with increasing salt concentration in agreement with the data of Ceres and Zlotnick (2002), as is shown in Fig. 3. A similar phenomenon can also be observed in micelles of charged surfactants (Gunnarsson et al., 1980).

With the critical capsid concentration defined, it is now straightforward to show that the fraction material absorbed in capsids, $f \equiv qc_q/c$, is a universal function of $c/c_*$, or, in other words, that $f = f(c/c_*)$ is independent of the temperature, the ionic strength and, in fact, the pH. From Eqs. 9, 10, and 15, it follows that

$$\frac{c}{c_*} = \frac{f^{1/q-1}}{2(1-f)^{q/q-1}}, \tag{16}$$

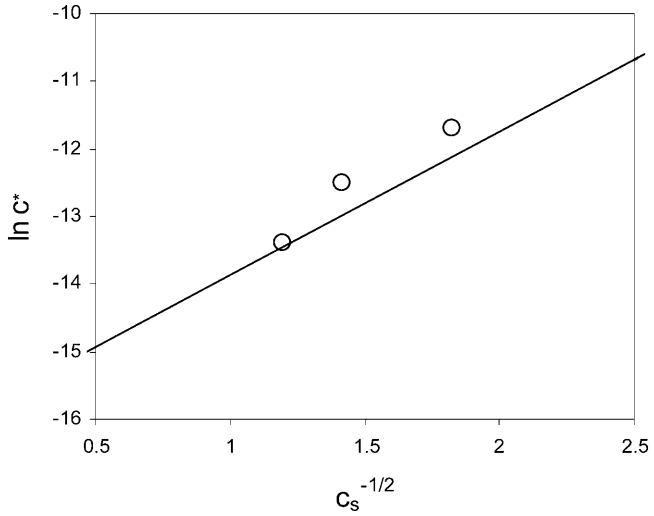



FIGURE 3 The logarithm of the critical capsid concentration, $\ln c_*$, versus the inverse square-root of the salt concentration, $1/\sqrt{c_S}$. The straight line gives our theoretical fit to the measured equilibrium constants of Ceres and Zlotnick (2002). The circles are estimates from the experimental association curves given in Fig. 1 *B* of Ceres and Zlotnick (2002).

where $q$ is again the aggregation number. In Fig. 4 we have plotted $f$ as a function of $c/c_*$ for the data given in Fig. 1 *B* of Ceres and Zlotnick (2002) for the three ionic strengths 0.3, 0.5, and 0.7 M, showing that our analysis is indeed pertinent and that there is indeed a close resemblance between capsid assembly and micellization.

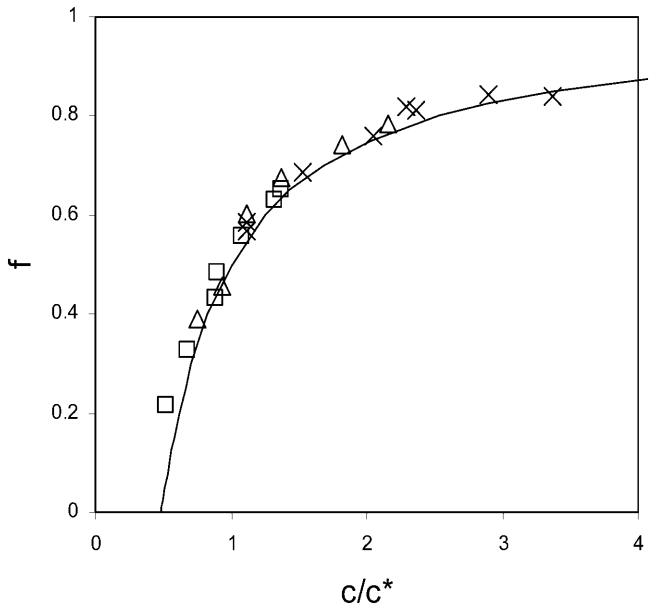



FIGURE 4 The fraction of material assembled into capsids, $f$, versus the overall concentration of dimer subunits $c$, scaled to the critical capsid concentration $c_*$. The symbols represent the data from Ceres and Zlotnick (2002) for samples at a temperature of 25°C. Crosses, $c_S = 0.7$ M; triangles, $c_S = 0.5$ M; and squares, $c_S = 0.3$ M. The drawn line is the universal aggregation curve, given by Eq. 16.

It is important to point out that if $q \gg 1$, Eq. 16 simplifies to $f \approx 1 - c_*/2c$ for $c \geq c_*/2$ and $f \approx 0$ for $c < c_*/2$, i.e., the crossover from the monomeric to the polymeric regime becomes quite sharp as in fact is also evident from Fig. 4. The aggregation curve is therefore relatively insensitive to the relative amounts of $T = 3$ (with $q = 90$) and $T = 4$ capsids (with $q = 120$). In the limit $q \to \infty$, the polymerization or capsidization transition becomes infinitely sharp, and equivalent to a phase transition reminiscent of the Bose-Einstein condensation of ideal bosons (Cuesta and Sear, 2002).

## DISCUSSION AND CONCLUSIONS

Zlotnick and co-workers have convincingly shown that the assembly of hepatitis B capsids can be understood in terms of a chemical equilibrium that is regulated by a combination of mass action, and the intrinsic propensity of the capsid core proteins to aggregate (Ceres and Zlotnick, 2002; Zlotnick, 1994). In this work, we believe to have made plausible that this propensity to aggregate is due to hydrophobic interactions, as in fact seems to be the case for many types of capsid (Ceres and Zlotnick, 2002), but that it is counteracted by screened Coulomb interactions between charged moieties on the dimer subunits.

Our suggestion that the stability of the capsids is intimately linked with the presence of charges on them implies that the capsid assembly must be pH-dependent, if only in principle. The reason is that to a lowest order of approximation the surface charge density is a function of the pH according to

$$\sigma \approx \frac{\sigma_b}{1 + 10^{pH - pK_b}} - \frac{\sigma_a}{1 + 10^{pK_a - pH}}, \qquad (17)$$

at least if all the ionizable groups behave independently. (However, this may be too crude an approximation. See e.g., Boström et al., 2003.) Here, $\sigma_a$ and $\sigma_b$ denote the maximum surface charge densities of acidic and basic groups. Eq. 17 is a consequence of the well-known Henderson-Hasselbalch equation (Stryer, 1980), where we have for simplicity lumped all acidic groups and all basic groups together, and assign an effective $pK_a$ for the former and an effective $pK_b$ for the latter. A tentative estimate for both pK values and both surface charge densities may be obtained from the primary sequence of the hepatitis B monomer subunit protein (Reddy et al., 2001) by presuming that all acidic and basic peptide residues are water-exposed. This allows one to estimate the pH for which $\sigma \approx 0$ and the capsid is approximately charge-neutral, that is, estimate its isoionic value denoted as $pI$. We find that $pI \approx 6$–$7$ for the $Cp149_2$ hepatitis B virus capsid proteins studied by Ceres and Zlotnick (2002). This implies that the surface charge density should increase with increasing pH above its neutral value, leading to a reduced stability of the virus capsids.

There is in fact experimental evidence to support this prediction, albeit not on the $Cp149_2$ protein, but on a closely

related variant of the hepatitis B virus protein known as HBeAg. Like the $Cp149_2$ protein, HBeAg does not carry the C-terminal protamine domain present in the core protein HBcAg, but it does have a N-terminal extension not present on the $Cp149_2$ protein (Wingfield et al., 1995). The HBeAg protein dimers readily form capsids too, albeit mainly of the $T = 3$ type, which with a dimer aggregation number of $q = 90$ are somewhat smaller in size than the $T = 4$ species formed by the HBcAg and $Cp149_2$ proteins. Wingfield and collaborators found that for a particular choice of ionic strength and HBeAg protein concentration, the amount of protein absorbed into capsids strongly decreases from >80% to practically nil when increasing the pH from 7 to 9.5 (Wingfield et al., 1995). This strongly suggests that a charging up of the capsids does indeed destabilize them, as predicted. See Eqs. 12–16.

Wingfield and collaborators also find, in accordance with our analysis, that the degree of association increases with increasing ionic strength (Wingfield et al., 1995), confirming once again the importance of electrostatic interactions to the self-assembly of capsids. The ionic strength data of Wingfield et al. (1995) can in fact be described almost quantitatively by the theory presented in this article, as Fig. 5 shows. Plotted is the fraction HBeAg proteins incorporated into capsids versus the concentration of sodium chloride at a fixed concentration of the protein. The fitting to the data is straightforward, once we realize that the shift of the critical capsid concentration with the concentration of 1:1 electrolyte obeys

$$\ln\frac{c_*}{c_*^0} = \frac{A_C\sigma^2\lambda_B\kappa_0^{-1}}{q-1}\left(\sqrt{\frac{c_S^0}{c_S}} - 1\right), \qquad (18)$$

where $c_*^0$ is a reference critical capsid concentration at the salt concentration $c_S^0$ and the corresponding Debye length $\kappa_0^{-1}$. The experiments were done in 50 mM Tris-HCl buffer at pH = 7 (Wingfield et al., 1995), implying that $c_S \approx 0.05$ + [NaCl] with [NaCl] the concentration of sodium chloride in M. (At neutral pH, almost all of the Tris is dissociated.) The buffer only significantly impacts upon the fitted capsid aggregation curve for salt concentrations below about 0.1 M.

We read off from Fig. 5 that for the given concentration of 0.5 g/l of HBeAg protein, $f = 1/2$ for $c_S^0 \approx 0.15$ M. The steepness of the capsidization curve is now determined by the quantity $A_C\sigma^2\lambda_B\kappa_0^{-1}(q-1)^{-1}$, which by fitting to the data of Fig. 5 we find to be equal to ~3.2. If we presume that all the capsids are of the $T = 3$ type, we find that this implies that there must be ~11 net charges per HBeAg dimer, which is somewhat less than we found for the $Cp149_2$ dimers. The slight discrepancy between theory and experiment at very low [NaCl] may be due to the influence of a net charge density that can only be presumed to be a constant at sufficiently high ionic strengths (Boström et al., 2003).

It is, in the light of our findings, tempting to speculate that the potentially large number of positive charges on the arginine-rich protamine domains of the viral HBcAg subunits strongly suppresses their assembly into capsids at near neutral pH. It may well be that the binding of negatively charged nucleic acids to the protamine domains causes the Coulomb repulsion between the core proteins to be sufficiently reduced to assist the assembly of the complete virus particles under physiological conditions (see also Hatton et al., 1992; Zlotnick et al., 1997). This, then, would provide a physical mechanism to prevent the formation of empty virus capsids in vivo. As many different globular viruses share a protamine domain (Bringas, 1997), this conclusion may extend to a large class of virus.

Another consequence of our findings is that viral assembly and viral disassembly (or uncoating) depends on the local physical environment. Indeed, ionic strength, pH, and temperature together determine the effective interaction between virus subunits and thereby, loosely speaking, whether the assembled or the uncoated state of the virus is stable. Different environments (in particular, pH) in different organelles in biological cells may switch one of either state on or off. This, in turn, could enable viruses to release their genetic material at the desired location only.

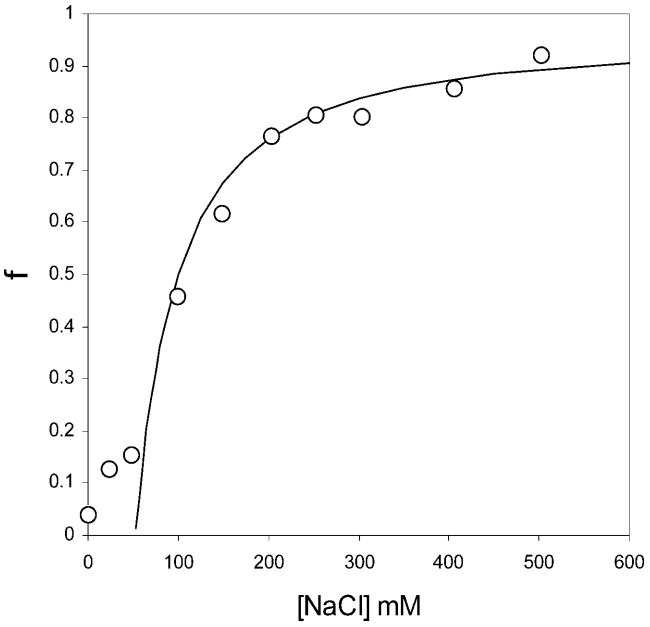



FIGURE 5 Fraction of material assembled into capsids, $f$, as a function of the concentration of NaCl in units of mM. The symbols represent the data of Wingfield et al. (1995) on aqueous solutions of HBeAg dimer proteins that assemble into $T = 3$ capsids with an aggregation number of $q = 90$. Concentration of protein 0.5 g/l at pH = 7.0 and at near room-temperature. The drawn line is theoretical fit using Eqs. 16 and 18.

We thank Ineke Braakman (Utrecht University) and Jan Groenewold (Delft University of Technology) for discussions and for critically reading the manuscript.

## REFERENCES


Boström, M., D. R. M. Williams, and B. W. Ninham. 2003. Specific ion effects: the role of co-ions in biology. *Europhys. Lett.* 63:610–615.

Bringas, R. 1997. Folding and assembly of hepatitis B virus core protein: a new model proposal. *J. Struct. Biol.* 118:189–196.

Broide, M. L., T. M. Tominc, and M. D. Saxowsky. 1996. Using phase transitions to investigate the effect of salts on protein interactions. *Phys. Rev. E.* 53:6325–6335.

Bruinsma, R. F., W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi. 2003. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* 90:248101.

Caspar, D. L. D. 1963. Assembly and stability of the Tobacco Mosaic virus particle. *Adv. Protein Chem.* 18:37–121.

Caspar, D. L. D. 1980. Movement and self-control in protein assemblies—quasiequivalence revisited. *Biophys. J.* 32:103–138.

Ceres, P., and A. Zlotnick. 2002. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry.* 41:11525–11531.

Ciferri, A. 2000. Supramolecular Polymers. Marcel Dekker, New York, NY.

Claesson, P. M., R. Kjellander, P. Stenius, and H. K. Christenson. 1986. Direct measurement of temperature-dependent interactions between nonionic surfactant layers. *J. Chem. Soc. Faraday Trans. I.* 82:2735–2746.

Conway, J. F., N. Cheng, A. Zlotnick, P. T. Wingfield, S. J. Stahl, and A. C. Steven. 1997. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature.* 386:91–94.

Cuesta, J. A., and R. P. Sear. 2002. Phase transition analogous to Bose-Einstein condensation in systems of noninteracting surfactant aggregates. *Phys. Rev. E.* 65:031406.

Curtis, R. A., C. Steinbrecher, M. Heinemann, H. W. Black, and J. M. Prausnitz. 2002. Hydrophobic forces between protein molecules in aqueous solutions of concentrated electrolyte. *Biophys. Chem.* 98:249–265.

Debye, P. J. W. 1949. Light scattering in soap solutions. *Ann. N. Y. Acad. Sci.* 51:575–592.

Gunnarsson, G., B. Jönsson, and H. Wennerström. 1980. Surfactant association into micelles. An electrostatic approach. *J. Phys. Chem.* 84:3114–3121.

Hatton, T., S. Zhou, and D. N. Standring. 1992. RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. *J. Virol.* 66:5232–5241.

Israelachvili, J. 1992. Intermolecular and Surface Forces. Academic Press, San Diego, CA.

Lauffer, M. A. 1966. Polymerization-depolymerization of Tobacco Mosaic Virus protein. VII. A model. *Biochemistry.* 5:2440–2446.

Levin, Y. 2000. Interfacial tension of electrolyte solutions. *J. Chem. Phys.* 113:9722–9726.

Lyulin, S. V., L. J. Evers, P. van der Schoot, A. A. Darinskii, A. V. Lyulin, and M. A. J. Michels. 2004. Effect of solvent quality and electrostatic interactions on the size and structure of dendrimers—Brownian dynamics simulation and mean-field theory. *Macromolecules.* 37: 3049–3063.

Matubayasi, N., H. Matsuo, K. Yamamoto, S. Yamaguchi, and A. Matuzawa. 1999. Thermodynamic quantities of surface formation of aqueous electrolyte solutions*1. I. Aqueous solutions of NaCl, MgCl2, and LaCl3. *J. Colloid Interface Sci.* 209:398–402.

Messina, R. 2002. Image charges in spherical geometry: application to colloidal systems. *J. Chem. Phys.* 117:11062–11074.

Morse, D. C., and S. T. Milner. 1995. Statistical mechanics of closed fluid membranes. *Phys. Rev. E.* 52:5918–5945.

Ninham, B. W., and V. Yaminsky. 1997. Ion binding and ion specificity: the Hofmeister effect and Onsager and Lifshitz theories. *Langmuir.* 13:2097–2108.

Odijk, T. 1994. Long-range attraction in polyelectrolyte solutions. *Macromolecules.* 27:4998–5003.

Onsager, L., and N. N. T. Samaras. 1934. The surface tension of Debye-Hückel electrolytes. *J. Chem. Phys.* 2:528–536.

Reddy, V. S., P. Natarajan, B. Okerberg, K. Li, K. Damodaran, R. T. Morton, C. L. Brooks, III, and J. E. Johnson. 2001. Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.* 75:11943–11947.

Reiss, H., W. K. Kegel, and J. Groenewold. 1996. Length scale for configurational entropy in microemulsions. *Ber. Bunsenges. Phys. Chem.* 100:279–295.

Stryer, L. 1980. Biochemistry. W. H. Freeman and Company, New York.

Tanford, C. 1980. The Hydrophobic Effect. Wiley, London, UK.

Toikka, G., R. A. Hayes, and J. Ralston. 1996. Surface forces between spherical ZnS particles in aqueous electrolyte. *Langmuir.* 12:3783–3788.

Trizac, E., L. Bocavet, M. Auboy, and H. H. v. Grünberg. 2003. Alexander's prescription for colloidal charge renormalization. *Langmuir.* 19:4027–4033.

vanVlijmen, H. W. T., S. Curry, M. Schaefer, and M. Karplus. 1998. Titration calculations of foot-and-mouth disease virus capsids and their stabilities as a function of pH. *J. Mol. Biol.* 275:295–308.

Verwey, E. J. W., and J. T. G. Overbeek. 1999. Theory of Stability of Lyophobic Colloids. Dover Press, Minneola, NY.

Villers, D., and J. K. Platten. 1988. Temperature dependence of the Interfacial tension between water and long-chain alcohols. *J. Phys. Chem.* 92:4023–4024.

Weast, R. C., M. J. Astle, and W. H. Beyer. 1984. Handbook of Chemistry and Physics. CRC Press, Boca Raton, FL.

Wingfield, P. T., S. J. Stahl, R. W. Williams, and A. C. Steven. 1995. Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. *Biochemistry.* 34:4919–4932.

Zlotnick, A. 1994. To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* 241:59–67.

Zlotnick, A., P. Ceres, S. Singh, and J. M. Johnson. 2002. A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *J. Virol.* 76:4848–4854.

Zlotnick, A., N. Cheng, J. F. Conway, F. P. Booy, A. S. Steven, S. J. Stahl, and P. T. Wingfield. 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry.* 35:7412–7421.

Zlotnick, A., N. Cheng, S. J. Stahl, J. F. Conway, A. C. Steven, and P. T. Wingfield. 1997. Localization of the C-terminus of the assembly domain of hepatitis B virus capsid protein: implications for morphogenesis and organization of encapsidated RNA. *Proc. Natl. Acad. Sci. USA.* 94:9556–9561.

# 5. Polymers

Gert Jan Vroege
*Van 't Hoff Laboratory*
*Utrecht University*

# 1. INTRODUCTION

As is well known polymers are long molecules consisting of simple building blocks, the monomers, generally connected through covalent bonds. A simple example is polyethene (figure 1) typically consisting of a 100 to 10000 ethene groups which after polymerization form single C–C bonds around which the polymer can more or less freely rotate. The thus formed chain is an example of a flexible polymer chain, which changes its direction on a length scale of 1 nm because of its rotational freedom around the C–C bonds. Therefore, this polymer is not elongated in solution but it forms a kind of randomly curled coil. Another example is DNA with its double–helix structure, in which 2 DNA–strands are connected through hydrogen bonds (figure 2). DNA can contain up to $10^{10}$ monomers, leading to a length of up to 1 m if it were to be completely stretched. This polymer is much stiffer, but it also changes its direction gradually (on a scale of 100 nm) through small fluctuations in bond angles and bond lenghts. At sufficient length stiff chains also form random coils (in absence of specific interactions, for instance with proteins which can induce very specific DNA conformations in vivo).

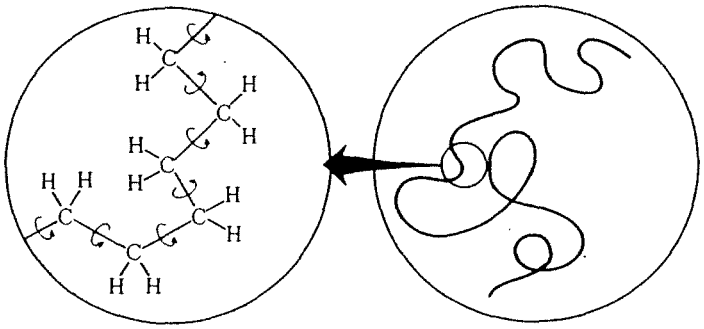



Figure 1. (a) The atomic structure of a polyethene molecule. (b) A schematic representation of a complete molecule. There is rotational freedom around each C–C bond: the molecule forms a long, flexible chain.

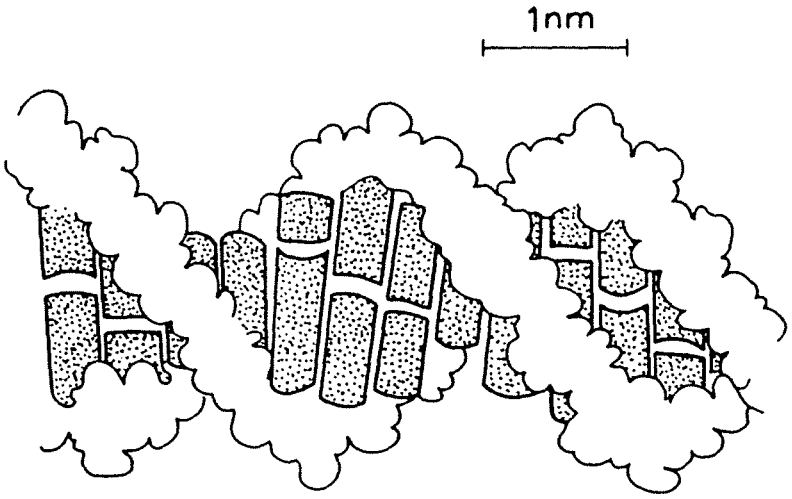



Figure 2: A double helix of DNA in its right-handed B form. Both saccharide–phosphate chains are connected via H-bridges between complementary base pairs.

Polymers can be found everywhere around us. The natural polymers comprise e.g. polysaccharides (like starch and cellulose), DNA and proteins (the last 2 also contain *information* through an alternation of different monomers; this may lead to very specific effects which we will not consider in these lecture notes). Semi-synthetic polymers entered on the scene about a hundred years ago with viscose (chemically modified cellulose, which was used to produce fibres). Totally synthetic polymers were developed in particular after the acceptation of Staudinger's hypothesis (in the twenties and thirties) that polymers consisted of covalently bonded, linear chains of monomers (until then a common notion was that polymers were colloidal aggregates of monomers). Staudinger received the 1953 Nobel prize in Chemistry. During the second world war polymer chemistry became more and more important, an important project being the development of artificial rubber (necessary for some countries because they were cut off from the supply of natural rubber). After the war the production of polymers expanded more and more and now forms a very important part of chemical industry.

Historically, physical chemistry played an important role in characterizing polymers (e.g. through osmotic pressure and viscosity measurements, light scattering and sedimentation), which led to the acceptation of the hypothesis of linear chains, and the concomitant developments of theories. An important name in this context is P.J. Flory (1974 Nobel Prize in Chemistry). In more recent years theoretical developments are increasingly performed by physicists applying general concepts of theoretical physics; for this P.G. de Gennes obtained the 1991 Nobel Prize in Physics.

In these lecture notes we aim at giving a universal description of simple linear flexible polymers. In contrast to synthetic chemists who consider polymers in a very specific way (viz. as built up from specific types of monomers coupled together with a specific type of chemical bond), within physical chemistry polymers are modelled with as few parameters as possible. An indispensable tool is statistical mechanics since one polymer chain consists of a very large number of units and already forms a statistical mechanical system in its own right. In many cases this leads to universal behaviour, which will be illustrated here with comparatively simple calculations.

From statistical mechanics we will mainly use 2 formulas by Boltzmann. The first expresses the entropy of a system in terms of the total number of states $W$ that the

system can assume (at fixed energy):

$$S = k_B \ln W \tag{1.1}$$

and the second the probability $P_i$ of a certain state $i$ with energy $E_i$ (at fixed temperature $T$):

$$P_i = \frac{e^{-E_i/k_B T}}{Z}$$

In this equation $k_B$ is Boltzmann's constant and $Z$ the partition function:

$$Z = \sum_i e^{-E_i/k_B T} \tag{1.2}$$

The formula can also be written as

$$Z = \sum_j G_j \, e^{-E_j/k_B T} \tag{1.3}$$

where the sum now runs over the different energies instead of the different states and the degeneracy $G_j$ represents the number of states of the same energy. The partition function $Z$ forms the connection with thermodynamics through the (Helmholtz) free energy $A$:

$$A = -k_B T \ln Z \tag{1.4}$$

# 2. IDEAL CHAINS

## 2.1. The freely jointed chain

When we want to study the universal properties of polymer chains it is useful to consider a simplified model. Here, so-called segments (often point particles) are connected by "bonds" with certain properties. In the simplest possible model, the so-called *freely jointed* model (figure 3), segments are connected through bonds of fixed length $b$ but with completely arbitrary mutual angles.

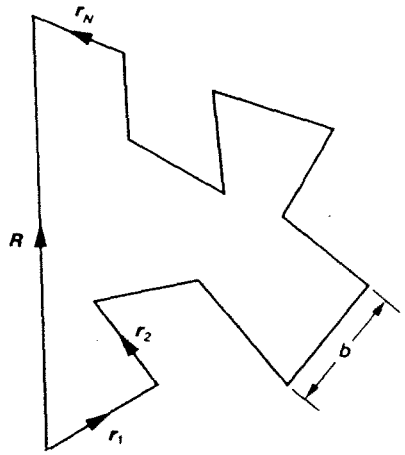



Figure 3. The freely jointed chain.

We can now represent the chain as a sequence of $N$ vectors $\mathbf{r}_i$, each with a length $b$. The total length of the (completely stretched) chain, also indicated as the *contour length,* is $L = Nb$. The vector connecting the endpoints of the conformation in space is:

$$\mathbf{R} = \sum_{i=1}^{N} \mathbf{r}_i$$

From this we can derive the average value of $\mathbf{R}$:

$$\langle \mathbf{R} \rangle = \sum_{i=1}^{N} \langle \mathbf{r}_i \rangle = 0$$

Since every bond vector $\mathbf{r}_i$ has an arbitrary direction, every average $\langle \mathbf{r}_i \rangle = 0$ and also the end–to–end vector $\mathbf{R}$ does not have a preferential direction. To characterize the size of the chain it is therefore more adequate to use the (square root of) the mean

square of $\mathbf{R}$:

$$\langle \mathbf{R}^2 \rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle = \sum_{i=1}^{N} \langle \mathbf{r}_i^2 \rangle + \sum_{i=1}^{N} \sum_{j \neq i} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \tag{2.1}$$

The last step is taking $\langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle = 0$ if $i \neq j$, since the direction of each bond vector is completely independent of the direction of other bond vectors. Hence we end up with

$$\langle \mathbf{R}^2 \rangle = Nb^2 \tag{2.2}$$

The result of (2.2) is very important. The square root of $\langle \mathbf{R}^2 \rangle$ forms a useful measure for the size of the chain. According to (2.2) the size is proportional to $N^{1/2}$, whereas the contour length is proportional to $N$: the chain is strongly curled in space. Note that there is no restriction within this model to place two segments at the same position in space; this is a characteristic of a so–called ideal chain.

## 2.2. The Gaussian distribution

The mean square end–to–end distance is only one property describing the conformation of a chain. It is possible to calculate the complete probability distribution of the end–to–end distance. Therefore we make use of the following equation for the probability $P(\mathbf{R}, N)$ to find an end–to–end vector $\mathbf{R}$ for a polymer of $N$ segments:

$$P(\mathbf{R}, N) = \langle P(\mathbf{R} - \mathbf{r}_N, N-1) \rangle_{\mathbf{r}_N} \tag{2.3}$$

This equation tells that the end–to–end distribution of a chain of $N$ segments can be determined by stepping back one segment within the chain (in space a step $\mathbf{r}_N$) and averaging over all possible steps $\mathbf{r}_N$. In the appendix a derivation is given showing that for large $N$ this leads to a differential equation for $P(\mathbf{R}, N)$:

$$\frac{\partial P}{\partial N} = \frac{1}{6} b^2 \Delta P \tag{2.4}$$

where $\Delta$ represents the Laplacian (note that we write $\mathbf{R} = (x, y, z)$):

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

This equation is identical to the well–known diffusion equation (Fick's second law)

$$\frac{\partial c}{\partial t} = D \Delta c \tag{2.5}$$

using the following correspondence:

$$\begin{aligned} \text{segment number } N &\longleftrightarrow \text{time } t \\ \text{position } \mathbf{R} &\longleftrightarrow \text{position } \mathbf{R} \\ \text{probability } P &\longleftrightarrow \text{concentration } c \\ \frac{1}{6} b^2 &\longleftrightarrow \text{diffusion coefficient } D \end{aligned} \tag{2.6}$$

An ideal chain can thus be compared to a diffusion problem of a particle starting to diffuse from the origin at time $t = 0$ with diffusion coefficient $D$. The fanciful

trajectory such a particle traverses in time (increasing $t$) is completely comparable to the conformation consecutive segments of an ideal polymer (increasing $N$) form within space. A larger typical steplength $b$ leads according to (2.2) to a larger mean square of the end–to–end distance; application of the correspondence (2.6) gives for the analogous diffusion problem:

$$\langle \mathbf{R}(N)^2 \rangle = Nb^2 \longleftrightarrow \langle \mathbf{R}(t)^2 \rangle = 6Dt$$

This is the well–known Einstein formula that tells that a diffusing particle (on average) moves with the square root of time (instead of linearly if the particle would always move in the same direction). Now it is also intuitively clear why (2.2) only applies to an ideal polymer: the polymer must have the opportunity to intersect itself. For a diffusing particle there is after all not a single impediment to return to the same place in space at a later time. In reality this is not possible for a polymer and this may lead to drastically different behaviour. As we will see later there are certain circumstances under which an ideal chain may be a good representation for a polymer.

The solution of (2.4) can be found by applying the correspondence (2.6) to the well–known solution of (2.5) for the analogous diffusion problem (or use Fourier transforms, see appendix 7.2):

$$P(\mathbf{R}, N) = \left( \frac{3}{2\pi N b^2} \right)^{3/2} \exp\left( -\frac{3\mathbf{R}^2}{2Nb^2} \right) \tag{2.7}$$

Here $\mathbf{R}^2 = \mathbf{R} \cdot \mathbf{R} = x^2 + y^2 + z^2$. This solution may be verified by substitution into the differential equation (2.4). Such a distribution function is called a Gaussian distribution. The form (2.7) is normalized:

$$\int P(\mathbf{R}, N)\, d\mathbf{R} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(\mathbf{R}, N)\, dx\, dy\, dz = 1$$

as can be easily verified using the integrals given in appendix 7.4. With this probability distribution also other properties than $\langle \mathbf{R}^2 \rangle$ can be easily calculated, like the standard deviation of $\mathbf{R}^2$ :

$$\sigma_{\mathbf{R}^2} = \sqrt{\langle \mathbf{R}^4 \rangle - \langle \mathbf{R}^2 \rangle^2} = \sqrt{2/3} N b^2 \tag{2.8}$$

We see that the deviation in the size of the polymer coil is of the same order of magnitude as the size itself: an ideal coil is a strongly fluctuating object.

## 2.3. The freely rotating chain

The so–called freely rotating chain is a more realistic model in the sense that there is now a fixed angle $\theta$ between consecutive steps (figure 4). The model is called freely rotating because the chain is free to rotate around each bond (though retaining fixed bond angles $\pi - \theta$).

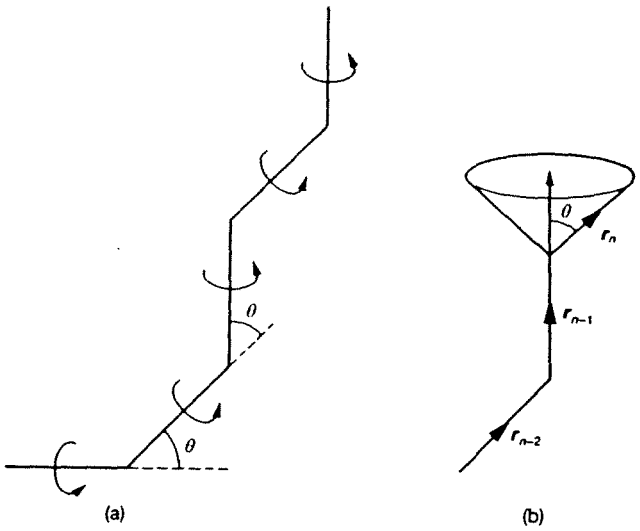



Figure 4. (a) The freely rotating chain. (b) The average of $\mathbf{r}_n$ at fixed $\mathbf{r}_{n-1}$ gives $\cos\theta\,\mathbf{r}_{n-1}$.

If we now average over all possible directions of a bond, there remains a factor $\gamma \equiv \cos\theta$ in the direction of the previous bond, so that

$$\langle \mathbf{r}_i \cdot \mathbf{r}_{i+1} \rangle = b^2\gamma$$

while at every next bond an additional factor $\gamma$ appears because of the averaging over the possible directions of that bond (see figure 4(b))

$$\langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle = b^2\gamma^{|i-j|}$$

We can use this to calculate the mean square end–to–end distance. To this end we further apply the approximation that most segments are located far from the ends of the chain ( if $N$ large) so that in (2.1):

$$\sum_{j\neq i} \langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \approx 2 \sum_{i-j=1}^{\infty} b^2\gamma^{|i-j|} = b^2 \frac{2\gamma}{1-\gamma} \cdot$$

where we used the summation formula of a geometric series. Substituted in (2.1) this ultimately gives:

$$\langle \mathbf{R}^2 \rangle \approx Nb^2 \frac{1+\gamma}{1-\gamma} = Nb^2 \frac{1+\cos\theta}{1-\cos\theta} \tag{2.9}$$

The most important result is that $\langle \mathbf{R}^2 \rangle$ is still proportional to $N$, albeit with a proportionality factor larger than for the freely jointed chain (as long as we keep $\theta < \pi/2$).

## 2.4. More general considerations

We now found for two simple models that $\langle \mathbf{R}^2 \rangle$ is proportional to $N$. This property has a much wider generality and is a consequence of the so–called *central limit theorem*

in mathematics. In a nutshell this theorem says that a variable (say $x$), itself a sum of a large number (say $N$) identical, independent, stochastic variables, is distributed according to a normal (=Gaussian) distribution $\exp(-x^2/2\langle x^2\rangle)$ and that $\langle x^2\rangle \propto N$. A polymer chain complies to this if it has a short–range memory (i.e. the position of a certain segment only depends on the positions of a limited number of neighbouring segments within the chain). Then, a chain has "forgotten" where it came from after a limited number of segments. The consequence is that also for a more complex and realistic polymer model the mean square end–to–end distance remains proportional to $N$ (for large $N$)

$$\langle \mathbf{R}^2 \rangle = N b_{\text{eff}}^2 \tag{2.10}$$

The proportionality constant obviously has a dimension (length)$^2$ and defines the effective step length $b_{\text{eff}}$. For the freely jointed chain we find (see equation (2.9))

$$b_{\text{eff}} = b\sqrt{\frac{1+\cos\theta}{1-\cos\theta}}$$

As could be expected the effective step length is larger than $b$ if $\theta < \pi/2$. In the remainder we generally write $b$ instead of $b_{\text{eff}}$. The concept of an effective segment length was first introduced by Kuhn. In order to be independent of the rather arbitrary division into segments (2.10) is now written as

$$\langle \mathbf{R}^2 \rangle = L l_K \tag{2.11}$$

where L is the length of the polymer if it were completely stretched and $l_K$ is called the *Kuhn length.*

Also the Gaussian distribution (2.7) remains valid in the case of more general short–range models and therefore forms one of the basic formulas of polymer theory. In the end it is not so very important how we exactly model a chain since all (short–range) models give the same results for large chain length. A widely used class of models are the so–called *lattice models*, where chain segments are placed on the lattice points of a space filling lattice (see later).

## 2.5. The entropy of an ideal coil

The end–to–end distribution $P(\mathbf{R}, N)$ is directly proportional to the number of possible realizations $W(\mathbf{R})$ of an ideal chain at a given end–to–end vector (and at a given number of segments $N$). From this we can derive an expression for the *entropy* of such an ideal chain via (1.1):

$$S(\mathbf{R}) = k_B \ln W(\mathbf{R}) = \text{cst} - \frac{3k_B}{2Nb^2}\mathbf{R}^2 \tag{2.12}$$

Later we shall use this formula a number of times. At this point we shall use it to show the analogy of a polymer coil with a spring. First we form a (Helmholtz free) energy from this entropy

$$A(\mathbf{R}) = U(\mathbf{R}) - TS(\mathbf{R}) = -TS(\mathbf{R}) = \text{cst} + \frac{3k_BT}{2Nb^2}\mathbf{R}^2 \tag{2.13}$$

Note that $U = 0$ for an ideal chain. The interesting point of this formula is its quadratic form similar to the potential energy of a spring. If the endpoints of a

polymer chain are brought apart, it experiences (on average) a retracting force

$$\mathbf{f} = -\frac{\partial A}{\partial \mathbf{R}} = -\frac{3k_BT}{Nb^2}\mathbf{R}$$

The spring constant is $3k_BT/Nb^2$. This entropic effect forms the basis for the elasticity of rubber, which consists of a network of interconnected polymer chains. Note that the above formula predicts that the spring constant of rubber increases with temperature (in strong contrast to most other materials). This is also found experimentally and is sometimes called the *Guch-Joule* effect. The formula tells further that the material can be stretched more easily if the number of segments between consecutive linking points in the network $N$ is larger.

## 2.6. Further models of a flexible polymer chain

Since different parts of an ideal chain do not have any interactions, a part of the chain can itself be described by a Gaussian distribution. We then have to use a generalized version $G_N(\mathbf{R}|\mathbf{R}')$ for the probability of a chain of $N$ segments starting at a position $\mathbf{R}'$ and ending at $\mathbf{R}$ (so $P(\mathbf{R}, N) = G_N(\mathbf{R}|0)$). Let us for instance consider a chain on which we also want to fix segment $\nu$ at position $\mathbf{R}''$. Since both parts of an ideal chain are independent the probability for such a configuration is then given by the product of both probabilities, $G_{N-\nu}(\mathbf{R}|\mathbf{R}'')G_\nu(\mathbf{R}''|\mathbf{R}')$. Integrating over all possible positions of the $\nu$th segment, gives back the original end–to–end distribution:

$$G_N(\mathbf{R}|\mathbf{R}') = \int G_{N-\nu}(\mathbf{R}|\mathbf{R}'')G_\nu(\mathbf{R}''|\mathbf{R}')\, d\mathbf{R}''$$

This relation can be verified easily using (7.4) and the convolution theorem in Fourier space.

A widely used model, the *standard Gaussian* o- *Gaussian bond model* retains this Gaussian property up to the level of each single bond (assuming the distribution is still valid if we set $N = 1$ in (2.7)). This is frequently justified since we saw that many results do not depend on the specific details of the local structure of the chain.

If we compare this with the previous section this also implies that each bond is a spring with spring constant $3k_BT/b^2$. This forms the basis of another model the *bead–spring model*, in which a polymer is modelled as a string of beads connected by springs. This model is especially used in describing polymer dynamics (Rouse/Zimm model).

The ultimate consequence of the Gaussian model is a continuous model where we do not have separate segments anymore, but a continuous line. This leads to path integrals.

## 2.7. Gaussian chains in an external field

Here we consider a slightly more general case than before by placing a polymer chain in an external field. We define $\varphi(\mathbf{R})$ as the energy a segment obtains at place $\mathbf{R}$ due to the interaction with this external field. In appendix 7.3 is shown that differential equation (2.4) then takes the form

$$\frac{\partial G}{\partial N} = \frac{1}{6}b^2\Delta G - \frac{\varphi(\mathbf{R})}{k_BT}G \qquad (2.14)$$

where we now formulate the differential equation in terms of $G = G_N(\mathbf{R}|\mathbf{R}')$. It is shown in the appendix that the general solution to this equation can be written as a so–called bilinear expansion:

$$G_N(\mathbf{R}|\mathbf{R}') = \sum_n \psi_n(\mathbf{R}')\psi_n(\mathbf{R}) \exp(-\lambda_n N) \tag{2.15}$$

where $\psi_n(\mathbf{R})$ and $\lambda_n$ are the eigenfunctions and eigenvalues of the following equation (very similar to the time–independent Schrödinger equation in quantum mechanics):

$$-\frac{1}{6}b^2\Delta\psi_n + \frac{\varphi(\mathbf{R})}{k_BT}\psi_n = \lambda_n\psi_n \tag{2.16}$$

The solution critically depends on the type of eigenvalues found. In some cases these may take a continuous range of values (a continuous spectrum), which corresponds to a scattering state in quantum mechanics. An example of this is the case with $\varphi(\mathbf{R}) = 0$, corresponding to a free particle in quantum mechanics. In appendix 7.3 is shown that (2.15) leads back to a Gaussian coil (as expected).

Another possibility is a discrete spectrum of eigenvalues. In that case one of the eigenfunctions may dominate the problem for long chains. Let the smallest eigenvalue be $\lambda_0 < \lambda_1, \lambda_2, \cdots$. If we make $N$ very large, the exponential term containing $\lambda_0$ will be by far the largest in (2.15) and will dominate the solution (*ground state dominance):*

$$G_N(\mathbf{R}|\mathbf{R}') \sim \psi_0(\mathbf{R}')\psi_0(\mathbf{R}) \exp(-\lambda_0 N) \tag{2.17}$$

This result seems reasonable for a long chain: both endpoints get completely uncorrelated. A state like this is called a *bound state.* The parallel with quantum mechanics gets even stronger if we calculate the segment density $c(\mathbf{R})$ for (2.17). We then have to integrate over all segment numbers $\nu$ in the chain that must be located at $\mathbf{R}$, irrespective of the position of the beginning $\mathbf{R}'$ and end of the chain (here taken as $\mathbf{R}''$):

$$c(\mathbf{R}) = \frac{\int_0^N d\nu \int d\mathbf{R}' \int d\mathbf{R}'' G_\nu(\mathbf{R}|\mathbf{R}')G_{N-\nu}(\mathbf{R}''|\mathbf{R})}{\int d\mathbf{R}' \int d\mathbf{R}'' G_N(\mathbf{R}''|\mathbf{R}')}$$

The numerator is a normalizing factor. Using (2.17) for large $N$ simply gives:

$$c(\mathbf{R}) \sim N\psi_0^2(\mathbf{R}) \tag{2.18}$$

So $|\psi_0(\mathbf{R})|^2$ is the probability to find one of the segments at position $\mathbf{R}$, much like in quantum mechanics.

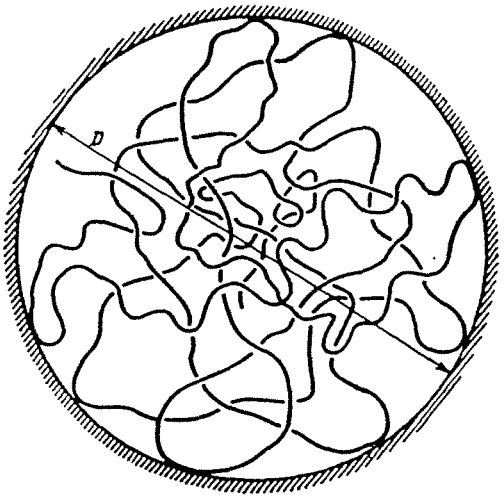



Figure 5. A polymer chain in a spherical cavity of diameter $D$.

An example of this situation is a polymer chain confined to a spherical cavity of diameter $D$ (see figure 5). For this spherical symmetry we can express the Laplacian in terms of the distance to the origin $R$:

$$\Delta \cdots = \frac{1}{R}\frac{d^2}{dR^2}(R\cdots)$$

If we now solve equation (2.16) with $\varphi(\mathbf{R}) = 0$ within the cavity, but all eigenfunctions $= 0$ outside (since the chain obviously cannot be there), we find as the lowest eigenvalue and (normalized) eigenfunction:

$$\lambda_0 = \frac{2\pi^2 b^2}{3D^2} \text{ and } \psi_0 = \frac{1}{R\sqrt{\pi D}} \sin\left(\frac{2\pi R}{D}\right)$$

Giving a segment density like in figure 6. This is an example of a chain in a globular state.

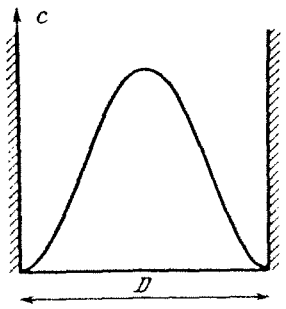



Figure 6. Segment density of a very long ideal polymer in a spherical cavity of diameter D.

## 2.8. Lifshitz entropy

For a polymer in a bound state it is possible to derive an expression for the entropy. Here, we only consider the situation of ground state dominance ($N$ large). The total partition function would now be:

$$Z = \int d\mathbf{R}' \int d\mathbf{R}'' G_N(\mathbf{R}''|\mathbf{R}')$$

since $G_N(\mathbf{R}''|\mathbf{R}')$ takes into account all chain configurations and Boltzmann weights in the external field and the integrals all possibilities of its beginning and end points. Substituting (2.17) simply gives

$$Z \sim \exp(-\lambda_0 N) \left( \int d\mathbf{R} \psi_0(\mathbf{R}) \right)^2$$

and a concomitant free energy from (1.4)

$$A \sim k_B T \lambda_0 N \tag{2.19}$$

Apart from unimportant end terms this expression is proportional to $N$ and is therefore an extensive property, like in a true macroscopic system.

Note that this expression still depends on the external field $\varphi(\mathbf{R})$, since $\lambda_0$ is the lowest eigenvalue of (2.16). Remarkably however, it is possible to eliminate the externally imposed field if we calculate the entropy of the chain. To show this we write

$$S = \frac{U - A}{T} = \int \frac{\varphi(\mathbf{R})}{T} c(\mathbf{R})\, d\mathbf{R} - \frac{A}{T}$$

Using (2.18) and (2.19) this reduces to

$$S = N \int \frac{\varphi(\mathbf{R})}{T} \psi_0^2(\mathbf{R})\, d\mathbf{R} - k_B \lambda_0 N$$

The integral in this expression can now also be obtained by taking (2.16) for $n = 0$, multiplying by $\psi_0(\mathbf{R})$ and integrating over $\mathbf{R}$:

$$-\frac{1}{6} b^2 \int \psi_0 \Delta \psi_0\, d\mathbf{R} + \frac{1}{k_B} \int \frac{\varphi(\mathbf{R})}{T} \psi_0^2(\mathbf{R})\, d\mathbf{R} = \lambda_0$$

Combining these last 2 equations eliminates both $\varphi(\mathbf{R})$ and $\lambda_0$:

$$S = N k_B \frac{1}{6} b^2 \int \psi_0 \Delta \psi_0\, d\mathbf{R}$$

In terms of the segment density $c(\mathbf{R})$ from (2.18) this can be written as

$$S[c(\mathbf{R})] = k_B \frac{1}{6} b^2 \int c^{1/2}(\mathbf{R}) \Delta c^{1/2}(\mathbf{R})\, d\mathbf{R} = const - k_B \frac{1}{6} b^2 \int (\nabla c^{1/2}(\mathbf{R}))^2\, d\mathbf{R} \tag{2.20}$$

This is the *Lifshitz entropy* of a single chain. We see that this entropy is connected to concentration gradients and therefore the spatial inhomogeneity of the segment distribution. We could say that spatial inhomogeneity is unfavorable for the entropy,

because in the presence of a concentration gradient, the chain is forced to bend in certain directions more often than in others. This restricts the total number of allowed conformations and therefore decreases the entropy.

When there are many chains, their conformational entropies add up and $c(\mathbf{R})$ represents the total concentration of segments of all chains. In fact an ideal translational term should be added, but this is very small since for every chain a very large number of segments $N$ are connected together to form only one translational unit. Therefore, the conformational entropy which as a rule is insignificant for low–molecular–weight substances (compared to the translational entropy) becomes predominant for many properties of polymers.

## 2.9. Self–consistent field method

The results in the last section are useful for the so–called self–consistent field (SCF) method for polymers. In statistical physics we try to describe systems in terms of a restricted number of macroscopic variables instead of specifying all positions and momenta of individual particles. In the limit of a very large system (thermodynamic limit) we expect that the free energy shows a very sharp minimum when varying the values of these macroscopic variables. This sharpness justifies the usual procedure of minimizing the free energy with respect to these macroscopic variables to find its equilibrium value. However, this neglects the possible influence of fluctuations.

In the case of our polymer system the form of (2.20) suggests to take $c(\mathbf{R})$ as our (quasi)macroscopic variables. The entropy $S[c(\mathbf{R})]$ is now a functional of $c(\mathbf{R})$ (i.e. it is a function of a function: its variables are the values of the segment density at *all* positions $\mathbf{R}$ in the system, but this already is an average over many microscopic configurations). The important point is that in (2.20) the external field $\varphi(\mathbf{R})$ has been eliminated, so its form is independent of the forces that cause this macroscopic state. The next step to obtain an SCF theory would be to add interactions between segments in the form of an energy term $U[c(\mathbf{R})]$, giving

$$A[c(\mathbf{R})] = U[c(\mathbf{R})] - TS[c(\mathbf{R})]$$

The equilibrium distribution $c_{eq}(\mathbf{R})$ is obtained by functional minimization of $A$ with respect to $c(\mathbf{R})$. Since in many polymer systems the segment concentrations are low, an often used approximation for $U[c(\mathbf{R})]$ is to take the corresponding expressions for disconnected segments (a non–ideal gas), although omitting the translational (ideal gas) term since segments are connected here.

# 3. NON–IDEAL CHAINS

Until now we only considered the short–range interaction between segments (i.e. interaction between segments closely together if measured along the chain contour). As soon as we also take into account long–range interactions (in this case indicating interaction between segments far apart if *measured along the chain*; by backfolding of the chain these segments can come closely together in space) the chain statistics can change drastically. For simplicity we consider a lattice model for a polymer (figure 7) where every lattice point has $z$ neighbouring positions at a distance $b$.

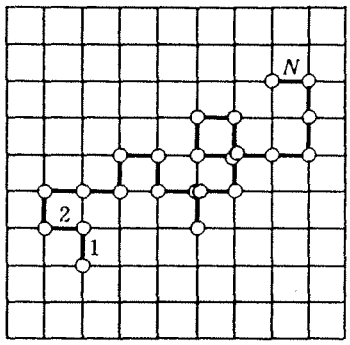



Figure 7. A lattice model for a polymer. The white cirkels are the segments and the bold line segments form the bonds.

The volume per lattice point is called $\nu_c$. If such a chain were ideal (no interaction) there would be no obstruction to placing several segments on one and the same lattice position. In that case each consecutive segment has $z$ possibilities to be placed and the total number of possibilities for a chain of $N$ segments would be $z^N$. It can be easily verified that $\langle \mathbf{R}^2 \rangle = Nb^2$ also applies in this case, so that (2.7) remains valid. The total number of configurations $W_0(R)$ at a given end–to–end distance $R$ is now proportional to:

$$W_0(R) \propto z^N 4\pi R^2 \exp\left(-\frac{3R^2}{2Nb^2}\right) \tag{3.1}$$

Since we do not write the number of configurations as a function of end–to–end vector $\mathbf{R}$ but in terms of end–to–end distance $R$, we get an extra factor of $4\pi R^2$ ($\propto$ the surface of a spherical shell of radius $R$). We may again formulate this problem in terms of a free energy $A_0(R)$ (or entropy, $S_0(R)$):

$$\frac{A_0(R)}{k_B T}\left(= -\frac{S_0(R)}{k_B}\right) = -\ln W_0(R) = \text{cst} - 2\ln R + \frac{3}{2Nb^2}R^2 \tag{3.2}$$

To get the most probable end-to-end distance we minimize this free energy by taking the derivative with respect to $R$ and equating it to 0:

$$-\frac{2}{R}+\frac{3}{Nb^2}R=0 \Rightarrow R_0^*=\sqrt{\frac{2}{3}}\sqrt{Nb^2} \tag{3.3}$$

This procedure gives us an ideal chain dimension $R_0^*$ (NB the procedure followed is not entirely correct but gives a reasonable estimation). We can also use this formula to estimate the average volume fraction of segments within the coil $\phi \approx N\nu_c/R_0^{*3} \approx N\nu_c/(b\sqrt{N})^3 \approx N^{-1/2}$; this obviously gets very small for large $N$.

## 3.1. An excluded volume chain

We now consider the simplest type of long-range interaction, where a once occupied lattice point cannot be occupied by a second polymer segment. In analogy with the diffusion problem this type of conformation is also called a SAW(=Self-Avoiding Walk). To describe the resulting conformation qualitatively we assume that the polymer coil swells to a different size $R$ but that the internal structure of the chain is retained (such that we can still apply expression (3.1)). We further assume that encounters between segments take place independently and that only pair interactions are important. We account for this by multiplying (3.4) by a correction factor $p(R)$ which represents the probability that a given conformation is allowed (i.e. we did not place one single segment on the same position as one of the others). The probability that, if we place a segment on a specific lattice point, this is already occupied by another segment, is given by the volume of 1 segment $\nu_c$ divided by the total volume of the coil (approximately $R^3$). There are $N(N-1)/2$ of this kind of possible (pair) contacts within a coil of $N$ segments, so that

$$p(R) \approx \left(1-\frac{\nu_c}{R^3}\right)^{N(N-1)/2} \approx \exp\left(\frac{N(N-1)}{2}\ln(1-\nu_c/R^3)\right) \approx \exp\left(-\frac{N^2\nu_c}{2R^3}\right)$$

Multiplying (3.1) by this factor we get the number of configurations with excluded volume:

$$W(R) \propto z^N 4\pi R^2 \exp\left(-\frac{3R^2}{2Nb^2}-\frac{N^2\nu_c}{2R^3}\right) \tag{3.4}$$

The free energy now gets

$$\frac{A(R)}{k_BT} = -\ln W(R) = \text{cst} - 2\ln R + \frac{3}{2Nb^2}R^2 + \frac{N^2\nu_c}{2R^3} \tag{3.5}$$

and a similar minimization procedure as for (3.3) now gives

$$-\frac{2}{R}+\frac{3}{Nb^2}R-\frac{3N^2\nu_c}{2R^4}=0$$

Combined with $R_0^*$ from (3.3) this leads to the following equation for the chain size $R^*$ of an excluded-volume chain (first given by Flory):

$$\left(\frac{R^*}{R_0^*}\right)^5-\left(\frac{R^*}{R_0^*}\right)^3 \approx \frac{\nu_c}{b^3}N^{1/2} \approx N^{1/2} \tag{3.6}$$

To emphasize we are performing a qualitative calculation we left out a factor of $\sqrt{243/128}$. For large $N$ we can neglect the cubic term and the solution is:

$$\frac{R^*}{R_0^*} \approx N^{1/10} \Rightarrow R^* \approx bN^{3/5} \tag{3.7}$$

$R^*/R_0^*$ is often called the *expansion factor* $\alpha$. In comparison to the ideal chain (3.3) the coil swells and it does so the stronger when the chain is longer. It is essential to realize the difference with a chain with short-range interactions: with short-range interactions the chain may also swell but in this case because of an increased effective step length $b_{\text{eff}}$, while the ideal dependence on $N$ persists. With a long-range excluded-volume interaction the exponent of $N$ changes. The above-described model is strongly simplified. More extensive calculations and computer simulations give for large $N$:

$$R^* \approx bN^{\nu} \text{ with } \nu \approx 0.588$$

which does not differ much from the Flory exponent 3/5 in (3.7).

## 3.2. A chain in solution

Up to now we only counted conformations to describe a polymer chain, i.e. we only took into account the entropy using formula (1.1). This was made possible by the fact that energy did not play a role. In actual life this never occurs since a polymer chain will always be suspended in a solvent, so different conformations may have different energies. We now account for this in a crude way by adding an (average) energy term $\overline{E}(R)$ to the entropic term in the free energy :

$$\begin{aligned} A(R) &= U(R) - TS(R) \Longrightarrow \\ \frac{A(R)}{k_B T} &\approx \frac{\overline{E}(R)}{k_B T} - \ln W(R) \end{aligned}$$

We can determine $\overline{E}(R)$ in the above-used lattice model by again viewing the polymer coil as an object of $N$ segments in a volume $R^3$ and hence a volume fraction $\phi(R) \approx N\nu_c/R^3$. The lattice points not occupied by the polymer are now occupied by a solvent molecule (see figure 8).

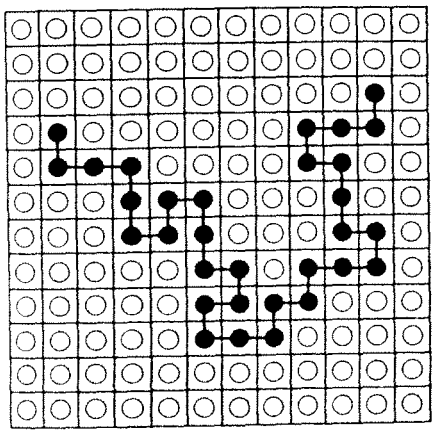


Figure 8. The lattice model of a chain in solution. Black cirkels represent polymer segments and white cirkels solvent molecules.

We now suppose only interaction between neighbouring lattice points, viz.

$$\begin{aligned} \text{polymer segment–polymer segment} &: -\varepsilon_{pp} \\ \text{polymer segment–solvent molecule} &: -\varepsilon_{ps} \\ \text{solvent molecule–solvent molecule} &: -\varepsilon_{ss} \end{aligned} \tag{3.8}$$

There are $Nz$ neighbouring points next to the polymer, each with a probability $\phi$ to be a polymer segment and $1-\phi$ to be a solvent molecule. The average energy of the coil is then:

$$\overline{E}(R) \approx Nz\left[\frac{1}{2}\phi(-\varepsilon_{pp}+\varepsilon_{ss})+(1-\phi)(-\varepsilon_{ps}+\varepsilon_{ss})\right]$$

The factor $1/2$ appears to avoid double–counting of polymer–polymer interactions and we have chosen the pure solvent as the reference state (for each contact with a polymer segment a solvent–solvent contact is broken). We can rewrite this as:

$$\begin{aligned} \overline{E}(R) &\approx \text{cst (independent of } R) - Nz\phi(R)\left[\frac{1}{2}(\varepsilon_{pp}+\varepsilon_{ss}-2\varepsilon_{ps})\right] \\ &\approx \text{cst} - \frac{N^2\nu_c}{R^3}z\left[\frac{1}{2}(\varepsilon_{pp}+\varepsilon_{ss}-2\varepsilon_{ps})\right] \\ \frac{\overline{E}(R)}{k_BT} &\approx \text{cst} - \frac{N^2\nu_c}{2R^3}2\chi \end{aligned} \tag{3.9}$$

The last step introduces the so–called *chi–parameter* $\chi$, which we consider more closely in the next section. The reason to write (3.9) as in the last line is clear if we compare this term with the last term of (3.5): the effect of the energetic interaction with the solvent has the same functional form as the excluded volume interaction between segments. We can therefore immediately use the results from the previous section after transforming

$$\nu_c \rightarrow \nu \equiv \nu_c(1-2\chi) \tag{3.10}$$

The Flory equation (3.6) now gives:

$$\left(\frac{R^*}{R_0^*}\right)^5 - \left(\frac{R^*}{R_0^*}\right)^3 \approx \frac{\nu}{b^3}N^{1/2} \tag{3.11}$$

The difference is that in this case the right–hand side is not necessarily large. This now depends on the value of $\chi$. We will differentiate between several regimes which we discuss in the next section.

## 3.3. Good, bad and ideal solvents

The chi–parameter in (3.9) is given by

$$\chi \equiv \frac{z\Delta\varepsilon}{k_BT} = \frac{z(\varepsilon_{pp}+\varepsilon_{ss}-2\varepsilon_{ps})}{2k_BT} \tag{3.12}$$

where $2\Delta\varepsilon$ represents the energy change at the formation of 1 polymer segment–polymer segment contact (see figure 9).

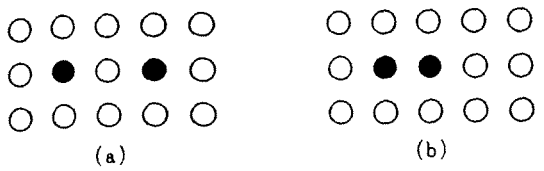



Figure 9. The effective interaction between two polymer segments. If the polymer segments (indicated by black cirkels), originally separated as in (a), are brought together, as in (b), the energy of the system decreases with an amount $2\Delta\varepsilon = \varepsilon_{pp} + \varepsilon_{ss} - 2\varepsilon_{ps}$.

A simple rule of thumb for (London–van der Waals) dispersion forces gives for the interaction energy between two molecules $i$ and $j$:

$$\varepsilon_{ij} = k\,\alpha_i\alpha_j$$

Proportionality constant $k$ is generally positive and the $\alpha$s are (segmental) polarizabilities (note that in the definition of the $\varepsilon$s (3.8) appears a $-$ sign: $k > 0$ implies attraction). Therefore, the energy difference $\Delta\varepsilon = \frac{1}{2}(\varepsilon_{pp}+\varepsilon_{ss}-2\varepsilon_{ps}) = \frac{1}{2}k\,(\alpha_p-\alpha_s)^2$ is $> 0$ in most cases. Consequently, $\chi$ is also usually positive. We differentiate between 3 regimes:

$\chi \ll 1$ This implies that the energy change at immersing a polymer segment in its solvent is much smaller than $k_BT$ (see (3.12)). We expect little influence from energy effects. This also appears from (3.10): $\nu \approx \nu_c$; the coil is swollen as in the previous section. In this case, the solvent is called a *good solvent*. When $\chi$ increases, energy effects are getting more important and the polymer coil starts to shrink.

$\chi = 1/2$ In this special case (3.10) gives $\nu = 0$. The Flory equation (3.11) now gives $R^* = R_0^* \approx bN^{1/2}$, so that the chain behaves like an ideal chain. This is caused by two opposing tendencies. The unfavourable energy change at immersing the polymer segments in solvent wants to contract the coil, while the excluded volume of the segments (and the higher entropy of an expanded coil) leads to swelling of the chain. At $\chi = 1/2$ these two effects just compensate. In this case the solvent is called an *ideal solvent*, a *theta solvent* or sometimes a *marginal solvent*. An important parameter to vary $\chi$ is the temperature. The temperature at which the coil behaves ideally, is called the *theta temperature* $\theta$. The term marginal solvent is connected to the fact that $\chi = 1/2$ is a value where the polymer remains only just soluble. At a slightly higher value of $\chi$ energy effects are going to dominate.

$\chi \gtrsim 1/2$ At these values the polymer changes its structure rather abruptly. The higher entropy of the coil structure is not sufficient anymore to compensate unfavourable energy effects, the solvent is expelled and the coil collapses (see figure 10). The chain now forms a rather compact structure, a so–called *globule* (an example is formed by the globular proteins), where $R^* \propto N^{1/3}$. At about the same time different globules in the solution tend to prefer being in mutual contact rather than with the solvent, so that the polymer precipitates (or does not dissolve). In this regime the solvent is called a *bad solvent*.

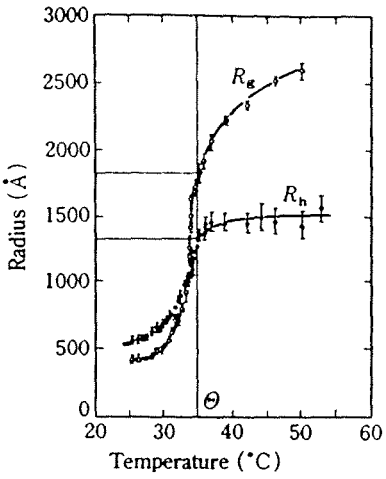



Figure 10. The coil–globule transition in a solution of polystyrene in cyclohexane. The radius of gyration $R_g$ and the hydrodynamic radius $R_h$ undergo a dramatic change when the temperature passes the $\theta$–temperature.

A last remark about the free energy used, (3.5) with transformation (3.10), is in order. When we look at the last term of (3.5) we can write it slightly differently in terms of the segment concentration $c \approx N/R^3$:

$$\frac{N^2\nu}{2R^3} \approx Nc\frac{1}{2}\nu$$

Written in this form the term strongly resembles the second virial term in the expansion of the free energy of a non–ideal gas as a series of concentration $c$. This is no coincidence since the above model describes the interaction between the polymer segments in terms of two–particle interactions and as if the segments can move independently within the coil. The factor $\frac{1}{2}\nu$ corresponds with the second virial coefficient. In the case of excluded–volume interactions only it is equal to $\frac{1}{2}\nu_c$, half the excluded volume of one segment as in the case of gases. We now see the close connection with the remarks made in section 2.9.

# 4. CONCENTRATED POLYMER SOLUTIONS

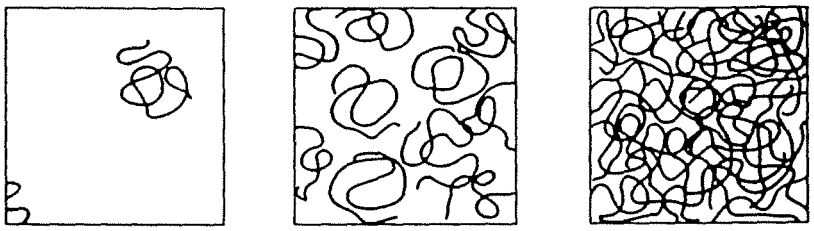



Figure 11. (a) A dilute polymer solution; (b) a solution at the overlap concentration $c^*$; (c) a more concentrated solution.

If we increase the concentration of a polymer solution separate coils will start to overlap at a certain point (see figure 11). The corresponding segment concentration is called the *overlap concentration* $c^*$. We can easily make an estimate of $c^*$ since at that point the total space will be completely filled by coils touching each other. This implies that $c^*$ is equal to the (average) segment concentration within 1 isolated coil:

$$c^* \approx \frac{N}{R^{*3}} \approx \frac{N^{1-3\nu}}{b^3} \tag{4.1}$$

Note that, $\nu$ lying between 1/2 and 3/5, $c^*$ can be very small for long polymers (large $N$). Polystyrene at a molecular weight of 1,000,000 can have its $c^*$ at 0.5 weight %. This means that polymer chains can be strongly entangled and will have a strong mutual interactions. The limit is formed by molten polymers (called a polymer melt), where a solvent is completely absent, and which are very important for industrial applications of plastics.

## 4.1. The Flory-Huggins approximation

A conceptually important description of polymer solutions is the well-known *Flory-Huggins approximation*. A derivation is given in Appendix 7.5. This theory is an example of a so-called mean-field theory, where polymer segments are assumed to be randomly distributed in space and the interaction is calculated on the basis of averaged concentrations in the system. Fluctuations in segment concentrations are completely neglected, although they can be large especially for polymers (if only because of the fact that segments are attached to each other). The Flory-Huggins approxximation is therefore better suited for concentrated solutions (far above $c^*$). Here we are mainly interested in qualitative effects of the degree of polymerization $N$.

We consider a system of $n_p$ polymers of $N$ segments and $n_s$ solvent molecules (occupying $\Omega = n_p N + n_s$ lattice positions) thus having a volume fraction $\phi = n_p N/\Omega$.

The (Helmholtz) free energy of mixing $A_m$ within the Flory–Huggins approximation is

$$A_m(\Omega,\phi) = \Omega k_B T\, f_m(\phi) = \Omega k_B T \left[\frac{1}{N}\phi\ln\phi + (1-\phi)\ln(1-\phi) + \chi\phi(1-\phi)\right] \quad (4.2)$$

The first two terms are the entropy of mixing. Here the factor $1/N$ is noticeable (in the "polymer" term), which is absent for low molecular weight liquids. The last term represents the internal energy of mixing analogous to the model considered previously for the single chain (see (3.8)) and contains the same chi-parameter

$$\chi \equiv \frac{z\Delta\varepsilon}{k_B T} = \frac{z(\varepsilon_{pp} + \varepsilon_{ss} - 2\varepsilon_{ps})}{2k_B T} \quad (4.3)$$

## 4.2. Phase separation in polymer solutions

To be able to predict from the Flory–Huggins free energy whether a polymer solution remains homogeneous or that phase separation occurs, we use a graphical method. If we plot the free energy per lattice point, $f_m(\phi)$ (see (4.2)), against the volume fraction $\phi$ we could for instance obtain a picture like figure 12.

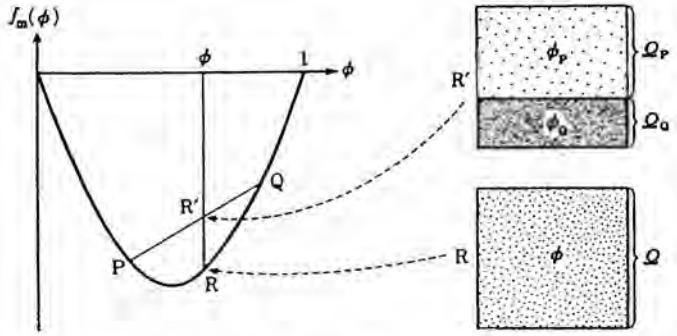



Figure 12. The free energy of mixing for a system without phase separation. Situation at R: a homogeneous phase with volume $\Omega$ and volume fraction $\phi$. (Imaginary) situation at R': separation in two phases with concentrations $\phi_P$, $\phi_Q$ and volumes $\Omega_P$, $\Omega_Q$.

If this system is homogeneous, the free energy per lattice point at volume fraction $\phi$ is given by the value of $f_m(\phi)$ at point R (here we consider all energies in units $k_B T$). However, if we would suppose that the system separates in 2 phases P and Q of volume fractions $\phi_P$ and $\phi_Q$, we first have to determine the volumes $\Omega_P$ and $\Omega_Q$ of both phases from the following conservation conditions of mass and volume:

$$\begin{aligned} \Omega_P\phi_P + \Omega_Q\phi_Q &= \Omega\phi \\ \Omega_P + \Omega_Q &= \Omega \end{aligned}$$

Its solution is:

$$\Omega_P = \frac{\phi_Q - \phi}{\phi_Q - \phi_P}\Omega \quad (4.4)$$

$$\Omega_Q = \frac{\phi - \phi_P}{\phi_Q - \phi_P}\Omega \quad (4.5)$$

Hence the free energy per lattice point of the phase–separated system is

$$\begin{aligned}\frac{A_m^{separated}}{\Omega k_B T} &= \frac{\Omega_P}{\Omega} f_m(\phi_P) + \frac{\Omega_Q}{\Omega} f_m(\phi_Q) \\ &= \frac{\phi_Q - \phi}{\phi_Q - \phi_P} f_m(\phi_P) + \frac{\phi - \phi_P}{\phi_Q - \phi_P} f_m(\phi_Q)\end{aligned}$$

A more precise consideration of this expression shows that its value (per lattice point) in figure 12 lies on the line connecting P and Q at a volume fraction of $\phi$, i.e. at R' (expressions (4.4) and (4.5) represent the lever rule for the relative volumes). In this case the free energy of the phase–separated system is higher than the homogeneous system. For an upward concave function like in figure 12, this is true for every possible way of separating the system in 2 phases and the homogeneous system is always the most stable.

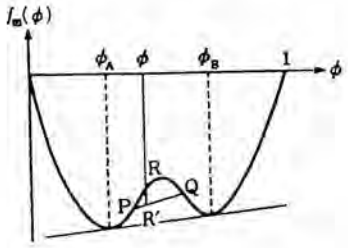



Figure 13. The form of the free energy of mixing for a case of phase separation. Solutions at $\phi_A < \phi < \phi_B$ find there lowest free energy if the system separates in 2 phases of concentrations $\phi_A$ and $\phi_B$.

A very different situation occurs when there is a convex part in the graph, like in figure 13. Here R' lies clearly below R. P and Q can be chosen in many different ways. However, the lowest possible point is situated on the double tangent to the curve and this is therefore the equilibrium state with the two coexisting phases as the tangent points (this well–known double–tangent construction also assures that both tangent points have equal osmotic pressure and chemical potentials).

In (4.2) the value of $\chi$ determines the form of the $f_m(\phi)$–curve. At low values of $\chi$ (usually at high temperatures $T$) its form is upward concave and the system remains homogeneous. At higher values of $\chi$ (low $T$) 2 minima appear and phase separation takes place. This is represented in figure 14 for different values of $T$.

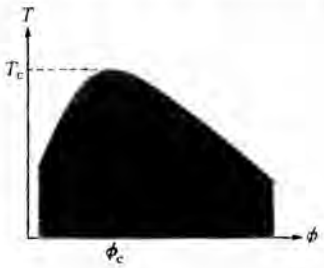



Figure 14. A typical phase diagram for a polymer solution. In the cross–hatched region phase separation occurs.

The $\chi$ value at which phase separation first ocurs can be found from the fact that both minima in figure 13 shift towards each other and merge at this specific value ($\chi = \chi_c$). This requires:

$$\frac{\partial^2 f_m}{\partial \phi^2} = 0 \text{ and } \frac{\partial^3 f_m}{\partial \phi^3} = 0$$

This point is called the *critical point* and can easily be determined from (4.2):

$$\begin{aligned} \phi_c &= \frac{1}{1+\sqrt{N}} \\ \chi_c &= \frac{1}{2}\left(1+\frac{1}{\sqrt{N}}\right)^2 \end{aligned}$$

If $N$ increases, these formulas predict that $\chi_c$ approaches $1/2$ so that the critical temperature $T_c = z\Delta\varepsilon/k_B\chi_c$ rises and the critical volume fraction decreases. This tendency is clearly visible in figure 15. However, no quantitative agreement with experiments is found. This is understandable, since in the neighbourhood of a critical point particularly strong concentration fluctuations take place, whence a mean-field theory can not be expected to give a good description.

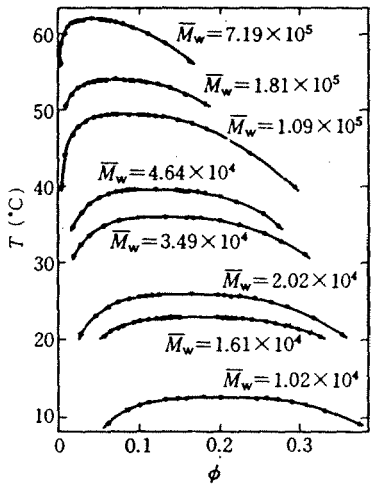



Figure 15. Coexistence curves for a solution of polystyrene in methylcyclohexane. The $\theta$-temperature in this system is 70.3°C.

A further indication of the failing of the Flory–Huggins theory is the osmotic pressure. This can be derived from the Helmholtz free energy:

$$\begin{aligned} \Pi &= \frac{k_B T}{\nu_c}\left[-\ln(1-\phi) + \frac{\phi}{N} - \phi - \chi\phi^2\right] \\ &\approx \frac{k_B T}{\nu_c}\left[\frac{\phi}{N} + \left(\frac{1}{2} - \chi\right)\phi^2 + \cdots\right] \end{aligned} \tag{4.6}$$

In figure 16 the experimental osmotic pressure is plotted for a number of molecular weights of the same polymer. Note that Van 't Hoff's law (the first term in (4.6)) is only reached at the lowest molecular weight (in this representation the curve should be

horizontal in that regime). At higher concentrations the second term dominates and does not depend on the molecular weight according to (4.6). This is in fact observed but the slope is larger than predicted by Flory–Huggins. This effect is attributed to fluctuations which we shall discuss later.

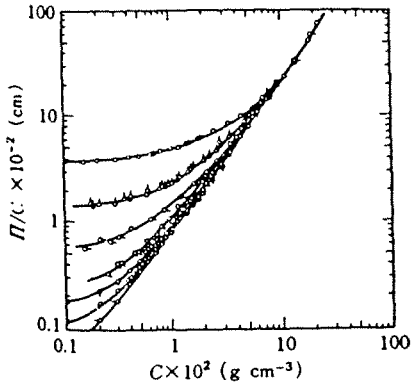



Figure 16. The concentration dependence of the osmotic pressure of poly($\alpha$–methylstyrene) molecules of different molecular weights dissolved in toluene. From the top to the bottom molecular weights of: $7 \times 10^4$, $20 \times 10^4$, $50.6 \times 10^4$, $7 \times 10^4$, $119 \times 10^4$, $182 \times 10^4$, $330 \times 10^4$, $747 \times 10^4$.

# 5. SCALING THEORY

We shall conclude these lecture notes with a more qualitative discussion, where we give an indication of a number of more modern ideas within polymer theory. If we look at the Gaussian distribution function (2.7) for a chain of $N$ segments with step length $b$, it is striking that $N$ and $b$ enter only as a fixed combination $Nb^2$. One could say that there exists a length scale $N^{1/2}b$ that determines the total distribution and therefore also all equilibrium properties on the level of the total chain. The same combination is found within the expression for the entropy of the chain (2.12). An ideal Gaussian coil has the particular property that its structure reproduces itself at different magnifications (see figure 17). This is the property of self–similarity of a so–called *fractal* object.

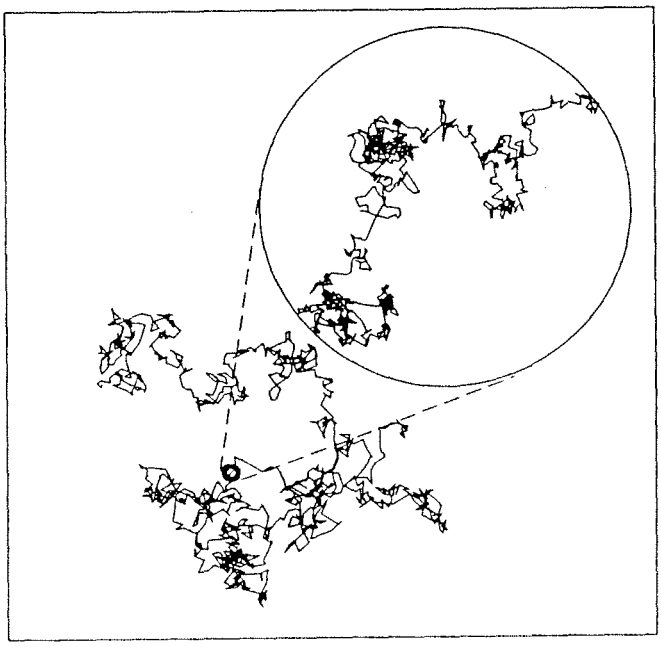


Figure 17. A computer simulation of a "random walk" of $10^6$ random steps. In the figure every $10^3$ steps are represented as one segment. The inset zooms in on one such segment and represents all steps. The structure of the chain at both levels is completely analogous (self–similar or fractal).

The picture is getting clearer when we ask ourselves whether the monomers within a polymer chain are really determined unequivocally. From a chemical point of view the monomer will of course be the smallest repetition unit within the chain, but we may as well combine several monomers and subsequently use these as the building

block of the polymer ending up with the same polymer. To be more concrete: a chain of $N$ segments and step length $b$ has an average end–to–end distance $N^{1/2}b$. But if we first combine $g$ segments into a new building block, the (average) distance the chain spans within each building block will be $b_g \equiv g^{1/2}b$ which now forms our new step length. On the other hand we only have $N_g \equiv N/g$ of these new building blocks. Our newly built coil is a random walk of the new building blocks (see figure 18) for which we determine the average end–to–end distance in the usual way: $N_g^{1/2}b_g$. The transformation may be summarized as follows:

$$\begin{aligned} \text{step length} &: b \longrightarrow b_g \ \left(\equiv g^{1/2}b\right) \\ \text{number} &: N \longrightarrow N_g \ (\equiv N/g) \\ \text{end–to–end distance} &: N^{1/2}b \longrightarrow N_g^{1/2}b_g \ \left(= N^{1/2}b\right) \text{ invariant!} \end{aligned}$$

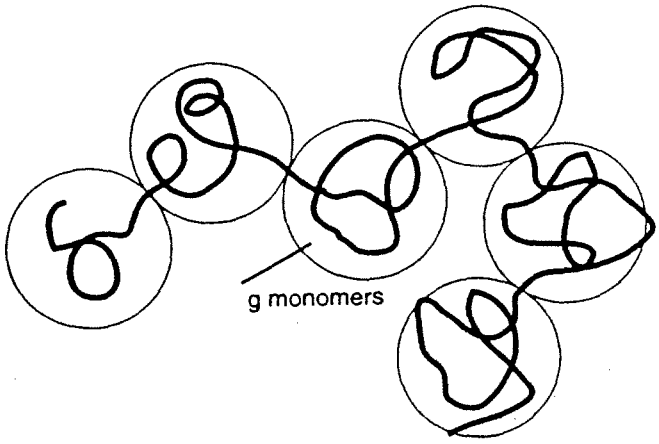



Figure 18. An arbitrary number of segments $g$ in a polymer chain may be thought of as a new segment.

The same story also holds for a non–ideal chain with end–to–end distance $N^{\nu}b$. P.G. de Gennes applied this type of ideas to determine all kinds of properties of polymers in a simple fashion. Consider for instance the situation that we put a polymer chain in a good solvent ($\nu = 3/5$) into a capillary of diameter $D$. Can a simple expression be given for the length $L$ which the polymer will occupy within the capillary? We have seen that the size of the building blocks of the polymer can be chosen at will. In this case de Gennes chose the step length equal to the capillary diameter: $bg^{3/5} = D$ so that $g = (D/b)^{5/3}$. The $N/g$ new building blocks of size $D$ will have an excluded–volume interaction like the original segments, so that they only fit into the capillary one behind the other. This gives as the total length of the chain within the capillary:

$$L \approx D\left(\frac{N}{g}\right) \approx ND^{-2/3}b^{5/3} \tag{5.1}$$

It is striking that this calculation is supported by much more involved calculations and simulations. Two limiting cases of the formula can be immediately checked. Firstly, if we consider a wide capillary with a diameter equal to the size of a coil outside the capillary, or $D = N^{3/5}b$. Substitution in (5.1) now gives $L \approx N^{3/5}b$,

hence in a wide capillary the chain is not distorted (for even larger values of $D$ (5.1) of course no longer applies). The second limit is a very narrow capillary of diameter $D = b$. Now the chain is prevented from folding back and can only lie completely stretched. Formula (5.1) indeed gives $L \approx Nb$. The behaviour of $L$ as a function of $D$ is schematically represented in figure 19.

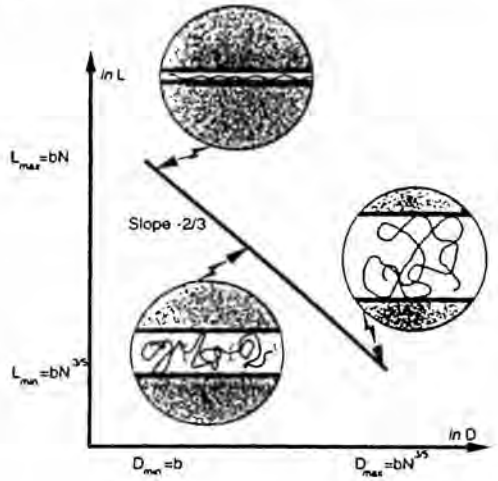



Figure 19. A double–logarithmic plot of length $L$ occupied by a chain in a capillary of diameter $D$ (see text).

As we have seen from (2.8) a polymer coil is a strongly fluctuating object. This kind of fluctuations is completely neglected in mean–field theory, since there only averaged segment concentrations are used. In scaling theory fluctuations are taken along. More formal scaling theory uses an analogy between the behaviour of systems near to a critical point in phase transitions (which also show large fluctuations) and polymer systems (where $1/N$ is comparable with the relative distance to the critical point). One of its ingredients is the above–described procedure of redefining the segments of a chain (also called renormalization). A typical feature in these theories is the prediction of all kinds of power laws containing the relative distance to the critical point. In that sense an expression like

$$R^* \approx bN^{\nu} \text{ where } \nu \approx 0.588$$

is such a power law where $1/N$ is raised to the power $-0.588$ (the critical exponent). On the basis of this type of theories a much simpler description of scaling laws (=power laws) is built, using simple physical considerations in a more intuitive way. Nobel prize laureate P.G. de Gennes has been the pioneer of this approach. In the next section we apply this description to semi–dilute solutions.

## 5.1. Semi–dilute solutions in a good solvent

As we have seen in chapter 4 coils start to overlap around

$$\phi^* = c^* b^3 \approx Nb^3/R^{*3} \approx N^{1-3\nu} \approx N^{-4/5}$$

(note that also $c^*$ is only determined by a global length scale $R^*$). Since every coil is a strongly fluctuating object, we also expect strong fluctuations in a semi–dilute solution near $\phi^*$. Therefore the mean–field description (4.6) is no longer applicable. The scaling expression for the osmotic pressure is now:

$$\Pi = k_B T \frac{c}{N} f(\phi/\phi^*) \tag{5.2}$$

We recognize the ideal law (Van 't Hoff's law) valid at very small $\phi$, multiplied by a function $f$ of the relative distance to the overlap concentration $\phi/\phi^*$. For $\phi \to \phi^*$ this function approaches 1 and at higher concentrations it follows a power law with exponent $m$:

$$f(x) \sim x^m \tag{5.3}$$

This form is justified by the above–mentioned more extensive theories. The symbol $\sim$ is used for scaling relations and implies that prefactors are left out. Note that $f(x)$ goes to 1 for $x \to 1$ (or $\phi \to \phi^*$ ). Exponent $m$ is now determined from an additional physical condition, i.e. that $\Pi$ can not depend on $N$ far into the semi–dilute region. An entangled collection of very long polymers contains after all a very small concentration of end points, so that variation in their number can only have a very small influence on the osmotic pressure. Combining the 3 previous equations then gives $m = 1/(3\nu - 1) = 5/4$ and

$$\Pi \sim \frac{k_B T}{b^3} \phi^{9/4} \tag{5.4}$$

This law corresponds to the region in figure 16 where all curves coincide (indeed independent of $N$) and shows a stronger dependence on $\phi$ than the mean–field result, which is also found experimentally.

The above derivation shows the typical structure used to derive a scaling law. The unknown quantity is written as a known law in the trivial regime (here Van 't Hoff's law in very dilute solutions) times a scaling function $f$. Now a characteristic parameter is assumed (here the segment concentration), which serves as an argument in the scaling function $f$, and the law is written analogously to equation (5.2). The scaling function $f$ has the form of a power law (5.3) and connects to the trivial regime. An additional physical argument (here the independence on $N$) finally fixes the value of the exponent.

A similar derivation can be given for the characteristic length scale $\xi$ in a semi–dilute solution:

$$\xi = N^\nu b f(\phi/\phi^*) \tag{5.5}$$

As we have seen before the characteristic length scale in the dilute regime is $N^\nu b$. We again suppose a power law:

$$f(x) \sim x^m \tag{5.6}$$

We use the same physical argument that the structure far into the semi–dilute solution does not depend on $N$. This now leads to $m = -\nu/(3\nu - 1) = -3/4$, whence we find

$$\xi \sim b\phi^{-3/4} \tag{5.7}$$

## 5.2. Physical picture of a semi–dilute solution

Both results (5.4) and (5.7) are very important to understand a semi–dilute solution. In this section we try to elaborate upon this picture. The *mean–field* result for long polymers in a semi–dilute solution in a good solvent ($\chi = 0$) is given by (4.6)

$$\Pi \approx \frac{k_B T}{\nu_c} \left[ \frac{\phi}{N} + \frac{1}{2}\phi^2 + \cdots \right] \sim \frac{k_B T}{b^3}\phi^2 \tag{5.8}$$

The osmotic pressure is mainly determined by interactions between segments (the ideal part of the entropy is very small because segments are connectedinto very large units). The square of $\phi$ in the above expression represents the probability of contacts (mainly pair contacts), viz. the probability $\phi$ to find 1 segment times the probability $\phi$ to find a second segment nearby.

*Scaling theory* (5.4) now gives the probability to find a second segment not as $\phi$ but as $w \sim \phi^{5/4}$: a smaller number. This comes about because around one segment in a polymer chain there is always a "cloud" of other segments (of the same chain) that reduces the probability of finding a second segment: at a given contact between 2 segments the surrounding segment clouds also have to come in contact leading to a stronger repulsion.

Given this probability $w \sim \phi^{5/4}$ of a contact between 2 segments we can immediately calculate how many segments $g$ will lie between 2 encounters of one specific chain with other chains:

$$g \sim \phi^{-5/4}$$

If we use this to calculate the ("end–to–end") distance between consecutive encounters in a good solvent this gives

$$g^{\nu} b \sim b\phi^{-3/4}$$

This "*mesh size*" of the polymeric network is equal to the characteristic length $\xi$ (5.7). A chain part of $g$ segments with total size $\xi$ was called a "*blob*" by de Gennes (see figure 20).

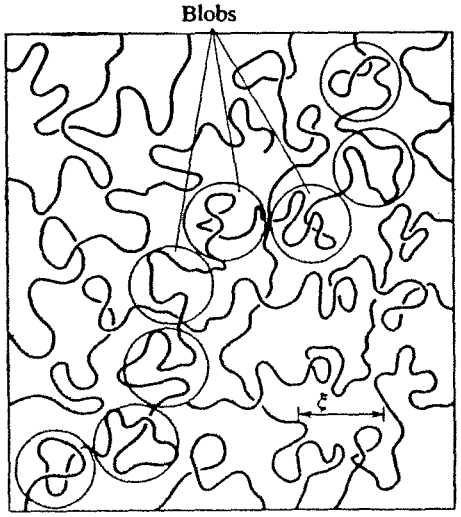



Figure 20. A realization of the conformations in a semi–dilute polymer solution as a network with mesh size $\xi$ or as a system of blobs.

Let us determine the segment concentration within such a blob. This is simple: there are $g$ segments in a volume $\xi^3$ which makes the concentration $g/\xi^3 \sim \phi/b^3$. However, this is equal to the average segment concentration $c$, so we conclude that blobs do touch each other but do not overlap. This picture is confirmed if we rewrite the osmotic pressure (5.4):

$$\Pi \sim \frac{k_B T}{b^3}\phi^{9/4} \sim k_B T \frac{c}{g} \sim \frac{k_B T}{\xi^3}$$

A semi–dilute polymer solution is an ideal gas of blobs! $\xi$ apparently also can be interpreted as a "*screening length*" for the excluded–volume interaction: after a chain comes into contact with another polymer chain it seems as if it forgets which segments are on the same chain and which on neighbouring chains. Beyond this length scale $\xi$ the chain will be ideal.

We can now calculate the end–to–end distance for 1 chain: there are $N/g$ blobs with "step length" $\xi$ that themselves form an ideal chain. This implies

$$R \sim \left(\frac{N}{g}\right)^{1/2} \xi \sim N^{1/2} b \phi^{-1/8} \tag{5.9}$$

This relation with the volume fraction is experimentally verified. Note that this expression approaches the completely ideal expression $R \to N^{1/2}b$ when $\phi \to 1$. The same is seen from (5.7): $\xi \to b$. This is caused by the fact that in a melt no concentration fluctuations are possible and there is already an encounter with a different chain after a single step, so that the chain loses its memory right away. The fact that a polymer chain in a melt behaves ideally is also called *Flory's theorem.*

# 6. SUMMARY

To end these lecture notes it is appropriate to give the following diagram from de Gennes' book, which summarizes many of the results we encountered:

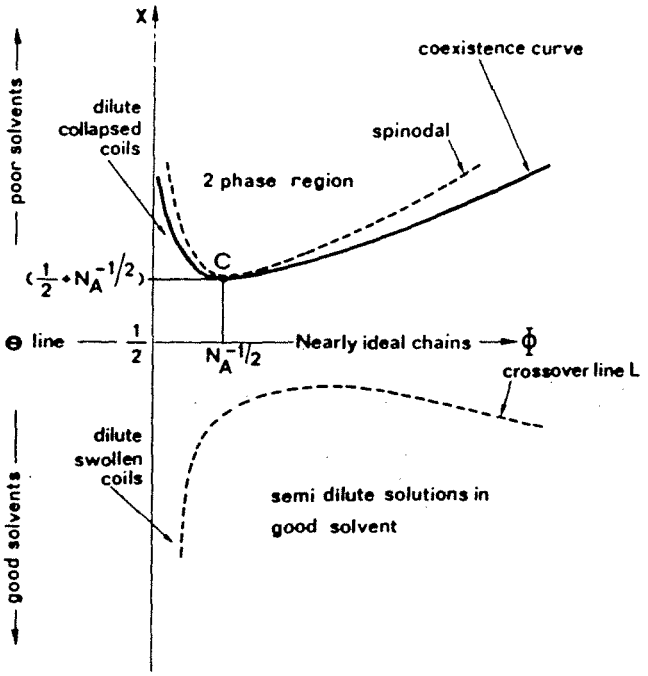



**Figure IV.8.**

Phase diagram for a polymer-solvent system. $\chi$ is the Flory interaction parameter, and $\Phi$ is the volume fraction occupied by the polymer. The condition $\chi = 1/2$ defines the Flory $\Theta$ temperature. In usual cases such as polystyrene-cyclohexane, $\chi$ is a decreasing function of the temperature $T$; high temperatures correspond to the lower part of the diagram.

# 7. APPENDIX

## 7.1. Derivation of the diffusion equation for an ideal chain

In order to derive from relation (2.3)

$$P(\mathbf{R}, N) = \langle P(\mathbf{R} - \mathbf{r}_N, N-1) \rangle_{\mathbf{r}_N} \tag{7.1}$$

an expression for $P(\mathbf{R}, N)$, we use the fact that $N$ is large. This implies that 1 is small with respect to $N$ and $\mathbf{r}_N$ is small with respect to $\mathbf{R}$, so that $P(\mathbf{R}, N)$ varies only gradually and it makes sense to make a Taylor expansion of the function on the right–hand side around $\mathbf{R} = (x, y, z)$ and $N$:

$$\begin{aligned} P(\mathbf{R} - \mathbf{r}_N, N-1) \approx\ & P(\mathbf{R}, N) + \frac{\partial P}{\partial N}(-1) + \sum_{\alpha=x,y,z} \frac{\partial P}{\partial \alpha}(-\mathbf{r}_{N,\alpha}) \\ & + \frac{1}{2} \sum_{\alpha=x,y,z} \sum_{\beta=x,y,z} \frac{\partial^2 P}{\partial \alpha \partial \beta}(-\mathbf{r}_{N,\alpha})(-\mathbf{r}_{N,\beta}) + \cdots \end{aligned} \tag{7.2}$$

Since this function actually depends on 4 variables $(N, x, y, z)$ with respect to which we expand, this simple Taylor expansion looks slightly terrifying (we only write down those terms that we ultimately need; i.e. to quadratic order for the derivatives with respect to $\mathbf{R}$ since the linear term turns out to cancel). We must now average this expression over all directions of $\mathbf{r}_N$, using the fact that there is no preferential direction for $\mathbf{r}_N$:

$$\begin{aligned} \langle \mathbf{r}_{N,\alpha} \rangle &= 0 \\ \langle \mathbf{r}_{N,\alpha} \mathbf{r}_{N,\beta} \rangle &= \langle \mathbf{r}_{N,\alpha} \rangle \langle \mathbf{r}_{N,\beta} \rangle = 0 \text{ als } \alpha \neq \beta \\ \langle \mathbf{r}_{N,x}^2 \rangle &= \langle \mathbf{r}_{N,y}^2 \rangle = \langle \mathbf{r}_{N,z}^2 \rangle = \frac{1}{3} \langle \mathbf{r}_N^2 \rangle = \frac{1}{3} b^2 \end{aligned} \tag{7.3}$$

Combining (7.1), (7.2) and (7.3) finally leads to to the following differential equation for $P(\mathbf{R}, N)$:

$$\frac{\partial P}{\partial N} = \frac{1}{6} b^2 \Delta P$$

where $\Delta$ indicates the Laplacian:

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Note that this derivation would in fact also be valid if we would not go back just one but several steps along the chain (although a small number compared to the total number of segments) as long as conditions (7.3) can be fulfilled. This is the case for a chain with a so–called short–range memory. Now $b$ represents an effective segment. This substantiates the statements of the opening section of 2.4.

## 7.2. Solution for the probability of a Gaussian chain

If we want to find the solution for (2.4), this can be easily done by using Fourier transformation

$$P(\mathbf{k}) \equiv \int e^{i\mathbf{k}\cdot\mathbf{R}} P(\mathbf{R})\, d\mathbf{R}$$

and its reverse

$$P(\mathbf{R}) \equiv \frac{1}{(2\pi)^3} \int e^{-i\mathbf{k}\cdot\mathbf{R}} P(\mathbf{k})\, d\mathbf{k}$$

Taking the Fourier transform of (2.4) gives a simple differential equation for its Fourier transform

$$\frac{\partial P(\mathbf{k}, N)}{\partial N} = -\frac{1}{6} b^2 \mathbf{k}^2 P(\mathbf{k}, N)$$

with the simple solution

$$P(\mathbf{k}, N) = \exp\left(-\frac{N b^2 \mathbf{k}^2}{6}\right) \tag{7.4}$$

which reduces to (2.7) on reverse transformation.

## 7.3. The differential equation for a Gaussian chain in an external field

Here we consider a slightly more general case than in section 7.1 by placing a polymer chain in an external field. We define $\varphi(\mathbf{R})$ as the energy a segment obtains at place $\mathbf{R}$ due to the interaction with this external field. We now have to modify (7.1) by taking into account the Boltzmann factor connected with this energy:

$$P(\mathbf{R}, N) = \langle P(\mathbf{R} - \mathbf{r}_N, N-1) \rangle_{\mathbf{r}_N} \exp(-\varphi(\mathbf{R})/k_B T)$$

In a Taylor expansion like (7.2) we must now also linearize this Boltzmann factor (this means that the segment energy must be much smaller than $k_B T$; however, this may always be achieved by our freedom of dividing the chain into more segments) giving

$$\frac{\partial P}{\partial N} = \frac{1}{6} b^2 \Delta P - \frac{\varphi(\mathbf{R})}{k_B T} P \tag{7.5}$$

Equation (7.5) is similar to a diffusion equation with an external field. In a slightly more general form (where a chain starts at point $\mathbf{R}'$ instead of the origin) $P$ is replaced by the so–called Green function $G = G_N(\mathbf{R}|\mathbf{R}')$, which has the character of a conditional probability: the probability of finding the end of a chain of $N$ segments at point $\mathbf{R}$ given that it starts at point $\mathbf{R}'$. $\Delta$ should be understood to be the second derivative to $\mathbf{R}$ (as opposed to $\mathbf{R}'$). The negative of this form is

$$-\frac{\partial G}{\partial N} = -\frac{1}{6} b^2 \Delta G + \frac{\varphi(\mathbf{R})}{k_B T} G \tag{7.6}$$

and bears a remarkable resemblance to the time–dependent Schrödinger equation for the wave function $\psi(\mathbf{R},t)$ (of a particle of mass $m$ in an external potential $V(\mathbf{R})$) in quantum mechanics

$$-i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\psi + V(\mathbf{R})\psi$$

with $N$ now in the role of an imaginary time $it/\hbar$.

We will now follow the lines of argumentation of quantum mechanics in obtaining a number of useful results for our equation (7.6). A standard solution method of a partial differential equation like (7.6), called *separation of variables*, assumes a solution of the form:

$$G = f(N)\psi(\mathbf{R}) \tag{7.7}$$

Substituting in (7.6) and rearranging gives

$$-\frac{1}{f}\frac{df}{dN} = \frac{1}{\psi}\left(-\frac{1}{6}b^2\Delta\psi + \frac{\varphi(\mathbf{R})}{k_BT}\psi\right)$$

of which the left–hand side only depends on $N$ and the right–hand side only on $\mathbf{R}$. Since this would imply that both sides can be varied independently. However, they should always be equal so we conclude that both sides must be equal to a constant, say $\lambda$:

$$\begin{aligned}\frac{df}{dN} &= -\lambda f \\ -\frac{1}{6}b^2\Delta\psi + \frac{\varphi(\mathbf{R})}{k_BT}\psi &= \lambda\psi\end{aligned}$$

The solution of the first equation is simply

$$f(N) = c\exp(-\lambda N) \tag{7.8}$$

The second equation is an eigenvalue equation completely analogous to the *time–independent Schrödinger equation*:

$$-\frac{\hbar^2}{2m}\Delta\psi + V(\mathbf{R})\psi = E\psi$$

so we must now find the eigenvalues $\lambda_n$ and eigenfunctions $\psi_n(\mathbf{R})$ for this equation, obeying

$$-\frac{1}{6}b^2\Delta\psi_n + \frac{\varphi(\mathbf{R})}{k_BT}\psi_n = \lambda_n\psi_n \tag{7.9}$$

Here we review a number of properties of the eigenfunctions of such an equation, for simplicity restricting ourselves to the case that the eigenfunctions are real. It is possible to construct a complete set of orthonormal eigenfunctions, with the property

$$\int \psi_n\psi_m \, d\mathbf{R} = \delta_{nm} \tag{7.10}$$

($\delta_{nm} = 1$ if $n = m$ and 0 otherwise). This can be easily proved for different eigenvalues $\lambda_n \neq \lambda_m$ using (7.9):

$$\begin{aligned}(\lambda_n - \lambda_m)\int \psi_n\psi_m \, d\mathbf{R} &= \int (\psi_m\lambda_n\psi_n - \psi_n\lambda_m\psi_m)\, d\mathbf{R} \\ &= -\frac{1}{6}b^2\int (\psi_m\Delta\psi_n - \psi_n\Delta\psi_m)\, d\mathbf{R} \\ &= 0\end{aligned}$$

The last step is obtained by integration by parts. Since $\lambda_n \neq \lambda_m$ this proves that $\int \psi_n \psi_m \, d\mathbf{R} = 0$. Further proof can be found in books about quantum mechanics.

Combining (7.7), (7.8) and the solutions to (7.9) gives:

$$G_N(\mathbf{R}|\mathbf{R}')_n = c_n(\mathbf{R}')\psi_n(\mathbf{R}) \exp(-\lambda_n N)$$

where $c_n(\mathbf{R}')$ indicates that the integration constant may be different for every $n$ and can also depend on $\mathbf{R}'$. Since (7.6) is a homogeneous, linear equation, linear combinations of its solutions are also solutions:

$$G_N(\mathbf{R}|\mathbf{R}') = \sum_n c_n(\mathbf{R}')\psi_n(\mathbf{R}) \exp(-\lambda_n N)$$

which is the general solution, since the $\psi_n$s form a complete set. This expression can be further simplified since a chain starting at $\mathbf{R}'$ must have the same probability of ending at $\mathbf{R}$ as a chain starting at $\mathbf{R}$ to end at $\mathbf{R}'$: the expression must not change upon interchanging $\mathbf{R}$ and $\mathbf{R}'$. This implies

$$G_N(\mathbf{R}|\mathbf{R}') = \sum_n a_n \psi_n(\mathbf{R}')\psi_n(\mathbf{R}) \exp(-\lambda_n N)$$

In the limit of very short chains ($N \to 0$) of this expression, we must require that the chain ends at the same position where it starts

$$\lim_{N \to 0} G_N(\mathbf{R}|\mathbf{R}') = \sum_n a_n \psi_n(\mathbf{R}')\psi_n(\mathbf{R}) = \delta(\mathbf{R} - \mathbf{R}')$$

Multiplying with $\psi_m(\mathbf{R})$, integrating over $\mathbf{R}$, and applying (7.10) gives $a_m = 1$ for all $m$. So our final expression is:

$$G_N(\mathbf{R}|\mathbf{R}') = \sum_n \psi_n(\mathbf{R}')\psi_n(\mathbf{R}) \exp(-\lambda_n N) \tag{7.11}$$

If we allow complex eigenfunctions we have to insert a complex conjugate $*$:

$$G_N(\mathbf{R}|\mathbf{R}') = \sum_n \psi_n^*(\mathbf{R}')\psi_n(\mathbf{R}) \exp(-\lambda_n N) \tag{7.12}$$

We can check this formula for the known case of $\varphi(\mathbf{R}) = 0$. Then the solutions of (7.9) are:

$$\begin{aligned} \psi_{\mathbf{k}}(\mathbf{R}) &= \exp(i\mathbf{k} \cdot \mathbf{R}) \\ \lambda_{\mathbf{k}} &= \frac{1}{6} b^2 \mathbf{k}^2 \end{aligned}$$

valid for every value of $\mathbf{k}$. This means that there is a continuous spectrum of eigenvalues and the sum over $n$ in (7.11) reduces to an integral over $\mathbf{k}$:

$$G_N(\mathbf{R}|\mathbf{R}') = \int \exp(i\mathbf{k} \cdot (\mathbf{R} - \mathbf{R}')) \exp(-\frac{1}{6} b^2 \mathbf{k}^2 N) \, d\mathbf{k}$$

which is the Fourier transform of the usual Gaussian distribution.

## 7.4. Gaussian integrals

Integrals containing Gaussian functions $\exp(-Ax^2)$ can be derived in a simple way from the following standard integral

$$I(A) \equiv \int_{-\infty}^{+\infty} e^{-Ax^2} dx = \sqrt{\frac{\pi}{A}}$$

By taking its derivative with respect to $A$ we obtain integrals of the product with even powers of $x$, e.g.

$$\int_{-\infty}^{+\infty} x^2 e^{-Ax^2} dx = -\frac{dI(A)}{dA} = -\frac{d}{dA}\sqrt{\frac{\pi}{A}} = \frac{1}{2}\frac{\pi^{1/2}}{A^{3/2}}$$

and by taking higher derivatives we generate the higher even powers. Odd powers of $x$ give 0 since the integrand is odd in that case:

$$\int_{-\infty}^{+\infty} x^{2n+1} e^{-Ax^2} dx = 0$$

Using these formulas, averages can be calculated quite easily, e.g.

$$\langle x^2 \rangle = \frac{\int_{-\infty}^{+\infty} x^2 e^{-Ax^2} dx}{\int_{-\infty}^{+\infty} e^{-Ax^2} dx} = \frac{1}{2A}$$

From this last relation we can express $A$ in terms of $\langle x^2 \rangle$. A normalized Gaussian distribution (in 1 dimension) is therefore also written like

$$P(x) = \sqrt{\frac{A}{\pi}} \exp(-Ax^2) = \frac{1}{\sqrt{2\pi \langle x^2 \rangle}} \exp(-x^2/2\langle x^2 \rangle)$$

## 7.5. The Flory-Huggins approximation

In a concentrated solution polymer chains are interpenetrating, so that a description at the level of individual chains does not apply. Here, we use the same lattice model as in 3.2 to determine the free energy in such a concentrated system. We follow the derivation of M. Doi in his *Introduction to Polymer Physics* (Clarendon Press, 1996). We assume that $n_p$ polymers each occupy $N$ consecutive lattice positions and the remainder of the $\Omega$ lattice positions are occupied by $n_s = \Omega - n_p N$ solvent molecules (each occupying 1 lattice position). Therefore, the polymer volume fraction in the system is $\phi = n_p N/\Omega$. If we now write:

$$Z = \sum_i \exp(-E_i/k_B T) \approx W \exp(-\overline{E}/k_B T)$$

the Helmholtz free energy is approximated as:

$$\frac{A}{k_BT} = -\ln Z \approx \frac{\overline{E}}{k_BT} - \ln W \tag{7.13}$$

So, like we did for the single chain, we should determine the total number of configurations of the system $W$ and its average energy $\overline{E}$. For the average energy we again assume random mixing and interaction between adjacent lattice positions only, like in (3.8). In total the number of nearest-neighbour contacts is $\Omega z/2$, each partner in a contact having a probability $\phi$ to be a polymer segment and $(1-\phi)$ to be a solvent molecule. This gives:

$$\overline{E}(\Omega,\phi) \approx -\frac{\Omega z}{2}\left[\varepsilon_{pp}\phi^2 + \varepsilon_{ps}\phi(1-\phi) + \varepsilon_{ps}(1-\phi)\phi + \varepsilon_{ss}(1-\phi)^2\right] \tag{7.14}$$

The total number of polymer configurations $W$ is harder to establish. Let us place the polymer chains on the lattice segment by segment. The first segment of the first polymer can be placed in $\Omega$ ways and each following segment at approximately $z-1$ positions with respect to the previous one. This gives $w_1$, the number of realizations for the first polymer:

$$w_1 = \Omega(z-1)^{N-1}$$

For the $j+1^{\text{st}}$ polymer to be placed the number of possibilities to place its first segment is already less, viz. $\Omega - Nj$, and for each of its next segments we must account for the probability that a lattice point is already occupied (for simplicity we assume that this is the same for every segment within one polymer, viz. $(1-Nj/\Omega)$). This leads to the following expression for the number of realizations for the $j+1^{\text{st}}$ polymer, $w_{j+1}$,

$$w_{j+1} \approx (\Omega - Nj)\left[(z-1)\left(1-\frac{Nj}{\Omega}\right)\right]^{N-1} \approx w_1\left(1-\frac{Nj}{\Omega}\right)^N$$

The total number of ways to place $n_p$ polymers on the lattice is therefore

$$W = \frac{1}{n_p!}\prod_{j=1}^{n_p} w_j$$

The factor $n_p!$ corrects for the fact that the polymer molecules are indistinguishable. The logarithm of $W$ can be calculated simply by transformation to an integral:

$$\begin{aligned}
\ln W &= \sum_{j=1}^{n_p} \ln(w_j/j) \\
&= \int_0^{n_p} \mathrm{d}j \left[\ln\left(\Omega(z-1)^{N-1}\left(1-\frac{Nj}{\Omega}\right)^N\right) - \ln j\right] \\
&= \Omega\left[\frac{\phi}{N} - \frac{\phi}{N}\ln\frac{\phi}{N} - (1-\phi)\ln(1-\phi) + \phi\left(\ln(z-1)-1\right)\right]
\end{aligned} \tag{7.15}$$

Combining (7.13), (7.14) and (7.15) now gives an expression for the free energy of a system of volume fraction $\phi$ occupying $\Omega$ lattice points, $A(\Omega,\phi)$. In general, the free

energy of mixing $A_m$ is used, from which the free energy of the components before mixing has been subtracted

$$A_m(\Omega,\phi) = A(\Omega,\phi) - A(\Omega\phi,1) - A(\Omega(1-\phi),0)$$

This now leads to the celebrated *Flory–Huggins* expression:

$$A_m(\Omega,\phi) = \Omega k_B T \left[\frac{1}{N}\phi\ln\phi + (1-\phi)\ln(1-\phi) + \chi\phi(1-\phi)\right] \qquad (7.16)$$

Note that in this case the usual solute term of the entropy of mixing $\phi\ln\phi$ is multiplied by a factor $1/N$ and the energy of mixing $\chi\phi(1-\phi)$ contains the same chi–parameter as the energy for a single coil (3.9):

$$\chi \equiv \frac{z\Delta\varepsilon}{k_B T} = \frac{z(\varepsilon_{pp} + \varepsilon_{ss} - 2\varepsilon_{ps})}{2k_B T}$$

## 7.6. Literature

These lecture notes are largely based upon a number of standard books about polymer theory. A modern book on a very basic level is:

- A.Yu. Grosberg/A.R. Khokhlov, Giant Molecules (1997, Academic Press, San Diego)

Several concepts are described in a relatively simple way in:

- M. Doi, Introduction to Polymer Physics (1996, Clarendon Press, Oxford)

Two classical books by Flory are:

- P.J. Flory, Principles of Polymer Chemistry (1953, Cornell University Press, Ithaca)
- P.J. Flory, Statistical Mechanics of Chain Molecules (1969, Interscience Publishers, New York)

Another older but useful book is:

- H. Yamakawa, Modern Theory of Polymer Solutions (1971, Harper and Row, New York)

The classical book about scaling theories is:

- P.G. de Gennes, Scaling Concepts in Polymer Physics (1979, Cornell University Press, Ithaca)

The first chapters of the following book give a short introduction to static properties:

- M. Doi/S.F. Edwards, The Theory of Polymer Dynamics (1986, Clarendon Press, Oxford)

Finally, the Russian school of polymer physics is very well described in:

- A.Yu. Grosberg/A.R. Khokhlov, Statistical Physics of Macromolecules (1994, AIP Press, New York)

# 6. DLVO theory & Measurement of Interaction Forces

Alfons van Blaaderen
*Soft Condensed Matter, Debye Institute*
*Utrecht University*

# Contents

## 1. THE DLVO POTENTIAL

It should always be remembered that the origin of forces between colloidal particles is molecular of origin. And although we will soon turn, as one of the many approximations necessary to arrive at results, to a continuous description of matter, it is instructive to write down the different contributions to the interaction potential $V$ between two molecules:

$$V = \text{overlap repulsion (1)} + \text{elctrical multipole - elctrical multipole (2)} + \text{elctrical multipole - induced elctrical multipole (3)} + \text{dispersion (4)} \quad (1)$$

Clearly, all these interactions can in principle be calculated exactly by solving the quantum mechanical Schrödinger equation for the two molecules under consideration. Such a calculation would give both the absolute intramolecular energies and charge distributions. However, this equation can only be solved in (very) simple cases and it is more convenient to interpret the total interaction potential as the sum of the contributions given in Eq. (1). Each contribution depends on molecular properties that can be derived from the Schrödinger equation as well, but some of these terms can also be expressed in a more intuitive (semi) classical form.

Contribution (1) can only be understood in quantum mechanical terms as it represents the repulsion that occurs when the electron clouds of two closed-shell molecules start to overlap. In order to really let that happen the Pauli exclusion principle forbids more than 2 electrons in the same molecular orbital and thus forcing the electrons into excited state orbitals upon approach. This produces a strong increase in energy with the repulsion being approximately proportional to the square of the overlap and increases very strongly as the separation between the molecules decreases. This effect determines a molecule's size ($\sigma$), and also a distance of closest approach in a continuum description of matter. Usually, this overlap repulsion is described as a power law in the distance: $V(r) = (\sigma/r)^n$. For $n = 12$, we obtain the repulsive term in the Lennard-Jones potential and for $n = \infty$, we are in the limit of hard spheres.

Electrical multipole-electrical multipole interactions (2) occur between two molecules that posses net charges or an asymmetrical distribution of electrons or nuclei. This term can be (approximately) described by classical electrostatics.

Electrical multipole-induced multipole interactions (3) occur between a molecule with a permanent electrical multipole and a polarizable molecule, a term which can be described semi classical.

Also the dispersion term (4) can only be described in quantum mechanical terms. It represents the coupling between a spontaneous dipole-induced dipole interaction.

In Eq. (1) there is formally a fifth term that describes charge-transfer interactions between two molecules if they are very close together and if one of the molecules acts as a donator of electrons to the accepting molecule with an electron deficiency. For colloidal interactions this term is unimportant and we will not consider it further.

In the following section we will give some (simplified) expressions for the terms 1-4 from Eq. (1) in the case of two neutral molecules. The mono-pole moments that will be part in the case of charged molecules will not be considered but will be deferred to Section 1.4 and 1.5. Taken together these three simplified terms constitute the Van der Waals forces: the Keesom interactions, the Debye interactions and the London interactions.

### 1.1 Van der Waals forces between two molecules

Before turning to two spheres consisting of so many molecules that we will take their distribution as continuous, we will first consider the Van der Waals forces between two polar and polarizable molecules. As pointed out above, we will only keep simplified expressions. In terms of Eq. (1), we will only consider the second order terms in the full multipole expansion; the mono pole terms are zero because the molecules are uncharged.

But first a few words about the origin of the names associated with these interactions that will turn out all to have an $1/r^6$ dependence on the intermolecular distance $r$.

It was already Newton in 1686 in his *Principia*, as pointed out by Sparnaay [13], who discussed the attraction between two molecules separated by a distance $r$, in terms of a force, proportional to $r^{-n}$, where $n > 4$. He did of course not know the origin of such a force, but could show that if $n$ would be smaller than four the interaction energy of a molecule with a large plate would turn out to be infinitely large. Van der Waals made in 1873 another major step when he treated the equation of state of a gas in his famous thesis [14] and separated out the short-range repulsive forces (term 1 in Eq. (1)), resulting in the excluded volume term ($b$), from the long range attractive forces that are described by the constant $a$:

$$\left(P + \frac{n^2 a}{V^2}\right) \cdot (V - nb) = nRT \tag{2}$$

This is how the term Van der Waals forces came into being and in the beginning of the previous century several workers sought an explanation for these long range forces. The three contributions that turned out to be the most important are named after their inventors Keesom, Debye and London forces.

Keesom interactions are Boltzmann-averaged interactions between two permanent dipoles (term 2 in Eq. (1)). It follows from simple electro statics that two dipoles placed head to toe are in their lowest energy configuration which is given by:

$$V(r) = -\frac{m_1 m_2}{2\pi\varepsilon_o r^3} \tag{3}$$

with $m$ the dipole moment of the molecules and $\varepsilon_0$ the dielectric permittivity of vacuum. To give some idea of the magnitudes of these interactions: two opposite elementary charges separated by a distance of 0.1 nm gives a dipole moment $m$ = (0.1 nm) x ($1.6 \times 10^{-19}$ C) = $1.6 \times 10^{-29}$ C.m = 4.8 D. The Debye is often used for dipole moments and equals $3.336 \times 10^{-30}$ C.m. Permanent dipole moments occur in asymmetric molecules and thus not in single atoms. Water has a dipole moment of 1.85 D. Two dipoles with $m$ = 1 D in their lowest energy configuration have in vacuum an energy of $k_BT$ (with $k_B$ Boltzmann's constant and $T$ the absolute temperature) at a separation distance of 0.36 nm. These figures indicate that dipole-dipole interactions in liquids, where the interactions are reduced by a factor $\varepsilon$ the relative dielectric constant (80 for water), are not strong enough to lead to substantial mutual alignment. Therefore, the interaction energies between two dipolar molecules

can be (Boltzmann) averaged rotationally in order to arrive at the following free energy $w(r)$ (see e.g. [7, 5] for the derivation):

$$w(r) = -\frac{m_1^2 m_2^2}{3(4\pi\varepsilon\varepsilon_o)^2 k_B T r^6} \tag{4}$$

From the above it will be no surprise that the interaction between a polar molecule with a polarizable other molecule with polarizability $\alpha$ also has to be (Boltzmann) orientationally averaged leading to the so-called Debye or induction interactions:

$$w(r) = -\frac{m_1^2 \alpha_2}{(4\pi\varepsilon\varepsilon_o)^2 r^6} \tag{5}$$

Or, more generally if we consider two polarizable molecules with two dipole moments:

$$w(r) = -\frac{m_1^2 \alpha_2 + m_2^2 \alpha_1}{(4\pi\varepsilon\varepsilon_o)^2 r^6} \tag{6}$$

In order to get an intuitive idea of the origin and magnitudes of the (electronic) polarizability $\alpha$, let us imagine a one-electron atom whose electron (charge $e$) circles the nucleus at a distances $R$, which also defines the radius of the atom. If under the influence of an external electric field $E$ the electron orbital is shifted by a distance $l$ from the original orbit around the nucleus (see Figure 1), then we get for the induced dipole moment:

$$m_{\text{ind}} = \alpha.E = l.e \tag{7}$$

The *external* force on the electron due to the field $E$ is given by:

$$F_{\text{ext}} = e.E \tag{8}$$

which must be balanced at equilibrium by the attractive force between the displaced electron orbit and the nucleus and is given by the Coulombic force $e^2/4\pi\varepsilon_o R^2$ projected along the direction of the field (see Figure 1). The *internal* restoring force thus becomes:

$$F_{\text{int}} = \frac{e^2}{4\pi\varepsilon_o R^2} \cdot \sin(\theta) \approx \frac{e}{4\pi\varepsilon_o R^3} \cdot m_{ind} \tag{9}$$

At equilibrium $F_{\text{ext}} = F_{\text{int}}$ and thus

$$m_{\text{nd}} = 4\pi\varepsilon_o R^3 E = \alpha.E \tag{10}$$

whence we obtain for the polarizability:

$$\alpha = 4\pi\varepsilon_o R^3 \tag{11}$$

The unit of (the electronic) polarizability is therefore $4\pi\varepsilon_o$ x (volume) or $C^2.m^2.J^{-1}$ and it is of the order of $4\pi\varepsilon_o$ x (molecular radius)$^3$. For example, water has $\alpha/4\pi\varepsilon_o = 1.48 \times 10^{-30}$ $m^3$, which would lead to a molecular size of 0.114 nm, where 0.114 nm is about 15 % less than the real radius of a water molecule (0.135 nm) [5].

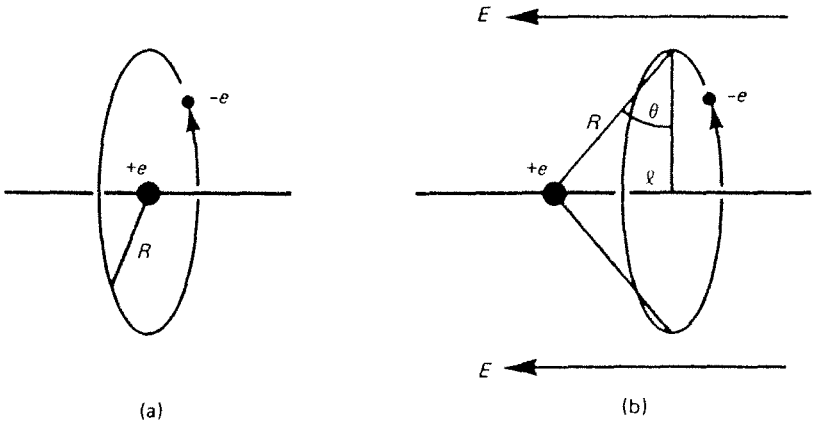



**Figure 1** Induced dipole in a one-electron atom: a) no external electric field, b) in an external electric field with magnitude $E$ which shifts the orbital a distance $l$ from the nucleus, so that the induced dipole moment is $m_{ind} = l.e = \alpha.E$, where the polarizability is given by: $\alpha = 4\pi\varepsilon_o R^3$ [5].

The London or dispersion forces are conceptually the most difficult contribution to the Van der Waals forces because they are of quantum mechanical origin. They could therefore only be derived after the advent of this theoretical framework. Intuitively, their origin may be understood from realizing that for a non polar molecule the time average of its dipole moment might be zero, at every instant there exists a finite dipole moment because the electronic charge and the nuclear proton charge do not reside at the same position in space. This instantaneous dipole moment can induce by polarization a dipole in a nearby neutral molecule (or in other words the two fluctuating dipoles will couple). The resulting interaction between the two dipoles gives rise to an instantaneous attraction between the nonpolar molecules with a time average that is not zero. For getting even a semi quantitative feeling of this interaction we take an example from Israelachvili's book on interaction forces [5].

Let us consider the dispersion interactions between two Bohr atoms. In the Bohr atom one electron is orbiting a proton. The smallest distance (that there *is* a smallest distance and its size follow only from quantum mechanics) between the electron and the proton is known as the first Bohr radius $a_o$. At this radius the Coulomb energy of the system, $e^2/4\pi\varepsilon_o a_o$ is equal to $2h\nu$, or,

$$a_o = e^2/2(4\pi\varepsilon_o)\, h\nu = 0.053 \text{ nm} \tag{12}$$

where $h$ is the Planck constant and $\nu$ the orbiting frequency of the electron. For a Bohr atom, $\nu = 3.3 \times 10^{15}$ $s^{-1}$ and thus $h\nu = 2.2 \times 10^{-20}$ J. This is the energy of an electron in its first Bohr radius and equals the energy to ionize the atom, i.e., the first ionization potential $I$. As argued above the Bohr atom has no permanent dipole moment, but at

any moment the instantaneous dipole moment $m$ is given by: $m = a_o.e$ whose field will polarize a nearby other neutral apolar atom giving rise to an attractive interaction that is completely analogous to the dipole-induced dipole interaction discussed above. The energy of interaction is therefore given by (see Eq. (6)):

$$w(r) = -u^2\alpha/(4\pi\varepsilon_o)^2 r^6 = -(a_o.e)^2\alpha/(4\pi\varepsilon_o)^2 r^6 \tag{13}$$

here $\alpha$ is the polarizability of the second Bohr atom, which from Eq. (11) is approximately given by $4\pi\varepsilon_o\, a_o^3$. Using this expression for $\alpha$ and Eq. (12) gives:

$$w(r) \approx -\alpha^2\, h\nu/(4\pi\varepsilon_o)^2 r^6 \tag{14}$$

Apart from a numerical factor Eq. (14) is equal to the result that London derived in the 1930's using quantum mechanical perturbation theory. London's expression for the dispersion interaction between two different molecules is:

$$w(r) = -\frac{3}{2}\frac{\alpha_1\alpha_2}{(4\pi\varepsilon_o)^2 r^6}\cdot\frac{h\nu_1\nu_2}{(\nu_1+\nu_2)} = -\frac{3}{2}\frac{\alpha_1\alpha_2}{(4\pi\varepsilon_o)^2 r^6}\cdot\frac{I_1 I_2}{(I_1+I_2)} \tag{15}$$

Although London's equation has of course been followed by more exact, and more complicated, expressions (see section 1) it gives fairly accurate numbers, though somewhat on the low side compared with more rigorous values. The 'derivation' of Eq. (14) also demonstrates that although the dispersion forces arise from quantum mechanical effects (which gave the strength of the instantaneous, but fluctuating dipole moments), the interaction itself can still be seen as essentially electrostatic. The name *dispersion* force stems from the relation of the forces to the dispersion of light in the visible and UV part of the spectrum (as exemplified by $\nu$ in Eq. (15)). A more thorough account of molecular Van der Waals forces can be found in [5, 15].

### 1.2 Van der Waals forces between two spheres: Hamaker approach

All contributions to the Van der Waals forces, induction, orientation and dispersion, have the same functional form and can be taken together as:

$$w_{VDW}(r) = -C_{VDW}/r^6 = -[C_{ind} + C_{orient} + C_{disp}]/r^6 \tag{16}$$

where the constants $C$ can be taken from Eqs. (4), (6) and (15). From tabulations of the constants given in Eq. (16) it becomes clear that in almost all cases the dispersion forces are the most important, except for small polar molecules like water.

The simplest approach to calculate the Van der Waals forces between macroscopic bodies (macroscopic in the sense that we will treat them as consisting of continuous matter) is to assume that the molecular contributions given above are pairwise additive. (Which as it turns out is for most substances not such a bad approximation). We can than obtain interaction energies from simple integration an approach first taken by a Dutch physicist called Hamaker [16]. The interaction energy $dU_{12}$ between two infinitesimal volume elements $dV_1=dx_1\, dy_1\, dz_1$ and $dV_2=dx_2\, dy_2\, dz_2$ inside bodies 1 and 2 respectively becomes then:

(17)

$$dU_{12} = \frac{-C_{12}\rho_1\rho_2 dV_1 dV_2}{\left\{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2\right\}^3}$$

where the ρ's are the number densities of the molecules in the material 1 and 2. It is now a matter of algebra (or a numerical calculation) to obtain the interaction energies between the two bodies. An interesting feature of the 6th power dependence is that the resulting energies are scale invariant. (That means that two colloids at a separation of 10% of their distance, say 10 nm, have the same interaction energy as two apples a few cm apart.). We will first present the interaction energy $U_{12}$ between two infinite half-spaces at a distance $h$ (because the half-spaces are infinite the energy is given per area of the half-spaces):

$$U_{12} = \frac{-A_{12}}{12\pi} \cdot \frac{1}{h^2} \tag{18}$$

where we have combined (for historical reasons) $2\pi C_{12}\rho_1\rho_2$ to form the Hamaker constant $A_{12}$. (There is a very nice anecdote connected to this derivation (see [17])). The integration has turned the short range $r^{-6}$ dependence into a quite long range $h^{-2}$ dependence! Integration of Eq. (17) between two spheres of size $2R$ gives [16]:

$$U_{12} = \frac{-A_{12}}{12} \cdot \left( \frac{1}{x^2 - 1} + \frac{1}{x^2} + 2\ln\left(\frac{x^2 - 1}{x^2}\right)\right) \tag{19}$$

where $x = r/(2R)$ is the reduced distance between the spheres.

Equations like Eq. (18) are still of not too much use in the description of interactions between two colloidal particles, because these are always dispersed in a dispersion medium and Eq. (18) describes interactions in vacuum. The very nice feature of the pairwise summation assumption is that it leads in a straightforward way to the interactions between two bodies 1 and 3 dispersed in a medium 3. The same principle behind the derivation therefore holds also for other forces that can be summed in a pairwise fashion (like gravity) and the principle behind it is therefore also called the principle of Archimedes (see Appendix A). From this principle it follows directly that by immersion of the two bodies into a third medium not the geometrical terms are changed but instead only the Hamaker constant according to the following equation (see Appendix A):

$$A_{123} = A_{13} + A_{22} - A_{12} - A_{23} \tag{20}$$

Similarly as for the case which Archimedes made famous with his outcry of 'Eureka', the combined Hamaker constant can both be positive and negative. For interactions between two similar bodies Eq. (20) reduces to:

$$A_{121} = A_{11} + A_{22} - 2A_{12} \tag{21}$$

If we look at Eq. (16) it can be shown that $A_{121}$ can be positive only and equals $A_{212}$ indicating that Van der Waals forces between two bodies of the same material are always attractive! As mentioned for most substances the most important contribution

(certainly for apolar substances) to the Hamaker constant is given by the dispersion interaction. From Eq. (15) it follows that approximately:

$$A_{12}^2 \approx A_{11} \cdot A_{22} \tag{22}$$

thus Eq. (21) and Eq. (20) become:

$$A_{121} \approx \left(\sqrt{A_{11}} - \sqrt{A_{22}}\right)^2 \tag{23}$$

and

$$A_{123} \approx \left(\sqrt{A_{22}} - \sqrt{A_{11}}\right) \cdot \left(\sqrt{A_{22}} - \sqrt{A_{33}}\right) \tag{24}$$

which is negative if $A_{11} > A_{22} > A_{33}$ or $A_{11} < A_{22} < A_{33}$. This situation occurs most often when a liquid (2) wets a solid surface (1) in air (3). Then $A_{33} \approx 0$ and often $A_{11} > A_{22}$. This explains for instance why liquid helium will be attracted by a wall of a vessel, and form a relative thick film on the container wall. Because of its low viscosity it can even flow out of the container (see for theory about wetting phenomena the lectures by Blokhuis).

Although a lot of approximations have been made these kind of combining relations are quite useful to estimate Hamaker constants between different materials from known values. To give just two examples: the Hamaker constant for a $CaF_2$-helium-air interaction would be predicted by Eq. (24) to be: $A_{123} \approx (\sqrt{7.2}-\sqrt{0.057})(0-\sqrt{0.056})\text{x } 10^{-20} = -0.58 \text{ x } 10^{-20}$ J. A more rigorous value gives $-0.59$ x $10^{-20}$ J. Similarly, for a quartz-octane-quartz system Eq. (23) gives: $A_{121} \approx (\sqrt{6.3}-\sqrt{4.5})^2\text{x } 10^{-20} = 0.15 \text{ x } 10^{-20}$ J compared to 0.13 x $10^{-20}$ J [5]. Estimates can also be made from the following approximate values for interactions in water:
$A_{121}$: (30-10 for metals, 3-1 for oxides and halides, ~ 0.30 for hydrocarbons) x $10^{-20}$ J.

### 1.3 Van der Waals forces between two spheres: Modern Theory

For many situations in colloid science the approximate formulae derived in the previous section are accurate enough. However, it is clear that many approximations were made and that there are many places/levels where improvements can be made.

In a book dedicated to him on the occasion of his retirement Overbeek remarks that is was perhaps his major discovery that he realized that the London-Van der Waals forces must show retardation at separations larger than the London wave length [18]. In the same heuristic way as the dispersion forces were introduced above, it is not hard to understand where such a retardation comes from. It was assumed above that the dipole moments that were present at each instant in time could induce a dipole in the other molecules instantaneously. It is clear that this can not be correct and in reality such an induced interaction can not travel faster than with the speed of light. This means that if the distance between the molecules becomes so large that the dipole moment in the first molecule has already changed, the interaction with the induced dipole moment is reduced. Overbeek asked Casimir whether he saw a way to a theoretical treatment of this problem and Casimir and Polder succeeded in formulating a quantum mechanical theory for these effects in the case of two metals [19]. It turns out that in the retarded limit the $1/r^6$ potential is reduced to a faster decaying $1/r^7$ functional form. (Note that the other contributions to the Van der Waals forces do not show this retardation).

Using advanced quantum field theory combined with statistical mechanics, Lifshitz [20] and later Dzyaloshinskii [21], derived the general case including dielectric materials. In this theory non-additivity effects and retardation are all treated in a natural way. Because the results were so complex it took a long way for them to reach the complex fluids community. This only happened after Van Kampen and others, along the same treatment as was used by Casimir and Polder, had rederived the essential equations using a much simpler (semi classical) approach. It goes very roughly along the heuristic lines indicated above about the coupling of electromagnetic fluctuations traveling from the one body to the other, but is cast completely in terms of the dielectric responses of the bulk materials. Quite a lot of researchers subsequently simplified the equations to such an extent that it is possible to calculate Hamaker constants from just a few relevant dielectric properties of the materials under consideration. Here we will just refer to the standard books on this subject [15, 4, 5, 7, 8, 10]. All geometrical equations of the previous section are obtained as limiting cases though, and the material properties like the Hamaker constant, are not derived from molecular but bulk dielectric properties.

### 1.4 Overlap of flat double layers: Debye-Hückel approximation

The interactions between a charged surface and the distribution of ions, both from added salt (we will limit ourselves to so-called $z$-$z$ symmetric salts) and counterions will be treated within the self-consistent mean-field Poisson-Boltzmann approach. In this approach the Poisson equation, which follows directly from the Coulomb law of electrostatics, is combined with the statistic mechanical Boltzmann equation in order to obtain the distribution of ions in the potential field of the plate with a certain surface charge. We will treat the distribution of the ions along one plate first, after Gouy and Chapman who were the first to do this along these lines, and subsequently calculate the free energy difference of bringing two such plates together.

The surface charge is supposed to be of uniform density and the ions are treated as point objects with no size, embedded in a solvent which is approximated as a continuous dielectric medium. The surface charge on the plate sets up a potential $\psi$ in the solvent, which depends on the distance from the plate and which puts the free ions in the solution at an electrical potential energy. Conversely, the distribution of ions determines the local potential. The iterative or self-consisted way in which these dependencies are expressed is a combination of the Poisson equation, which relates the local charge density to the local potential, and the Boltzmann equation which describes the probability of finding an ion at a certain (electrical) free energy. The Poisson equation is given by:

$$\mathrm{div}(\mathrm{grad}\,\psi) = \nabla^2\psi = \frac{-\rho}{\varepsilon\varepsilon_o} \tag{25}$$

here $\rho$ is the local volume density of charge, i.e., the number of charges per unit volume:

$$\rho = \sum_i n_i z_i e \tag{26}$$

where the summation is over all the species of ion present with valency $z_i$ and number density $n_i$, here $z_i$ includes the sign of the charge.

The Boltzmann equation reads for an ion at potential $\psi$:

$$n_i = n_{oi} \exp(-w_i / k_B T) \tag{27}$$

where $w_i$ represents the work done in bringing an ion $i$ up from the bulk, where its number density is $n_{oi}$, to a point in the double layer where the potentials is $\psi$. The amount of work $w_i$ is now approximated as $w_i = z_i e\psi$. In other words the only work done that is taken into account is the electrical work done on or by the ion when it is brought into the double layer. This ignores work done to move other ions away or create a hole in the solvent, or any effect the ion has on the structure of the solvent or the distribution of the other ions. As stated above the ion is simply treated as a point charge.
Combining Eqs. (25)-(27) gives the Poisson-Boltzmann equation (PB):

$$\nabla^2\psi = \frac{-1}{\varepsilon\varepsilon_o}\sum_i n_{oi} e z_i \exp(-z_i e\psi / k_B T) \tag{28}$$

If the electrical energy is small compared to the thermal energy ($|z_i e\psi| < k_B T$) it is possible to expand the exponential in this non-linear equation. Neglecting all but the first to terms:

$$\nabla^2\psi = \frac{-1}{\varepsilon\varepsilon_o}\left(\sum_i z_i e n_{oi} - \sum_i z_i^2 e^2 n_{oi}\psi / k_B T\right) \tag{29}$$

The first summation must be zero because of charge neutrality in the bulk solution, thus:

$$\nabla^2\psi = \left(\frac{\sum z_i^2 e^2 n_{oi}}{\varepsilon\varepsilon_o k_B T}\right)\cdot\psi = \kappa^2\psi \tag{30}$$

where

$$\kappa = \left(\frac{e^2\sum z_i^2 n_{oi}}{\varepsilon\varepsilon_o k_B T}\right)^{1/2} \tag{31}$$

This linearization of the PB equation is called the Debye-Hückel (DH) approximation because it was used by these workers in their theory of strong electrolytes. Eq. (30) can be solved easily and one obtains:

$$\psi = \psi_o \exp(-\kappa\cdot x) \tag{32}$$

with $\psi_o$ the surface potential. The quantity $\kappa^{-1}$ the *Debye-Hückel screening length* plays an important role in the theory of the double layer. From Eq. (32) it can be seen why, as in this approximation it determines the extension of the double layer and the region at which the potential around a colloid is different from that of the bulk where it is zero. Apart from some fundamental constants, $\kappa$ depends only on the temperature and the ionic strength $I$ defined as: $I = (1/2)\sum c_i z_i^2$ where $c_i$ is the ionic concentration

in mol/l. A useful rule of thumb is obtained by filling in the constants for water at room temperature (25 °C):

$$\kappa = 3.288\sqrt{I}\ [\mathrm{nm}^{-1}] \tag{33}$$

Thus for a solution of 1:1 electrolyte at $10^{-3}$ M the double layer thickness is 9.6 nm. At $10^{-6}$ M it is 304 nm.

For flat plates the non-linear PB equation can be solved analytically using a mathematical trick. However, in these notes we are more concerned with spherical double layers so we will refer to the literature for this result [2, 4, 9]. Without giving the full result it is easy to see that even in the case of high surface potentials at large distances $\psi$ becomes small and the long distance tail of the potential takes on the exponential form given by Eq. (32), but now with a different scaling factor which is determined by the full solution. Because an other particle will only sample the not too high potential regions the exponential form is often used.

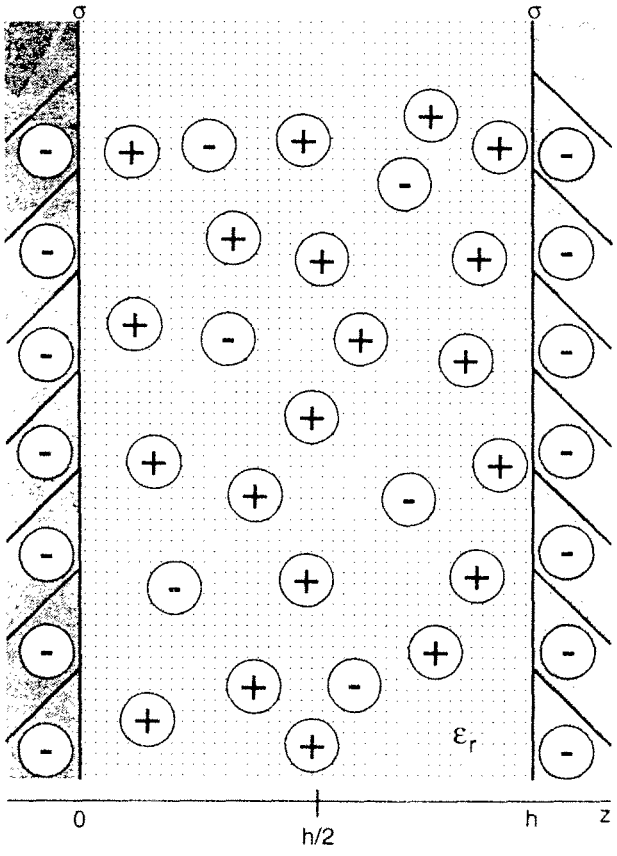



**Figure 2** Two interacting negatively charged surfaces separated by *h* experience a repulsive force as they are pushed together.

Before turning to the much more difficult problem of a double layer around a sphere, we will first give one thermodynamic path of how to obtain the amount of work done if two flat double layers are brought in each-others neighborhood and start to overlap, i.e., the free energy as a function of distance between two infinitely large spheres. In the case of overlapping double layers the mechanism of surface charge regulation becomes of importance. There are several mechanisms by which a surface can obtain and regulate a surface charge, some of the most important are: 1) A very small imbalance in the amount of crystal lattice anions or cations (e.g., as in the case of AgI crystals in water). 2) Surface dissociation. 3) Crystal lattice defects. 4) Surface absorption of ionic species. The problem of how the surfaces regulate their charge on the surface on overlap of two double layers is complicated and also depends on the mechanism of charge generation and maybe even the speed of approach. This problem has not yet been solved unambiguously [9]. However, it can be shown that

real systems will lie within the borders of a constant surface charge upon approach and a constant surface potential, for the moment we will assume the latter.

One other subject that is still to this date a matter of controversy in the literature is the experimental determination of the surface charge or potential. Again going into this would lead us much too far astray, but is should be mentioned that a lot of experimental observations are not yet explained satisfactorily [7, 9].

Lets now turn to the work required to bring two plates from infinity to a separation distance $x$ in the limit where we can use the DH approximation. There are several ways to obtain this free energy difference of double layer overlap $V_{dlo}$ in the case of flat plates [1-2], here we will only go into one, called the 'force method' by Overbeek [2]. If the plates are brought together reversibly all the forces on the system should cancel because a reversible path goes through equilibrium states. We can use this balancing of forces to recognize that the free energy difference can be obtained by integrating the osmotic pressure $\Pi$ difference (compared to the bulk) at the midplane between the plates from infinity to the separation distance $h$:

$$V_{dlo} = -\int_{\infty}^{h} \Pi \, \mathrm{d}\, x \tag{34}$$

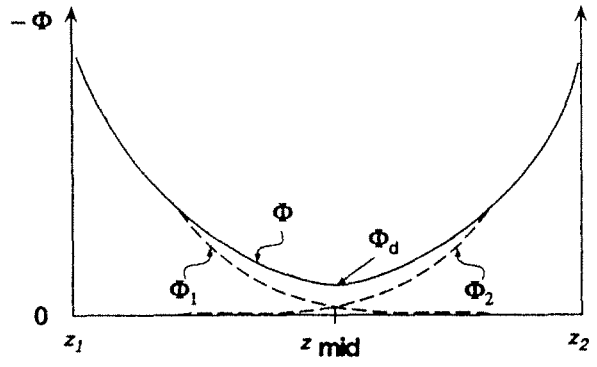



**Figure 3** Superposition approximation for two similarly charged surfaces. If the surface charges on the surfaces on the plates are thought to be unaffected by the overlapping double layers, the resulting charge distribution is the sum of the two and by the linearity of the Poisson equation the resulting potential at the midplane will also be the sum of the unperturbed potentials, which can be related to an osmotic pressure at the midplane by the Boltzmann equation.

The rationale behind Eq. (34) is that symmetry dictates that at the midplane the electrical forces on the ions are equal and thus there is no excess charge at this position as well, both because the potential has a minimum here. In the PB approach there are no correlations between the ions so the osmotic pressure of the non interacting ions is simply given by the 'ideal gas' value:

$$\Pi = k_B T \left( [\Sigma\, n_i]_{\text{midplane}} - [\Sigma\, n_i]_{\text{infinity}} \right).$$

In the linearized DH approximation of small potentials one can take the potential at the midplane as twice the potential of the single plate potential as given by Eq. (32), see Figure 2. This leads to the following double layer overlap potential:

$$V_{dlo} = 2\varepsilon\varepsilon_o \kappa \psi_0^2 \exp(-\kappa \cdot h) \tag{35}$$

Similarly as stated above, in the limit of large separations also the curves with high surface potential will adopt a limiting form similar to Eq. (35), again with a different 'apparent' surface potential, see Figure 4. It should be remarked that even in the

'simple' two plate geometry the general case can not be treated analytically anymore and all kind of approximate formula's or numerical schemes have to be used (see for more details [2, 9]).

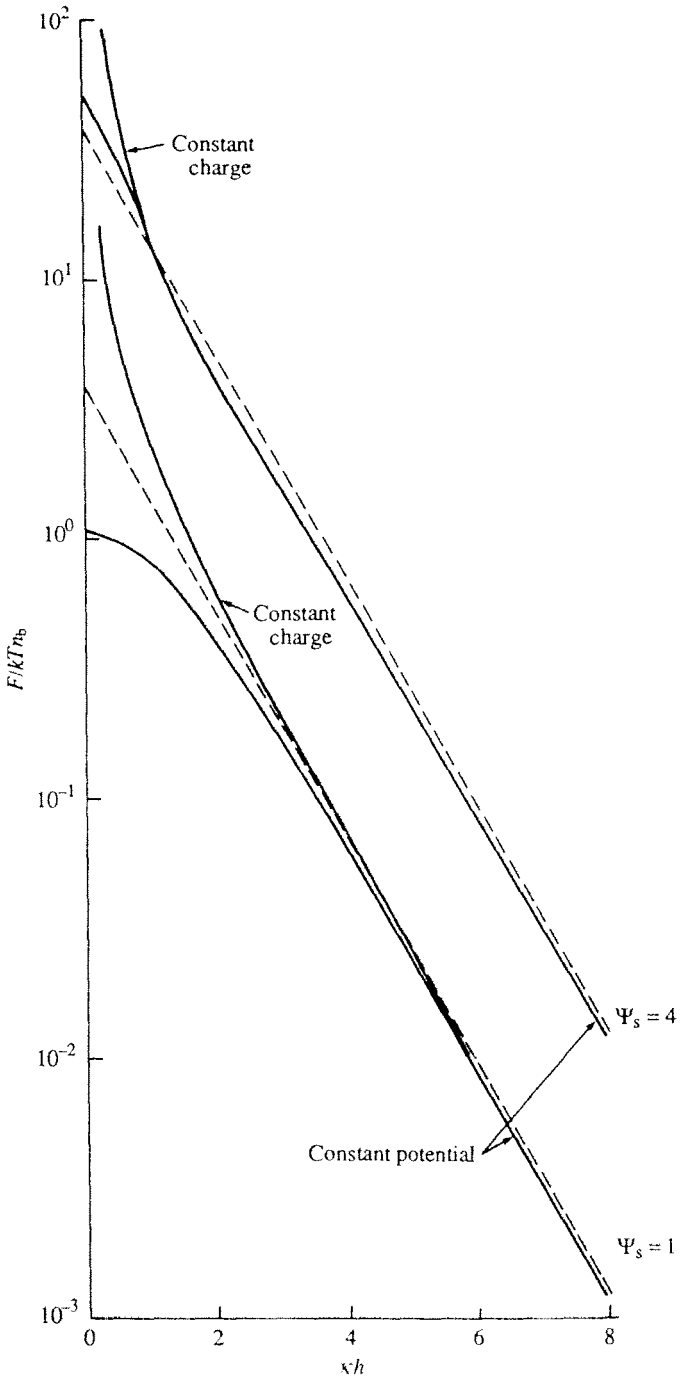



**Figure 4** Two identical overlapping flat double layers immersed in an ionic solution. Results are shown in reduced units and as force curves for two different surface potentials and assuming both a constant charge and constant surface potential. The curves are calculated using the exact theory also indicated are approximate results along the lines of Eq. (35). At large separations the approximate results are quite good [4].

## 1.5 Double layer overlap between two spheres

It will be no surprise that in the much more awkward geometry of two spheres analytical results are even harder to obtain. Not even for a single double layer around a sphere are there analytical results that hold for a reasonable range of parameters and

semi-empirical equation's have been proposed [9]. Only in the case of the DH approximation one obtains an analytical result. Because of the symmetry of the problem the PB equation can best be given in spherical coordinates. In these coordinates the Laplace operator takes the form:

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\psi}{\mathrm{d}r}\right)=\Delta\psi=\frac{-1}{\varepsilon\varepsilon_o}\sum_i n_{oi}z_i\exp(-z_i e\psi/k_BT) \tag{36}$$

linearizing the exponential ($e^{-x}\approx 1-x$) leads to (compare with Eq. (30)):

$$\Delta\psi=\kappa^2\psi \tag{37}$$

Solving Eq. (37) gives a screened Coulomb or Yukawa potential:

$$\psi=\psi_o\frac{a}{r}\exp[-\kappa(r-a)] \tag{38}$$

where $a$ is the sphere size.

At double layer overlap the situation is even worse. Only for very thin double layers compared to the sphere size it is possible to use the so-called Derjaguin approximation. This approximation comes down to using the results obtained between flat plats to derive equations for large interacting bodies. For interactions between two colloidal spheres the limit of very thin double layers is totally uninteresting. However, because the Derjaguin approximation can also be used with other potentials and is useful in interpreting measurements done with the surface force measurements it is given in Appendix B.

For smaller values of $\kappa a$ (< 5) the Derjaguin procedure breaks down. Verwey and Overbeek have shown that for low surface potentials and if an error of up to 40% can be tolerated approximate formulae in the spirit of Eq. (35) (i.e., by taking sums of potentials) can be derived resulting in again a Yukawa or screened Coulomb form:

$$V_{dlo}=\pi\varepsilon\varepsilon_o\ a\psi_0^2\frac{\exp[-\kappa a(x-1)]\cdot}{x} \tag{39}$$

here $x = r/a = r/2R$.

### 1.6 Summation of forces

The combination of the forces resulting from the overlap of two double layers treated theoretically within the assumptions behind the PB equation and the Van der Waals forces constitute the DLVO potential:

$$V_{DLVO}=V_{VdW}+V_{dlo}+V_{Pauli} \tag{40}$$

In this equation we have added the forces that determine the closest distant of approach of two colloids through a strong and steep repulsion caused by the Pauli exclusion principle of electrons in filled orbitals (see Eq. (1)). This repulsion is also quite naturally taken up into a closest distance of approach in the Van der Waals

forces that would otherwise diverge. For two spheres of the same size and using the approximations for the interactions as used in these notes the DLVO potential becomes:

$$V_{DLVO} = \pi\varepsilon\varepsilon_o \, a\psi_0^2 \frac{\exp[-\kappa a(x-1)]}{x} + \frac{-A_{12}}{12}\left(\frac{1}{x^2-1} + \frac{1}{x^2} + 2\ln\left(\frac{x^2-1}{x^2}\right)\right) \tag{41}$$

One can find numerous forms of DLVO equations, even within the PB approach, it is therefore important to look at which of the many approximations have been used and for what situations the equations are valid.

In a schematic way the different potential shapes, depending on the constants, Eq. (41) can give rise to are given in Figure 5.

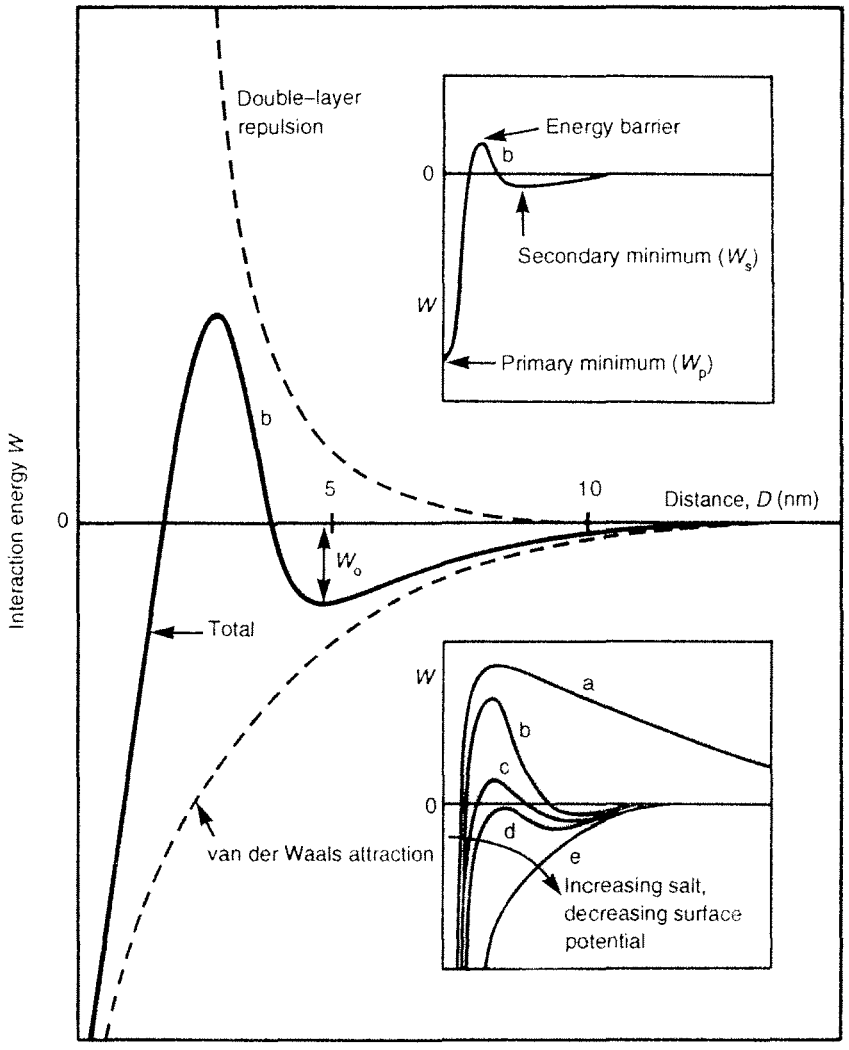



**Figure 5** Schematic interaction potentials between two charged colloidal spheres according to Eq. (41) [5].

## 1.7 Deviations from DLVO

Most of the shortcomings of the Van der Waals forces have already been discussed and with sufficient effort they can at the moment be calculated to great accuracy. The most serious problems arise from the fact that the Lifshitz treatment is a continuum approach and thus if effects that are a consequence of the molecular discreteness are involved the theory fails. For most colloidal interactions there are no indications though that these effects are very important.

There are more (known) problems with the PB description of the double layer interaction potentials. Already mentioned is that in this equation the finite size of the ions and correlations among them are completely neglected. Finite size corrections were realized quickly and the 'Stern' layer of closest approach of hydrated ions is one of the earliest examples to remedy this neglect. It has not been until the advent of computer simulations, however, that both new theoretical improvements and the accuracy of the PB approach could be explored fully. It goes too far to discuss these matters in length here (see e.g. Refs. cited in [7, 22]). One of the most important conclusions is that for 1-1 electrolytes at not too high surface potentials and not too close separations the PB equation gives a fair to good description. (It is also relevant in preparation of what follows to remark that between two equal surfaces no attractions have been observed). If the coupling between ions becomes stronger, like for higher valency ions, PB breaks down and qualitatively different behavior is observed [7, 22].

It also seems that equations that are derived under the DH approximation are relatively useless because in practice surface potentials are often higher than ca. 25 mV. However, it is already discussed that this assumption also gives a good description at larger separations if an adjusted surface charge or potential is used. Furthermore, it has been shown theoretically that the DH approximation results in a description that is thermodynamically consistent and can be derived as a limiting case within the framework of liquid state theories (see e.g., the lectures of Briels). This thermodynamic consistency is *not* achieved by solutions obtained from the full non-linear PB equations! Both theoretical work and computer simulations have given additional justification to the use of potentials of the Yukawa form. For instance, Alexander *et al.* have shown through calculations of salt and counterion profiles in a spherical Wigner-Seitz cell that also for strong interactions between the colloids an effective Yukawa pair interaction is obtained if the volume fraction of the particles is not too high. The charge must be renormalized and the double layer thickness adjusted compared to the DH value [23]. Furthermore, Löwen and Kramposthuber have shown from *ab initio* theory that screened Coulomb potentials can be used in many instances as good approximations, but that both the screening length and effective charge have to be adjusted. Moreover, it turns out that in these strongly interacting systems the effective surface charge and screening length become dependent on the phase as well [24].

Now that we have some understanding of the ideas, approximations and limitations behind the DLVO potential(s) it is time to see what experimental methods have been developed over the years to measure these interactions in a direct way. A development which recently has been speeded up, partially from new input from the related field of biology inspired physics and from recent experimental findings that will be described in Section 3.

## 2. MEASURING INTERACTION POTENTIALS

In this Section we will limit ourselves to a qualitative description of *direct* methods to measure interaction forces. With '*direct*' is meant that the measurements give force-distance curves in an unambiguous way. This excludes methods like osmotic pressure measurements, where only information of a thermodynamic nature is obtained, which can not be directly translated into a force law, or methods that determine only a certain aspect of the force distance relationship like adhesion measurements or coagulation studies. We also exclude potential measurements that rely on an inversion of structural information obtained through scattering studies. Although this method is in principal direct in the above sense it is clear that, largely because of experimental limitations, the inversion procedure is 'ill-defined' and does not give unambiguous potentials (see e.g., [25, 26] and refs. cited).

### 2.1 Surface Force Apparatus (SFA)

The first attempts to measure both double-layer and Van der Waals forces started in the early fifties in Russia by Derjaguin [27] and in the Netherlands by Overbeek [28] and their coworkers. These studies were performed with set-ups that were essentially similar to what would later be called the surface force apparatus. However, the forces measured were limited to the retarded regime at large surface-to-surface separations because of the roughness of the fused quartz and glass surfaces that were used. It was not until the end of the sixties that Tabor and Winterton started using cleaved mica which made it possible to measure force-distance curves down to molecular separations [29]. Native mica crystallizes in layer structures which can be easily cleaved providing a clay-like atomically smooth surface over a large area, ideally suited for force measurements both in vacuum annd in liquids. It is quite remarkable though that just such a 'trivial' matter of obtaining atomically smooth surfaces over large areas has held up the measurement of forces at non-retarded and important range of distances for quite a number of years.

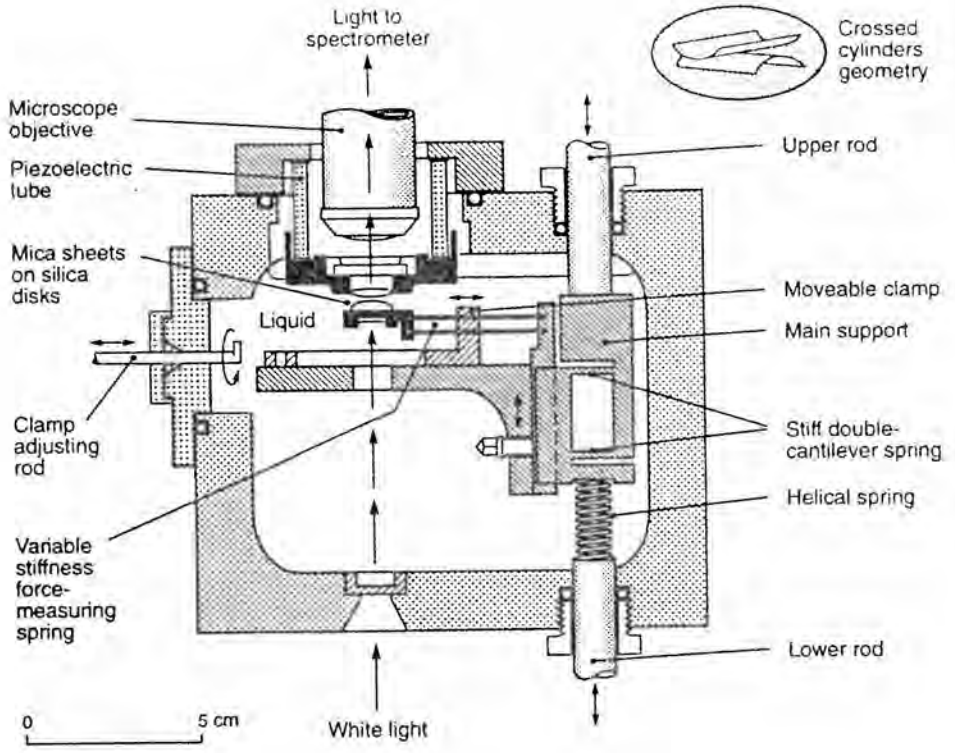



**Figure 6.** Surface Force Apparatus (SFA) with which force-distance curves can be obtained between molecularly smooth mica sheets with Å (0.1 nm) resolution and a force sensitivity around 10 nN ($10^{-8}$ N) [5].

The modern version of the surface force apparatus is conceptually simple and is depicted in Figure 6 [5]. The separation between the mica surfaces, which are glued onto two quartz pieces with a radius of curvature of 1 cm and are silvered on the back with a partially reflecting silver film, can be determined with an accuracy of 1 Å (0.1 nm). This high accuracy is reached by analyzing the interference that results from multiple reflections of white light by the silver layers on the mica in a spectrometer. Measuring the distance between the surfaces with Å accuracy is only half of the story, it is also necessary to have this accuracy in positioning the surfaces. This is accomplished by a three-stage mechanism. The smallest scale displacements (between 1-10 Å) are achieved by means of a piezoelectric tube that can translate the upper mica surface. In piezo electric materials an electric field can cause the crystal lattice to expand or contract, e.g., by about 1 nm per volt that is applied across the surface of the cylinder wall. Positioning of the lower mica surface on the 1 nm level is achieved by a two spring construction where the difference in stiffness or spring constant between the stiff double-cantilever spring and the helical spring attached to the lower micrometer rod. Therefore, a displacement of the lower micrometer results in a nm displacement of the lower mica surface, in the absence of forces between the mica surfaces. After calibration of this positioning system the actual displacement of this surface can be measured from the interference between the crossed mica surfaces. The difference in displacement can be converted to a force by using the calibrated spring constant of a force measuring spring. The second, lower mica surface is attached to such a force measuring spring of which the stiffness can be varied (by a factor of 1000) by moving a clamp. In this way both attractive *and* repulsive forces can be measured with a sensitivity of about 10 nN ($10^{-8}$ N). Finally, an upper rod can move

the whole section of springs over distances between 1 μm and 1 cm, but it is not used during the actual force measurement. Within the Derjaguin approximation (Appendix B), which is clearly a very good approximation between these macroscopic surfaces, the ratio of the measured force and the radius of curvature, $F/R$, equals $2\pi U$, where $U$ represents the interaction potential per unit area. Physically, this equivalence means that at a distance $h$ between the curved mica surfaces an average is measured of all interaction forces larger than $h$. The result of this averaging, the total force equals the sum of the forces between the surface segments from infinity to $h$, is the energy at $h$.

In vacuum the forces between the mica sheets are solely due to Van der Waals forces. After all kind of corrections, like change of curvature of the mica sheets due to deformation of the glue by the strong attractive forces, the measured force curves come to within 10 per cent of the calculated curves using the Lifshitz approach over the full range of separations [5]. Immersed in water the mica surfaces obtain a negative surface charge by the dissociation of potassium ions and a double-layer is formed. In this way the DLVO theory was tested at different ionic strengths and potential determining ions (see Figure 8 and Figure 9).

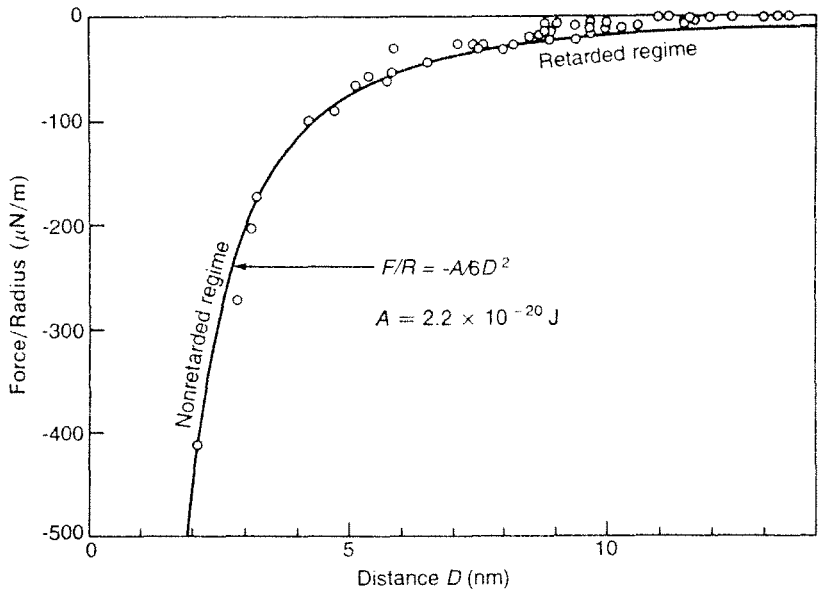



**Figure 7** Attractive van der Waals forces between two curved mica surfaces measured in an aqueous electrolyte solution with the SFA [5]. The measured non-retarded Hamaker constant is $2.2 \times 10^{-20}$ J. Retardation effects become apparent at distances above 5 nm.

At this date the SFA has been quite important in the measurement of all kinds of forces, besides Van der Waals forces and double-layer forces [30, 31], examples are: capillary forces, solvation forces [32], adhesion forces [33], 'hydration' forces [34, 35], depletion forces [36, 37], steric repulsion forces [38, 37], or special 'gel-like short-range' forces on silica [39, 40], and forces that were hardly considered before the experiments, like attractive hydrophobic forces [41, 5] and oscillatory structural forces [5]. All these forces are not elaborated on here, because under conditions under which they are measured it is clear where the assumptions underlying the DLVO potential are not met and/or why the description breaks down. It should be mentioned here that recently these tests are not limited anymore to just one kind of surface, mica in water. For instance, the mica can be used as a substrate to adsorb a thin film of some other material, for example, lipid monolayers, metal films, proteins etc. (see for

Refs. in [5]). In the case of opaque materials a capacitance method replaces the optical technique for measuring distances with similar overall accuracy [42]. Furthermore, Horn *et al.* have finally found a way to make silica surfaces smooth enough so that they could be used in the SFA as well down to molecular levels [43]. The trick is to blow small glass bubbles very fast from the melt so that the surface tension can keep the surface molecularly smooth.

Recently, the SFA measurements have also been extended to the measurement of dynamic interactions and time-dependent effects like the viscosity of liquids in very thin films, measurements which are outside the scope of these lecture notes.

Under circumstances where the assumptions and limitations of the PB approach are met and at distances not too close to contact, no important deviations, that were not expected see Section 1.5, have been reported see e.g. Figs. 7-9.

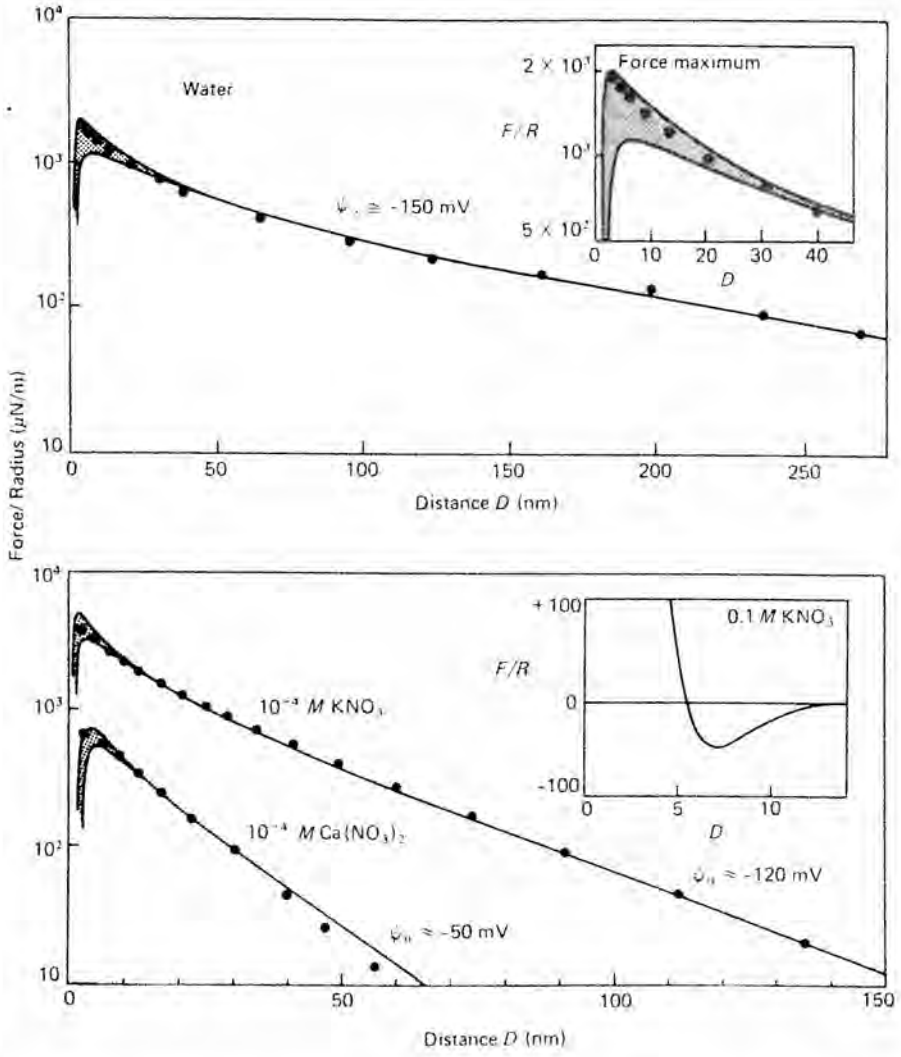



**Figure 8** Measured double layer and Van der Waals forces between two curved mica sheets in the SFA in dilute salt solutions [5]. Curves are calculated using a Hamaker constant of 2.2 x $10^{-20}$ J (see Figure 7) and the constant charge and potential limits are drawn.

The most serious critique against the SFA apparatus measurements is that it is never clear how well the macroscopic surfaces are representative for the surfaces of colloidal particles was somewhat alleviated by measurements with the SFA between

two surfaces at which colloidal spheres were adsorbed [44]. Of course this kind of approach goes at the cost of accuracy in the distance determination and the geometry becomes less well defined. Also the critique that the time scales of approach are completely different than in the colloidal domain is not alleviated. With the measurement of depletion potentials, where it is also assumed in the theories that one of the colloidal entities is much larger than the other, the measured potentials are probably close to 'true' colloidal potentials. For instance, recent measurements in which depletion forces were measured at a high volume fraction of microemulsion droplets [36]. In this case a true potential of mean-force was measured showing maxima and minima in the force curves as a consequence of the structuring of the droplets by the high volume fraction and the presence of a wall.

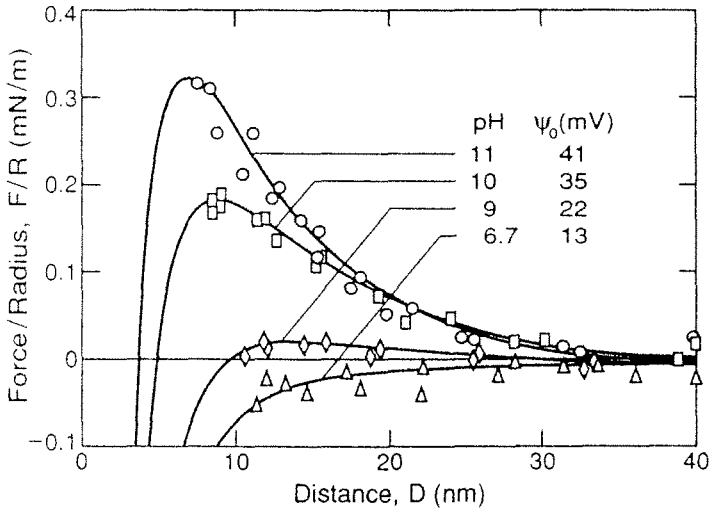



**Figure 9** DLVO potentials measured between two sapphire surfaces in $10^{-3}$ M NaCl solutions at different pH. Full curves have been calculated using the potentials shown and a Hamaker constant of 6.7 x $10^{-20}$ J [5].

## 2.2 Atomic Force Microscopy (AFM)

The development of the atomic force microscope was a direct consequence of the development of the scanning tunneling microscope (STM), but can in retrospect also be seen as a natural continuation of the ideas behind the SFA. The most important difference is that the forces are not measured between two macroscopic bodies, but between a fine tip and a surface [45]. The tip radii can be as small as one atom or larger than 1 μm. Because of this reduction in size compared to the SFA the spring constants need to be much smaller and the displacements of these springs need to be measured still with high accuracy. Already, spring constants as small as 0.5 $N.m^{-1}$ are used and displacements as small as 0.01 nm can be accurately measured. A schematic diagram of a typical set up is given in Figure 10. Position detection of the spring is achieved by reflecting a laser off the back of this spring onto a position-sensitive detector. The tip is moved over the surface by a piezo-scanner (not shown) and the tip deflection is used in a feedback loop operated via another piezoelectric tube to maintain a constant force between the tip and surface by changing the height of the tip. With the tip signal from the feedback loop as a function of position an image of

the surface can be created. Or with the known spring constant force versus distance curves can be obtained.

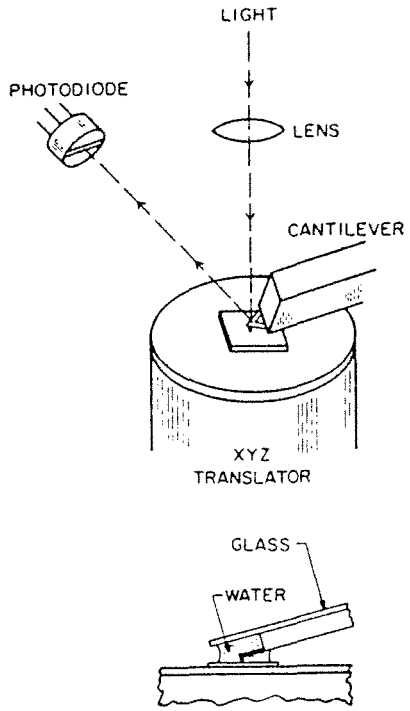



**Figure 10** Schematic depiction of an Atomic Force Microscope (AFM). Spring constants as small as 0.5 $N.m^{-1}$ can be combined with position determination of the spring of 0.01 nm to allow for the measurements of forces smaller than 100 pN ($10^{-10}$ N) [7].

Although 'atomic forces' can in principal be measured with the AFM we are here more interested in the extension of this technique to the measurement of forces

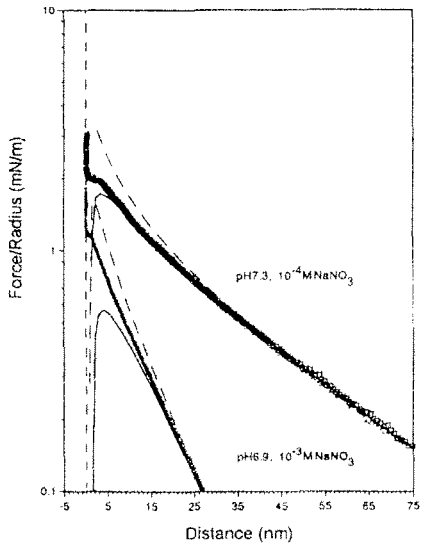



**Figure 11** DLVO potentials as measured between a silica sphere (size 5 μm) and a glass plate in water with added salt measured with an AFM. Surface potentials were taken to be equal for the sphere and plate at values of -105 mV and -42 mV [50].

between a particle in the colloidal size range and a wall. Ducker *et al.* [46] and Butt [47] were among the first to do this between a silica sphere glued to the AFM tip and a flat glass surface in aqueous salt solutions out to surface separations of 60 nm. Similar measurements with silica particles were done by Meagher [48]. Li *et al.* extended the method in the sense that they measured the potential between two polystyrene particles [49] of which one was glued to the tip and one was stuck to a glass surface. All these measurements found fair agreement with potentials derived using the PB approach (some used linear equations others like Butt and Li *et al.* numerically solved the full non-linear equations). Excellent agreement was found by Hartley *et al.* [50] who not only measured the force-distance curves between silica spheres and mica and silica surfaces, but also independently determined the $\xi$ potential of the spheres by electrophoresis measurements and of the surfaces by streaming potential measurements [9]. An example of their careful and accurate measurements is given in Figure 11.

## 2.3 Total Internal Reflection Microscopy (TIRM)

Prieve and coworkers [51, 52, 53] developed a method to measure the interaction force between a colloidal particle and a wall, which is referred to as total internal reflection microscopy (TIRM). The name is somewhat confusing as it was already used as a general microscopy technique where imaging relied on an evanescent field. When light is incident upon an interface from the more optically dense side (for instance glass) at an angle exceeding the critical or Brewster angle, total internal reflection occurs producing an evanescent wave in the less dense medium (e.g., air or water). If the interface is smooth no light is transmitted normal to the interface into the less dense medium. Imperfections the size of the wavelength or larger will scatter light and appear bright against a dark background. This way of imaging is used to inspect optical surfaces or as a special contrast technique in biology and is also abbreviated TIRM. Prieve *et al.* made use of the fact that the irregularity in the evanescent field could also be a colloidal particle. If such a particle will have a sufficiently large refractive index and is brought within a few thicknesses of the evanescent wave it can start scattering and thus transfer energy from the evanescent field into a propagating wave travelling away from the dense material. This situation is called 'frustrated total internal reflection' and Chew *et al.* [54] solved the Mie scattering problem of a single dielectric sphere by an evanescent wave. This solution can be used to accurately determine the height of a colloidal particle close to a wall.

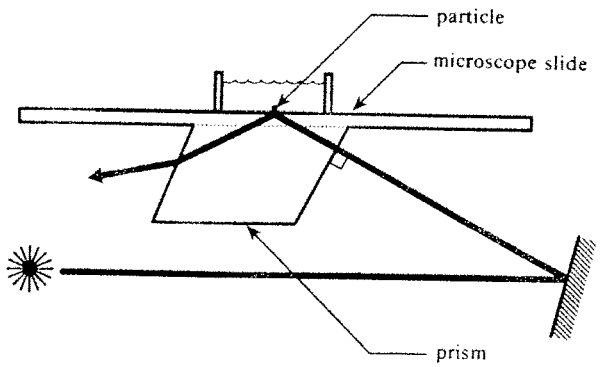



**Figure 12** Schematic diagram showing the scattering cell for total internal reflection microscopy (TIRM) measurements on colloidal particles; the (photometric) microscope is not shwon [51].

In this method a relatively large colloidal particle, e.g., a silica sphere with a diameter of 10 μm is allowed to sediment until the gravitational pull and the repulsive interaction forces with the wall are of even strength. Because of the Brownian motion, the particle will not be stationary but sample positions around this equilibrium point. These distances are measured as mentioned above and the potential between the particle and the wall can be evaluated around this equilibrium point by assuming a Boltzmann distribution of the distances.

Brown *et al.* extended the technique by combining it with optical tweezers (see Appendix C) which give additional control over where the equilibrium position of the particle could be placed and thus were the potential could be sampled [55, 56, 57]. In addition the tweezers made it possible to measure the absolute separation distance between the sphere and the reflecting surface.

Interaction potentials of polystyrene latex spheres of different size and at different ionic strengths were measured carefully by Bike *et al.* [58, 59]. They did not use tweezers, but were able to get absolute distance measurements. An example of the potentials, which were in good agreement with DLVO potentials, they obtained is given in Figure 13.

Dynamic measurements of the hindered diffusion of the colloid can also quite easily be made by feeding the light to a correlator [51, 60]

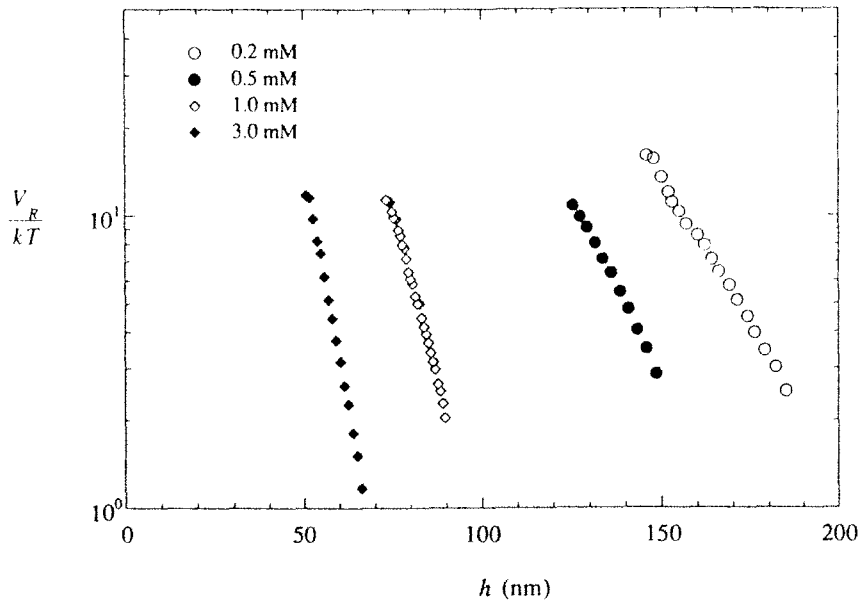



**Figure 13** Double-layer potential as measured with TIRM of 15 μm diameter latex sphere above glass as function of the ionic strength [59].

## 2.4 Direct Imaging

The above mentioned methods always involve one component that is not of a colloidal size. Quite recent methods, almost all relying in some way or another on direct imaging of the particles, finally make it possible to measure force-distance curves between two colloidal particles and have the potential to be used even in the very relevant concentrated regime. However, to this date only a handful of measurements have been reported between two colloidal spheres in 3D in the very or

semi-dilute regime where only a few particles are interacting. It is expected though, that it is only a matter of time before the assumption of pair-wise additivity of colloidal interaction potentials can be tested experimentally under concentrated conditions. In the following we will briefly discuss the papers that have appeared already.

The first papers [61, 62, 63], appeared in 1994 and the interaction potentials were extracted by using digital video analysis to obtain particle coordinates and an analysis based on the radial distribution function, $g(r)$, to obtain the pair-wise inter particle interaction potential $U(r)$ through the relationship:

$$U(r) = -k_B T \ln(g(r)) \tag{42}$$

The radial distribution function essentially gives the chance of observing a particle pair with a separation distance $r$, see the lectures by Briels. In general $U(r)$ is not the effective pair potential, but the so-called potential of mean-force. This potential describes the potential of a pair of particles in the presence of the interactions with all the neighboring particles as background. Only in the limit of 'infinite' dilution when only isolated pairs of particles interact one obtains the effective pair potential.

Fraden *et al.* analyzed the interactions between two spheres that were confined to an almost two-dimensional (2D) plane by the double layer repulsion of two glass walls separated by only several micron [61, 64]. The only way they could get good statistics was to measure at finite concentrations where more than two particles were interacting at the same time. In order to correct for these effects and obtain a true pair potential they used Brownian dynamics computer simulations in an iterative procedure. Carbajal-Tinocco *et al.* used the same experimental procedure as Fraden *et al.* and also got similar results [65]. They measured also potentials at higher 2D concentrations and used the Ornstein-Zernike equation together with several closures to obtain the effective pair potential from the measured potential of mean force (see the lectures of Briels). These results will be discussed more fully in Section 3.4.

Versmold *et al.* analyzed their potentials at sufficiently low density that three body interactions could be neglected and they analyzed their separations at least 2 micron away from the glass walls [63]. To get good statistics more than 20.000 video pictures were analyzed. (However, they do not describe how it was possible to get accurate distance measurements, because an ordinary microscope was used and the particles were not confined in any way.)

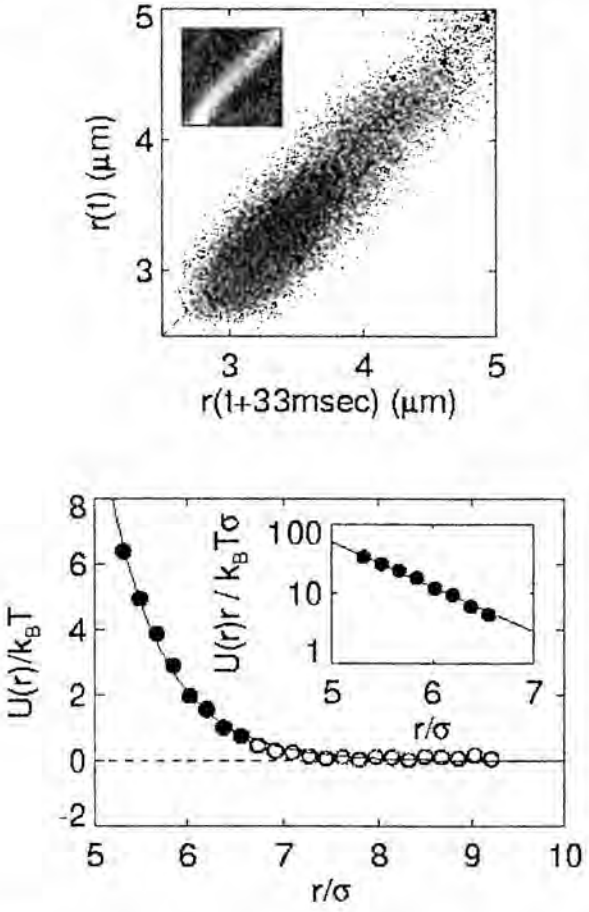



**Figure 14** Measurement of pair potential from direct imaging and optical tweezers manipulation [66]. Top: Distribution of particle separations initially at $r(t)$ which evolved to $r(t + 33\ ms)$. Deviations from the line are the result of interactions between the spheres. The density of points does not reflect the probability of finding the spheres at a certain separation, but rather the frequency of which they were put there with the tweezers. Inset shows a histogram of the same data set but now normalized as a propagation matrix. Bottom: sample potential curve between two latex spheres in water. Fit is to a 'charge-regulated' DLVO potential [66, 67].

Also in 1994, Grier *et al.* presented interaction potentials between two charged colloids in water far enough in the bulk to exclude any wall effects [62]. They measured the potential by analyzing the statistics of the displacements of the colloids in a set time interval after they were first put at a specific initial separation with laser tweezers (see Appendix C) [66]. In this way they were sure to measure only pairs and still got good statistics relatively easily. There were also no problems to determine the true separations, because the tweezers confined the particles first in the focal plane. To give some numbers: around 20.000 images of sphere pairs taken in 1/30 sec intervals suffice to measure an interaction potential with a 60 nm spatial resolution and 0.2 $kT$ energy sensitivity over a range of 6 μm [66]. A sample measurement from their paper on the experimental details of the determination of the pair potentials is given in Figure 14 [66]. (It should be mentioned here that some aspects of the procedure Grier *et al.* use in their analysis are unclear [67]).

Of more recent date are measurements performed by Sugimoto *et al.* who did not use Eq. (42) to obtain $U(r)$, but instead used the potential well created in the center of the optical traps (see Appendix C) to obtain the pair potential of latex spheres in water [68]. (Although not the potential of interest for these lectures, depletion forces

induced by polymers between a latex sphere and a glass wall have also been measured with optical tweezers [69]).

The ability to measure interaction forces between *two* colloidal particles both in the bulk of a dispersion and in confinement, have already provided very interesting results, even though almost all were limited to the very dilute regime. These will be discussed in the Section 3.4.

In this section two papers that make use of magnetic forces to measure colloidal force-distance curves are also worth mentioning, despite the fact that this method is at the moment limited to emulsion droplets with quite special properties. The emulsion droplets are filled with a ferrofluid, which can be magnetized in a magnetic field. The magnetic dipolar interactions can subsequently be used to 'handle' the particles in similar ways as can be done with the optical tweezers. Bibette *et al.* used the dipolar interactions to have the ferrofluid droplets, which interact without the magnetic fields as charged particles, self-organize in strings [70]. By changing the field strength the equilibrium distance between the magnetic dipoles could be balanced against the double-layer repulsion. By calibrating the dipole moments and measuring the inter particle distances in the chains through the Bragg reflection (but this could have also been done by imaging) force curves in agreement with DLVO were obtained. Weitz *et al.* used a variation of this method to measure attractive interactions as well. In order to do this they changed the geometry by forcing the particles in a 2D layer between two glass plates [71]. In this geometry the dipoles are repulsive. They obtained the attractive forces drawing the particles together after the magnetic field was switched off by analyzing the (stationary) velocity of the particles as a function of distance and converting this to force vs. distance with the known drag coefficient.

## *Appendix A* The Principle of Archimedes: Effect of the suspension medium

We will calculate the effect of the suspension medium on the Van der Waals forces between two bodies, but the principle used is quite general and we will only use the additivity assumption. In order to calculate the interactions between the two bodies 1 and 2 dispersed in 3 we consider the thermodynamic path depicted in Figure 25. In the initial state bodies 1 and 2 are immersed in 3 but infinitely far apart. We can regard the molecules 3 in the place where body 1 will finally be as constituting a 'ghost-body' 3 as depicted. Now lets remove both body 1 and 3 from the medium to vacuum. The change in free energy for this step is:

$$\Delta F = -F_1 - (F_3 - V_{33}(D) + V_{32}(D)) \tag{44}$$

where $D$ is the distance from 'body 3' to body 2 and $F_i$ is the interaction energy of the isolated body $i$ with a universe of medium 3. The energy change in removing body 3 not just $-F_3$ but $-(F_3 - V_{33}(D) + V_{32}(D))$ as its environment is not just pure 3. The energy $V_{32}(D) - V_{33}(D)$ represents the change in the interaction energy of body 3 with its environment, when the molecules of 3 that would have occupied the position of body 2 are replaced by body 2. In the second step body 3 and body 1 are placed back into the medium but with their positions changed. The free energy change of this second process is:

$$\Delta F' = F_3 + (F_1 - V_{13}(D) + V_{12}(D)) \tag{45}$$

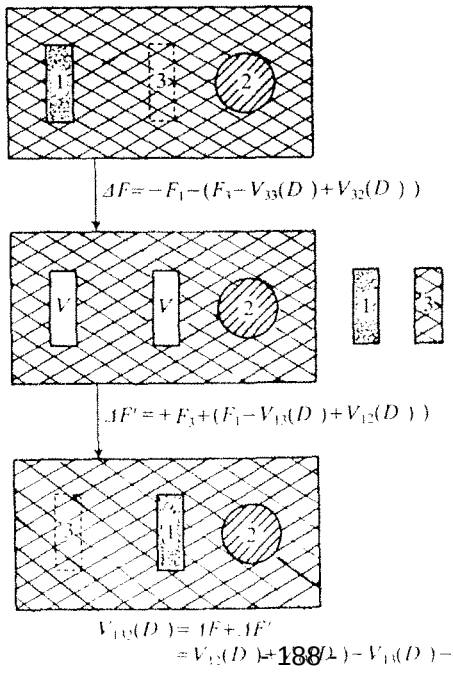
The interaction energy $V_{123}(D)$ of bodies 1 and 2 at separation $D$ immersed in medium 3 is given by

$$V_{123}(D) = \Delta F' + \Delta F = V_{12}(D) + V_{33}(D) - V_{13}(D) - V_{32}(D) \tag{46}$$

Because of the pairwise summation method: $V_{kj}(D) = -A_{kj}V(D)$ where $V(D)$ is a positive function only of the geometry of the system and independent of the nature of bodies 1 and 2 and $A_{kj}$ is the vacuum Hamaker constant. Thus Eq. (20) is obtained in which

$$A_{123} = A_{13} + A_{22} - A_{12} - A_{23} \tag{47}$$

Figure 25 Thermodynamic path for calculating the interaction energy between two bodies 1 and 2 immersed in a third medium 3 [6].



## *Appendix B* Derjaguin Approximation

Derjaguin realized that when surfaces are uniformly curved, their (attractive) force relates to the interaction energy between two planar surfaces. As an example demonstrating this principle (see [7]) we will consider the interaction (here taken to be Van der Waals) between a sphere with radius $R$ at separations $h + R$, which will be taken to be very small, from a half-space as given in Figure 26. Thus we assume in the derivation that $R >> h$. Then we can calculate the force $F$ acting on the sphere by considering the interaction energy $U$ at two positions displaced by an infinitesimal distance d$h$:

$$F = \frac{U(h) - U(h + \mathrm{d}h)}{\mathrm{d}h} \tag{48}$$

When $h/R$ is small, we need only to consider the shell with thickness d$h$ (in the $z$ direction) closest to the half-space; the shell on the other side of the sphere makes a negligible contribution to the force because of the large separation of the interaction potential contribution.

At a lateral distance $r$ from the point of closest approach of the sphere, the shell is at a distance $h_l$ from the surface. From $h_1$ to $h_1 + \mathrm{d}h$ there exists a circular strip with volume $2\pi r\,\mathrm{d}r\,\mathrm{d}h$ . From simple geometry it follows that $h_1 = h + r^2/2R$ as long as $R >> r$. We know the interaction energy per unit volume as $V_{12}(h)$. As in Eq. (48) we obtain the total force by integrating over $r$ so that

$$\begin{aligned} F_{12} &= \int_0^{R\equiv\infty} 2\pi\rho V_{12}(h_1)\,\mathrm{d}r \\ &= 2\pi\rho R\int_h^{\infty} V_{12}(h_1)\,\mathrm{d}h_1 \end{aligned} \tag{49}$$

Where we have used the geometrical relation between $h_l$ and $r$ and extended the upper integration limit because we assume $V_{12}(h)$ is negligible for $h_l$ of order $R$. If we compare Eq. (49) with the interaction energy between two planar surfaces, Eq. (18). The final integration step leading to Eq. (18) amounts to an integration over planar sheets of thickness d$z_2$ so that

$$U_{12}(h) = area\int_0^{\infty} \rho V_{12}(z)\,\mathrm{d}z \tag{50}$$

with $V_{12}$ is the interaction energy between a molecule and a half-space. The integrals in Eqs. (55) and (50) are the same except for the integration variables and thus

$$F_{12}(h) = 2\pi R\frac{U_{12}(h)}{area} \tag{51}$$

as long as radius $R$ is large enough compared to the range of $V_{12}(h)$.

In the derivation only the additivity assumption is used, not the interactions self, therefore this result is quite general as long as the potentials are additive and the range is short compared to $R$. Similar equations can be derived for two crossed cylinders with radii $R_1$ and $R_2$:

$$F(h) = 2\pi\sqrt{R_1 R_2}\frac{U(h)}{area} \tag{52}$$

And for two spheres:

$$F(h) = 2\pi\frac{R_1 R_2}{R_1 + R_2}\frac{U(h)}{area} \tag{53}$$

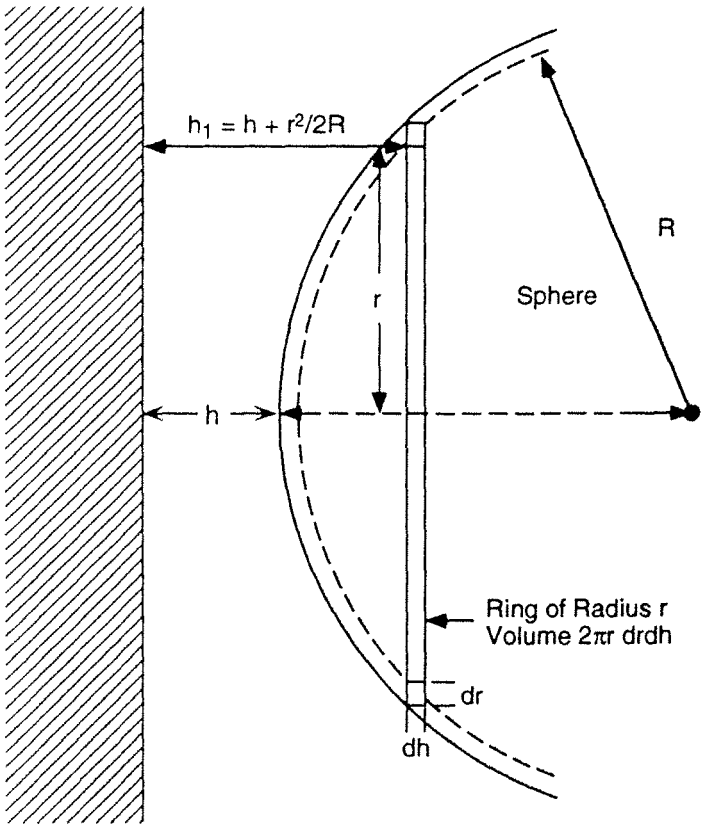



**Figure 26** Calculating the force between a sphere and a half-space.

## *Appendix C* Optical Tweezers or Single-beam gradient optical traps

The ability to measure forces in the pN range over nm distances with light was initiated by Ashkin in the early 70's when he showed that it is possible to move and trap atoms and dielectric particles with a highly focused laser beam [134]. The particle refractive index, $n_p$, has to be higher than that of the dispersion medium, $n_m$, or, $m = n_p/n_m$, has to be larger than 1. There are several regimes, depending on the ratio of the particle diameter, $a$, and the wavelength in the medium, $\lambda$, and $m$, in which these forces can be calculated or approximated.

The first regime of electromagnetic radiation interacting with a dielectric sphere is called the Rayleigh regime. A Rayleigh scatterer is a particle with a size much smaller than the wavelength, $a << \lambda$. In this regime the particle can be replaced by an effective dipole moment. If we consider the most simple optical trap formed by focusing a laser beam with a Gaussian intensity profile across the beam (e.g., like the fundamental spatial mode of a laser, the $TEM_{00}$ mode) propagating along the $z$ direction and with polarization in the $x$ direction than we can distinguish for this scattering regime two kinds of forces acting on the particle. These are the gradient and scattering forces. The gradient force tends to pull the particle into the region of highest intensity thus minimizing the energy of the dielectric sphere in the electromagnetic field. In the transverse directions the gradient is due to the Gaussian intensity profile. In the $z$ direction the presence of a focal point creates the gradient. The second force, the scattering force, is due to radiation pressure and destabilizes along the $z$-direction by pushing the particle out of the trap. There is, however, a small region where the gradient force exceeds the radiation pressure, thus defining the trapping region. Although most particles of interest will not fall in the Rayleigh regime, it is only for this regime that analytical results can be given. We will present these formula's, because they give at least a feeling for the relevant parameters of the problem. Under the above mentioned assumptions the gradient force, $\mathbf{F}_{grad}$ on a particle is given by [135]:

$$\mathbf{F}_{grad} = \frac{4\pi a^3}{c}\left(\frac{m^2-1}{m^2+2}\right)\cdot\nabla I \tag{54}$$

here $c$ is the velocity of light in vacuum, $I$ the light intensity. The scattering force is given by [135]:

$$\mathbf{F}_{scat} = \frac{I}{c}\frac{128\pi^5 a^6}{3\lambda^4}\left(\frac{m^2-1}{m^2+2}\right)^2 n_p \mathbf{z} \tag{55}$$

The most important conclusions, which are valid in the other regimes as well, that can be drawn from these equations is that there is an unequal dependence of the forces on the sphere size and refractive index and that there is an *optimal* radius and refractive index difference for trapping of spheres.

On the other extreme are particles that are (much) larger than $\lambda$. Here one enters the regime of geometrical optics and the forces can be calculated (numerically) by using ray optics and summing over all directions of the highly focused light. Again the forces can be decomposed in a trapping gradient contribution and a destabilizing

scattering contribution [136]. These calculations can be summarized by stating that the optimal refractive index difference is close to $m$ = 1.6. (In this regime the results are not depending on the radius anymore). It is likely that the ray optics calculations still give a reasonable estimate at particle sizes of about 5λ [136].

In the intermediate regime, the most relevant for most applications of the tweezers, the particle size is of the order of the wavelength ($m$ is relatively large) and the interactions with light are called Mie scattering. In this regime the forces are by far the most difficult to calculate and for the geometry of a highly focused light beam these calculations have not yet been done! For Mie scattering in the case of a plane wave of incidence, it is still possible to find an analytical solution, although in the form of a slowly converging series. For most other particle forms or a strongly converging light beam this is not the case.

The development of optical tweezers and very sensitive position detection has benefited a lot from researchers from the biophysical community. In this field there

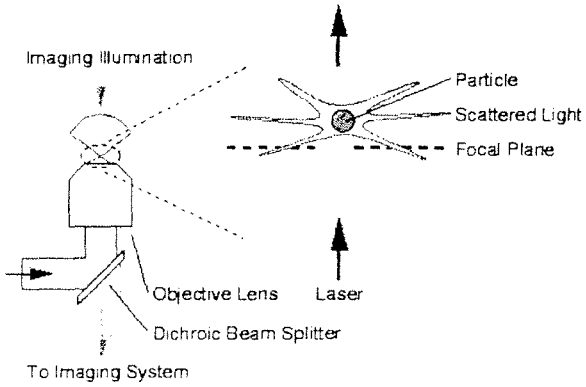



**Figure 27** Schematic diagram of a single beam optical trap or a pair of optical tweezers [140].

have already been quite a number of successful measurements of very small forces (~ 1 pN) on bio-molecules and a lot of quite advanced manipulations with the optical tweezers. To build up a simple set of tweezers is not hard and well described in the literature [137, 138, 139]. Recently, tweezers are finding their way to the colloid community as well [140].

## REFERENCES


1 B. V. Derjaguin, L. Landau, *Acta Physiochim.*, **14**, 633 (1941).
2 E. J. Verwey, J. Th. Overbeek, *Theory of the Stability of Lyophobic Colloids*, Elsevier, New York (1948).
3 *Ordering and Phase Transitions in Charged Colloids*, eds. A. K. Arora and B. V. R. Tata, VCH Publishers, New York (1996).
4 W. B. Russel, D. A. Saville, W. R. Schowalter, *Colloidal Dispersions*, Cambridge Univ. Press, Cambridge (1991).
5 J. Israelachvili, *Intermolecular & Surface Forces*, Academic Press, San Diego, 2nd ed. (1991).
6 R. J. Hunter, *Foundations of Colloid Science, Volume I and II*, Clarendon Press, Oxford (1989).
7 D. F. Evans, H. Wennerström, *The Colloidal Domain, where physics, chemistry, biology and technology meet*, VCH Publishers, New York (1994).
8 J. Lyklema, *Fundamentals of Interface and Colloid Science*, Academic Press, Volume I, II, III, IV, V (1995).
9 R. J. Hunter, *Zeta potential in colloid science*, Academic Press, London (1981).
10 P. C. Hiemenz, *Principles of Colloid and Surface Chemistry*, Marcel Dekker, 2nd Ed., New York, (1986).
11 M. J.Vold, R. D. Vold, *Colloid and Interface Chemistry*, Addison-Wesley, Reading (1983).
12 D. H. Napper, *Polymeric Stabilization of Colloidal Dispersions*, Academic Press, London (1983).
13 J. M. Sparnaay, *J. Colloid Interface Sci.*, **91**, 307 (1984).
14 Van der Waals, Thesis
15 J. Mahanty, B. W. Ninham, *Dispersion Forces*, Academic Press, New York (1976).
16 H. C. Hamaker, *Physica*, **4**, 1058 (1937).
17 K. J. Mysels, P. C. Scholten, *Langmuir*, **7**, 209 (1991).
18 J. Th. G. Overbeek, *Fifty Years Integration of Forces*, Utrecht University (1981).
19 H. B. G. Casimir, D. Polder, *Phys. Rev.*, **73**, 360 (1948).
20 E. M. Lifshitz, *Sov. Phys. JETP.*, **2**, 73 (1956).
21 I. E. Dzyaloshinskii E. M. Lifshitz, L. P. Pitaevski, *Adv. Phys.*, **10**, 165 (1961).
22 B. Jönsson, T. Äkesson, C. E.Woodward in *Ordering and Phase Transitions in Charged Colloids*, eds. A. K. Arora and B. V. R. Tata, VCH Publishers, New York (1996), Chapt. 11.
23 S. Alexander, P. M. Chaikin, P. Grant, G. J. Morales, P. Pincus, D. Hone, *J. Chem. Phys.*, **80**, 5776 (1984).
24 H. Löwen, G. Kramposthuber, *Europhysics Letters*, **23**, 673 (1993).
25 R. Rajagopalan, K. Srinivasa Rao, *Phys. Rev. E.*, **55**, 4423 (1997).
26 R. Rajagopalan in *Ordering and Phase Transitions in Charged Colloids*, eds. A. K. Arora and B. V. R. Tata, VCH Publishers, New York (1996), Chapter 13.
27 B. V. Derjaguin, A. S. Titijevskaia, I. I. Abrikossova, A. D. Malkina, *Disc. Far. Soc.*, **18**, 24 (1954).
28 J. Th. G. Overbeek, M. J. Sparnaay, *Disc. Far. Soc.*, **18**, 12 (1954).
29 D. Tabor, R. H. S. Winterton, *Proc. Roy. Soc. Lond. A*, **312**, 435 (1969).

30 J. N. Israelachvili, G. E. Adams, *J. Chem. Soc. Faraday Trans. 1*, **74**, 975 (1978).
31 J. N. Israelachvili, *Faraday Discuss. Chem. Soc.*, **65**, 20 (1978).
32 H. K. Christenson, *J. Chem. Soc. Faraday Trans. 1*, **80**, 1933 (1984).
33 J. N. Israelachvili, P. M. Mcguiggan, *J. Mater. Res.*, **5**(10), 2223 (1990)
34 J. N. Israelachvili, H. Wennerström, *Nature*, **379**, 219 (1996).
35 S. Marcelja, *Nature*, **385**, 2689 (1997).
36 J. L. Parker, P. Richetti, P. Kékicheff, S. Sarman, *Phys. Rev. Lett.*, **68**, 1955 (1992).
37 M. Ruths, H. Yoshizawa, L. J. Fetters, J. N. Israelachvili, *Macromolecules*, **29**, 7193 (1996).
38 J. N. Israelachvili, M. Tirrell, J. Klein, Y. Almog, *Macromolecules*, **17**, 204 (1984).
39 J. P. Chapel, *J. Colloid Interface Sci.* **162**, 517 (1994).
40 G. Vigil, Z. Xu, S. Steinberg, J. N. Israelachvili, *J. Colloid Interface Sci.* **165**, 367(1994).
41 H. K. Christenson in *Modern Approaches to Wettability: Theory and Applications*, M. E. Schrader, G. Loeb, eds., Plenum Press, New York (1992).
42 J. N. Israelachvili, P. M. Mcguiggan, *J. Mater. Res.*, **5**(10), 2223 (1990) and Refs. cited.
43 R. G. Horn, D. T. Smith, W. Haller, *Chem. Phys. Lett.*,**162**, 404 (1989).
44 D. Atkins, P. Kékicheff, O. Spalla, *J. Colloid Interface Sci.* **188**, 234 (1997).
45 P. K. Hansma, V. B. Elings, O. Marti, C. E. Bracker, *Science*, **242**, 209 (1988) and refs. cited.
46 A. Ducker, T.J. Senden, R.M. Pashley, *Nature*, **353**, 231 (1991).
47 H.-J. Butt, *Biophys. J.*, **60**, 777 (1991).
48 L. Meagher, *J. Colloid Interface Sci.* **152**, 293 (1992).
49 Y. Q. Li, N. J. Tao, J. Pan, A. A. Garcia, S. M. Lindsay, *Langmuir*, **9**, 637 (1993).
50 P. G. Harley, I. Larson, P. J. Scales, *Langmuir*, **13**, 2207 (1997).
51 D. C. Prieve, F. Luo, F. Lanni, *Faraday Discuss. Chem. Soc.*, **83**, 222 (1987).
52 Dennis C. Prieve, Nasser A. Frej, *Langmuir*, **6**(2), 396 (1990).
53 D. C. Prieve, S. G. Bike, Nasser A. Frej, *Faraday Discuss. Chem. Soc.*, **90**, 209 (1990).
54 H. Chew, D. S. Wang, M. Kerker, *Appl. Opt.*, **18**, 2679 (1979).
55 M.A. Brown, A.L. Smith, E.J. Staples, *Langmuir*, **5**(6), 1319 (1989).
56 M.A. Brown, E.J. Staples, *Faraday Discuss. Chem. Soc.*, **90**, 193 (1990).
57 M.A. Brown, E.J. Staples, *Langmuir*, **6**, 1260 (1990).
58 Scott G. Flicker, Stacy G Bike,. *Langmuir*, **9**(1), 257 (1993).
59 Scott G.Flicker, Jennifer L. Tipa, Stacy G. Bike, *Journal of colloid and interface science*, **158**(2), 317 (1993).
60 S. Tanimoto; H. Matsuoka, H. Yamaoka, *Colloid Polym. Sci.*, **273**, 1201 (1995).
61 G. M. Kepler, S. Fraden, *Phys. Rev. Lett.*, **73**, 356 (1994).
62 J.C. Crocker, D.G. Grier, *Phys. Rev. Lett.*, **73**, 352 (1994).
63 K. Vondermassen, J. Bongers, A. Mueller, H. Versmold, *Langmuir*, **10**, 1351 (1994).
64 S. Fraden, G. M. Kepler, *Langmuir*, **10**(8), 2501 (1994).

65 M. D. Carbajal-Tinoco, F. Castro-Román; J. L. Arauz-Lara, *Phys. Rev. E.*, **53**, 3745 (1996,).
66 J.C. Crocker, D.G. Grier, *Journal of colloid and interface science*, **179**, 298 (1996).
67 The problem referred to is the influence of hydrodynamic interactions on the displacements measured by Grier et al. in the 1/30 s. They claim to measure equilibrium properties, but this aspect of the measurement is not explained well in their papers and is not understood by the author of these notes.
68 T. Sugimoto, T. Takahashi, H. Itoh, S. Sato, A. Muramatsu, *Langmuir*, **13**, 5528 (1997).
69 Y.N. Ohshima, H. Sakagami, K. Okumoto, T. Tokoyoda, T. Igarashi, K.B. Shinktaku, S. Toride, H. Sekino, K. Kabuto, I. Nishio, *Phys. Rev. Lett.*, **78**(20), 3963 (1997).
Crocker,
70 F.L. Calderon, T. Stora, O. M. Monval, P. Poulin, J. Bibette, *Phys. Rev. Lett.*, **72**(18), 2959 (1994).
71 P. Poulin, V. Cabuil, D. A. Weitz, *Phys. Rev. Lett.*, **79**(24), 4862 (1997).
72 S. Inoué, K. R. Spring, *Video Microscopy*, 2nd Ed., Plenum Press, New York (1997).
73 *Confocal Microscopy*, Wilson Ed., Academic Press, London (1990).
74 A. van Blaaderen, P. Wiltzius, *Science*, **270**, 1177-1179 (1995).
A. van Blaaderen, R. Ruel, P. Wiltzius, *Nature*, **385**, 321 (1997).
W. K. Kegel, A. van Blaaderen, *Science*, **287**, 290-293 (2000).
75 N. Ise, *Faraday Discuss. Chem. Soc.*, **90**, 62 (1990).
76 N. Ise, *Faraday Discuss. Chem. Soc.*, **90**, 182 (1990).
77 N. Ise, K. Ito, H. Matsuoka, H. Yoshida, in *Ordering and Phase Transitions in Charged Colloids*, eds. A. K. Arora and B. V. R. Tata, VCH Publishers, New York (1996), Chapt. 5.
78 H. Yoshida, J. Yamanaka, T. Koga, N. Ise, T. Hashimoto, *Langmuir*, **13**, 5528 (1998).
79 N. Ise, T. Okubo, Y. Hiragi, H. Kawai, T. Hashimoto, M. Fujimura, A. Nakajima, H. Hayashi, *J. Am. Chem. Soc.*, **101**, 5836 (1979).
80 N. Ise, T. Okubo, K. Yamamoto, H. Kawai, T. Hashimoto, M. Fujimura, Y. Hiragi *J. Am. Chem. Soc.*, **102**, 7901 (1980);
*J. Chem. Phys.*, **78**, 541 (1983);
*J. Chem. Phys.*, **81**, 3294 (1984)
81 J. G. Daly, R. Hasting, *J. Phys. Chem.*, **85**, 294 (1981)
82 F. Grüner, W. P. Lehmann, *J. Phys. A: Math. Gen.*, **15**, 2847 (1982).
83 A. Arora, R. Kesavamoorthy, *Solid State Commun.*, **54**, 1047 (1985).
84 K. Ito, H. Nakamura, N. Ise, *J. Chem. Phys.*, **85**, 6136 (1986)
85 K. Ito, H. Nakamura, H. Yoshida, N. Ise, *J. Am. Chem. Soc*, **110**, 6955 (1988).
86 H. Yoshida, K. Ito, N. Ise, *J. Chem. Soc. Faraday Trans.*, **87**, 371 (1991) and Refs. cited.
87 A. Kose, M. Ozaki, K. Takano;Y. Kobayashi, S. Hachisu, *J. Colloid Interface Sci.* **44**, 330 (1973).
88 N. Ise, H. Matsuoka, K. Ito, H. Yoshida, *Discuss. Faraday Soc.*, **90**, 153 (1990).
89 R. Kesavamoorthy, M. Rajalakshmi, C. B. Rao, *J. Phys. Condens. Matter*, **1**, 7149 (1989).

90 A. K. Arora, B. V. Tata, A. K. Sood, R. Kesavamoorthy, *Phys. Rev. Lett.*, **60**, 2438 (1988).
91 B.V.R. Tata, M. Rajalakshmi, A. K. Arora, *Phys. Rev. Lett.*, **69**, 3778 (1992).
92 T. Palberg, M. Würth, *Phys. Rev. Lett.*, **72**, 786 (1994).
93 B.V.R. Tata, A. K. Arora, *Phys. Rev. Lett.*, **72**, 787 (1994).
94 S. Dosho, N. Ise, K. Ito; et al. *Langmuir*, **9**, 394 (1993).
95 K. Ito; H. Yoshida; N. Ise, *Chem. Letters*, **1992**, 2081 (1992).
96 K. Ito; H. Yoshida; N. Ise, *Science*, **263**, 66 (1994).
97 H. Yoshida, E. Yamahara, P. V. Rajamani, N. Ise, *Phys. Rev. Lett.*, **78**(13), 2660 (1997).
98 B.V.R. Tata, A. K. Arora *Phys. Rev. Lett.*, **72**, 787 (1994).
99 K. Ito, T. Muramoto, H. Kitano, *J. Am. Chem. Soc*, **117**, 5005 (1995).
100 T. Muramoto, K. Ito, H. Kitano, *J. Am. Chem. Soc*, **119**, 3592 (1997).
101 F. Ghezzi, J. C. Earnshaw, *J. Phys.: Condens. Matter*, **9**, L517 (1997)
102 J. Ruiz-Garzia, R. Gamez-Corrales, B. I. Ivlev, *Physica A.*, **236**, 97 (1997).
103 A. E. Larsen, D. G. Grier, *Nature*, **385**, 230 (1993).
104 A. E. Larsen, D. G. Grier, *Phys. Rev. Lett*, **76**, 3862 (1996).
105 J.C. Crocker, D.G. Grier, *Phys. Rev. Lett.*, **77**, 1897 (1996).
106 B.V.R. Tata, A.K. Arora, *Phys. Rev. Lett.*, **75**, 3200 (1995).
107 N. Ise, J. Yamanaka, *Phys. Rev. Lett.*, This comment was never published, see [108]
108 D. G. Grier, J. C. Crocker, *Phys. Rev. Lett.*, Reply never published. However the text can be read at: http://rainbow.uchicago.edu/~grier/
109 I. Sogami, N. Ise, *J. Chem. Phys.*, **81**, 6320 (1984).
110 J. Th. G. Overbeek, *J. Chem. Phys.*, **87**, 4406 (1987).
111 C. E. Woodward, *J. Chem. Phys.*, **89**, 5140 (1988).
112 M. V. Smalley, *Molec. Phys.*, **6**, 1251 (1990).
113 I. Sogami, T. Shinohara, M. V. Smalley, *Molec. Phys.*, **74**, 599 (1991); **76**, 1 (1992).
114 S. Levine, D. G. Hall, *Langmuir*, **8**, 1090 (1992).
115 J. Th. G. Overbeek, *Molec. Phys.*, **80**, 685 (1993).
116 R. Ettelaie, *Langmuir*, **9**, 1888 (1993).
117 I. Sogami, M. V. Smalley, *Molec. Phys.*, **85**, 869 (1995).
118 M. V. Smalley, *Langmuir*, **11**, 1813 (1995).
119 M. V. Smalley in *Ordering and Phase Transitions in Charged Colloids*, eds. A. K. Arora and B. V. R. Tata, VCH Publishers, New York (1996), Chapt. 12.
120 D. G. Hall, *Langmuir*, **12**, 4308 (1996).
121 K. S. Schmitz, *Langmuir*, **12**, 1407 (1996).
122 K. S. Schmitz, *Langmuir*, **12**, 3828 (1996).
123 K. S. Schmitz, *Langmuir*, **13**, 5852 (1996).
124 B. V. R. Tata, Akhilesh K. Arora, M. C. Valsakumar, M. C., *Phys. Rev. E.*, **47**, 3404 (1993).
125 M. Medina-Noyola, D. A McQuarrie, *J. Chem. Phys.*, **73**, 6279 (1980).
126 D. A. McQuarrie, *Statistical Mechanics*, Harper & Row, New York (1973).
127 M. D. Carbajal-Tinocco, D. G. Grier, submitted to *J. Chem. Phys.*
128 P. Kékicheff, O. Spalla, *Phys. Rev. Lett.*, **75**(9), 1851 (1995).
129 O. Spalla, L. Belloni, *J. Chem. Phys.*, **95**, 7689 (1991).
130 O. Spalla, L. Belloni, *Phys. Rev. Lett.*, **74**(13), 2515 (1995).

131 L. Belloni, O. Spalla, *J. Chem. Phys.*, **107**(2), 465 (1997).
132 R. van Roij, J-P. Hansen, *Phys. Rev. Lett.*, **79**(16), 1382 (1997).
133 A. Delville, *Langmuir*, **12**, 2605 (1996)
134 A. Ashkin, *Phys. Rev. Lett.*, **24**, 156 (1970).
135 A. J. Simon, Ph.D. Thesis, University of Chicago (1992).
136 A. Ashkin, "Forces of a Single-Beam Laser Trap on a Dielectric Sphere in the Ray Optics Regime", in "Methods in Cell Biology: Laser Tweezers in Cell Biology", Volume 55, M. P. Sheetz ed., Academic Press, San Diego, 1998, pp 1-25.
137 "Methods in Cell Biology: Laser Tweezers in Cell Biology", Volume 55, M. P. Sheetz ed., Academic Press, San Diego, 1998, pp 1-25.
138 K. Svoboda, Ch. R. Schmidt, B. J. Schnapp, S.M. Block, *Nature*, **365**, 721 (1993).
139 K. Visscher, S. P. Gross, S. M. Block, *IEEE J. Selected Topics Quantum Electronics*, **2**(4), 1066 (1996).
140 D. G. Grier, *Current Opinion in Colloid & Interface Science*, **2**(3), 264 (1997).
141 X. Chu, D. T. Wasan, *J. Colloid Interface Sci.* **184**, 268 (1996).

## REFERENCES FEBRUARY 1998 – FEBRUARY 2002

141 O. I. Vinogradova, G. E. Yakubov and H. J. Butt, Forces between polystyrene surfaces in water-electrolyte solutions: Long-range attraction of two types? J. Chem. Phys. **114**(18), 8124 (2001).
142 J. Lyklema, H. P. van Leeuwen and M. Minor, *DLVO-theory, a dynamic re-interpretation.* Adv. Colloid Interface Sci. **83**(1-3), 33 (1999).
143 N. V. Churaev, *The DLVO theory in Russian colloid science.* Adv. Colloid Interface Sci. **83**(1-3), 19 (1999).
144 B. W. Ninham, *On progress in forces since the DLVO theory.* Adv. Colloid Interface Sci. **83**(1-3), 1 (1999).
145 W. Wu, R. F. Giese and C. J. van Oss, *Stability versus flocculation of particle suspensions in water - correlation with the extended DLVO approach for aqueous systems, compared with classical DLVO theory.* Colloid Surf. B-Biointerfaces **14**(1-4), 47 (1999).
146 D. Leckband and S. Sivasankar, *Forces controlling protein interactions: theory and experiment.* Colloid Surf. B-Biointerfaces **14**(1-4), 83 (1999).
147 P. Attard, *Recent advances in the electric double layer in colloid science.* Curr. Opin. Colloid Interface Sci. **6**(4), 366 (2001).
148 M. Dijkstra, *Computer simulations of charge and steric stabilised colloidal suspensions.* Curr. Opin. Colloid Interface Sci. **6**(4), 372 (2001).
149 T. J. Senden, *Force microscopy and surface interactions.* Curr. Opin. Colloid Interface Sci. **6**(2), 95 (2001).
150 N. Ise, T. Konishi and J. Yamanaka, *X-Ray scattering study of ionic colloidal crystals.* Curr. Opin. Colloid Interface Sci. **6**(2), 126 (2001).
151 N. Ise, T. Konishi and B. V. R. Tata, *How homogeneous are "homogeneous dispersions"? Counterion- mediated attraction between like-charged species.* Langmuir **15**(12), 4176 (1999).
152 A. K. Arora and B. V. R. Tata, *Interactions, structural ordering and phase*

*transitions in colloidal dispersions.* Adv. Colloid Interface Sci. **78**(1), 49 (1998).
153 B. V. R. Tata, *Colloidal dispersions and phase transitions in charged colloids.* Curr. Sci. **80**(8), 948 (2001).
154 J. P. Hansen and H. Lowen, *Effective interactions between electric double layers.* Annu. Rev. Phys. Chem. **51**, 209 (2000).
155 L. Belloni, *Colloidal interactions.* J. Phys.-Condes. Matter **12**(46), R549 (2000).
156 D. G. Grier, *When like charges attract: interactions and dynamics in charge-stabilized colloidal suspensions.* J. Phys.-Condes. Matter **12**(8A), A85 (2000).
157 H. Lowen, M. Watzlawek, C. N. Likos, M. Schmidt, A. Jusufi and A. R. Denton, *Phase transitions in colloidal suspensions and star polymer solutions.* J. Phys.-Condes. Matter **12**(8A), A465 (2000).
158 V. Vlachy, *Ionic effects beyond Poisson-Boltzmann theory.* Annu. Rev. Phys. Chem. **50**, 145 (1999).
159 C. N. Likos, *Effective interactions in soft condensed matter physics.* Phys. Rep.-Rev. Sec. Phys. Lett. **348**(4-5), 267 (2001).
160 M. Kardar and R. Golestanian, *The "friction" of vacuum, and other fluctuation-induced forces.* Rev. Mod. Phys. **71**(4), 1233 (1999).
161 E. Ruckenstein, *Attraction between identical colloidal particles caused by collective electrostatic repulsion.* Adv. Colloid Interface Sci. **75**(3), 169 (1998).
162 J. P. Hansen, D. Goulding and R. van Roij, *Effective interactions between charged colloidal particles: Repulsion, attraction and phase separation.* J. Phys. IV **10**(P5), 27 (2000).
163 M. Bostrom, D. R. M. Williams and B. W. Ninham, *Specific ion effects: Why DLVO theory fails for biology and colloid systems - art. no. 168103.* Phys. Rev. Lett. **8716**(16), 8103 (2001).
164 E. J. Verwey, J. Th. Overbeek, *Theory of the Stability of Lyophobic Colloids*, Dover, New York (2000).
165 T. J. Senden, *Force microscopy and surface interactions.* Curr. Opin. Colloid Interface Sci. **6**(2), 95 (2001).
166 J. Yamanaka, H. Yoshida, T. Koga, N. Ise and T. Hashimoto, *Reentrant solid-liquid transition in ionic colloidal dispersions by varying particle charge density.* Phys. Rev. Lett. **80**(26), 5806 (1998).
167 G. C. de Leon, J. M. Saucedo-Solorio and J. L. Arauz-Lara, *Colloidal interactions in partially quenched suspensions of charged particles.* Phys. Rev. Lett. **81**(5), 1122 (1998).
168 G. C. de Leon and J. L. Arauz-Lara, *Static structure and colloidal interactions in partially quenched quasibidimensional colloidal mixtures.* Phys. Rev. E **59**(4), 4203 (1999).
169 J. A. Weiss, A. E. Larsen and D. G. Grier, *Interactions, dynamics, and elasticity in charge-stabilized colloidal crystals.* J. Chem. Phys. **109**(19), 8659 (1998).
170 E. R. Dufresne and D. G. Grier, *Interactions, dynamics, and elasticity in charge-stabilized colloidal crystals (vol 109, pg 8659, 1998).* J. Chem. Phys. **110**(17), 8845 (1999).
171 T. M. Squires and M. P. Brenner, *Like-charge attraction and hydrodynamic interaction.* Phys. Rev. Lett. **85**(23), 4976 (2000).
172 S. H. Behrens and D. G. Grier, *Pair interaction of charged colloidal spheres near a charged wall .* Phys. Rev. E **6405**(5), 0401 (2001).
173 J. C. Crocker, J. A. Matteo, A. D. Dinsmore and A. G. Yodh, *Entropic attraction*

*and repulsion in binary colloids probed with a line optical tweezer.* Phys. Rev. Lett. **82**(21), 4352 (1999).
174 B. V. R. Tata and N. Ise, *Monte Carlo study of structural ordering in charged colloids using a long-range attractive interaction.* Phys. Rev. E **58**(2), 2237 (1998).
175 D. G. Grier and J. C. Crocker, *Comment on "Monte Carlo study of structural ordering in charged colloids using a long-range attractive interaction".* Phys. Rev. E **61**(1), 980 (2000).
176 D. G. Grier, *Colloids: A surprisingly attractive couple.* Nature **393**, 621 (1998).
177 W. R. Bowen and A. O. Sharif, *Long-range electrostatic attraction between like-charge spheres in a charged pore.* Nature **393**(6686), 663 (1998).
178 J. C. Neu, *Wall-mediated forces between like-charged bodies in an electrolyte.* Phys. Rev. Lett. **82**(5), 1072 (1999).
179 J. E. Sader and D. Y. C. Chan, Long-range electrostatic attractions between identically charged particles in confined geometries and the Poisson- Boltzmann theory. Langmuir **16**(2), 324 (2000).
180 J. E. Sader and D. Y. C. Chan, *Long-range electrostatic attractions between identically charged particles in confined geometries: An unresolved problem.* J. Colloid Interface Sci. **213**(1), 268 (1999).
181 Y. X. Qian and W. R. Bowen, *Long-range electrostatic interaction between a charged wall and two similarly charged colloidal spheres at low surface potentials,* Journal of Colloid and Interface Science **213**, 316 (1999).
182 W. R. Bowen and A. O. Sharif, *Hydrodynamic and colloidal interactions effects on the rejection of a particle larger than a pore in microfiltration and ultrafiltration membranes,* Chemical Engineering Science **53**, 879 (1998).
183 W. R. Bowen, A. N. Filippov, A. O. Sharif, *et al., A model of the interaction between a charged particle and a pore in a charged membrane surface,* Advances in Colloid and Interface Science **81**, 35 (1999).
184 W. R. Bowen and A. O. Sharif, *Long-range electrostatic attraction between like-charge spheres in a charged pore (vol 393, pg 663, 1998),* Nature **402**, 841 (1999).
185 J. J. Gray, B. Chiang, and R. T. Bonnecaze, *Colloidal particles - Origin of anomalous multibody interactions,* Nature **402**, 750 (1999).
186 W. R. Bowen and A. O. Sharif, *Long-range electrostatic attraction between like-charge spheres in a charged pore (vol 393, pg 663, 1998),* Nature **402**, 841 (1999).
187 R. van Roij, M. Dijkstra and J. P. Hansen, *Phase diagram of charge-stabilized colloidal suspensions: van der Waals instability without attractive forces.* Phys. Rev. E **59**(2), 2010 (1999).
188 P. B. Warren, *A theory of void formation in charge-stabilized colloidal suspensions at low ionic strength.* J. Chem. Phys. **112**(10), 4683 (2000).
189 D. Y. C. Chan, *Density functional theory of charged colloidal systems - art. no. 061806.* Phys. Rev. E **6306**(6), 1806 (2001).
190 D. Y. C. Chan, P. Linse and S. N. Petris, *Phase separation in deionized colloidal systems: Extended Debye-Huckel theory.* Langmuir **17**(14), 4202 (2001).
191 B. Beresford-Smith, D. Y. C. Chan and D. J. Mitchell, *The electrostatic interaction in colloidal systems with low added electrolyte.* J. Colloid Interface Sci. **105**, 216 (1985).

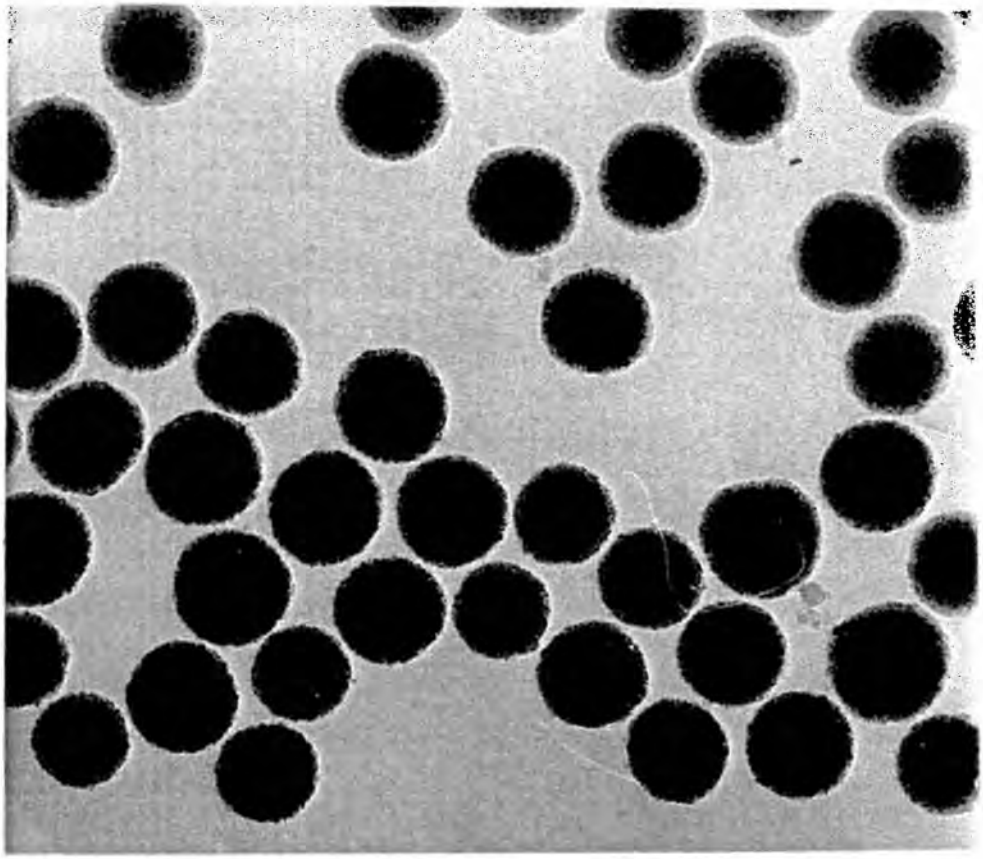
192 G. Stell, *Criticality and Phase-Transitions in Ionic Fluids.* J. Stat. Phys. **78**(1-2), 197 (1995).
193 H. Lowen and E. Allahyarov, *The role of effective triplet interactions in charged colloidal suspensions.* J. Phys.-Condes. Matter **10**(19), 4147 (1998).
194 D. Goulding and J. P. Hansen, *Attraction between like-charged colloidal particles induced by a surface: A density-functional analysis.* Europhys. Lett. **46**(3), 407 (1999).
195 R. Tehver, F. Ancilotto, F. Toigo, J. Koplik and J. R. Banavar, *Absence of many-body effects in interactions between charged colloidal particles.* Phys. Rev. E **59**(2), R1335 (1999).
196 E. Allahyarov, I. D. D'Amico and H. Lowen, *Effect of geometrical confinement on the interaction between charged colloidal suspensions.* Phys. Rev. E **60**(3), 3199 (1999).
197 P. Linse and V. Lobaskin, Electrostatic attraction and phase separation in solutions of like-charged colloidal particles. Phys. Rev. Lett. **83**(20), 4208 (1999).
198 P. Linse, Structure, phase stability, and thermodynamics in charged colloidal solutions. J. Chem. Phys. **113**(10), 4359 (2000).
199 P. Linse and V. Lobaskin, Electrostatic attraction and phase separation in solutions of like-charged colloidal particles. J. Chem. Phys. **112**(8), 3917 (2000).
200 V. Lobaskin, A. Lyubartsev and P. Linse, *Effective macroion-macroion potentials in asymmetric electrolytes.* Phys. Rev. E **6302**(2), 0401 (2001).
201 I. Larson and P. Attard, *Surface charge of silver iodide and several metal oxides. Are all surfaces nernstian?* J. Colloid Interface Sci. **227**(1), 152 (2000).

# 7. Colloid synthesis and characterization

Albert Philipse
*Van 't Hoff Laboratory, Debye Institute*
*Utrecht University*

# Mastercourse Synthesis of Complex Nanostructures

## Aspects of Colloid Synthesis



**A. P. Philipse**
**Van 't Hoff Laboratory Physical and Colloid Chemistry**
**Padualaan 8, 3584 CH Utrecht**
**January 2005**

**Acknowledgement**

The following text is a manuscript for a chapter from a book[1] to appear in 2005. The author is very grateful to Mrs. Marina Uit de Bulten and Mieke Lanen for their help in the preparation of this manuscript, and to Prof. J. Lyklema for his editorial work and many suggestions. Karel Planken, Chantal Vonk and Natalie Zuiverloon are thanked for their input for sections 2.4b and 2.4d.

[1] The full reference is: A.P. Philipse, 'Particulate Colloids: Aspects of Preparation and Characterization' in J. Lyklema (ed.) *Fundamentals of Colloids and Interface Science,* Vol. IV (Elsevier, 2005, in print)

## 2 PARTICULATE COLLOIDS: ASPECTS OF PREPARATION AND CHARACTERIZATION

**Albert Philipse**

# 2 PARTICULATE COLLOIDS: ASPECTS OF PREPARATION AND CHARACTERIZATION

**ALBERT PHILIPSE**

## 2.1 Introduction

Dispersions of inorganic colloids have been prepared and processed since the very beginning of human technology. Already around 7000 BC, about 4000 years before the invention of the wheel, the Near East produced complicated ceramic shapes, which manifested a thorough practical knowledge of concentrated clay dispersions and their processing. Such knowledge is still indispensable in the fabrication of traditional ceramics, such as pottery. The desired outcome of shaping techniques, such as the slip casting of clay dispersions, critically depends on the skilful preparation of colloidal suspensions. Important parameters are the shape and size distribution of particles, their concentration and state of aggregation, which is controlled by ionic strength and polymeric additives. Optimization of these parameters is often a laborious trial and error process and, so, it is not surprising that details of industrial preparation are usually either patented or kept confidential. Another impressive and historical example of dispersion preparation underlies the very pages on which this text is written. Papermaking[1)] starts with the degradation of wood chips to an aqueous suspension of cellulose fibres with a large percentage of fibres with dimensions in the colloidal size range. Inorganic particles, in the form of silica or bentonite sols, are added to improve the quality and rate of papermaking, a process which comprises the filtering and drying of the mixture of fibres and sol particles on a wire. Dried sheets run out of a papermaking machine at a rate of a few hundred metres per minute, or even faster, and any slight change in the properties and composition of the starting dispersions may have a disastrous effect on this very rapid process.

Paper also reminds us of other colloidal fluids, such as paints and ink, with roots nearly as ancient as those of ceramic suspensions. The example of ink preparation by the Egyptians for writing on papyrus is well known[2)]. The Roman author Vitruvius

---

[1)] S.G. Mason, *Tappi* **33** (1950) 440; R.B. McKay (Ed.), *Technological Applications of Dispersions*, Marcel Dekker (1994).

[2)] K. Beneke, *Zur Geschichte der Grenzflächenerscheinungen*, Verlag Reinhard Knof, (1995).

(born around 100 BC) mentions in his *De Architectura*[1] the deposition of soot on a wall and its manufacture to ink by mixing it with gum (resin). This is an early reference to steric stabilization of inorganic colloids, in this case carbon particles in water. The carbon colloids function as pigments giving the ink its colour. Many other pigments in printing ink, paints and plastics are found in the form of finely ground inorganic oxides or hydroxides. Iron oxides (see also sec. 2.4d) such as red haematite ( $\alpha$-$Fe_2O_3$ ), dark brown maghemite ( $\gamma$-$Fe_2O_3$ ) and black magnetite ( $Fe_3O_4$ ), were widely applied in ancient painting[2] and still belong to the most important pigments[3].

In paintmaking we probably find the earliest examples of colloid preparation that goes beyond the mere processing of natural materials. The ancient Egyptians, for example, knew how to synthesize the green pigment verdigris, mainly composed of $Cu(OH)_2$, and a silicate pigment known as Egyptian blue with $CaCuSi_4O_{10}$ as the main component[4]. Also, for these synthetic materials grinding or milling must have been required to obtain the desired pigment dispersion. For direct precipitation of inorganic particles in a liquid phase, or at least reports thereof, we have to make a leap in history. The Flemish chemist van Helmont (1577-1644) fused silica sand with excess alkali to form so-called waterglass and discovered that silica was recovered by treating the waterglass with acid[5]. Interestingly, waterglass is still a major source for the preparation of silica particles and gels. The method, as will become clear in this chapter, is also a didactic illustration of many aspects of particle formation (see also sec. 2.4a).

Another earlier documented example of inorganic colloid synthesis is that of the pigment Prussian Blue (iron (III) hexacyanoferrate (II)). It was discovered in 1710[5] that when solutions of potassiumferrocyanide and ferric chloride are mixed, deep blue particles precipitate instantaneously[6]. This beautiful classroom demonstration of colloid formation raises a number of questions for the attentive student, as soon as it is realized that precipitates are actually sub-visible colloidal particles or agglomerates thereof (a by no means trivial insight). What determines the sizes of the colloidal particles and how can they be controlled? What factors determine the growth rate of particles and why is it that nucleation is sometimes extremely fast and sometimes extremely slow? How can we characterize and control the size distribution of particles? What other methods are suitable to monitor properties of the dispersed colloids?

---

[1] Vitruvius, *On Architecture*, edited and translated by F. Granger, Harvard Univ. Press (1931-34). Two volumes.

[2] W.J. Russell, *Ancient Egyptian Pigments*, *Nature* **49** (1894) 374.

[3] A. Giltes, *Eisenoxid Pigmente* in; *Pigmente*, Ullmann's Enzyklopädie der Technischen Chemie, 3 Aufl., Band **13**, Verlag Chemie (1951-70).

[4] K. Volke, *Kolloidchemie im Altertum*, Akademie der Wissenschaften der DDR, Forschungsinstitut für Aufbereitung Freiberg (1989).

[5] W.H. Brock, *The Fontana History of Chemistry*, Fontana Press (1992).

[6] Prussian Blue colloids are a true classic: Selmi studied their precipitation as early as 1847. Renewed interest was sparked by their magnetic properties, see S. Choudhury, N. Bagkar, G.K. Dey, H. Subramanian, and J.V. Yakhmi, *Langmuir* **18** (2002) 7409.

Such questions motivate us to study in this chapter several aspects of preparation (sec. 2.2) and characterization (sec. 2.3) of mainly inorganic colloids. The aim is to provide a brief introduction, comprising some basic principles, useful facts and characterization methods, together with references for the reader to pursue a topic in much more depth than is possible or desirable in this chapter. The focus will be on colloids with (approximately) spherical shapes, which simplifies the treatment, and is also reasonable in view of the widespread study of colloidal spheres. Nevertheless, an occasional reference will be made to anisotropic particles to do some justice to the inorganic colloids in nature and industry with sometimes quite extreme aspect ratios (fig. 2.1), as in the clay dispersions referred to earlier.

## 2.2 Preparation

Insoluble substances, such as metals and their oxides, do not disperse spontaneously in water, so work is needed to bring them into a dispersed colloidal state. One strategy is prolonged *milling* and fracturing of minerals in a solution of stabilizing surfactants or polymers until a colloidal system is obtained. A drawback is the broad variety in

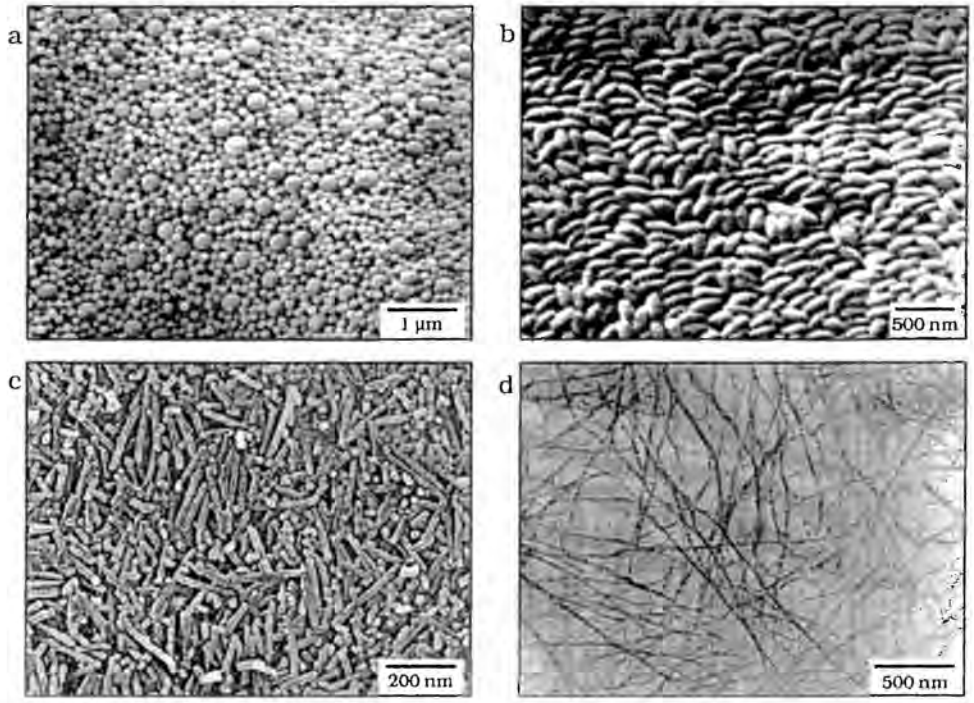



**Figure 2.1.** Examples of random packings of inorganic model colloids with increasing shape anisotropy: (a) silica spheres, (b) haematite spindles, (c) boehmite-silica rods, and (d) imogolite fibres. (Sources, see ref. [1].)

---

[1] Pictures redrawn from D.M.E. Thies-Weesie, A.P. Philipse, *J. Colloid Interface Sci.* **162** (1994) 470 (a); D.M.E. Thies-Weesie, A.P. Philipse, and S. Kluijtmans, *J. Colloid Interface Sci.* **174** (1995) 211 (b); M.P.B. van Bruggen, *Langmuir* **14** (1998) 2245 (c); and G.H. Koenderink, S. Kluijtmans, and A.P. Philipse, *J. Colloid Interface Sci.* **216** (1999) 429 (d).

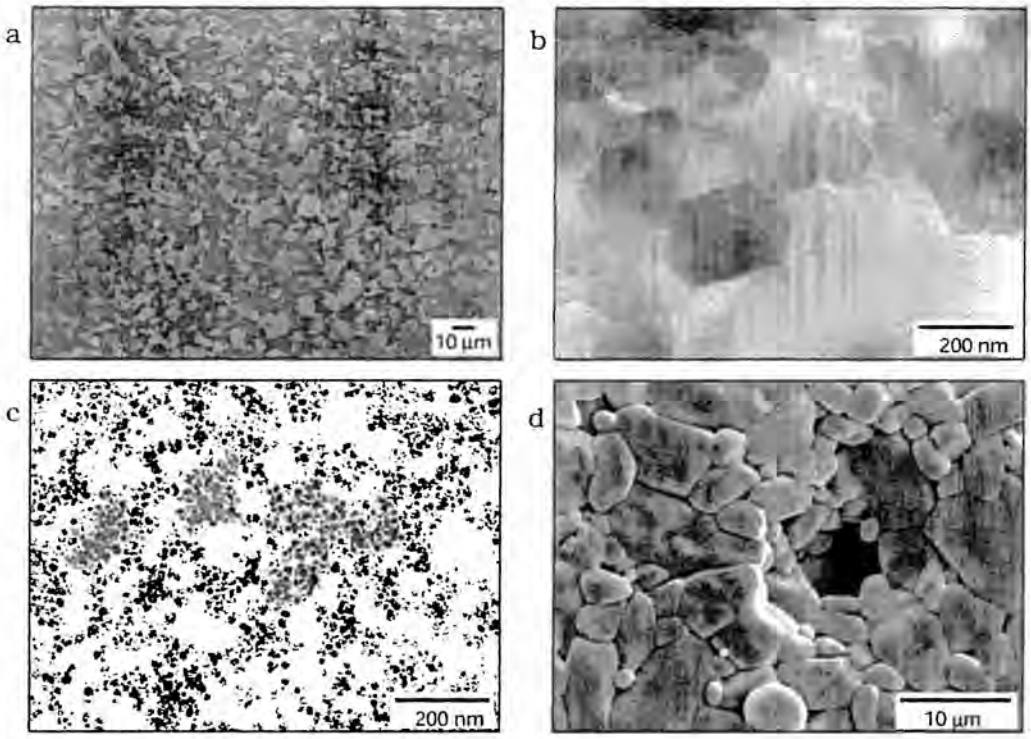



**Figure 2.2.** Some examples of industrial inorganic colloids: (a) cordierite particles prepared by milling used in refractories, (b) kaolinite platelets used in porcelain, (c) magnetite particles in ferrofluids (courtesy of Diona Bica, Timişoara, Romania) and (d) alumina grains in sintered ceramics.

shapes and sizes of the final colloids (fig. 2.2). To achieve better control of the morphology of colloidal particles, a *condensation* (or precipitation) method is preferred. Here, the colloidal state is approached from a molecular solution in which solute molecules are made to precipitate or polymerize into large units. The distinction between the two methods, milling and precipitation, can be illustrated by using a glass beaker in the preparation of a silica sol as the starting material. Glass largely consists of amorphous silica and hardly dissolves in water of pH ~ 7. So, to bring the material of the beaker into a colloidal state, we could fracture it and treat the glass pieces in a ball mill in water until a sol is obtained. The condensation alternative would be first to dissolve the glass pieces in a strongly alkaline sodium hydroxide solution to obtain waterglass, which is then diluted to a low weight percent of soluble silica and acidified to neutral pH, during which colloidal silica will precipitate (see also sec. 2.4a). Acidification is needed here to achieve a sufficiently large supersaturation of dissolved silica, exploiting the fact that the solubility of silica strongly decreases below pH ~ 10. The glass milling produces a polydisperse sol, whereas silica polymerization in a waterglass solution can be controlled to yield silica particles with a narrow size distribution in what can be counted as one of the classic sols of inorganic colloid

chemistry[1].

In principle, any substance can be brought into colloidal dispersion via precipitation in a supersaturated solution. All that is needed is a method to achieve a sufficiently large supersaturation of the desired material to induce homogeneous nucleation (sec. 2.2b) and prevent or control heterogeneous precipitation (sec. 2.2f). Of course, measures must be taken to ensure colloidal stability of the growing particles, such as increasing the particle surface charge (keeping the pH far away from the isoelectric point) or adding a stabilizing protecting polymer. A high supersaturation can, for instance, be achieved by a chemical reaction which produces a poorly soluble substance. A classic example[2] is bubbling hydrogen sulphide through a saturated arsenic trioxide solution to produce an arsenic trisulfide sol

$$As_2O_3 + 3H_2S \rightarrow As_2S_3 \downarrow + 3H_2O \qquad [2.2.1]$$

Other strategies involve mixing two soluble salts:

$$AgNO_3 + KI \rightarrow AgI \downarrow + KNO_3 \qquad [2.2.2]$$

the reduction of a metal salt to produce metal colloids

$$H_2PtCl_6 + BH_4^- + 3H_2O \rightarrow Pt \downarrow + 2H_2 \uparrow + 6HCl + H_2BO_3^- \qquad [2.2.3]$$

$$Ag_2O + H_2 \rightarrow 2Ag \downarrow + H_2O \qquad [2.2.4]$$

and the hydrolysis of metal salts to form oxides or hydroxides

$$2FeCl_3 + 3H_2O \rightarrow Fe_2O_3 \downarrow + 6HCl \qquad [2.2.5]$$

$$AlCl_3 + 2H_2O \rightarrow AlOOH \downarrow + 3HCl \qquad [2.2.6]$$

Precipitation can also be induced by a change in temperature, pH or solvent composition. For example, when water is added to a sulphur solution in ethanol, sulphur particles precipitate because sulphur has a much lower solubility in water[3]. Metal alkoxides are increasingly used as alternatives for inorganic salts in colloid synthesis. The alkoxides easily hydrolyze to reactive monomers, which polymerize to form discrete particles or gels (networks of particles). The archetypical example is silicium tetraethoxide or tetraethoxysilane (TES), which hydrolyzes as

$$Si(OR)_4 + nH_2O \rightarrow Si(OR)_{4-n}(OH)_n + nROH \qquad [2.2.7]$$

---

[1] R.K. Iler, *The Chemistry of Silica*, John Wiley (1979).

[2] For this and other classical examples, see *Colloid Science* **I**, *Irreversible Systems*, H.R. Kruyt, Ed., Elsevier (1952).

[3] This is the so-called Von Weimarn sulphur sol.

where *R* is an ethoxy or other alkoxy group. Partially hydrolyzed TES molecules polymerize via condensation reactions such as

$$(OR)_3Si\text{-}OH + HO\text{-}Si(OR)_3 \rightarrow (OR)_3Si\text{-}OSi(OR)_3 + H_2O$$
$$(OR)_3Si\text{-}OR + HO\text{-}Si(OR)_3 \rightarrow (OR)_3Si\text{-}O\text{-}Si(OR)_3 + ROH \qquad [2.2.8]$$

Such condensation reactions, depending on the reaction conditions (see e.g. the silica synthesis in sec. 2.4a), may under well-controlled conditions lead to well-defined silica spheres or networks and gels of aggregated small particles. For inorganic colloid syntheses, such control of particle size and structure is the exception rather than the rule, and is based on in-depth studies as illustrated by Iler's classic study on silica[1].

### *2.2a Size control*

Dispersed systems, in which all particles have the same or nearly the same size, have always attracted the attention of colloid science. Such *monodisperse* (also referred to as *homodisperse* or *isodisperse*) sols may be of practical importance; colloidal crystals in photonic materials require uniform particles, and semiconductor colloids in the nanometer size range have specific optical properties, which are very sensitive to particle size[2]. The sizes of silver halide colloids for so-called photographic emulsions need to be controlled to less than about 5% to optimize their photographic properties[3]; a demand which implies tight control of particle nucleation and growth. However, for many practical suspensions, such as in paints or ceramic processing, a modest polydispersity is not a serious problem, and is sometimes even beneficial. For example, the random packing density of spheres mixtures is greater than that of monodisperse particles and, consequently[4], the viscosity of the mixtures is generally below the viscosity for monodisperse spheres at the same volume fraction. Thus, manipulating the size distribution may be helpful for the processing and densification of sols of ceramic particles.

One academic motivation for monodispersity is its requirement of a critical test for theories of colloidal systems or thermal systems in general. Thermodynamically speaking, colloids are nothing but giant molecules but their large sizes allow studying, for example, their (thermo)dynamics via light scattering techniques or microscopy. Preferably there is only one particle size, or a very narrow size distribution, in the sol to keep theory and data interpretation manageable. James Clark Maxwell, unaware of the existence of isotopes, argued[5] that the monodispersity of atoms could only be

---

[1] R.K. Iler, *loc. cit.*
[2] See for example C.B. Murray, C.R. Kagan, and M.G. Bawendi, *Ann. Rev. Mater. Sci.* **30** (2000) 545.
[3] I.H. Leubner, *Current Opinion Colloid & Interface Sci.* **5** (2000) 151.
[4] For the viscosity as a function of volume fraction, see secs. 6.8 and 6.10.
[5] J.C. Maxwell, *Nature* **8** (1873) 437.

secured by the Lord himself. The need for divine intervention on the colloidal scale may be disputable, but the preparation of large quantities of nearly identical inorganic colloids is certainly a demanding task, largely due to unavoidable thermal fluctuations in a precipitating solution as will be explained later. The importance of monodispersity was already clearly perceived from the beginnings of colloid science, as witnessed in the work of Jean Perrin[1] on the verification of Einstein's theory for Brownian motion and his demonstration of the thermodynamic equivalence between colloids and molecules referred to above. Einstein derived that the average mean square displacement, $\langle r^2 \rangle = 6Dt$, of a colloidal particle in time $t$ is determined by the translational diffusion coefficient

$$D = kT / f \qquad [2.2.9]$$

which expresses that Brownian motion, driven by the thermal energy $kT$, is counteracted in a liquid by the hydrodynamic friction factor $f$. Einstein's results are valid for particles of arbitrary shape[2] but, of course, for an experimental test $f$ must be specified. The obvious choice is the Stokes friction factor, namely $f = 6\pi\eta a$, which is valid for a hard sphere of radius $a$ in a solvent with viscosity $\eta$. Thus, the diffusive displacements of monodisperse spheres with known radii provide a test of the Einstein equations without any adjustable parameter. The well-known outcome of this test by Perrin[3] is often considered as the first decisive evidence for the existence of molecules[4]. Perrin realized that this evidence was as strong as his colloids were monodisperse and, so, he and his co-workers undertook a laborious fractional sedimentation procedure to obtain a few hundred milligrams of uniform resin spheres from an initial weight of a kg of gamboge or mastic. This substance was dissolved in methanol and then precipitated by dilution in a large volume of water, resulting in monodisperse fractions of emulsions of spherical particles with a wide variation in size between these fractions.

Fractional sedimentation, which in Perrin's case took several months, is not a very practical procedure. An interesting alternative is the addition of non-adsorbing polymers, which cause a depletion attraction (see secs. V.1.8 an 9) with strength depending on the particle size. The repeated, size-selective, phase separations may produce quite uniform emulsions[5]. Nevertheless, if possible we would like to avoid fractionation altogether. Realizing that nature provides a very limited source of monodisperse colloids, at least with respect to inorganic particles, we need to understand the essential aspects underlying preparation of uniform particles by precipitation from a solution.

---

[1] J. Perrin, *Les Atomes*, Alcan (Paris) (1913).
[2] A. Einstein, *Ann. Phys.* **17** (1905) 549.
[3] J. Perrin, *Ann. Chim. Phys. (8)* **18** (1909) 5.
[4] M. Kerker, *J. Chem. Educ.* **51** (1974) 764.
[5] J. Bibette, *J. Colloid Interface Sci.* **147** (1991) 474.

We note here that the triad in Perrin's approach, namely the preparation of well-defined colloids, the characterization of their size distribution (dispersity), and their eventual application to investigate a physical problem, has served as a model strategy ever since. An example of such an application is the study of concentrated monodisperse sols to be discussed in chapter 5.

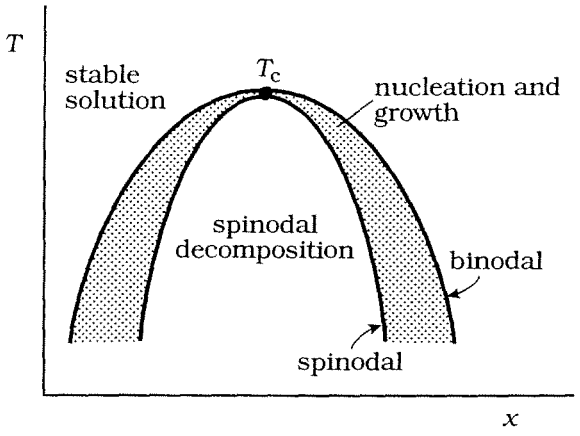



**Figure 2.3.** Schematic phase diagram for a solution, which becomes supersaturated upon cooling; $x$ is the solute mole fraction and $T$ is the temperature.

### *2.2b Homogeneous precipitation*

If a substance becomes less soluble by a change of some parameter, such as the temperature decrease in fig. 2.3, the solution may enter a metastable state on crossing the binodal in the phase diagram. In the metastable region, the formation of small precipitates or nuclei initially increases the Gibbs energy; thus, demixing by nucleation is an activated process, occurring at a rate, which is extremely sensitive to the extent of penetration in this metastable region, as will be discussed in 2.2c. In contrast, when we quench the solution into the unstable region on crossing the spinodal (fig. 2.3), there is no activation barrier to form a new phase. This is the so-called *spinodal decomposition* (briefly alluded to at the end of sec. I.2.19) in which a spongy phase is formed with a characteristic wavelength[1] rather than the collection of particulate colloids formed by nucleation and growth. The morphological contrast is illustrated by fig. 2.4 showing a labyrinth-like silica structure, resulting from spinodal decomposition in a cooling silicate melt[2] compared with discrete silica spheres prepared by nucleation and growth in a silica precursor solution (Stöber synthesis, see sec. 2.4a). A slow rise in supersaturation by slowly changing temperature or pH in fairly dilute solutions favours the formation of particulate colloids, because we then avoid a deep quench in the phase diagram (unless we are close to the critical point). Well-known examples are the slow precipitation of silica particles in aqueous silicate solutions at near-neutral pH and the nucleation of sulphur colloids upon addition of water to a sulphur solution in ethanol (see also sec. 2.4).

---

[1] J.W. Cahn, *Trans. Metall. Soc. of AIME* **242** (1968) 166.

[2] H. Xihuai, *J. Non-Cryst. Solids* **112** (1989) 58; S.G.J.M. Kluijtmans, J.K.G. Dhont, and A.P. Philipse, *Langmuir* **13** (1997) 4976.

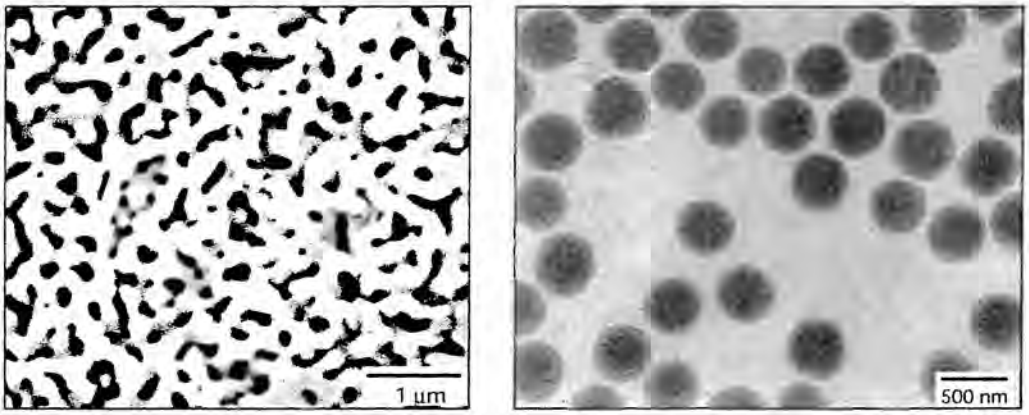



**Figure 2.4.** Left: a spongy structure of amorphous silica (so-called porous glasses), prepared by spinodal decomposition of a silica-containing melt. Right: amorphous silica spheres, formed by nucleation and growth in a solution (see sec. 2.4a). The spheres have been imaged *in situ* by cryogenic electron microscopy (see sec. 2.3 a).

We will briefly recapitulate[1)] the thermodynamics of homogeneous nucleation, i.e. particle formation in a solution with one solute only, a topic initiated in sec. I.2.23d.

Classical nucleation theory is based on an approximate macroscopic description according to which a precipitating particle (later referred to as a *nucleus* or *cluster*) is considered to consist of a bulk phase, containing $N_i^S$ molecules and a shell with $N_i^\sigma$ molecules of type $i$ (fig. 2.5). The particle is embedded in a solution containing dissolved molecules $i$. The volume of this solution is large as compared with that of the particle, so that the former acts as the surroundings of the latter. The Gibbs energy of the particle consists of a bulk part and a surface part

$$G^S = \mu_i^S N_i^S + \gamma A \qquad [2.2.10].$$

This follows from [I.A3.8], except that the amounts of substance and the chemical potentials are now written in terms of molecules rather than moles. The surface tension is taken as a constant and, for lack of better insight, equated to its bulk value, which is hardly measurable anyway, see sec. III.1.13. Implicit is the assumption that the size of the particle is large enough to ignore its influence on $\gamma$. Unlike the equilibrium state underlying [I.A3.8], characterized by equality of $\mu_i$ throughout, we now consider a non-equilibrium situation in which the solution is supersaturated: the activity $a_i > a_i(\text{sat})$. As a result, transfer of molecules takes place. We compute the change $\Delta G^S$ upon the transport of a small number $N_i$ of molecules from the solution to the particle. Obviously, this consists of two contributions

---

[1)] For an extensive treatment see F.F. Abraham, *Homogeneous Nucleation Theory*, Academic Press, (1974) and P.G. Debenedetti, *Metastable liquids; Concepts and Principles*, Princeton University Press, (1996). The last author also discusses spinodal decomposition and the still poorly understood transition from nucleation at very high supersaturation – deep into the metastable region – to spinodal decomposition.

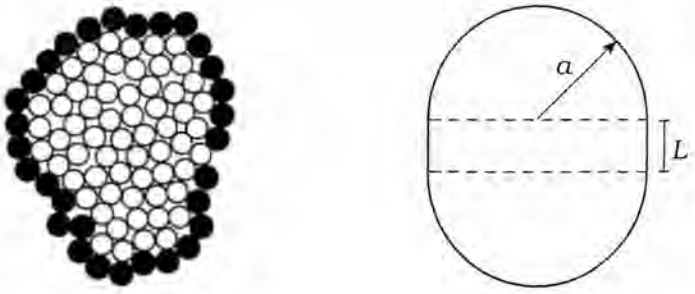



**Figure 2.5.** In classical nucleation theory a nucleus (left) is modeled by a droplet composed of bulk molecules and surface molecules, which have a higher free energy per molecule than the bulk. The nucleus is not necessarily spherical and is modeled here (right) by a spherocylinder.

$$\Delta G^{S} = \Delta G^{S}(\text{bulk}) + \Delta G^{S}(\text{surface}) \qquad [2.2.11]$$

Of these, the first is negative (it is the driving force), the second is positive (work has to be carried out against the expansion of the interface). We have, upon withdrawing $N$ molecules from the solution, transferring them to the bulk of the particle,

$$\Delta G^{S}(\text{bulk}) = -N_i\left[\mu_i^{L} - \mu_i^{L}(\text{sat})\right] \qquad [2.2.12]$$

where the superindex L refers to the solution. From this

$$\Delta G^{S}(\text{bulk}) = -N_i kT \ln\left[a_i / a_i(\text{sat})\right] \qquad [2.2.13]$$

which can also be written as

$$\Delta G^{S}(\text{bulk}) = -N_i kT \ln S \qquad [2.2.13a]$$

after introducing the *supersaturation ratio* $S$ as

$$S \equiv a_i / a_i(\text{sat}) \qquad [2.2.14]$$

Regarding $\Delta G^{S}(\text{surface})$, we can say that the surface area $A$ is proportional to $(N_i^{S})^{2/3}$ with a proportionality constant $\beta$ depending on the shape of the nucleus. Hence the Gibbs energy increase caused by the transfer is

$$\Delta G^{S}(\text{surface}) = \gamma\beta N_i^{2/3} \qquad [2.2.15]$$

Combination gives

$$\Delta G^{S} = -NkT \ln S + \gamma\beta N^{2/3} \qquad [2.2.16]$$

where we have omitted the subindex i because there is no confusion. We shall use [2.2.16] as an integrated equation, i.e. with $N = N^{S}$, but omit the superscript for typographical reasons. For relatively small clusters the surface area term dominates, whereas $\Delta G$ as a function of $N$ only starts to decrease due to the bulk term beyond a

critical value $N^*$ (see fig. 2.6). This critical cluster size follows from the condition $d\Delta G / dN = 0$

$$(N^*)^{1/3} = \frac{2\gamma\beta}{3kT \ln S} \qquad [2.2.17]$$

which can be used to rewrite the Gibbs energy for formation of a cluster as

$$\Delta G = A\gamma\left[1 - \frac{2}{3}\left(\frac{N}{N^*}\right)^{1/3}\right] \qquad [2.2.18]$$

This form is independent of the shape of the cluster and equally holds, for example, for crystalline cubes and amorphous spheres. The maximum in the Gibbs energy is

$$\Delta G^* = \frac{1}{3}A^*\gamma; \qquad A^* = \beta(N^*)^{2/3} \qquad [2.2.19]$$

This maximum is the activation barrier in the formation of colloidal particles by homogeneous nucleation in a supersaturated solution or vapour. Note that the (reversible) work needed to form the surface of the critical cluster equals $A^*\gamma$ and that the maximum in $\Delta G$ is only one third of this value because bulk is also formed. This expression for a critical cluster explains why a high supersaturation favours the formation of small colloids; a large $S$ pushes the critical size $N^*$ to smaller values and simultaneously lowers the activation barrier (fig. 2.6). A decrease in the interfacial tension $\gamma$ between colloid and solution, for example by adsorption of surfactants, has, according to [2.2.17] and [2.2.19], a similar effect. This is understandable since a low $\gamma$ cannot compete with the spontaneous bulk formation driving the precipitation, unless the clusters are very small.

Colloidal particles, of course, often do not precipitate as well-defined spheres, which is why we left the cluster shape unspecified via the parameter $\beta$ introduced in [2.2.15]. As a specific example of a non-spherical precipitate, we consider a cylinder of length $L$, capped at both ends by a hemisphere of radius $a$ (fig. 2.5). The number of molecules in the spherocylinder with volume $V$ equals $N = V/v_m$, where $v_m$ is the molecular volume. The Gibbs energy for the formation of the spherocylinder is

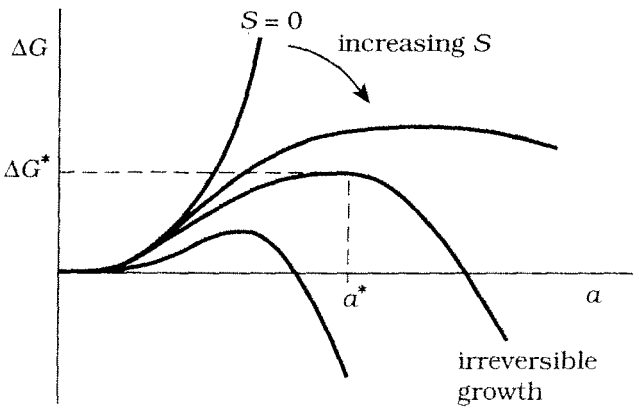



**Figure 2.6.** Sketch of [2.2.16] for nucleation and growth of a spherical precipitate of radius $a$ in a solution with supersaturation ratio $S$.

$$\Delta G = 4\pi a^2\left(1+\frac{L}{2a}\right)\gamma - (4/3)\pi a^3\left(1+\frac{3L}{4a}\right)\frac{kT\ln S}{v_m} \qquad [2.2.20]$$

For a sphere $\Delta G$ will always pass through a maximum when the radius is large enough (fig. 2.6), but increasing the length of the spherocylinder does not necessarily produce spontaneous growth at some point. We find

$$\frac{\partial \Delta G}{\partial L} = 0 \qquad \text{for} \qquad a^* = \frac{2v_m\gamma}{kT\ln S} \qquad [2.2.21]$$

and that this derivative is positive for $a > a^*$ and negative only for $a < a^*$. So, a cylinder can only grow spontaneously in length above a certain critical diameter, which is another consequence of the competition between surface and bulk effects. If the cylinder is too thin, there is insufficient increase in bulk volume to compensate for the increase in surface area caused by the lengthwise growth. By the same token, a disc or plate only grows spontaneously above a certain thickness. Nevertheless, even if $\Delta G$ is negative for lengthwise growth the spherocylinder can always lower its free energy further by reorganizing itself into a sphere which, after all, is the shape with minimum surface area for a given cluster volume; amorphous colloids are spherical – at least in their equilibrium form. Nucleation and sustained growth of non-spherical colloids requires clusters with various faces, which differ in solubility and surface tension, in other words, small crystals[1]. In the further discussion of nucleation and growth in sections 2.2.d-f, we focus for simplicity on non-crystalline spheres and therefore take the limit $L \to 0$ in [2.2.20]

$$\Delta G = 4\pi a^2\gamma - (4/3)\pi a^3\frac{kT\ln S}{v_m} \qquad [2.2.22]$$

with a maximum given by

$$\Delta G^* = \frac{4}{3}\pi(a^*)^2\gamma \qquad a^* = \frac{2v_m}{kT}\frac{\gamma}{\ln S} \qquad [2.2.23]$$

The results in this section for the energetics of nucleation are based on a description, which at first sight leaves much to be desired. Nuclei cannot become arbitrarily small without the macroscopic treatment at some point breaking down[2], which is why [2.2.16] contains the inconsistency that $\Delta G$, the excess Gibbs energy relative to unassociated molecules, does not actually reduce to zero for $N = 1$. Further, any internal degrees of freedom of clusters, and their translational entropy are not included in [2.2.16]. Finally, the cluster surface is entirely characterized by only one

---

[1] For crystals with faces $i$, each having an area $A_i$ and surface tension $\gamma_i$, each face contributes $A_i\gamma_i/3$, to the activation Gibbs energy, so the form [2.2.19] remains valid, see R. Defay, I. Prigogine, A. Bellemans, and D.H. Everett, *Surface Tension and Adsorption*, Wiley (1960).

[2] Debenedetti, *loc. cit.*

surface tension, whereas non-spherical crystalline precipitates may have more than one interfacial tension owing to different crystallographic orientations of the particle surface[1]. (As noted before, shape anisotropy does not change the form of the activation energy [2.2.19]). Granted that only one $\gamma$ suffices to evaluate the activation barrier in [2.2.23], its interpretation is still problematic. Usually $\gamma$ is equated to the surface Gibbs energy of a planar interface at phase co-existence. Thus, $\gamma$ in [2.2.23] is taken to be independent of the activity of molecules in the solution (i.e. the supersaturation ratio S). A numerical evaluation of the activation energy for crystal formation in a hard-sphere fluid by computer simulation[2] shows that the classical expression [2.2.22] is essentially correct, but that the value of $\gamma$ needs to be adjusted to obtain agreement between [2.2.22] and the numerical results. Extrapolation of the effective $\gamma$ to zero supersaturation yielded[2] the expected surface tension at phase coexistence, but as these surface tensions are experimentally hardly accessible, quantitative predictions from [2.2.22] are in many cases at best conjectural.

### 2.2c *Precipitation kinetics*

In the precipitation kinetics of colloids in a metastable solution[3], we can, in accordance with fig. 2.6, distinguish two regimes. When the colloidal particle is significantly larger than the critical size, it is in the regime of irreversible growth with kinetics to be discussed later. First, we consider the initial regime where small particles struggle with their own solubility to pass the Gibbs energy barrier $\Delta G^*$. This passage is called a *nucleation event*, which for simplicity we will define as the capture of *one* molecule by a critical cluster, assuming that after this capture the cluster enters the irreversible growth regime upon which a new colloid is born. This assumption, of course, neglects the finite probability that supercritical clusters may also dissolve. For an estimate of the nucleation rate, however, this simple picture is sufficient. Hence, the number $I$ of colloids which per second come into existence is proportional to $c_m$ and $c^*$

$$I = k c_m c^* \qquad [2.2.24]$$

where $k$ is a rate constant; $c_m$ and $c^*$ are the concentrations of single, unassociated molecules and critical clusters, respectively. Note that [2.2.24] predicts second-order reaction kinetics because of our choice to consider only encounters between a critical cluster and one molecule as the rate-determining events. To quantify $I$, we first evaluate the frequency at which molecules encounter a spherical cluster of radius $a$ by diffusion, following in essence Smoluchowski's diffusion model for coagulation kinetics

---

[1] See for example A.C. Zettlemoyer (Ed.), *Nucleation*, Marcel Dekker (1969).

[2] S. Auer, D. Frenkel, *Nature* **409** (2001) 1020; *Nature* **413** (2001) 711.

[3] For in-depth studies on various inorganic colloids the work of de Bruyn and co-workers is recommended reading. See for example, J. Dousma, P.L. de Bruyn, *J. Colloid Interface Sci.* **64** (1978) 154; H.A. van Straten, B. Holtkamp, and P.L. de Bruyn, *J. Colloid Interface Sci.* **98** (1984) 342; M.J.M. van Kemenade, P.L. de Bruyn, *J. Colloid Interface Sci.* **118** (1987) 564.

(see sec. 4.3). The diffusion flux $J$ of molecules through any spherical envelope of radius $r$ is, according to Fick's first law,

$$J = 4\pi r^2 D \frac{dc(r)}{dr} \qquad [2.2.25]$$

where $D$ is the molecular diffusion coefficient relative to the sphere positioned at the origin at $r = 0$. Each molecule that reaches the sphere surface irreversibly attaches to the insoluble sphere, and we assume that the concentration $c_m$ of molecules in the liquid far away from the sphere remains constant

$$c(r = a) = 0 \qquad c(r \to \infty) = c_m \qquad [2.2.26]$$

For these boundary conditions [2.2.25] yields

$$J = 4\pi D a^* c_m \qquad [2.2.27]$$

if it is assumed that $J$ is independent of $r$, that is, if the diffusion of molecules towards the sphere has reached a stationary state. Such a state is approached by the concentration gradient around a sphere in a time of order $a^2/D$ needed by molecules to diffuse over a sphere diameter. Assuming that sphere growth is a sequence of stationary states, we can identify the nucleation rate $I$ as the flux $J$ multiplied by the concentration $c^*$ of spheres with critical radius $a^*$

$$I = 4\pi D a^* c_m c^* \qquad \left[\mathrm{s^{-1}\,m^{-3}}\right] \qquad [2.2.28]$$

The concentration $c^*$ may be evaluated as follows[1]. Since the reversible work to form a cluster out of $N$ molecules is the $\Delta G$ from fig. 2.6, the Boltzmann distribution

$$c(N) = c_m \exp[-\Delta G/kT] \qquad [2.2.29]$$

determines the equilibrium concentration of clusters composed of $N$ molecules. Applying this result to clusters with a critical size, we find on substitution in [2.2.28] for the nucleation rate

$$I = 4\pi D a^* c_m^2 \exp[-\Delta G^*/kT] \qquad \Delta G^* = (4\pi/3)(a^*)^2 \gamma \qquad [2.2.30]$$

where $\Delta G^*$ is the height of the nucleation barrier; the exponent may be identified as the probability (per particle) that a spontaneous fluctuation will produce a critical cluster. The use of an equilibrium Boltzmann distribution in a nucleation flux is perhaps unexpected[2], but one can think of a distribution of subcritical clusters from

---

[1] For an extensive discussion see Debenedetti *loc.cit.*

[2] In the thermodynamics of reversible coagulation an expression can be derived for the distribution of aggregate size which is very similar to [2.2.29]. See D.H. Everett, *Basic Principles of Colloid Science, Roy. Soc. Chem.* (1994).

which critical clusters are removed as soon as they capture additional molecules. Each removal is compensated by the insertion of an equivalent number of single molecules into the metastable bulk solution. In this manner, one can define a steady state nucleation rate for a given supersaturation[1]. Equation [2.2.30] shows that the nucleation rate is extremely sensitive to the value of $a^*$ and, thus, to the supersaturation via [2.2.23]. The maximum nucleation rate at very large supersaturation, the pre-exponential kinetic factor in [2.2.30], is of the order

$$I \sim \frac{kT}{\eta} \quad c_m^2 \qquad [2.2.31]$$

as follows from substitution of the Stokes-Einstein diffusion coefficient $D = kT/6\pi\eta a^*$, where we neglect the size difference between molecules and critical clusters. For an aqueous solution at room temperature with a molar concentration $c_m = 10^{-3}$ M, we find a maximal nucleation rate of order $10^{29}\ m^{-3}\,sec^{-1}$. A decrease in supersaturation to values around $S \approx 5$ suffices to reduce this astronomical rate to practically zero. For silica precipitation in dilute, acidified waterglass solutions (see sec. 2.2e), the supersaturation is in order of magnitude close to $S \approx 5$ and nucleation may take hours to days. For comparison, the industrial, continuous precipitation of the highly insoluble silver halide colloids[2], the basis of classical photographic materials, occurs at a supersaturation, which generally exceeds $S \sim 10^6$.

The kinetics of precipitation in a homogeneous solution is notoriously difficult to assess within better than an order of magnitude because of uncertainties in, for example, the interfacial tension that are strongly amplified in [2.2.30]. Nevertheless, the trend predicted from [2.2.30] is qualitatively correct. Within a narrow range of supersaturation after crossing the binodal in fig. 2.3 the rate of homogeneous precipitation increases from negligibly small to astronomically large. In practice, however, the increase is limited because experimental nucleation rates often go through a maximum at sufficiently high supersaturation[3]. In concentrated solutions, the assumption of freely diffusing molecules underlying the pre-exponential factor in [2.2.30] breaks down, though reduced diffusivity is unlikely to be the sole cause of any maximum in the precipitation rate. At high solute concentrations, long-time self-diffusion admittedly will vanish but for nucleation only local rearrangements of molecules are required, which may be feasible up to (and possibly even including) close-packing densities. Another factor of importance is that, as already noted in sec. 2.2b, the interfacial Gibbs energy $\gamma$ is actually not a constant. Simulations of absolute nucleation rates show that,

---

[1] R. Becker, *Theorie der Wärme*, Springer Verlag (1978).

[2] I.H. Leubner, *Current Opinion in Colloid & Interface Sci.* **5** (2000) 151, reviews nucleation models for silver halides.

[3] P. Pusey, in *Liquid, Freezing and Glass Transition* J.P. Hansen, D. Devesque, and J. Zinn-Justin, Eds. 763-931, North Holland (1991).

in any case for hard-sphere fluids[1], the maximum in the nucleation rate is indeed primarily due to an increase of $\gamma$ with supersaturation. This increase diminishes the probability that a critical cluster will form on account of [2.2.30]. So, any quantitative prediction for the nucleation rate must at least take this change in $\gamma$ into account.

The reader may have noticed that [2.2.30] is very similar to the classical *Becker-Döring* result[2] for homogeneous nucleation in a vapour (see sec. I.2.23). The difference is the form of the pre-exponential kinetic factor, which is obtained here using a diffusion model instead of kinetic gas theory. Consequently, the result [2.2.31] is equivalent to Smoluchowski's expression for the rate of diffusion-controlled coagulation of identical spheres in the initial state of coagulation (see sec. 4.3). In Smoluchowski's treatment, incidentally, there is no activation barrier because of the assumption that colloids irreversibly stick whenever they happen to collide by Brownian motion. However, when attractions are at a level of weakness such that colloidal clusters[3] can be disrupted by the thermal energy, the existence of a critical aggregate size can be expected with a rate of formation similar to [2.2.30].

### 2.2d *Particle growth and polydispersity*

When no precautions are taken, precipitation from a supersaturated solution inevitably produces polydisperse colloids because nucleation of new particles and further particle growth overlap in time. This overlap is a consequence of the statistical nature of the nucleation process; near the critical size particles may grow as well as dissolve. To narrow down the initial size distribution as much as possible, nucleation should take place in a short time, followed by equal growth of a constant number of particles. La Mer[4] pointed out that this can be achieved by rapidly creating the critical supersaturation required to initiate homogeneous nucleation after which particle growth lowers the saturation sufficiently to suppress new nucleation events. It should be noted that La Mer's scheme rests on the extreme sensitivity of homogeneous nucleation rates to supersaturation. An instance of La Mer's scheme is found in the double-jet precipitation of silver halide colloids, in which $AgNO_3$ and NaBr solutions are simultaneously added to an agitated gelatin solution. Here, the number of newly formed crystals quickly reaches a constant value and further addition of reagents causes only further growth of fairly monodisperse cubic crystals[5]. Another option is to add nuclei (seeds) to a solution with a subcritical supersaturation as when silica particles are added to a saturated aqueous silicate solution (heterogeneous nucleation,

---

[1] S. Auer, D. Frenkel, *Nature* **413** (2001) 711.
[2] R. Becker, *loc.cit.*
[3] For reversible coagulation see also J. Groenewold, W.K. Kegel, *J. Phys. Chem.* **B105** (2001) 11702.
[4] V.K. La Mer, R.H. Dinegar, *J. Am. Chem. Soc.* **72** (1950) 4847.
[5] J.S. Wey, R.W. Strong, *Photogr. Sci. Eng.* **21** (1977) 14; C.R. Berry, *Photogr. Sci. Eng.* **18** (1974) 4.

see sec. 2.2e). The advantage of this seeded growth technique is that the final particle size can be influenced by the concentration of seed particles.

A fortunate consequence of particle growth is that in many cases the size distribution is *self-sharpening*. We will illustrate this effect for colloidal spheres of radius $a$, which irreversibly grow by the uptake of molecules from a solution according to the rate law[1)]

$$\frac{da}{dt} = k_0 a^n \qquad [2.2.32]$$

where $k_0$ and $n$ are constants. This growth equation leads either to spreading or sharpening of the relative size distribution, depending on the value of $n$, as can be demonstrated as follows. Consider at a given time $t$ any pair of spheres with arbitrary size from the population of independently growing particles. Let $1+\varepsilon$ be their size ratio such that $a(1+\varepsilon)$ and $a$ are the radius of the larger and smaller sphere, respectively. The former grows according to:

$$\frac{d}{dt} a(1+\varepsilon) = k_0 a^n (1+\varepsilon)^n \qquad [2.2.33]$$

which can be combined with growth equation [2.2.32] for the smaller sphere to obtain the time evolution of the size ratio:

$$\frac{d\varepsilon}{dt} = k_0 a^{n-1}\left[(1+\varepsilon)^n - (1+\varepsilon)\right] \qquad \varepsilon \geq 0 \qquad [2.2.34]$$

Clearly, the relative size difference $\varepsilon$ increases with time for $n > 1$, in which case particle growth broadens the distribution. For $n = 1$ the size ratio between two spheres remains constant, whereas for $n < 1$ it monotonically decreases in time. Since this decrease holds for any pair of particles in the growing population, it follows that for $n < 1$ the relative size distribution is self-sharpening, a conclusion also drawn by other authors[2)]. It should be noted that what applies to the growth kinetics of two spheres also holds for two sufficiently sharp distributions. Thus, [2.2.34] also describes the time evolution of the relative distance of two peaks in a bimodal size distribution. These two peaks are much easier to monitor in time than the width of a single size distribution, which is why growth of a binary sphere mixture is a convenient source of experimental information on kinetic mechanisms, as has been demonstrated for latex[3)] and silica[4)] dispersions.

---

[1)] The concentration of molecules is incorporated here in the rate constant $k_0$ and may depend on time because of a generating chemical reaction. Such dependence does not alter the effect of exponent n on the polydispersity because $k_0$ is the same for all particles.

[2)] J.Th.G. Overbeek, *Adv. Colloid Interface Sci.* **15** (1982) 251.

[3)] E.B. Bradford, J.W. vanderHoff, and T. Alfrey Jr., *J. Colloid Interface Sci.* **11** (1956) 135.

[4)] A. van Blaaderen, J. van Geest, and A. Vrij, A., *J. Colloid Interface Sci.* **154** (1992) 481.

The requirement $n<1$, for self-sharpening, is in practice a realistic one. For example, when the growth rate is completely determined by a slow reaction of molecules at the sphere surface, we have

$$\frac{da^3}{dt} = k_0 a^2 \qquad [2.2.35]$$

implying that $da/dt$ is a constant, so $n=0$. The opposite limiting case is growth governed by the rate at which molecules reach a colloid by diffusion. The diffusion flux for molecules with a diffusion coefficient $D$, relative to a sphere centred at the origin at $r=0$, is given by [2.2.25]. We assume that the saturation concentration is maintained at the particle surface, neglecting the influence of particle size on $c$(sat) (the Kelvin effect, see sec. 2.2e), and keeping the bulk concentration of molecules constant[1)]

$$c(r=a) = c(\text{sat}) \qquad c(r\to\infty) = c(\infty) \qquad [2.2.36]$$

For these boundary conditions, the stationary (i.e. $r$-independent) flux towards the sphere equals (see [2.2.27]):

$$J = 4\pi Da[c(\infty) - c(\text{sat})] \qquad [2.2.37]$$

showing that the rate at which the colloid intercepts diffusing molecules is proportional to its radius and not to its surface area. Suppose every molecule contributes a volume $v_m$ to the growing colloid, then for a homogeneous sphere the volume increases at a rate

$$\frac{d}{dt}\frac{4}{3}\pi a^3 = Jv_m \qquad [2.2.38]$$

which on substitution of [2.2.37] leads to

$$\frac{da}{dt} = Dv_m[c(\infty) - c(\text{sat})]a^{-1} \qquad [2.2.39]$$

with the typical scaling $a^2 \sim t$ as expected for a diffusion-controlled process. Thus, the exponent in [2.2.32] for diffusion-controlled growth is $n=-1$, and consequently the relative width of the size distribution decreases in time. This conclusion is based on a diffusion flux, which assumes a steady-state diffusion of molecules towards colloids, which grow independently from each other. Reiss[2)], however, has shown that also when these assumptions are invalid, diffusional growth still sharpens the size distribution.

Diffusion-controlled growth of a homogeneous sphere was first studied by Langmuir[3)], who introduced a formula very similar to [2.2.39], albeit for the evaporation of

---

[1)] A decrease in $c_\infty$ due to exhaustion of a finite bulk is treated in A. Philipse, *Colloid Polym. Sci.* **266** (1988) 1174.

[2)] H. Reiss, *J. Chem. Phys.* **19** (1951) 482.

[3)] I. Langmuir, *Physical Rev.* **12** (1918) 368.

a sphere for which the derivative in [2.2.39] is negative. Langmuir used a diffusion model to explain the evaporation rate of millimetre-sized iodine spheres in quiet air. He found that the rate of weight loss of the spheres confirmed diffusion control, and obtained from the rate a reasonable value for the diffusion coefficient of iodine molecules in air. Equation [2.2.39] is also useful to estimate colloidal growth rates. Molecular diffusion coefficients in water at 25°C are of the order $D \approx 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and taking a typical volume fraction $v_m[c(\infty) - c(\mathrm{sat})] = 0.01$ of reactive molecules we find from [2.2.39] that for diffusion-controlled growth the surface area increases in time as $\mathrm{d}a^2/\mathrm{d}t \approx 20(\mu\mathrm{m})^2\,\mathrm{s}^{-1}$. This implies a nearly instantaneous growth of submicron colloids, which indeed is observed in, for example, the precipitation of magnetite (see section 2.4d). Whenever particle growth is much slower, the kinetics may be determined by a slow reaction step at the surface of the colloid, or by the slow production of precipitating molecules via a chemical reaction as in the case of sulphur sols (see sec. 2.4b).

This is not the place for in depth refinement[1] of diffusion-controlled kinetics beyond a flux of the form [2.2.37], but we cannot totally ignore the involvement of charged species in the precipitation of inorganic colloids. Hence, an electrostatic interaction may be present between the growing colloids and the molecules they consume, which will either enhance or retard the growth, depending on whether colloids and monomers attract or repel each other. From the classic studies of Kramers[2] and Debye[3] on diffusion in a force field, we can infer that the diffusion coefficient $D$ of the monomers in the diffusion flux $J$ has to be replaced by an effective coefficient of the form

$$D_{\mathrm{eff}} = \frac{D}{a\int_a^\infty \mathrm{e}^{-u(r)/kT} r^{-2}\mathrm{d}r} \qquad [2.2.40]$$

where $u(r)$ is the interaction energy between molecule and colloid. The same type of integral, incidentally, appears in the theory of slow coagulation in sec. 4.3b. Suppose the molecules are ions with charge $ze$ and that the colloidal sphere has a surface potential $\psi^o$. To obtain an upper estimate of the effect of the ion-colloid interaction on the growth kinetics, we consider the low salt limit where the interaction is unscreened. Then $u(r)$ is obtained from Coulomb's law as

$$\frac{u(r)}{kT} = u_o \frac{a}{r} \qquad u_o = \frac{ze\psi^o}{kT} = zy^o \qquad [2.2.41]$$

where $u_o$ is the colloid-ion contact interaction energy and $y^o = e\psi^o/kT$, as before.

---

[1] D.F. Calef, J.M. Deutch, *Ann. Rev. Phys. Chem.* **34** (1983) 493.
[2] H.A. Kramers, *Physica* **7** (1940) 284.
[3] P. Debye, *Trans. Electrochem. Soc.* **82** (1942) 265.

Thus, this Coulombic interaction [2.2.40] yields

$$D_{\text{eff}} = D\frac{zy^{o}}{e^{zy^{o}} - 1} \qquad [2.2.42]$$

So, for colloids that have to grow by a diffusion flux of like-charged ions, the growth kinetics is slowed down exponentially by the Coulombic repulsion; when $\psi_0$ = 75 mV the effective diffusion coefficient for divalent ions is about 0.01 $D$. Added salt screens the colloid-ion interaction and, therefore, moderates the influence of $y_0$ on the growth kinetics.

The interaction between monomers and the growing colloid, within the approximations underlying [2.2.42], does not change the growth equation [2.2.39] and, hence, does not affect the conclusion that diffusional growth sharpens the size distribution. We will investigate whether this conclusion still holds when we drop the assumption that the growing sphere is a homogeneous object of constant mass density. It is well known that diffusional growth may produce heterogeneous structures with an internal density profile. A familiar example[1] is the precipitation of silica at low pH, where ramified clusters are formed rather than the fully condensed $SiO_2$ particles at alkaline pH. The difference is due to the low reactivity of silanol groups towards condensation at acid pH, which obstructs the densification of a cluster. Suppose a monomer volume fraction profile $\phi(x)$ is present in the growing colloid, where $x$ is the distance to its centre. Then, the rate of growth is, instead of [2.2.39], given by

$$\phi(a)\frac{da}{dt} = Dv_{\text{m}}[c(\infty) - c(\text{sat})]a^{-1} \qquad [2.2.43]$$

because each monomer contributes a volume $v_{\text{m}}/\phi(a)$ to the growing colloid upon arrival at its surface at $x = a$. When this volume contribution increases with the colloid radius, i.e. when the average mass density of the colloid decreases, the large particles in the size distribution have a gain in growth rate. This scenario will occur for the fractal clusters produced by *diffusion-limited aggregation*[2,3] (DLA). Precipitation by DLA forms an interesting, purely kinetic contrast to classical nucleation and growth, where the excess surface Gibbs energy provides the nucleation barrier, as well as the driving force for further growth by ripening (see sec. 2.2e). The kinetics of fractal growth will be treated in sec. 4.5c; here, a further completion of [2.2.43] will suffice. Consider monomers with volume $p^3$ which diffuse towards a single spherical cluster with total radius $a$. The number of monomers, $N$, in the cluster scales as[4]

---

[1] R.K. Iler *loc. cit.*

[2] T.A. Witten Jr., L.M. Sander, *Phys. Rev. Lett.* **47** (1981) 1400.

[3] P. Meakin, *Faraday Discuss. Chem. Soc.* **83** (1987) 1.

[4] L.G.B. Bremer, *Fractal Aggregation in Relation to Formation and Properties of Gels*, Ph.D. thesis, Wageningen Agricultural University, The Netherlands (1992).

$$N \sim \left(\frac{a}{p}\right)^{d_f} \qquad [2.2.44]$$

where $d_f$ is the fractal dimensionality. The average monomer volume fraction in the cluster is accordingly

$$\langle\phi\rangle \sim \frac{Np^3}{a^3} \sim \left(\frac{a}{p}\right)^{d_f-3} \qquad [2.2.45]$$

assuming, as usual in DLA models, that the monomers are spheres. However, the cluster may also be composed of randomly oriented fibers or platelets. They significantly reduce the average density of a cluster, but do not necessarily change its fractal dimension[1]. The local volume fraction at a distance $x$ from the cluster centre is

$$\phi(x) \sim \frac{d_f}{3}\left(\frac{x}{p}\right)^{d_f-3} \qquad [2.2.46]$$

as can be checked by its substitution into the definition of the average density

$$\langle\phi\rangle = \frac{3}{4\pi a^3}\int_0^a \phi(x) 4\pi x^2 dx \qquad [2.2.47]$$

Thus, the volume fraction at the edge of the cluster is

$$\phi(a_c) \sim \frac{d_f}{3}\langle\phi\rangle \qquad [2.2.48]$$

which on substitution into [2.2.43] leads to the following scaling of the growth rate of the outer radius of the cluster

$$\frac{da}{dt} \sim a^{2-d_f} \qquad a \sim t^{1/d_f-1} \qquad [2.2.49]$$

For $d_f = 3$, we recover the familiar square-root time dependence of diffusional growth of a homogeneous sphere. A fractal dimensionality, $d_f < 3$, enhances the growth rate, but as long as $d_f > 1$, the form of $da/dt$ is such that self-sharpening will occur. Clusters with a fractal dimensionality $d_f = 1$ are rather unlikely for colloidal growth as they should consist of straight spikes of length $a$ growing from a common source into a radial direction, such that the mass increases linearly with the radius ($N \sim a$ in [2.2.44]). Since for three-dimensional DLA, the fractal dimensionality certainly exceeds unity, a value as high as $d_f = 2.49$ has been reported[2], it follows that self-sharpening will occur for diffusional growth of both homogeneous and heterogeneous clusters. Two comments should be made here. First, we have disregarded the 'fingered' surface structure of a fractal cluster by assuming that a monomer will stick whenever it arrives

---

[1] Heterogeneous structures and gels of inorganic fibres are reviewed in A. Philipse and A. Wierenga, *Langmuir* **14** (1998) 49.

[2] M. Fleischmann, D.J. Tildesley, and R.C. Ball, *Fractals in the Natural Sciences*, Princeton University Press (1990).

at a distance $a$ from the cluster centre, i.e. the cluster surface is uniformly sticky. Taking structural details into account may change the scaling relations [2.2.49]. Second, consistent with the approach discussed elsewhere in this chapter we look at clusters growing independently in a bulk. In a later stage of growth, it is clear that aggregation of fractal clusters themselves becomes the kinetically dominating event in the formation of large aggregates and space-filling gels, a topic extensively treated elsewhere.

The size distribution resulting from the precipitation and growth of inorganic particles is often (and also here) regarded as a purely kinetic phenomenon. The colloids simply stop growing when supersaturation has dropped sufficiently. The size distribution may further change in time due to cluster-cluster aggregation and Ostwald ripening (see sec. 2.3e) and, of course, by coagulation due to Van der Waals forces, but there is no evolution towards a thermodynamically stable size distribution. That, at least, is the classical view for inorganic colloidal dispersions. It is well known, however, that other dispersed systems exist, such as microemulsions, in which the dispersed phase is actually in a state of thermodynamic equilibrium. This is a consequence of very low interfacial tensions owing to the adsorption of surfactants. There is no *a priori* reason why inorganic colloids could not be thermodynamically stable due to adsorbed layers, which puts a Gibbs energy penalty on further decrease of the surface area. Various authors have alluded in the past to this possibility[1], and recent experiments on the preparation of silver halide sols[2] provide clear examples of inorganic colloids with thermodynamic control of the size distribution. Understanding this control and its occurrence is undoubtedly important for further improvement and extension of preparation methods for well-defined colloids.

### *2.2e Particle solubility and Ostwald ripening*

For colloidal spheres of given radius $a$ and surface tension $\gamma$, there is one solute concentration $c(a)$ at which the colloids have a critical size and reside at the Gibbs energy maximum in fig. 2.6. This concentration, $c(a)$, also called the equilibrium solubility of the colloids, follows from [2.2.23], which in the context of solubility usually is written as

$$\ln[c(a)/c(\text{sat})] = 2\gamma v_m / a^* kT \qquad [2.2.50]$$

which is a result known as the *Gibbs-Kelvin equation*; $c(\text{sat})$ is the equilibrium solubility of a flat surface as we already encountered before, see [I.2.23.24]. Note that we have replaced the activities in the supersaturation ratio $S$ by concentrations, so [2.2.50] is valid only for dilute solutions. The increase in solubility for small spheres

---

[1] J.Th.G. Overbeek, *Faraday Discuss. Chem. Soc.* **65** (1978) 7.
[2] I.L. Mladenovic, W.K. Kegel, P. Bomans, and P.M. Frederik, *J. Phys. Chem.* **B107** (2003) 5717.

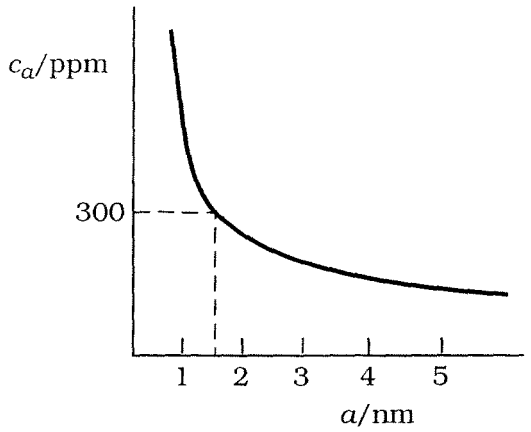



**Figure 2.7.** Equilibrium solubility of a sphere with radius $a$ according to the Gibbs-Kelvin equation [2.2.50]. For silica particles with a radius around $a$ = 1 nm, the solubility in neutral water is about 300 ppm (see text for discussion).

(or in the vapour pressure of small droplets) is equivalent to an enhanced, excess Laplace pressure $\Delta p = 2\gamma/a$, which can be recognized in [2.2.50]. An important feature of the Gibbs-Kelvin equation is its generality; its form does not depend on the assumption of amorphous spheres characterized by only one surface tension. For the equilibrium solubility of a crystal, the equation applies to each of the faces of the crystal[1)]

$$\frac{\gamma_1}{r_1} = \frac{\gamma_2}{r_2} = \ldots \frac{\gamma_i}{r_i} = \frac{kT}{2v_m} \ln \frac{c(a)}{c(\text{sat})} \qquad [2.2.51]$$

Here, $\gamma_i$ is the surface tension of face $i$, which is at a distance $r_i$ from the centre of the crystal; only faces of the equilibrium polyhedron are taken into account and corners and edges are ignored. In these *Wulff relations*, the ratio $\gamma_i / r_i$ for a crystal face plays the same role as $\gamma/a$ for a spherical droplet. Alternatively, one can also interpret [2.2.51] as describing polydisperse spheres with different surface tensions such that they are all in equilibrium with the same solution with concentration $c(a)$. The Wulff relations fix the equilibrium shape of the crystal since in the equilibrium form there exists one centre such that these relations are satisfied.

The increased solubility according to [2.2.50], also referred to as the *Gibbs-Kelvin effect*, is not easy to quantify for sols of inorganic particles as their surface tension is difficult to determine experimentally (see sec. III.1.13) and has to be indirectly obtained from adsorption data. For an order of magnitude estimate of the Gibbs-Kelvin effect, we consider the case of amorphous silica in water (fig. 2.7). Iler[2)] reports the range $\gamma$ = 0.05-0.1 N/m, corresponding to an equilibrium radius in the range

$$a^* \ln[c(a)/c(\text{sat})] \approx 1\text{-}2 \text{ nm} \qquad [2.2.52]$$

---

[1)] See for a derivation and further discussion: R. Defay, I. Prigogine, A. Bellemans, and D.H. Everett, *Surface Tension and Adsorption*, Wiley (1966).

[2)] R.K. Iler, *loc. cit.*

at $T$ = 298 K, for a molar silica volume of $v_m$ = 27.2 $cm^3$/mol. At neutral pH the bulk solubility $c$(sat) of silica is about 100 ppm, which implies that in an aqueous waterglass solution containing $c$(a) = 300 ppm of soluble silica the equilibrium radius is in the range $a^*$ = 0.9-1.8 nm. At alkaline pH the bulk solubility of silica rapidly increases to a value of $c$(sat) = 300 at about pH = 10. Thus, in the 300 ppm waterglass solution at pH = 10 the critical particle size of [2.2.23] tends to infinity, so precipitation of silica particles in this case is extremely unlikely. The expected significant solubility of silica particles in the nanometer size range is indeed observed in practice, as is the drastic effect of pH. It is not a coincidence, but a consequence of the Gibbs-Kelvin effect, that in aqueous silica sols particle radii are usually at least a few nm and that the pH of the sols is usually set below 10.

The interfacial tension in the Gibbs-Kelvin equation [2.2.50] is, of course, not a fixed parameter and may be intentionally decreased to enhance particle solubility. A classical example is AgI, which is hardly soluble in water. However, when excess silver or (in particular) iodide ions are added the solubility increases dramatically as manifested by the formation of complexes or small silver iodide clusters, which apparently have a reduced interfacial tension. It has been shown that these small particles (up to ~1 nm in size) form spontaneously[1]. Clearly, such reversible dissolution of inorganic salts is a possibility whenever ions or other species strongly adsorb on the material in question. Modification of a particle surface, of course, may also *decrease* the solubility or rather reduce the rate of particle dissolution due to a protective layer of insoluble material. Even much less than monolayer coverage may be sufficient for this purpose as in the case of the solubility of silica, which is drastically reduced by adsorption of aluminate species in minute quantities[2].

Returning to the Gibbs energy maximum in fig. 2.6, we note that it presents an unstable equilibrium, which can be maintained only for critical particles of exactly the same size. For polydisperse particles (with the same surface tension), there is no single, common equilibrium solubility; particles either grow or dissolve. Clearly, the largest particles have the strongest tendency to grow owing to their low solubility. This coarsening of colloids, i.e. the decrease of specific surface area in time, is also known as *Ostwald ripening* and it is an important ageing effect, which may occur in any polydisperse system of sufficiently small particles. It is observed in emulsions and aqueous sols, as well as colloidal metal catalysts in a high temperature gas (decrease of catalytic activity in time). An obvious consequence of Ostwald ripening is loss of specific surface area, which may proceed quite rapidly for small, highly soluble particles. Illustrative examples are aqueous sols of nanometer-sized silica particles (fig. 2.8), which immediately after preparation undergo a rapid decrease in surface area on a time scale of hours to days, followed by a much slower decay, which may continue to

---

[1] I.L. Mladenovic et al., *loc. cit.*
[2] R.K. Iler, *loc. cit.*

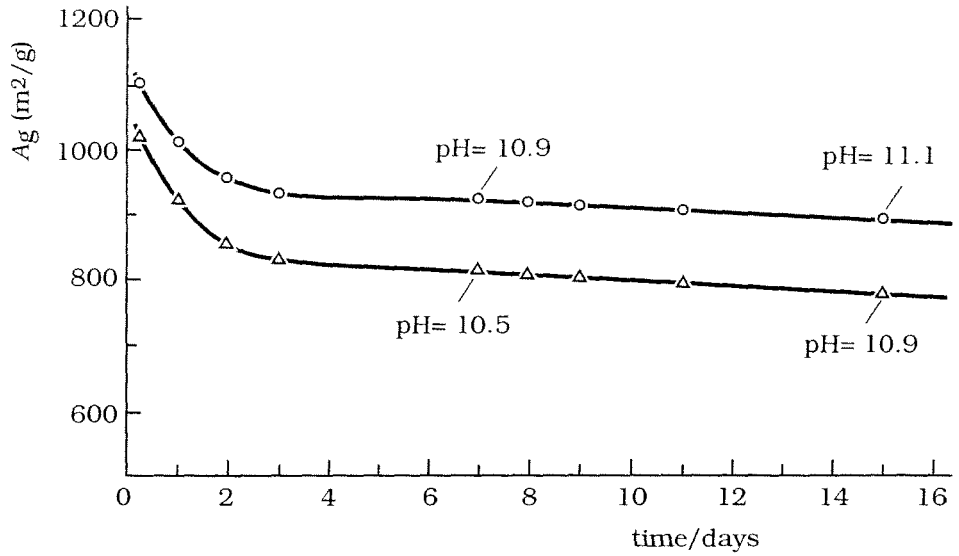



**Figure 2.8.** Aqueous sols of very small silica particles, freshly prepared by acidifying a waterglass solution, exhibit a rapid, initial decrease in specific surface area (measured following the Sears method) due to Ostwald ripening. On a time scale of months the surface area saturates at a value typically in the range of 700-800 $m^2/g$. (Courtesy of Kenneth Larsson and Bo Larsson, EKA Chemicals, Sweden).

stabilize at 700-800 $m^2\,g^{-1}$ after several months. The area is still high and on TEM pictures very small clusters can be seen. Ageing is accompanied by a slight increase of pH. The behaviour of small, highly soluble particles, as in fig 2.8, illustrates that in the condensation method there is actually no sharp distinction in time between 'sol preparation' and 'ageing.' In the initial precursor solution, the specific surface area decreases as soon as precipitates start to grow. The surface area, of course, is a macroscopic quantity, which loses its meaning for particles with radii down to molecular size. Note, however, that we introduce a precipitate surface area $A$ via [2.2.11] and, therefore *de facto*, also a specific surface $A_g$ which continuously decreases upon the growth of the particles of fig. 2.6.

We will now briefly outline the kinetics of dissolution and ripening. In a polydisperse sol, the bulk concentration $c(\infty)$ is not constant, but slowly decaying in time due to the gradual disappearance of small, soluble particles. At any moment in time there is one sphere radius $a_0$, which is in metastable equilibrium with the bulk concentration

$$c(\infty) = c(\text{sat})\exp\left[\frac{2\gamma v_m}{a_0 kT}\right] \qquad [2.2.53]$$

where $c(\text{sat})$ is again the equilibrium solubility of a flat surface. If the *local* solute concentration near a sphere with radius $a_i$ is also fixed by the Gibbs-Kelvin equation, the steady state diffusion flux for sphere $i$ is

$$J_i = 4\pi D a_i c(\text{sat})\left(\exp\left[\frac{2\gamma v_m}{a_0 kT}\right] - \exp\left[\frac{2\gamma v_m}{a_i kT}\right]\right) \qquad [2.2.54]$$

It is clear that spheres with radii $a_i < a_0$ dissolve because $J < 0$, whereas for $a_i > a_0$, the particles grow. The average particle radius and, of course, the critical radius $a_0$ increase in time, so that the exponents in the diffusion flux can be linearized at a later stage of the ripening process. In that case, we can write for the growth (or for the dissolution rate) of sphere $i$ the approximate result

$$\frac{d}{dt}a_i^3 = 6Da_i c(\text{sat})\frac{\gamma v_m^2}{kT}\left[\frac{1}{a_0} - \frac{1}{a_i}\right] \qquad [2.2.55]$$

One limiting case of Ostwald ripening allows for a simple analytical solution, namely monodisperse spheres with radius $a$, from which dissolved matter is deposited on very large particles or a flat substrate. If this substrate controls the bulk concentration, $a_0$ is infinitely large and consequently

$$\frac{da^3}{dt} = -6Dc(\text{sat})\frac{\gamma v_m^2}{kT} \qquad [2.2.56]$$

Thus, for this case the particle volume decreases at a constant rate. To go beyond such a bidisperse model and evaluate the time evolution of a continuous size distribution of spheres, growing and dissolving according to [2.2.55], is a demanding task, dealt with in the classical studies of Lifshitz and Slezov[1] and Wagner[2] (*LSW theory*). We quote the essential results, referring to reviews for more discussion on the principles[3] and applications[4] of the LSW theory. The assumptions in this theory are the same as those underlying [2.2.55]: there is only transport due to diffusion, the sphere solubility is so low that the Gibbs-Kelvin equation can be linearized and there is no interaction between the spheres other than that their growth rates are coupled by the average bulk concentration. The LSW theory predicts for large times the asymptotic result

$$\frac{d\langle a\rangle^3}{dt} \sim \frac{8}{9}Dc(\text{sat})\frac{\gamma v_m^2}{kT} \qquad [2.2.57]$$

i.e. in a late stage of the ripening process, the average particle size increases as $t^{1/3}$. Further, the supersaturation correspondingly falls as $t^{-1/3}$ and the number of spheres as $t^{-1}$. A remarkable finding of the LSW theory is that due to Ostwald ripening the size distribution approaches a certain universal, time-independent shape, irrespective of the initial distribution. The LSW theory appears to work well for emulsions[4]; for

---

[1] I.M. Lifshits, V.V. Slezov, *Zhur. Eksp. Teor. Fiz.* **35** (1958) 479. The names are also transcribed as Lifshitz and Slyezov or Slyozov, for instance in *J. Phys. Chem. Solids* **19** (1961) 35.
[2] C. Wagner, *Z. Elektrochem.* **65** (1961) 581.
[3] W. Dunning, *Particle Growth in Suspensions*, A.L. Smith, Ed., Academic Press (1973).
[4] P. Taylor, *Adv. Colloid Interface Sci.* **75** (1998) 107.

inorganic particles, a comparison with experimental data is less straightforward[1]. It is, in any case not correct to use the $t^{-1/3}$ scaling as the *general* hallmark for Ostwald ripening in view of the restrictive validity of the LSW theory, also see sec. V.8.3b. For example, close to the nucleation stage when many highly soluble particles are present, linearization of the Gibbs-Kelvin equation and the assumption of non-interacting particles will be invalid. Another important factor is the topology, which for grain growth in a polycrystalline material (fig. 2.2d) or bubble growth in a foam[2], is obviously very different from the isolated spheres in the LSW theory.

### *2.2f Seeded nucleation and growth*

So far, we have assumed that particles nucleate and grow in a solution of only one solute. In practice, strict homogeneous precipitation is difficult to realize because of the omnipresence of contaminants, dust, motes and irregularities on the vessel wall, which may act as nucleation sites for the new phase. This so-called *heterogeneous nucleation* may have a dramatic effect on the kinetics as can be observed after opening a bottle of beer or champagne when carbon dioxide bubbles rapidly nucleate on the glass surface. See also sec. V.7.2a. The reader may wish to verify the effect of adding extra nucleation sites in the form of sugar or sand grains. Heterogeneous nucleation, however, is not necessarily a nuisance. Actually, it is an important strategy to decrease size polydispersity. This was first exploited by Zsigmondy[3] who used the extremely fine Faraday gold sol[4] as a seed solution for the preparation of quite monodisperse gold colloids. The seed can also differ chemically from the precipitating material, leading to the formation of core-shell colloids. Of the many examples, we mention the growth of silica on gold cores[5], and other inorganic particles[6] and the preparation of core-shell semiconductor particles[7]. Such well-defined composite colloids are increasingly important in materials science, in addition to their use in fundamental studies.

The efficiency of seeds or a container wall to catalyze nucleation is due to the reduction of the interfacial Gibbs energy of a precipitating particle. As a simple but illustrative example[8] we consider a phase $\alpha$, which nucleates as a spherical cap of radius $a$ on a flat seed substrate $\beta$ immersed in a liquid $L$. The cap wets the substrate with a contact angle $\theta$ as shown in fig. 2.9. As in the case of homogeneous nucleation

---

[1] W. Dunning, *loc. cit.*

[2] N. Rivier in: D. Bideau and A. Hansen, Eds., *Disorder and Granular Media*, North Holland (1993).

[3] R. Zsigmondy, P.A Thiessen, *Das Kolloide Gold*, Akad. Verlag. Leipzig (1925).

[4] M. Faraday, *Phil. Trans. Royal Soc.*, **147** (1857) 145. Faraday prepared gold particles with a diameter around 3 nm by reduction of a gold salt with phosporus in ether.

[5] L.M. Liz-Marzán, M. Giersig, and P. Mulvaney, *Langmuir* **12** (1996) 4329.

[6] F. Caruso, *Adv. Mater.* **13** (2001) 11.

[7] H. Weller, *Quantized semiconductor particles*, *Adv. Mater.* **5** (1993) 88.

[8] For a detailed treatment of heterogeneous nucleation, including other nucleus shapes, see the chapters by R.A. Sigsbee and A.G. Walton in *Nucleation*, A.C. Zettlemoyer (Ed.) Marcel Dekker (1969).

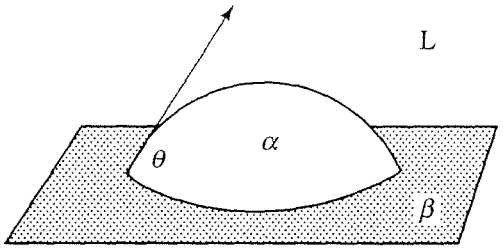



**Figure 2.9.** Heterogeneous nucleus of substance $\alpha$ in the form of a spherical cap on a planar seed $\beta$ immersed in a liquid $L$. The precipitating phase $\alpha$ partially wets the seed with a contact angle $\theta$.

in 2.2.b, it is assumed here that a nucleus is a macroscopic piece of structureless[1] bulk matter to which equilibrium thermodynamics can be applied. The Gibbs energy change due to the formation of the cap in fig. 2.9 then equals

$$\Delta G_{het} = 2\pi a^2(1-\cos\theta)\gamma_{\alpha L} + \pi a^2 \sin^2\theta(\gamma_{\alpha\beta} - \gamma_{\beta L}) - \frac{4}{3}\pi a^3 \frac{kT\ln S}{v_m} f(\theta)$$
$$f(\theta) = \frac{1}{4}(1-\cos\theta)^2(2+\cos\theta) \qquad [2.2.58]$$

This expression comprises the surface area with interfacial tension $\gamma_{\alpha L}$ between $\alpha$ and the liquid, the interfacial area between $\alpha$ and the substrate times the difference $\gamma_{\alpha\beta} - \gamma_{\beta L}$, and the decrease in Gibbs energy due to the volume of the cap. The geometrical factor $f(\theta)$ is the ratio of the volume of the spherical cap in fig. 2.9 to that of a sphere with the same radius. The interfacial tensions in [2.2.58] are related by Young's equation, which is valid when $\theta$ is the contact angle for equilibrium with respect to the horizontal force components

$$\gamma_{\beta L} = \gamma_{\alpha\beta} + \gamma_{\alpha L}\cos\theta \qquad [2.2.59]$$

Substituting this result into [2.2.58], we find from the condition $d(\Delta G_{het})/da = 0$ that the critical radius equals

$$a^*_{het} = \frac{2v_m}{kT}\frac{\gamma_{\alpha L}}{\ln S} \qquad [2.2.60]$$

which is the same as the critical radius in [2.2.23] for homogeneous precipitation in the absence of a substrate. The energy barrier can be written in terms of the homogeneous energy barrier $\Delta G^*_{hom}$ in [2.2.22]

$$\Delta G^*_{het} = \Delta G^*_{hom} f(\theta) \qquad [2.2.61]$$

showing that $f(\theta)$ quantifies the catalytic effect of the substrate. The presence of this substrate does not change the critical radius of the sphere, but only reduces the Gibbs energy maximum due to complete or partial wetting by the newly precipitated phase.

---

[1] We disregard here any effect of crystal structures and their (miss)match in epitaxial growth, see Zettlemoyer *loc. cit.*

Note that for complete dewetting, when $\theta = 180°$ and $f(\theta) = 1$, the Gibbs energy maximum equals $\Delta G^*_{hom}$; then, precipitation proceeds as if no seeds or substrates are present. For any contact angle in the range $0 \leq \theta < 180°$, the substrate lowers the activation energy for nucleation because $0 \leq f(\theta) < 1$. If the contact angle is nearly zero, it will be impossible to maintain supersaturation in the presence of the substrate. In view of the strong dependence of the nucleation rate on the activation energy in [2.2.30], it is clear that seeds may speed up precipitation kinetics considerably. The heterogeneous nucleation rate $I_{het}$ will have a form similar to that of the homogeneous rate $I_{hom}$ in [2.2.30], and the pre-exponential factor [2.2.30] will remain the same in order of magnitude. Thus

$$I_{het} \sim I_{hom} \exp\left[\Delta G^*_{hom}(1 - f(\theta))/kT\right] \qquad [2.2.62]$$

For contact angles $\theta \leq 30°$, $f(\theta)$ is practically zero and the nucleation rate is enhanced by a factor of $\exp(\Delta G^*_{hom}/kT)$.

The independence of the critical size of the presence of a substrate in [2.2.60] is perhaps unexpected and, in any case, unlikely to be a general feature of heterogeneous nucleation. Here, this independence is a consequence of geometry assumed in fig. 2.9; the radius of curvature of the cap on the smooth, flat substrate is the same as for the spherical nucleus in a homogeneous solution. Therefore, neither the solubility nor the critical radius changes because these are fixed by the curvature. When the cap is deformed, or when the substrate is curved or structured, the situation is evidently more complicated. Steps and kinks on the substrate may act as active sites because they enable more of the surface of the nucleus to be in contact with the substrate, which lowers its surface excess Gibbs energy. One extreme case is a cavity in the substrate, which allows maximum contact for a (non-cylindrical) nucleus as discussed elsewhere[1]. See also fig. V.7.5.

Such a cavity is actually a simple example of a template, which may direct the morphology of a growing cluster. For homogeneous nucleation in a solution, the possibilities for controlling particle shape are very limited. There is, admittedly, an impressive variety of methods for synthesizing anisotropic colloids in a bulk solution[2], but in most cases the outcome of a method can rarely be anticipated. Particle shape is sensitive to various parameters, such as pH, temperature, reactant concentrations, nature of anions and organic additives that may block certain faces. Thermodynamics, of course, only provides us with an equilibrium shape for crystals according to the Wulff relations [2.2.51]. However, growth of the equilibrium form is rarely encountered[3] and the effect of the experimental parameters mentioned earlier on the growth kinetics of

---

[1] D.R. Uhlmann, B. Chalmers in *Nucleation Phenomena*, D.E. Gushee, Ed., *Am. Chem. Soc. Publ.* (1966).
[2] E. Matijević, *Chem. Mater.* **5** (1993) 412.
[3] W. Dunning, *loc. cit.*

the faces of a polyhedral particle is hard to predict. So, in this respect heterogeneous nucleation and growth on (or in) a host is an attractive alternative to achieve shape control. Potential hosts, such as micelles and other self-assembled structures, are reviewed elsewhere[1,2].

### *2.2g Comminution and other preparation methods*

In condensation methods, colloidal particles are prepared from molecular species, whereas in *dispersion methods* the colloidal size range is reached by breaking down a macroscopic phase into progressively smaller parts. A well-known example of the latter is emulsification, which comprises the dispersion of one liquid in the presence of another (see sec. V.8.2). Sometimes shaking or stirring suffices to obtain an emulsion, and in other instances strong mechanical forces from 'colloid mills' are needed. Only when the interfacial tension between the two liquids is very low may the thermal motion of the molecules provide the energy required for emulsification, a well-known practical example being the spontaneous emulsification of agricultural chemicals in water. Mostly, this situation leads to microemulsions.

Emulsification is also an ingredient of increasing importance for the preparation of inorganic colloids; a liquid reactant is emulsified and then polymerized to form a solid particle[3]. This strategy is being used, among others, for the synthesis of very small, monodisperse silica particles[4]. Instead of an emulsion, one can also start from an aerosol; here, the dispersion step is the formation of airborne liquid droplets, which contain some inorganic precursor. In spray drying (an important method in the ceramic industry), such droplets are dried in a flow of hot air to produce inorganic powders. In other aerosol methods water is not removed, but deliberately added to obtain inorganic particles, as in the case of titanium (IV) ethoxide aerosols that react with water vapour to yield spherical, amorphous titania colloids.

One can say that from the viewpoint of inorganic colloid synthesis, emulsification and aerosol methods are nothing but condensation methods, except that the 'reactor' has been formed by a dispersion technique. The dispersion of inorganic material itself, also called comminution[5], is the process of mechanical fracture in a ball mill. Such a mill is a rotating cylindrical vessel, containing inorganic materials (crystals, aggregates of particles) and tungsten carbide or alumina balls. The ease with which a mineral or clay can be ground depends on the surface tension $\gamma$ and the mechanical strength of

---

[1] M. Pileni, *Nature Materials* **2** (2003) 145.
[2] J.H. Adair, E. Suvaci, *Curr. Opin. Colloid Interface Sci.* **5** (2000) 160.
[3] The procedure reminds of monomer droplets which polymerize to latex colloids by the addition of a monomer-soluble initiator. See e.g. R. Buscall, T. Corner, and J.F. Stageman, *Polymer Colloids*, Elsevier (1985).
[4] K. Osseo-Asare, reviews microemulsion-mediated synthesis of inorganic colloids in the nanometer range in *Handbook of Microemulsion Science and Technology*, P. Kumar and K. Mittal, Eds., Marcel Dekker (1999) 549.
[5] See the review by De Castro and Mitchell, mentioned in sec. 2.5a.

the solid. The minimum work needed to break a column of material of cross-sectional area $A$ equals the surface excess Gibbs energy $\Delta G = 2\gamma A$. If the newly formed surfaces are immersed in a solution, any adsorption will lower the surface tension according to the Gibbs adsorption equation [I.2.13.8]

$$d\gamma = -RT\,\Gamma d\ln a \qquad [2.2.63]$$

where $\Gamma$ is the surface concentration, $a$ the activity of the adsorptive. Thus, in view of the Gibbs isotherm, we can understand why comminution is carried out on solids submerged in a solution rather than in air. In aqueous solutions, the species adsorbed on inorganic surfaces include ions which form an electrical double layer. Alternatively, often surfactants or polyelectrolytes are added to further reduce the tendency of particles to adhere to one another after comminution. However, most of the energy input is not invested in increasing the interfacial area, but dissipated as heat in the agitation process which must achieve a very high energy density close to the particles to break them down. The applied stresses must overcome the mechanical strength of the particles. There are several ways to achieve that. Agglomerates can be dispersed by impact on a surface (of the vessel or of balls that are added), or the particles are forced to undergo pressurization and decompression in rapid cycles. Brittle particles are better dispersed by the impact mechanism, elastic particles rather by shear. A variety of mills are commercially available, both for dry and for wet milling.

It may be added that the Van der Waals attraction is also reduced by the solution, in comparison to dry powders in air or another gaseous atmosphere. The notorious 'caking' of dry powders, incidentally, is partly caused by capillary forces when the powders are actually not dry enough. A similar attraction is observed for colloids in a liquid mixture, with one component preferentially wetting the colloids.

The comminution method produces particles with a broad distribution in shape and size and, in general, the (relative) distribution becomes wider for a longer milling time. The lower end of the size distribution is a particle size of O($\mu$m) for a typical milling time of hours to days. A much higher degree of dispersion is difficult to achieve because very high shear forces are required to fracture solids in the sub-micron range. Milling, nevertheless, is frequently applied in industrial practice for the preparation of dispersions, not only to break up particle aggregates but also to simultaneously and intensively mix particles and polymeric additives prior to processing the dispersions.

### *2.2h Separation and fractionation techniques*

Preparation of a sol is usually followed by a separation procedure, which may serve various purposes. Distillation is the obvious method to remove volatile components, such as ammonia from a silica alcosol (see sec. 2.4a). It may also be used to concentrate sols (at reduced pressure) or to transfer particles to other solvents. Dialysis and electrodialysis are employed to remove low-molecular compounds and contaminating electrolytes, which migrate out of the sol across a semipermeable membrane into a

liquid reservoir. A simple but effective dialysis setup is a flexible cellulose bag, filled with sol and suspended in a flow of demineralized water. (Such membranes, incidentally, are not chemically inert because cellulose fibres may hydrolyze.) Often the sol level in the cellulose tube rises due to the increasing osmotic pressure. The loss of electrolyte also manifests itself in a viscosity increase due to the electroviscous effects, see sec. 6.9b. Solute diffusing across the cellulose membrane may also stem from dissolving colloids because of the Gibbs-Kelvin effect. For example, prolonged dialysis of silica sols produces a notable weight loss due to the continuous removal of soluble silica, which promotes dissolution of small particles.

Dialysis against a salt solution is equivalent to ion exchange. Ion exchange resins are available for simultaneously exchanging cations for $H^+$ and anions for $OH^-$. These 'mixed-bed' resins are used for preparing nearly salt-free sols with a large Debye length. To separate the colloids themselves from the liquid phase, either filtration or sedimentation is required. These two methods will be discussed in some more detail; for filtration below, for sedimentation in 2.3d.

In a filtration process, the colloids are separated from the suspension by their accumulation on a filter or membrane, which is permeable for solvent and low-molecular species[1]. Liquid flow is driven by a pressure difference, which for vertical filtration is due to the weight of the liquid itself, plus a piston or external gas pressure. The liquid transport through the membrane and the growing packing of colloids, with typical microstructures as in fig. 2.1, is an example of flow in a porous medium. Thus, important trends, such as why small particles are difficult to separate, can be explained in terms of d'Arcy's law [I.6.4.36] for viscous flow of an incompressible fluid through a porous medium, which we write here as

$$\boldsymbol{u} = -\frac{B}{\eta}\nabla p \qquad [2.2.64]$$

Here $\boldsymbol{u}$ is the average flow velocity of the liquid with viscosity $\eta$ in a medium with permeability $B$, driven by an average hydrostatic pressure gradient $\nabla p$. The porous medium in filtration is the layer of deposited colloids, which grows in time. Thus, the draining liquid experiences an increasing drag, which retards the filtration rate, as is observed in processes such as slip casting of inorganic sols and water purification. To quantify this retardation, consider a sol with solid volume fraction $\phi$, which forms a filter cake with volume fraction $\phi_c$ and thickness $L$. Conservation of particle volume implies for the growth rate $\mathrm{d}L/\mathrm{d}t$ of the cake:

$$u = \frac{\phi_c - \phi}{\phi}\frac{\mathrm{d}}{\mathrm{d}t}L(t) \qquad [2.2.65]$$

Using d'Arcy's law [2.2.64], we can eliminate $\boldsymbol{u}$ as $-(B/\eta)(\Delta p/L(t))$, which can be

[1] For an illustrative example on poly(styrene) colloids, see K. Bridger, M. Tadros, W. Leu, and F. Tiller, *Sep. Sci. and Techn.* **18** (1983) 1417.

substituted in [2.2.65] and integrated to obtain the *filtration law*, which quantifies the rate at which colloids are separated

$$L^2(t) = \frac{2\phi}{\phi_c - \phi}\frac{B}{\eta}\Delta p\, t \qquad [2.2.66]$$

where $\Delta p$ is the absolute pressure drop across the cake. Hence, the colloid deposit grows according to $L \sim t^{1/2}$ and the liquid velocity decreases as $v \sim 1/t^{1/2}$. The scaling $L \sim t^{1/2}$ reminds one of diffusional growth, which, however, is clearly not the case here; particles are convected by the flowing liquid and diffusion is neglected. The square-root time dependence of the layer thickness $L$ is due solely to an increase in time of the hydrodynamic resistance. The absolute filtration rate is determined by the specific surface area $A_g$ of the colloids (see sec. 2.3c), which, according to the *Kozeny-Carman* (KC) relation (see also [I.6.4.41]) enters the liquid permeability via:

$$B = \frac{1}{C}\frac{(1-\phi_c)^3}{\phi_c^2}A_g^{-2} \qquad [2.2.67]$$

Here, $C$ is the Kozeny constant[1], which for a random sphere packing is about $C \approx 5$. The KC relation is very useful because a value of $C \sim 5$ also holds for dense packings

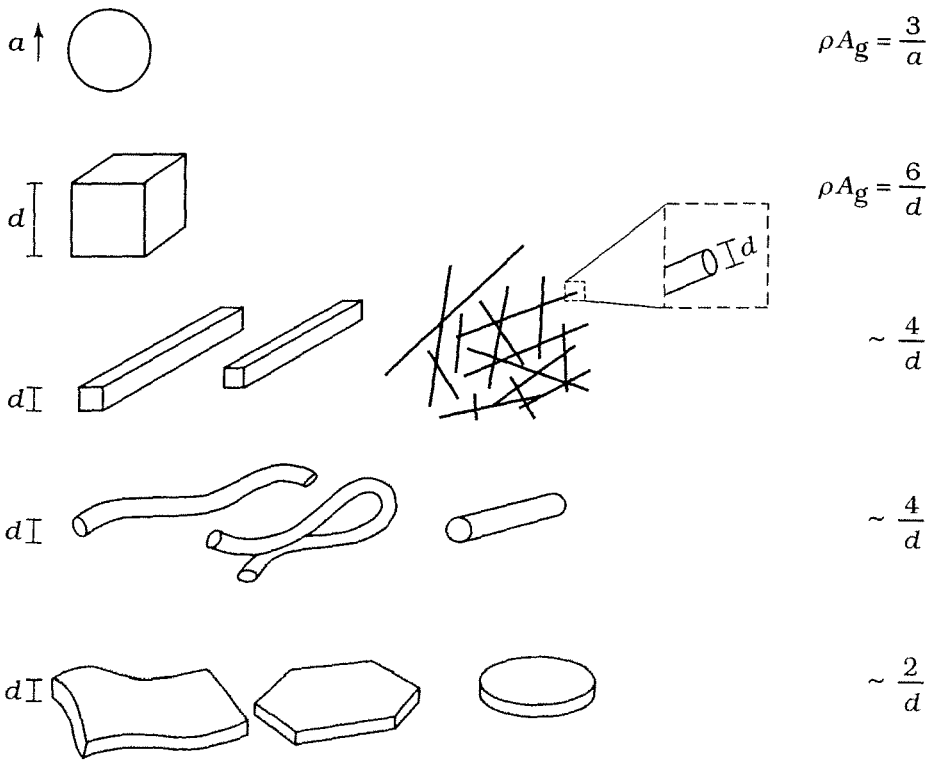



**Figure 2.10.** Specific surface area $A_g$ related to $a$ or $d$ for a variety of particle shapes. For high aspect ratios, $A_g$ only depends on the smallest dimension $d$.

[1] J. Kozeny, *Sitzber. Akad. Wissensch., (Wien) (IIa)* **136** (1927) 271.

of non-spherical particles (including fibers[1]), as well as mixed particle sizes[2,3]. The specific surface area (fig. 2.10), therefore, provides a reasonable estimate of $B$ and, consequently, of the filtration rate of a particular dispersion of particles.

We note here that $C$ may vary with the concentration $\phi$ of the filter cake. For example, substituting in [2.2.67] values for $B$ and $\phi$ from an extensive study on flow in fibrous media[4], we find that the Kozeny constant starts to decrease significantly below $\phi \sim 0.1$. For $\phi > 0.15$, however, $C$ values scatter around $5 \pm 1$, so indeed [2.2.67] with $C \approx 5$ is sufficiently accurate for dense filter cakes of particles of the type shown in fig. 2.1a-c.

It is sometimes stated that the KC relation is purely empirical, but this is not quite correct. The pore geometries in a filter cake are clearly too complicated to allow an exact solution of the Stokes equations for viscous flow. The KC relation is an approximation for this solution at the cost of introducing a constant $C$, which in most cases is indeed empirical. The KC scaling itself, however, is a consequence of Poiseuille-like flow as can be seen as follows[4]. Consider a packing of particles (as in fig. 2.1) with solid volume fraction $\phi_c$ and surface area $A$. We can assign a characteristic length $d_p$ to the pore space, with volume $V_{por}$ between the particles, which scales as

$$d_p \sim \frac{V_{por}}{A} = \frac{1-\phi_c}{\phi_c}\frac{1}{A_g} \qquad [2.2.68]$$

where $A_g$ is the specific surface area of the particles, here[5] defined as the surface area per particle volume. The liquid permeability has the dimensions of a length squared (for example, a tube radius squared in the case of the Hagen-Poiseuille law, see sec. I.6d sub (2)). Hence, the permeability of the pore space scales as

$$B_{pore} \sim \left(\frac{1-\phi_c}{\phi_c}\right)^2 A_g^{-2} \qquad [2.2.69]$$

In a filtration experiment, however, we do not measure fluid flow in the pores, but the flow rate averaged over the whole filter cake, including the solid phase. Since inside a solid the flow rate is zero, the overall permeability that determines the filtration rate is

$$B \sim \frac{(1-\phi_c)^3}{\phi_c^2} A_g^{-2} \qquad [2.2.70]$$

which is the KC scaling in [2.2.67]. It is clear that $B$ and, hence, the filtration rate, is determined by the smallest dimension of a particle, e.g. the thickness of a platelet,

---

[1] E.J. Wiggins, W.B. Campbell, and O. Maass. *Can. J. Res.* **B17** (1939) 318.
[2] P.C. Carman, *Trans. Inst. Chem. Eng.* **15** (1937) 150; *J. Soc. Chem. Ind.* **57** (1938) 225.
[3] G.W. Jackson, D.F. James, *Can. J. Chem. Eng.* **64** (1986) 364.
[4] D.M.E. Thies-Weesie, A.P. Philipse, *J. Colloid Interface Sci.* **162** (1994) 470.
[5] In [2.3.16] and elsewhere $A_g$ is in $m^2$ per gram.

which determines the specific surface area (fig. 2.10). Particles with all dimensions in the micron range are easy to separate by filtration, whereas any dimension in the nanometer range necessitates very high pressures. We refer here to the filtration of stable sols; aggregation of small particles to large clusters will enhance the filtration rate[1]. Sedimentation is not a suitable alternative to separate stable particles, which are too small for filtration because settling and filtration rate have the same particle size dependence, as will become clear in section 2.3 d on sedimentation.

An interesting aspect, lastly, of the filtration of colloidal suspensions is that the density of the particle deposit strongly depends on particle shape. For example, for rigid fibers or rods[2] with high aspect ratio $L/d >> 1$, the packing density of randomly oriented particles asymptotes towards zero as $\phi_c \sim d/L$. Thus, for an *a priori* estimate of permeability and the filtration rate on the basis of the KC-scaling, this shape effect must be taken into account.

### *2.2i Surface modification*

Surface modification is the deliberate attachment of (macro)molecules to the surface of an inorganic colloid to change its physical properties or chemical functionality. This change, of course, is only permanent if the attachment cannot be undone by thermal motion. Such surface modification occurs either via a covalent bond or significant adsorption energy. The canonical example is a polymer shell, which stabilizes colloidal cores in a medium where they otherwise would coagulate, as in the case of soot colloids in aqueous ink, which are protected by arabic gum. This polymeric stabilization and its entropic and enthalpic origin will be discussed in detail in sec. V.1. Here, we will only briefly address some general aspects of surface modification, referring to the literature for detailed examples[3].

Surface modification seems at first sight a straightforward procedure: identify reactive groups on the inorganic colloid surface, choose a molecule with a suitable chemical linker, work out the chemical reaction conditions and perform the reaction itself. For the hydroxyl groups on metal(hydr)oxide particles, one could think of a linkage to carboxylic acids or alcohols, or a modification using reactive siliconalkoxides. The surface silanol groups of silica, for example, react under mild conditions with so-called silane coupling agents (SCAs), i.e. siliconalkoxides with one alkoxide replaced by a functional organic group[3]. SCAs were initially designed to intermediate between inorganic surfaces and an organic matrix (hence, their name), but they are also very suitable for the *in situ* modification of colloids in a sol[4]. A major advantage is

---

[1] The filtration rate can be used to monitor the effectiveness of polymeric flocculants. An entrance to relevant literature is J. Gregory, A.E.I. de Moor, *ACS Symp. Ser.* **240** (1984) 445.
[2] S.R. Williams, A.P. Philipse, *Phys.Rev.* **E 67** (2003) 051301.
[3] E.P. Plueddeman, *Silane Coupling Agents* (Plenum) (1991); D.E Leyden, Ed., *Silanes, Surfaces and Interfaces*, Gordon and Breach (1986). Also see sec. 2.4.
[4] A. Philipse, A. Vrij, *J. Colloid Interface Sci.* **128** (1989) 121.

the large variety of functional moieties, which can be attached to the particle surface, and even buried inside particles when the processes of silica synthesis and modification are mixed[1].

SCAs hydrolyze to reactive silanols, which graft themselves onto silica via formation of siloxane linkages. As an illustration, immerse a hydrophilic glass slide for about 30 minutes (or even less) in a solution of typically one percent of SCA[2] in ethanol, with some acid or base added to facilitate hydrolysis of the SCA. Next, rinse the slide with pure ethanol to remove free SCAs and dry it in hot air. The slide is now poorly wetted by water as a manifestation of its surface modification. In a similar manner, one obtains hydrophobic sand or the functionalized silica grains used in affinity-chromatography. For the particles in a silica sol, the surface modification chemistry is basically the same as for macroscopic silica, but an important additional challenge arises and that is to avoid coagulation during the modification procedure.

Once reactive oligomers or polymers attach to a colloidal core, the core-shell particle behaves as one kinetic unit with an average kinetic energy of $(3/2)kT$. This energy has to be weighed against the replacement of a large number of solvent molecules by the adsorbed species. Even a very small Gibbs energy penalty per replacement may suffice to produce aggregates that do not break apart by thermal motion. Such aggregation can also be induced by minute changes in the nature or composition of the solvent, a subtle effect that is often difficult to predict or to explain afterwards. The fact is that any small change in composition involves a large number of low-molecular species, with a net enthalpy change that easily compensates the entropy loss due to aggregation of large colloids. Thus, the image of colloids coated by reactive molecules while diffusing around in an inert, neutral solvent background is clearly inappropriate. All molecular interactions must, in principle, be accounted for, a challenge which we cannot meet yet.

One obvious counterexample to this neutral background is any solvent adsorption on (modified or unmodified) colloids. Water adsorption on silica is well known, see sec. 3.13b, but polar organic solvents such as dimethylformamide or triethylphosphate also adsorb in significant amounts on bare silica particles, often sufficient to prevent their coagulation. One could make a case that accurate characterization of a colloidal dispersion includes measurement of the immersion enthalpy of particles in their solvent, see sec. II.2.3d. Surface modifiers have to compete with solvent adsorption, which will lower the grafting density[3].

We have indicated several reasons why transfer of charged colloids in polar liquids to modified particles in stable organic sols is a tricky process, which often has to be

---

[1] A. van Blaaderen, A. Vrij, *Langmuir* **8** (1992) 2921.

[2] Take a coupling agent with a hydrophobic group, such as the TPM in sec. 2.4a.

[3] A.M. Nechifor, A.P. Philipse, F. de Jong, J.P.M. van Duynhoven, R.J.M. Egberink, and D.N. Reinhoudt, *Langmuir* **12** (1996) 3844.

optimized by trial and error[1]. Small particles, it should be noted, also have a kinetic disadvantage, because the number densities of nanometer-sized particles are high. Therefore, any coagulation will occur rapidly, since the coagulation rate is proportional to the square of the number density. For modified, stable colloids, of course, the small particle size becomes a benefit in view of the many functional groups per gram.

Lastly, one attractive option, which should be mentioned, is the simultaneous synthesis and modification of inorganic colloids by their nucleation and growth in the presence of the modifying agent, which also influences (and perhaps even controls) the particle size. Examples are the formation of small magnetic particles[2] by thermolysis of metalcarbonyl-precursors in surfactant solutions, and the synthesis of extremely small gold colloids by reduction of gold salts in the presence of silane coupling agents[3].

### *2.2j Other methods*

Preparation methods mentioned so far certainly do not exhaust the routes for obtaining colloidal sols. Metal colloids, for example, can also be formed by electrical disintegration methods. Here, an arc is passing between electrodes under water, vaporizing electrode material to a metal gas that subsequently condenses into particles of colloidal dimensions. The formation of aqueous metal sols by electrical dispersion techniques was pioneered by Bredig and others[4] but has been replaced by more convenient alternatives such as the reduction of metal salts or the thermal decomposition of metal-carbonyl compounds and metal ions complexed by chelating agents (for example, triethanolamine)[5]. Another preparation strategy already explored in the early days of colloids science, however, has been more lasting, and even evolved into a separate branch of materials research, often referred to as *sol-gel* processing[6]. Thomas Graham[7] not only coined the term 'colloids' but also the terms 'sol' and 'gel' to denote, respectively, the initial and final state in the coagulation of a liquid dispersion to a space-filling solid-like material. Graham studied this transition for silica, alumina and other inorganic substances in water as well as ethanol. He found, for example, that water-glass in alcohol may change from an 'alcosol' to an 'alcogel' with nearly the same volume at already very low silica concentrations (see also section 2.4a). Sol-gel transitions and other coagulation phenomena became a classical topic of colloid science; the field of sol-gel *processing* is broader and aims to cover the whole route from a liquid sol via gelation, drying and sintering to the final solid state, usually

---

[1] C. Pathmamanoharan, PhD. thesis, Utrecht, The Netherlands (1998).
[2] T.W. Smith, D. Wychick, *J. Phys. Chem.* **84** (1980) 1621.
[3] P.A. Buining, B.M. Humbel, A.P. Philipse, and A.J. Verkleij, *Langmuir* **13** (1997) 3921.
[4] G. Bredig, *Z. Angew. Chem.* **11** (1898) 951; T. Svedberg, *Die Methoden zur Herstellung Kolloider Lösungen Anorganischer Stoffe*, Verlag von Theodor Steinkopff (1909).
[5] For an update on electrical methods see Delplancke's review in sec. 2.5.
[6] C.J. Brinker, G.W. Scherer, *Sol-Gel Science*, Academic Press (1990).
[7] T. Graham, *Phil. Roy. Soc. London* **151** (1861) 183; *J. Chem. Soc.* August (1864) 618.

a ceramic material[1]. The process may start with a precursor solution of aqueous salts or metal-organic compounds; in particular metal alkoxides are widely used in sol-gel research (sec. 2.4 provides examples of both types of precursors). In addition to molecular precursors, sol preparation may also employ dry powders synthesized by vapour-phase methods at the high temperatures produced by a furnace, flame, plasma or laser. Advantages are the high purity of the powders, and the possibility of atomic-scale mixing in the vapour phase.

Metal alkoxide precursors are convenient because they readily react with water to form metal (hydr)oxides at room temperature. In a limited number of cases (silica, titania and zirconia) monodisperse spheres are formed[2]. These amorphous spheres have a significantly lower mass density than the corresponding bulk oxide and contain solvent as well as residual organic groups. The specific surface area is generally much greater than expected from particle dimensions measured with TEM, and at least for silica it is well known that the spheres noticeably shrink when exposed to the vacuum in an electron microscope. These features clearly show that hydrolysis of metal alkoxides does not produce massive spheres but rather spherical sponges having sufficient internal cross-linking to maintain their shape, and sufficient porosity to allow permeation of solvent and small molecules.

The variety of preparation schemes in sol-gel processing involving either precursors or powders, is enormous, see the extensive literature survey in ref. [1]. Sol-gel literature is, consequently, often a useful information source on inorganic colloid synthesis, and the chemistry of hydrolysis and condensation of metal ions in solution[3] or reactions involving metal alkoxide precursors[1]. However, what has been said in section 2.2f on the limited predictability and control of particle size and morphology remains true. We may rightfully look with some envy to the controlled colloid formation in biomineralization[4]. Examples are the single crystals of magnetite ($Fe_3O_4$) and other minerals made by bacteria, the monodisperse ferrihydrate ($5Fe_2O_3.9H_2O$) colloids in the iron-storage protein ferritine, and the beautiful silica structures sculped by diatoms[5]. Further study of the still poorly understood in vivo preparation methods used by organisms may provide new ideas for man-made colloids.

## 2.3 Characterization

After synthesizing a colloidal dispersion and performing the required purification or separation techniques, as described in the previous section, we wish to characterize

---

[1] U. Schubert, N. Hüsing, *Synthesis of Inorganic Materials*, Wiley-VCH (2000); C.J. Brinker, G.W. Scherer, *loc. cit.*

[2] C.J. Brinker, G.W. Scherer, *loc. cit.*

[3] For this intricate chemistry see J.P. Jolivet, *Metal Oxide Chemistry and Synthesis; from Solution to Solid State*, Wiley (2000).

[4] S. Mann, J. Webb, and R. Williams, *Biomineralization*, VCH (1989).

[5] L. Addadi, S. Weiner, *Angew. Chem. Int. Ed. Engl.* **31** (1992) 153.

the colloidal particles. Nowadays, sophisticated techniques are available to investigate colloids in nanometric detail, in real as well as reciprocal space (*vide infra*). A discussion of all measurable parameters, with techniques to match, requires an encyclopaedia, so we only present a selection, referring for a more extensive coverage to the provided literature entries.

### *2.3a Visual observations and microscopy*

A great deal of information can already be obtained from visual inspection of a sol, aided by a torch or small laser. Some trends, to which no doubt exceptions may be found, are listed below, starting with optical properties.

*(i) Colours*[1)]

For colloids, which do not absorb light at visible wavelengths, the turbidity is only due to light scattering. A *bluish* appearance in this case is due to Rayleigh scattering of particles with a typical diameter on the order of 100 nm or smaller. This bluish *Tyndall effect* can be clearly observed for dilute dispersions of latex particles and several metal (hydr)oxide colloids, such as boehmite and silica. A *milky white* appearance may be due to anything that shortens the mean free path of photons in the dispersion: large particle size, high refractive index and high colloid concentration. Multiple scattering is easy to demonstrate as it spreads an incoming narrow beam of laser light. A white appearance sometimes manifests aggregation; the bluish Tyndall effect for small aluminum hydroxide or silica colloids changes to white turbidity when the particles coagulate. Inspection of a (either stirred or shaken) sol with a light beam between crossed polarizers reveals *optical birefringence* when the dispersed particles have an anisotropic (plate or rod-like) shape. This birefringence, a mosaic texture of dark and light patches of sol regions with different optical axes, is quite spectacular for concentrated sols of tungsten oxide platelets, and can also be observed for vanadium oxide or boehmite fibres (see sec. 2.4c). Optical birefringence is caused by particles, which align in a shear flow, and when stirring is stopped the mosaic pattern usually rapidly decays by rotational Brownian motion. For very concentrated, strongly interacting particles a permanent birefringence may result from the inability of the plates or fibres to reorient.

Colour effects due to absorption are too numerous to discuss here. Identification of particle composition on the basis of absorption is not always straightforward; witness, for example, the variety in yellow, brown and red colours of the iron (hydr)oxide colloids[2)]. Another important issue is the particle size dependence of absorption spectra

---

[1)] How informative colours and turbidity can be in assessing particle sizes is illustrated well for the case of polymer latex dispersions in E.I. Franses, L.E. Scriven, W.G. Miller and H.T. Davis, *J. Am. Oil. Chem. Soc.* **60** (1983) 1029. For particle sizes in surfactant systems the authors even present a diagnostic guide based only on perceptions of transparency and colour.

[2)] U. Schwertmann, R.M. Cornell, *Iron Oxides in the Laboratory*, VCH (1991)

and the quantum size effect extensively discussed elsewhere[1]. A well-known observation here is the blue shift caused by coagulation of an initially red sol of stable gold particles.

*(ii) Settling*

When particles settle significantly within a few days it is worthwhile to estimate the effective Stokes radius, which would produce the order of magnitude of the observed settling rate. If this radius is much larger than the expected colloid size, either this expectation is wrong or the colloids are aggregating (or both). A sharp interface between sedimenting suspension and supernatant does not necessarily imply monodispersity (see further 2.3d). The sediment on the bottom should also be observed when the vessel is tilted; stable colloids tend to flow like a liquid (be it very slowly when the particles are densely packed), whereas aggregated particles form sediments or gels with a yield stress. Stable colloids produce *diffraction colours* when they form ordered sediments with spacing on the order of optical wavelengths. Well-known examples are the colloidal crystals observed in sediments of repulsive spheres, but rod-like particles may also produce Bragg reflections, such as in the so-called *Schiller layers* of $\beta$-FeOOH rods, which may settle into a smectic structure[2]. In the latter case the visual appearance is a dark brown sediment, which exhibits specular reflection with iridescent colours depending on the angle of reflection.

With a ruler one can already easily estimate an informative number, namely the particle volume fraction $\phi_c H\phi/h$ in sediment with height $h$ formed in a suspension with height $H$ and initial volume fraction $\phi$. A density around $\phi_c \sim 0.63$ is expected for randomly packed, hard spheres. A higher density of, say $\phi_c \gtrsim 0.70$, betrays crystalline ordering or polydispersity since mixtures generally pack more densely than monodisperse spheres. A significantly lower sediment density manifests strongly attractive particles, particles with high aspect ratios, or both. Attractions obviously favour the formation of ramified sediments[3], but it is in particular the combination of fractal structures and high aspect ratio particles, which accounts for the very low densities observed in settled, flocculated clay[4] and fiber suspensions[5].

*(iii) Stability*

Instability of colloidal dispersions with respect to aggregation or phase separation is often easy to detect. Shaking a dilute, unstable sol usually produces visible specs of aggregated particles, which stick to the glass surface; a stable sol (de)wets the surface as a homogeneous fluid. For concentrated, aggregated sols, one observes the mean-

---

[1] H. Weller, *loc. cit.*
[2] Y. Maeda, S. Hachisu, *Colloids Surf.* **6** (1983) 1.
[3] H. Sonntag, K. Strenge, *Coagulation Kinetics and Structure Formation*, Plenum Press (1987).
[4] R. Buscall, *Colloids Surf.* **5** (1982) 269.
[5] A. Philipse, A. Wierenga, *loc. cit.*

dering structures on the glass surface characteristic of buttermilk. It should be noted that stirring or shaking might considerably enhance the rate of coagulation due to an autocatalytic effect. The largest particle clusters are the most efficient in capturing particles in a shear flow and, therefore, grow fastest. This is *orthokinetic coagulation*, to which we return in sec. 4.5b. That is why stirring a sol is not without risk; it may coagulate a sol, which has marginal stability. Exposing a sol sample to high shear rates in a rheometer is a severe test for stability because any small floc, which is 'harmless' in a quiescent disperion, will grow rapidly in the shear field in an autocatalytic fashion. For the influence of particle interaction on sol rheology, see sec. 6.13.

The onset of coagulation or phase separation sometimes announces itself clearly by the so-called *critical opalescence, i.e.* a strong increase in the light scattering on approach of a critical point due to the occurrence of large fluctuations in density, and, hence, in refractive index. Whenever such fluctuations can be observed in a gently shaken sol (their texture is reminiscent of the flow-induced birefringence mentioned earlier), it is pretty sure that the sol will gel or phase-separate soon thereafter.

Observing what happens when an acid or base is added to a charge-stabilized dispersion is always informative. The pH at which a sol coagulates will in general be at its isoelectric point (i.e.p.), though there is the notorious counter example of silica, which is often quite stable at its i.e.p. of about 2 but rather coagulates near pH ~ 8, see sec. 3.13b. The sediment volume of settled flocs is expected to reach its maximum at the pH where particles most strongly attract each other. The charge sign of colloids at various pH values can be checked simply by inserting the poles (or Pt-wires connected to them) of a battery and observing at which pole deposition takes place.

*(iv) Rheology*

The viscous and elastic properties of suspensions will be dealt with in chapter 6. Here we only mention some easy visual checks. Very concentrated stable dispersions, as in sediments of filter cakes, display shear thickening, which makes them hard to process. Squeezing such a filter cake, we notice that it falls dry due to *dilatancy*. When we observe shear thinning, i.e. lowering in effective viscosity when stirring or shaking a dispersion, the colloids are attractive, a hypothesis that can be checked by observing the increasing viscosity when leaving the dispersion quiescent for a while. The origin is the breakdown and reestablishment of coagulate networks. The latter process may take some time. Air bubbles are convenient markers for viscoelasticity. When quickly rotating a vessel, the bubbles are slightly out of phase with the oscillations of the fluid, an effect that can be clearly observed with a bottle of salad dressing. A gradual trapping of air bubbles accompanies the growth in yield stress in a gelling dispersion. When a concentrated dispersion gradually turns into a stiff gel with a high yield stress, as in the case of commercial silica sols on a time scale of months, a low-frequency response is heard upon gently tapping the vessel containing a gel, known as a *ringing gel*.

*(v) Microscopy*

Optical microscopy is, of course, a valuable extension of visual inspection. One can, for example, observe gel structures, growing clusters in a phase separation or using polarized light, liquid crystals or tactoid formation of anisotropic colloids. To observe colloids in bulk, confocal microscopy[1] is a versatile method, which in the footsteps of Perrin (sec. 2.2a), is used to study Brownian motion in concentrated dispersions[2]. This method, however, is not (yet) a routinely available characterization technique, but rather belongs to the category of research tools, which falls beyond the scope of this text. This category also comprises the rapidly expanding field of *scanning probe* microscopy of colloids in the nanometer size range[3]. The microscopy, without which no characterization of colloids is complete, is, of course, electron microscopy[4].

*Transmission electron microscopy* (TEM) is employed to determine sizes, size distributions and particle shapes. The number-average particle size and the spread around this average can be used to predict averages found from other techniques, as explained in appendix 1. Particle sizes can be made absolute, in principle, by adding calibrated latex spheres to the dispersion. The average colloid-latex number ratio on the TEM grid provides a rough estimate of the initial colloid number density. TEM has the disadvantage that the samples have to be dried (which may produce aggregation) and subsequent exposure to a high vacuum may distort or shrink the particles. Therefore, one should be very careful about drawing too many conclusions about the colloid structure in the wet state from TEM images. Particle topography can be imaged with *scanning electron microscopy* (SEM). The additional advantage of SEM is elemental analysis by energy dispersive X-ray analysis[4]. Also here, sample preparation and high vacuum exposure may give rise to artefacts. Two techniques that are presently in a state of development should be mentioned that circumvent this problem.

*Environmental scanning electron microscopy* (ESEM) images colloids, which are kept in an environmental chamber in a water vapour atmosphere with adjustable pressure and temperature. There is no need to coat the particles with conducting film as in conventional SEM. In this way, hydrated colloids can be characterized in their native state, as has been shown for latex spheres as well as inorganic particles[5].

*Cryogenic TEM* images a vitrified film prepared by a fast temperature quench (usually in liquid ethane) of a liquid dispersion film. In principle, vitrification preserves

---

[1] T. Wilson, *Confocal Microscopy*, Academic Press (1990).
[2] W.K. Kegel, A. van Bladeren, *Science* **287** (2000) 290.
[3] See e.g. A. ten Wolde, Ed., *Nanotechnology, The Netherlands Study Centre for Technology Trends* (1998), and B. Bhushan, Ed., *Springer Handbook of Nanotechnology*, Springer (2004).
[4] For a useful literature entrance, also for the various types of optical and scanning probe microscopies and their applications to colloidal dispersions see: E. Kissa, *Dispersions: Characterization, Testing and Measurement, Surfactant Series* **84**, Marcel Dekker (1999).
[5] R.H. Ottewill, A.R. Rennie, Eds., *Modern Aspects of Colloidal Dispersions*, Kluwer (1998).

the particle distribution and morphology of the structures in the liquid film[1,2]. Inorganic, iron colloids with radii as small as 2 nm have been imaged in this way[3]. Cryo-TEM characterization is without doubt an important complement to scattering techniques. The latter have the advantage of probing very large numbers of particles, in a 3-dimensional bulk, on a variety of length scales. Cryo-TEM studies fewer particles in a quasi 2-dimensional film, but directly visualizes any structure formation, shape and size details, which are usually difficult to obtain unambiguously from scattering data in reciprocal space.

### 2.3b *Light scattering*

To characterize colloids with scattering techniques, visible wavelengths as well as neutrons and X-rays are employed. The choice of the wavelength is determined by the length of scales to be probed, but also by the (complex) refractive index of the colloids. For example, concentrated colloidal dispersions are usually too opaque for classical light scattering methods to apply[4]. Light scattering, nevertheless, is a versatile characterization method for many colloid and polymer solutions. Restriction of the treatment to light is further justified by the fact that scattering by other radiation is not fundamentally different[5], so that the general form of the equations to be derived remain applicable, *mutatis mutandis*.

Light scattering is extensively treated in several excellent reviews[5,6,7]. The aim here is to give a brief description restricted to the characterization of homogeneous, non absorbing spherical particles in a dilute suspension. For a discussion on light- and other types of scattering in concentrated systems, see chapter 5. Elsewhere[8], the principles outlined here are generalized to spheres of variable composition, rods, polymers etc. The topic was introduced in chapter I.7.

*(i) Static light scattering (SLS)*

When the refractive index $n_1$ of a colloid differs from the index $n_2$ of the solvent, the electric field of an incident light beam induces an oscillating dipole in the colloid, which causes scattering of light in all directions. We assume that the electric vector of the incident light with wavelength $\lambda$ is polarized perpendicular to the scattering plane, and we detect the scattered photons with the same polarization at an angle $\theta$. A sphere

---

[1] Y. Talmon, *Ber. Bunsenges. Phys. Chem.* **100** (1996) 364.
[2] P.M. Frederik, W.M. Busing, *J. Microscopy* **144** (1986) 215.
[3] K. Butter, P. Bomans, P. Frederik, G. Vroege, and A. Philipse, *Nature Materials* **2** (2003) 88.
[4] Multiple scattering, however, is exploited in diffusive wave scattering methods, see E. Pike, J. Abbiss, *Light Scattering and Photon Correlation Spectroscopy*, Kluwer (1997).
[5] M. Kerker, *The Scattering of Light and other Electromagnetic Radiation*, Academic Press (1969).
[6] C. Tanford, *Physical Chemistry of Macromolecules*, Wiley (1961).
[7] B.J. Berne, R. Pecora, *Dynamic Light Scattering*, Wiley (1976).
[8] K.S. Schmitz, *An Introduction to Dynamic Light Scattering by Macromolecules*, Academic Press (1992).

of radius $a$ behaves as a Rayleigh (point) scatterer when $a/\lambda << 1$, producing a scattered intensity $I$ at a distance $r$ from the sample (see also [I.7.7.8]).

$$I = \frac{16\pi^4 a^6}{r^2 \lambda^4} \left( \frac{n_1^2 - n_2^2}{n_1^2 + n_2^2} \right)^2 \qquad [2.3.1]$$

This is Rayleigh's famous equation. $I$ is independent of the scattering angle, and increases with $a^6$, because the scattered field amplitude (the square root of the intensity $I$) is proportional to the polarizability and, hence, to the volume of the sphere. Note also the well-known Rayleigh law $I \sim \lambda^{-4}$, which accounts for the blue sky and also for the bluish appearance of sols of small particles. For a sphere with radius $a$ comparable with $\lambda$, the electric fields scattered from different regions of the sphere have different phases. The resulting interference decreases the intensity measured by the detector relative to the intensity $I$, according to [2.3.1]. This reduction can be accounted for by a factor $P(q)$,

$$I_{RGD} = I \times P(q) \qquad [2.3.2]$$

Here, 'RGD' refers to the *Rayleigh-Gans-Debye* limit (sec. I.7.8d)

$$(n_1 - n_2)a/\lambda << 1 \qquad [2.3.3]$$

where the incident light is hardly distorted by the sphere. Further, $P(q)$ is the *form factor*, which for homogeneous spheres is given by

$$P(q) = \left[ \frac{3[\sin(qa) - qa\cos(qa)]}{(qa)^3} \right]^2 \qquad [2.3.4]$$

Note that the form factor is normalized such that $P(0) = 1$, whereas $P(q) < 1$ for finite values of the scattering vector

$$q = (4\pi/\lambda)\sin(\theta/2) \qquad [2.3.5]$$

The zero values of the form factor occur when $\tan(qa) = qa$ with roots $qa = 4.493$, 7.725, etc. So, the location of minima in the angular scattering profile directly provides the sphere radius. This determination is only accurate in the *RGD* limit for sufficiently monodisperse spheres because polydispersity washes out the details of the form factor. Conversely, one can conclude from sharp intensity minima that the spheres under study must be quite monodisperse. It should be noted that even for monodisperse spheres, the form factor minima may fade due to multiple scattering if the sol is not sufficiently dilute. A well-known method to determine a sphere radius at small scattering angles is to employ the so-called *Guinier approximation*[1]

---

[1] A. Guinier, G. Fournet, *Small-angle Scattering of X-Rays*, Academic Press (1955).

$$P(q) = \exp\left[-qa^2/5\right] \qquad [2.3.6]$$

Equation [2.3.4] reduces to [2.3.6] for sufficiently small values of $qa$. Note that this method only requires relative intensities because the particle radius is obtained from the initial slope of a Guinier plot of $\ln I(q)$ against $q^2$. The *Guinier radius*, obtained from [2.3.6], is quite sensitive to deviations from monodispersity because of the strong $a$-dependence in [2.3.1 and 6]. The average scattered intensity for polydisperse non-interacting spheres in the Guinier region is proportional to

$$I_{RGD} \sim \left\langle a^6 \exp\left[-q^2a^2/5\right]\right\rangle \qquad [2.3.7]$$

where the angular brackets denote a number average as defined in 2.3f. We can rewrite [2.3.7] to

$$I_{RGD} \sim \langle a^6\rangle \exp\left[-q^2a_G^2/5\right] \qquad a_G^2 = \langle a^8\rangle/\langle a^6\rangle \qquad [2.3.8]$$

where the occurrence of the so-called Guinier radius $a_G$ shows that large spheres contribute heavily to the averaged SLS particle size. The apparent radius can be converted to the number-averaged radius $\langle a\rangle$

$$a_G^2 \cong \left(1+13s_a^2\right)\langle a\rangle^2, \quad \text{for } s_a^2 << 1 \qquad [2.3.9]$$

Here, $s_a$ is the relative polydispersity defined in app. 1, which is assumed to be small. This assumption also requires the absence of particle aggregates and contaminants, such as dust and air bubbles, which strongly contribute to the scattered intensity at small $q$. If these requirements are met, Guinier plots may be extrapolated to $q = 0$ to obtain the molar mass $M$ of the colloids. For $N$ identical scatterers in a total scattering volume $V$, we can rewrite [2.3.1] to

$$I = M\frac{9\pi^2 Vc}{r^2\lambda^4 N_{Av}\rho^2}\left[\frac{n_1^2 - n_2^2}{n_1^2 + 2n_2^2}\right]^2 \qquad [2.3.10]$$

where $c$ is the weight concentration of the particles with mass density $\rho$ and $N_{Av}$ is Avogadro's number. This form shows that the forward scattering intensity indeed provides a molecular mass since the other parameters can, in principle, be measured. This method, of course, necessitates absolute scattering intensities; Huglin explains calibration procedures[1]. The effect of polydispersity on the absolute SLS intensity at $q = 0$ can be found from a generalization of [2.3.10]. The molecular mass is a weight average and the corresponding apparent radius equals $a_{app}^3 = \langle a^6\rangle/\langle a^3\rangle$, which can be simplified further using the moment expansion from appendix 1.

Inspection of [2.3.10] makes clear that the scattering intensity can be reduced by

---

[1] M.B. Huglin, *Light Scattering from Polymer Solutions*, Academic Press (1972).

lowering the optical contrast, i.e. the difference between the refractive indices $n_1$ of the colloids and $n_2$ of the solvent. For perfectly homogeneous particles the scattered intensity vanishes at zero contrast, whereas inhomogeneous colloids have a residual scattering intensity even when the solvent matches their average refractive index. Measuring the intensity as a function of the solvent refractive index, also referred to as *contrast variation*[1], is useful to characterize the internal structure of colloidal particles. Contrast variation is, in particular, sensitive to refractive index changes at the surface of the colloids due to, for example, grafted or adsorbed polymers, as discussed in detail elsewhere[2].

*(ii) Dynamic Light Scattering (DLS)*

Particle characterization by SLS relies on time-averaged scattering, so it is immaterial whether the suspended particles are stationary or not. To characterize the Brownian motion of the colloids by a determination of the diffusion coefficient in [2.2.10], one can employ the fluctuations of scattered light in time using dynamic light scattering (DLS). The essence of a DLS experiment has been explained in secs. I.7.6c,d, 7 and 8, see also fig. I.7.10. Here we briefly summarize the method. The coherent light of a laser, illuminating a periodic grid, produces a static diffraction pattern on a screen, but when the grid is replaced by a colloidal suspension the pattern changes continually. We observe flickering bright spots due to constructive interference of light scattered by individual colloids and dark patches manifesting destructive interference. The time-dependent intensity fluctuations in this *speckle* pattern are, of course, caused by perpetual Brownian motion, and it is clear that somehow the dynamics of these fluctuations contain information on the translational diffusion coefficient $D$. One way to harvest this information is to determine the *time correlation function* $C_f(t)$ of the scattered light field[3]. This function comprises a characteristic time $t_q$ needed for a significant change in the speckle pattern for a given value of the wave vector $q$. For times $t << t_q$, the intensity pattern has not significantly decayed, whereas, for $t >> t_q$ the speckles are uncorrelated. The characteristic time is defined as

$$t_q = 1/Dq^2 \qquad [2.3.11]$$

which can be interpreted as the typical fluctuation time of the speckle pattern at the detector, or roughly the time taken by a particle to freely diffuse a distance $q^{-1}$, in accordance with Einstein's law for quadratic displacement by Brownian motion. For the simplest case of a sol of identical, non-interacting spheres a DLS experiment

---

[1] Contrast variation is also a versatile method for X-rays and neutrons, see e.g. R. Ottewill, in *Colloidal Dispersions*, J.W. Goodwin, Ed., *Roy. Soc. Chem.* (1982).
[2] A. van Helden, A. Vrij, *J. Colloid Interface Sci.* **76** (1980) 418.
[3] Time correlation functions were introduced in I.app. 11.

yields[1)]

$$C_f(t) = \exp\left[-t/t_q\right] \qquad [2.3.12]$$

i.e. a single exponential function of the correlation time $t$, which provides $t_q$ and ultimately the hydrodynamic sphere radius $a_h$ via the Stokes-Einstein equation, $D = kT/6\pi\eta a_h$. This radius is usually larger than the actual particle radius due to factors that slow down diffusion, such as the presence of an electrical double layer or adsorbed (solvent, surfactant or polymer) molecules. The friction factor $f$ determined via DLS measurements can also be used, in combination with ultracentrifugation, to determine the molar mass of the colloids. This classical procedure, further explained in sec. 2.3d, is quite general since the particle shape need not be specified. For polydisperse, non-interacting spheres, [2.3.12] is generalized to[2)]

$$C_f(t) = \frac{\langle I \exp[-t/t_q] \rangle}{\langle I \rangle} \qquad [2.3.13]$$

where the brackets denote an average over the distribution of particle radii. It is seen that $C_f(t)$ is now a sum of exponentials, each weighted by the intensity scattered by the pertinent species; for RGD spheres $I$ is given by [2.3.2]. When the distribution is so narrow that the delay times $t_q$ in [2.3.11] are close, one can expand the exponential about a mean value to find

$$\ln C_f(t) = -D_{app} q^2 t + O\left(q^4 t^2\right) + \ldots \qquad [2.3.14]$$

with an apparent diffusion coefficient

$$D_{app} = \frac{kT}{6\pi\eta a_h} \qquad a_h = \langle a^6 \rangle / \langle a^5 \rangle \qquad [2.3.15]$$

So, from a fit of the logarithm of the measured correlation time to [2.3.14], we obtain at small $t$ a hydrodynamic radius $a_h$, which can be converted to a number average using the moment expansion, discussed in appendix 1.

Our rudimentary sketch of DLS on dilute sphere suspensions neglects many complicated but important issues, such as particle interaction at finite concentration through long-range electric and hydrodynamic forces and scattering by non-spherical colloids or flexible polymers, which have extra terms in $C_f(t)$ due to rotational and internal motions. When it is known *a priori* (from electron microscopy) that the colloids under study are spherical and that they are non-interacting RGD scatterers, one obtains a hydrodynamic radius with an accuracy of a few percent, if all works well. A significant wave vector-dependence of the apparent diffusion coefficient obtained

---

[1)] K.S. Schmitz, *loc. cit.*; P. Pusey, R. Tough, in *Dynamic Light Scattering and Velocimetry*, R. Pecora, Ed., Plenum (1982).

[2)] P.G. Cummins, E.J. Staples, *Langmuir* **3** (1987) 1109.

from [2.3.14] may harbour a variety of aggregation or interaction effects, which are studied in various monographs[1].

### 2.3c *Surface area*

The specific surface area $A_g$ of colloidal particles is an important characteristic for many applications, for instance in catalysis and adsorption. It also determines the rate at which particles can be removed afterwards by filtration using [2.2.67]. For a particle with volume $V$ and mass density $\rho$, the specific area is defined as:

$$A_g = A/\rho V \qquad [2.3.16]$$

For a given amount of mass the sphere has the minimum surface to volume ratio; any shape deformation at constant volume increases $A_g$. For anisometric colloids, such as clay platelets or vanadiumpentoxide fibres, $A_g$ is largely determined by the particle thickness (see fig. 1.d). For sufficiently thin platelets or long fibres, the length (distribution) and detailed shape is irrelevant. Only when the relevant particle dimension in fig. 2.10 is inhomogeneous, will dispersity affect the value of $A_s$. We generalize [2.3.16] to

$$A_s = \frac{\sum_{i=1}^{m} N_i A_i}{\rho \sum_{i=1}^{m} N_i V_i} = \frac{\langle A \rangle}{\rho \langle V \rangle} \qquad [2.3.17]$$

where $N_i$ is the number of particles with area $A_i$; the brackets denote a number average. For polydisperse spheres, it follows that the specific surface area is given by

$$A_g = 3/\rho a_s \qquad a_s = \langle a^3 \rangle / \langle a^2 \rangle \qquad [2.3.18]$$

The apparent sphere radius $a_s$ can be estimated from the sphere dispersity as will be explained in app. 1. Liquid permeability measurements have been widely applied to determine $A_g$ for spheres[2] and non-spherical particles up to the extreme aspect ratios encountered in paper and fibrous media[3]. Carman introduced using [2.2.67] to obtain the surface area of powders, and found that the method is not affected in accuracy if the powder contains mixed sizes of particles and particles of irregular shape. A liquid permeability (or filtration rate) measurement, however, becomes impractical for colloids in the nanometer size range because this would require very high hydrostatic pressures, as discussed in 2.2h. One option here is to dry the sol and determine the surface area by the well-known BET method introduced in sec. II.1.5f. For very small particles, it should be noted that techniques of gelling and drying a sol generally

---

[1] K.S. Schmitz, *loc. cit.*; R. Pecora, *Dynamic Light Scattering: Application of Photon Correlation Spectroscopy*, Plenum (1983).
[2] D. Thies-Weesie et al., *loc. cit.*
[3] G. Jackson, D. James, *Can. J. Chem. Eng.* **64** (1986) 364.

produce some area loss by coalescence at particle-to-particle contacts and/or sintering.

For aqueous silica sols the specific surface area can be determined, following Sears[1], by measuring the amount of alkali adsorbed from solution as the pH is raised from 4 to 9. To increase the amount of adsorbed base and to eliminate any effect of unintentional small amounts of electrolytes in samples, the titration is carried out in a saturated sodium chloride solution (about 200 g NaCl/litre). Sears standardized this method using a number of silica powders with a specific surface area known from BET ($N_2$) adsorption, which allows a direct conversion of titer volume to specific surface area. Of course, for this conversion any other base-consuming species must be removed or corrected for. Silica is remarkably stable at such high ionic strength, as long as the pH is low and only when sufficient 0.1 N sodium hydroxide is added, such that pH ~ 8 sols start to slowly coagulate and become turbid (see sec. 3.13a). However, hydroxyl groups from the NaOH are still able to reach all surface silanol groups in the fresh particle aggregates, so coagulation does not affect the outcome of the Sears titration. The method, also used in fig. 2.8 to monitor the decrease of surface area in time due to Ostwald ripening, is routinely applied in industry as a rapid area check for freshly prepared sols[2]. It would be interesting to know whether sols of other inorganic colloids can also be rapidly characterized with a standardized acid-base titration.

The Sears method is just one member of a family of adsorption techniques to determine surface areas, employing adsorption of ions, nitrogen, water vapour and organic dyes, such as methylene blue. Some illustrations can also be found in sec. II.2.7c. Information on this family is easy to locate in the literature[3]. Mercury porosimetry for porous surfaces has been introduced in sec. II.1.6b; the method can also be used for particle size analysis[4].

Lastly, one often-overlooked aspect should be mentioned here and that is the effect of surface roughness, which is disregarded in the scaling $A_g \sim a^{-1}$. For compact spheres with a surface fractal dimensionality $d_f$, the specific surface area scales as[5]

$$A_g \sim \lambda^{2-d_f} a^{d_f-3} = \frac{1}{a}\left(\frac{a}{\lambda}\right)^{d_f-2} \qquad d_f \geq 2 \qquad [2.3.19]$$

where $\lambda$ is the diameter of the probing molecule, which is used to measure the surface area. The particle size dependence of the specific surface area is only reciprocal for a smooth object ($d_f = 2$), whereas the dependence is weaker for fractal surfaces and even disappears for $d_f = 3$. Soil particles form a well-documented example of the

---

[1] G.W. Sears, *Anal. Chem.* **28** (1956) 12, 1981. This is one of the oldest illustrations of a colloid titration.

[2] K. Andersson, B. Larsson, and E. Lindgren, *Silica Sols and Use of the Sols*, US Patent 5, 603, 805 (1997).

[3] S.J. Gregg, K.S.W.Sing, *Adsorption, Surface area and Porosity*, Academic Press (1982).

[4] D.M. Smith, D.L. Sternmer, *Powder Tech.* **53** (1987) 23.

[5] M. Borkovec, Q. Wu, G. Degovics, P. Laggner, and H. Sticher, *Colloids Surfaces* **A73** (1993) 65.

effect of surface roughness. Their surface dimension is close to $d_f = 2.4$, as follows from several independent surface area studies[1]. Clearly, for an accurate characterization of surface area and data interpretation, information on the surface structure is needed. Here it is convenient to employ SAXS, where at high magnitudes of the scattering vector $q$ the surface dimension follows from a log-log plot of scattering intensity versus $q$ (*Porod's law*[2]).

### *2.3d Sedimentation*

The settling of colloids under gravity or in a centrifuge is a rich (but surprisingly little consulted) source of information on their size, shape, and interactions. The equipment varies from an analytical ultracentrifuge, which records a sedimentation-diffusion equilibrium profile with a high resolution to a vessel for studying settling under gravity. In the latter case, the descent of the boundary between supernatant liquid and settling sol is measured. The observation of an initially sharp boundary, which gradually spreads in time, may manifest polydispersity, back-diffusion of the particles, or both. A boundary, which stays sharp, is consistent with the settling of monodisperse particles (with negligible diffusion), but certainly not proof of it. Even a polydisperse system may produce a sharp boundary due to a strong decrease of the settling rate with increasing concentration. Particles at the low concentration side of a boundary then catch up with the slower moving colloids in the high concentration region. The possibility of such a self-sharpening boundary necessitates additional tests before it can be concluded that a sample is monodisperse.

It is often thought that the presence of several sedimenting boundaries ('layered sedimentation') manifests a mixture of particles, which is fractionated during the sedimentation process. Layered sedimentation, however, may occur in any system due to small temperature gradients that induce convective rolls[3]. Convection is suppressed by letting the settling proceed in narrow tubes or capillaries, but in larger vessels convection must be expected to occur unless strict temperature control is applied. We also note that layered sedimentation may manifest a thermodynamic demixing (see chapter 5), in which gravity pulls different phases apart. If the descending boundary provides the sedimentation velocity $v(t)$ of non-interacting colloids, we can obtain their mass $m$ on the basis of Newton's second law

$$m\frac{\mathrm{d}}{\mathrm{d}t}v(t) = (m - m_0)g + f\,v(t) \qquad [2.3.20]$$

Here, $m_0$ is the mass of displaced solution or sol, $g$ is the acceleration of gravity and $f$ is the hydrodynamic friction factor. The effective colloid mass can also be written as

---

[1] M. Borkovec et al., *loc. cit.*

[2] For a review of small-angle scattering by fractal systems see P.W. Schmidt, *J. Appl. Crystallogr.* **24** (1991) 414.

[3] D. Mueth, J. Crocker, E. Esipov, and D. Grier, *Phys. Rev. Lett.* **77** (1996) 578.

$m - m_0 = m(1 - \rho_0 V_p)$ for colloids with partial specific volume $V_p$ in a solution with mass density $\rho_0$. For rigid (inorganic) colloids the inverse mass density usually is a good measure of the specific volume, but this is not so for drainable, porous particles or polymers. For a particle, which is initially at rest relative to the solvent, the solution of [2.3.20] is

$$v(t) = v[1 - \exp(-t/\tau)] \qquad \tau = m/f \qquad [2.3.21]$$

where

$$v = \frac{(m - m_0)g}{f} = \frac{m}{f}\left(1 - \rho_0 V_p\right)g \qquad [2.3.22]$$

is the stationary sedimentation velocity reached when the particle weight and the frictional force $fv(t)$ exactly balance. To find the effective colloid mass $m - m_0$ we need the friction factor, which can be obtained from the diffusion coefficient $D = kT/f$ measured in a separate, dynamic, light-scattering experiment. Then, the colloid mass follows from the *Svedberg equation*:

$$s = \frac{v}{g} = \frac{(m - m_0)D}{kT} \qquad [s] \qquad [2.3.23]$$

in which we have also introduced the *sedimentation coefficient* $s$, a mobility defined as the sedimentation velocity per unit of the applied acceleration, either from gravity or a centrifuge. The quantity $s$ is actually the viscous relaxation time of a particle with mass $m - m_0$, i.e. the time taken by the particle to dissipate its kinetic energy when the acceleration is switched off. This, of course, is very similar to the time $\tau$ in [2.3.21] needed to reach a stationary state. A typical value is $\tau \sim 5 \times 10^{-9}$ s for a silica particle with radius $a$ = 100 nm sedimenting in water. Hence, there is clearly no need to worry about inertia in a sedimentation experiment. The Svedberg equation [2.3.23] is valid for particles of arbitrary shape. Instead of a measured diffusion coefficient, one can also insert a theoretical friction factor in [2.3.22] when the shape of the colloids is known. Results are available for oblates[1], prolates, rods and a variety of other non-spherical particles. We only quote here the well-known outcome for spheres, also known as the Stokes value of the sedimentation coefficient

$$s = \frac{2}{9}\frac{\rho - \rho_0}{\eta}a^2 \qquad [2.3.24]$$

Here, $\eta$ is the solvent viscosity and $\rho - \rho_0$ the mass density difference between particle and solvent. When sedimentation coefficient *and* molecular mass are known, one directly obtains the friction coefficient of the colloidal particles. This provides only limited information about their shape. From the specific colloid volume $V_p$ we calculate a particle volume $mV_p$, and if we assume that the colloid is a sphere with

[1] J. Happel, H. Brenner. *Low Reynolds Number Hydrodynamics*, Prentice-Hall (1965).

radius $a$, Stokes law $f = 6\pi\eta a$ predicts the *minimum* value of the friction coefficient of the colloid in question. A larger, experimental, friction factor may be due to a hydration layer or a deviation from the spherical shape. The effect of the shape is modest for nearly spherical colloids; when a sphere is deformed at constant volume to become oblate or prolate, an aspect ratio of nearly 10 is needed[1] to increase the hydrodynamic friction by 50%. Since this increase is nearly the same for both shapes[1], it is clear that additional information is needed to extract a particle dimension or shape from the hydrodynamic friction factor.

*(i) Sedimentation-diffusion equilibrium*

Colloidal particles settle under the influence of gravity until a sedimentation-diffusion equilibrium is established. This equilibrium is the balance between the downward particle flux due to gravity and a back flux due to diffusion, which opposes the concentration gradient created by gravity. The equilibrium concentration profile $c(x)$ may also be seen as the isothermal balance between a gradient in osmotic pressure $\Pi$ and the particle weight per volume of sol

$$\frac{\mathrm{d}\Pi}{\mathrm{d}x} + c(x)(m - m_0)g = 0 \qquad [2.3.25]$$

Here, $x$ is the distance to the bottom of the vessel at $x = 0$. For ideal particles, for which van 't Hoff's law $\Pi = ckT$ applies, we find the exponential (or barometric) height distribution

$$c(x) = c_0 \exp\left(-x/l_g\right) \qquad l_g = \frac{kT}{(m - m_0)g} \qquad [2.3.26]$$

where $c_0$ is the particle concentration at the bottom and $l_g$ is the so-called *gravitational length*, which is a measure of the thickness of the profile ( $l_g$ is actually the average height of the colloids relative to $x = 0$ ). The equilibrium profile, in principle, provides the effective mass of the colloids. However, an accurate determination of the concentrations decay $c(x)$ is far from straightforward. Vessels should be rigorously thermostatted because the concentration profile is very susceptible to liquid convection. Convective rolls may induce layering or completely homogenize the sol, even for minute temperature gradients. Nearly inevitably, concentration effects also come into play because approaching the bottom of the vessel the concentration rises and at some point van 't Hoff's law may have to be replaced by a virial series, such as in [I.7.8.10]. At sufficiently high altitude, of course, the concentration profile approaches the exponential in [2.3.26]. For particles of known mass, it is possible to quantitatively investigate the concentration effects just mentioned. If we succeed to determine $c(x)$, the equation of state follows from [2.3.25] by the integration

---

[1] See K.E. van Holde, *loc. cit.* p. 81

$$\Pi_p = \Pi_h + (m - m_0) g \int_p^h c(x)\,dx \qquad [2.3.27]$$

Here, $h$ is an altitude that is sufficiently high for the pressure to obey van 't Hoff's law. By changing the integral's lower boundary $p$, the pressure $\Pi_p$ as a function of colloid-concentration $c_p$ is recovered. The main experimental challenge is to find a way to determine colloid number densities. For example, monitoring the optical turbidity as a function of height has the disadvantage that for higher concentrations the signal is non-linear in the colloid concentration. Piazza and co-workers[1] employed spheres with a crystalline anisotropy, of which the number can be counted by depolarized light scattering, allowing them to retrieve the equation of state over a wide concentration range. Such a quantitative characterization of particle interactions is unfortunately unfeasible in many practical cases, and even though the sedimentation-diffusion profile is equivalent to the osmotic pressure, it often only provides qualitative information. For example, when for monodisperse spheres the (visually observed) profile is much more extended than the gravitational length $l_g$, we can at least conclude that osmotic pressures are much larger than expected from van 't Hoff's law. This may be attributed to a charge on the colloids (see below) or to significant repulsive interactions. Attractions between the particles should shrink the equilibrium profile, though attraction between the colloids may also lead to voluminous non-equilibrium gels (sec. 6.14).

The sedimentation-diffusion equilibrium is quite sensitive to the dispersity, primarily because the particle mass enters into the Boltzmann exponent in [2.3.26]; small particles are pushed to high altitudes, whereas very large particles remain in the vicinity of $x = 0$. For spheres of species $i$ the ideal equilibrium distribution is

$$c_i(x) = c_{0,i} \exp\left(-x / l_{g,i}\right) \qquad c_{0,i} = N_{tot,i} / A l_{g,i} \qquad [2.3.28]$$

We employed here the normalization that the total number of particles $i$ in the height distribution in a vessel with cross-sectional surface area $A$ is

$$N_{tot,i} = A \int_0^\infty c_i(x)\,dx \qquad [2.3.29]$$

It is obvious from [2.3.28] that heavy particles (small $l_g$) contribute mostly to the concentration at the bottom, whereas the lighter ones (large $l_g$) dominate at high altitude. For non-interacting particles, the total number density decays exponentially, with a gravitational length that provides the number-averaged colloid mass

$$l_g = kT / \left(1 - \rho_0 V_p\right) \langle m \rangle g \qquad [2.3.30]$$

---

[1] R. Piazza, T. Bellini, and V. Degiorgio, *Phys. Rev. Lett.* **71** (1993) 4267.

a result that follows from summing the forces [2.3.25] for all species $i$. A profile of the total *weight* concentration yields the corresponding weight averaged mass $m_w = \langle m^2 \rangle / \langle m \rangle$, whereas application of the Schlieren optics[1] produces the $z$-average, $m_z = \langle m^3 \rangle / \langle m^2 \rangle$. As always, the type of average depends on the experimental method used to investigate a sample. For polydisperse colloids, the various averages may differ considerably (see also appendix 1), whereas their identity is a clear proof of monodispersity.

*(ii) Sedimentation of charged particles*

In comparison to uncharged colloids, fairly little is known about the SD-equilibrium of charged particles, although it is clear that charge effects may already be substantial for ideal colloids [2]. A striking example is shown by charged colloids at low external salt concentration. The Donnan osmotic pressure for *non*-interacting colloids in this case has the limiting form

$$\Pi = (z+1)c\,kT \qquad [2.3.31]$$

where $z$ is the number of free counterions produced by each colloid[3]. It is assumed here that the counterions dominate the external salt; when sufficient salt is added, the pressure gradually decreases to $\Pi = c\,kT$. On substitution of [2.3.31] in the force balance [2.3.25], we find

$$c(x) = c_0 \exp\left[-\left(x / l_g(1+z)\right)\right] \qquad [2.3.32]$$

showing a gravitational length which, compared with the uncharged state, has been increased by a factor $(1+z)$, which is quite substantial since $z$ may be of order 1000. The physical meaning of the $(1+z)$ term is that the practically weightless counterions tend to form a homogeneous distribution for entropic reasons, whereas the colloids are pulled down by gravity. Electroneutrality, however, couples colloids and counterions and the net result is an increase in the colloidal gravitational length. The 'entropic lift' due to counterions is actually equivalent to a homogeneous electric field, which is inevitably present in an equilibrium density profile of charged particles, and reduces the effective colloid mass, as discussed elsewhere in detail[2,4]. Thus, to determine the mass (i.e. the gravitational length) of charged colloids, sufficient salt should be added such that the Boltzmann profile reduces to [2.3.26].

---

[1] K.E. van Holde, R.L. Baldwin, *J. Phys. Chem.* **62** (1958) 734.

[2] R. van Roij, *J. Phys. Condensed Matter* **15** (2003) S3569; A.P. Philipse, *J. Phys. Condensed Matter* **16** (2004) S4051.

[3] Extensive discussions on the fraction of counterions that is free will follow in chapter III and V chapter 2.

[4] M. Raşa, A. Philipse, *Nature* **429** (2004) 860. See also R. van Roij, *loc. cit.*

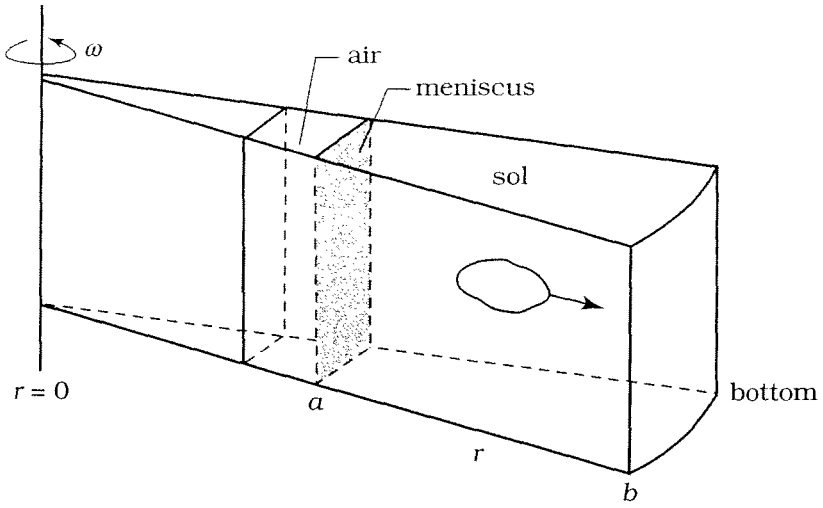



**Figure 2.11.** Schematic of an ultracentrifugation experiment (not to scale). The colloids move radially to the bottom of the sector-shaped cell with an apparent weight $(m - m_0)\omega^2 r$ at a distance $r$ from the axis, which rotates at an angular velocity $\omega$.

*(iii) Analytical ultracentrifugation*

Characterization of colloids via settling or sedimentation-diffusion equilibrium under gravity is only possible for a restricted class of particles, which have a suitable value of gravitational length in the range of mm to cm. Also, in view of the mentioned convection and detection problems, an analytical ultracentrifuge is an important, if not indispensable, characterization tool. There is an extensive, mainly biomolecular literature, on centrifugal analysis[1]. We will briefly discuss the methods to determine a colloid mass.

A spinning rotor exerts a centripetal force on the sedimentation cell, which is directed towards the rotation axis. The corresponding centripetal acceleration of the cell at a distance $r$ from this axis is $a = \omega^2 r$, where $\omega$ is the angular rotor velocity in radians per second. The colloids move towards the bottom of the cell (fig. 2.11), experiencing an effective weight increase, which is completely equivalent to an enhancement of the gravitational acceleration from $g$ to $\omega^2 r$; the colloids at some position $r$ cannot judge whether their weight is due to a centrifugal field or to gravitational pull. The Svedberg equation [2.3.23] remains, therefore, exactly the same, with the sedimentation coefficient $s = v/\omega^2 r$. The determination of $s$ is as follows. Suppose the boundary between sol and the supernatant moves at a rate $v = \mathrm{d}r_b/\mathrm{d}t$. Integration of $\omega^2 rs = \mathrm{d}r_b/\mathrm{d}t$ yields

$$\ln\frac{r_b(t)}{r_b(t_0)} = \omega^2 s\left(t - t_0\right) \qquad [2.3.33]$$

---

[1] *Analytical Ultracentrifugation* in *Biochemistry and Polymer Science*, S. Harding, A. Rowe, and J. Horton, Eds., *Roy. Soc. Chem.* (1992).

where $r_b(t)$ is the position of the boundary at time $t$. The sedimentation coefficient, therefore, follows from a graph of the logarithmic term in [2.3.33] versus $(t-t_0)$. The boundary, of course, does not remain infinitely sharp as it traverses the cell because of diffusional spreading. Then, the question is[1] which point should be used as $r_b$ in [2.3.33]. This point turns out to be the second moment of the curve for the concentration gradient

$$r_b^2 = \int r^2(\partial c/\partial r)\mathrm{d}r \Big/ \int (\partial c/\partial r)\mathrm{d}r \qquad [2.3.34]$$

where both integrations include the boundary, i.e. from a position in the homogeneous solvent to a position in the plateau region in the homogeneous sol. To determine the colloid mass from a sedimentation equilibrium profile, one uses a rotor speed, which is smaller than that used for a velocity experiment; packing of all colloids near the bottom of the cell has to be avoided. Instead, it is desirable to achieve a profile, which is sufficiently extended for data fitting, in particular of the dilute tail of the profile where colloidal interactions are insignificant. The ideal profile follows from the centrifugal force $F = (m-m_0)\omega^2 r$, corresponding to the potential energy of a colloid at position $r$

$$-\int_a^r F\mathrm{d}r = (m-m_0)\omega^2\frac{1}{2}(a^2-r^2), \qquad [2.3.35]$$

relative to the meniscus at $a$. The Boltzmann distribution for ideal particles is, therefore,

$$c(r) = c(a)\exp\left[\frac{r^2-a^2}{2\lambda^2}\right] \qquad \lambda^2 = \frac{kT}{(m-m_0)\omega^2} \qquad [2.3.36]$$

Note the analogy with the barometric height distribution [2.3.26]. The thickness of the profile, set by the length $\lambda$, can be adjusted by changing the rotor speed $\omega$. A graph of $\ln c(r)$ versus $r^2$ will yield the length $\lambda$ and, therefore, the effective mass of the colloids. This mass determination, which is in principle quite accurate, has been fruitfully (and frequently) checked for monodisperse biomolecules (proteins, viruses, DNA fragments); molecular masses generally match the values known from elemental compositions very well. The existence of extensive literature on data analysis and instrumental issues[2] shows that, nevertheless, for most investigators the analytical ultracentrifuge is anything but a simple black box, just as the engine of a car is for most drivers.

### *2.3e Other methods*

For practical dispersions such as paints or ceramic suspensions, their application will largely determine the choice of characterization techniques, in addition to those

---

[1] See for a pointed discussion: K.E. van Holde, *Physical Biochemistry*, Prentice Hall, (1971).
[2] S.E. Harding, A.J. Rowe, and J.C. Horton, *loc. cit.*

mentioned in previous sections. For inorganic colloids processed to an eventually dried compact, as in ceramic shaping techniques, one can largely appeal to the usual methods of powder technology. These methods include X-ray analysis to identify crystalline components, mercury-intrusion to measure porosities of 'green' or sintered bodies, and thermal analysis to investigate temperature-dependent properties. The last mentioned analysis comprises, among other things, differential thermal analysis (DTA) and thermal gravimetry (TG). DTA exposes a material to a controlled temperature increase as a function of time and records release or uptake of heat due to phase transitions (including melting points or melting trajectories), chemical reactions, and any other endothermic or exothermic process. TG monitors the weight of the sample in the course of the temperature-time scan and detects, for example, the loss of water which was adsorbed on particles or generated by condensation of hydroxyl groups, as are often found on oxidic materials.

A combination of DTA and TG is certainly also useful for inorganic model colloids, for example to determine the weight fraction of organic material due to a leftover of a surface modification (see sec. 2.i). The latter will produce an endothermic peak and simultaneous weight loss roughly in the range 400-600°C, the temperature range over which organic molecules are burnt off. In addition, physically adsorbed water will be detected as an endothermic loss already below 100°C, and release of water or other low molecular solvents at higher temperatures is indicative of porous colloids with internal silanol or alkoxy groups, as occur in the silica spheres prepared by the Stöber process. Exothermic peaks at temperatures around 1000°C or higher may manifest any of the many phase transitions found in alumina and silica containing (clay) materials[1]. Thus, DTA-TG, in combination with chemical analysis results for elemental percentages, contributes to a fairly complete material picture of colloidal particles. Spectroscopy (NMR, Infrared, etc.), of course, provides even more chemical detail on colloids and their surface coverage.

One important 'application' of model colloids is their use in critical test of theories. Then the primary concern is not so much knowledge of the chemical composition of colloids (useful as it may be), but rather the surface parameters which appear explicitly in the theory under study. For charged colloids these are in any case the double layer parameters. Two of such parameters offer themselves, the surface charge density $\sigma^o$ and the electrokinetic potential $\zeta$. The former follows from colloid titrations when the charge-determining mechanism is known, the latter from electrokinetics. Principles and elaborations can be found in various places of FICS, especially sec. I.5e (titration), chapter II.3 (composition of double layers), and chapter II.4 (measurement and interpretation of electrokinetic potentials). It is good to keep in mind that by titration and electrokinetics very different double layer parameters are measured. In fact, for a full characterization of the double layer composition both techniques should be simul-

[1] F.H. Norton, *Fine Ceramics*, Krieger (1987).

taneously applied to the same system. Only in this way can the composition of the inner part of the double layer be established. We note that the difference between the point of zero charge (p.z.c.) and the isoelectric point (i.e.p.) is a measure of specific adsorption (sec. II.3.8). In many cases only $\zeta$-potentials are available. Experience has shown that for situations of not too strong double layer overlap these potentials are satisfactory characteristics to be substituted in equations for the Gibbs energy of interaction (chapter 3).

For sterically stabilized particles, information about the amount and distribution of attached polymers is needed, see chapter V.1. Often one is interested in measuring a concentration dependence, which brings on a characterization problem that is often swept under the carpet, namely the issue of the specific particle volume, which here deserves some more discussion.

A theoretical concentration dependence is usually expressed in terms of particle volume fractions, whereas one measures, say, a diffusion coefficient or low-shear viscosity, as a function of colloid *weight* concentrations. How should they be converted to volume fractions? Clearly, a measurement of the mass density or specific volume of the colloids is needed. For rigid hard spheres, one option is to measure the intrinsic viscosity and to find the specific volume that produces agreement with Einstein's value of 2.5 for the coefficient of the volume fraction. (Viscosity is a better option here than sedimentation or diffusion, because the volume fraction enters on the level of single, non-interacting particles). Factors such as porosity, softness and surface charge of particles and deviations from the spherical shape, can be couched into an effective specific volume which matches the Einstein result. For further information on the viscosity of particulate matter, see secs. 6.9, 10 and 13. However, this procedure yields an effective *hydrodynamic* volume fraction which may be inappropriate for equilibrium measurements such as the osmotic pressure or the static structure factors from light scattering. To find thermodynamic volume fractions one can also choose the specific volume such that the colloidal hard spheres start to freeze at the theoretically expected volume fraction, a procedure which, of course, is only feasible for the limited class of colloids which form colloidal crystals. We note here that such crystals in principle produce the particle mass from the location in reciprocal space of Bragg peaks, analogous to the counting of atoms in a unit cell in X-ray diffraction.

Direct measurement of particle mass densities in solution by weighing dispersions as a function of concentration requires more material than is often available in the case of model colloids. Commercial equipment is available for this weighing on a small scale by measuring the resonance frequency of capillaries filled with dispersions. The latter method is very accurate for pure liquids, but for dispersions prone to uncertainties due to, among other things, the sensitivity to details of cleaning procedures of measuring cells. An alternative is using quartz crystal microbalances (QCM's). Sedimentation profiles from ultracentrifugation (see sec. 2.3d) provide the *buoyant* mass and therefore still require a separate specific volume measurement. For biomolecules centrifuga-

tion in a salt gradient is employed: the molecules stay suspended at a height at which their buoyant density is exactly matched by the salt solution. This method, which provides an accurate and well-defined thermodynamic specific volume, is suitable for mass densities below about 1.8 g/cm$^3$, the maximum density of the salt solution (usually $CsCl_2$). This mass-density range includes polymer colloids, but excludes many inorganic particles. The latter could be handled by measuring sedimentation velocities in solvent mixtures and extrapolation to zero velocity, a method which apparently has not been exploited yet. To conclude, the conversion of weight concentration to volume fraction (or particle number density) is usually not straightforward and needs to be made explicit in the characterization of colloids under study.

### *2.3f Size distributions*

Characterization of colloidal particles is incomplete without specification of their size distribution. For this, various options are available, including the ultracentrifuge, a method discussed by Harding et al.[1]. Advantages of light scattering methods include measurement speed and the very large number of particles that are sampled. The procedure, however, is far from simple. The main problem is the inversion of the measured field autocorrelation function [2.3.13] to obtain the intensity-weighted contribution of each particle species. This inversion has no unique solution when the measurements are contaminated by noise[2]. In addition, many subtleties in sample preparation and data analysis need to be addressed, as discussed extensively by Provder[3].

The direct determination of a large number of diameters by electron microscopy is accurate and simple, in particular for inorganic colloids, which usually maintain their integrity during drying on a grid and exposure to vacuum. For easily deformable latices or emulsion droplets, other techniques such as confocal microscopy may be used (see chapter V.8). Another useful (but yet little employed) option is *cryogenic electron microscopy*, a technique discussed briefly in 2.3a. When colloids are sufficiently small, say with radii below 100 nm, quite a large number of them can be simultaneously imaged in the glassy cryo-TEM film. When the colloids are repulsive due to surface charge or a polymer coating, which has a low contrast for TEM, one may observe clearly separated particle cores[4] (c.f. fig. 2.4), which form a convenient input for image analysis software; the S-distribution of fig. 2.11 has been obtained in this manner.

Though extensive single-particle imaging is the best option to obtain a reliable size distribution without a priori assumptions about the colloids, it is not always possible in practice or convenient for routine analysis. Often one relies on fractionation methods

---

[1] S.E. Harding, A.J. Rowe, and J.C. Horton, *loc. cit.*

[2] P.G. Cummins, E.J. Staples, *Langmuir* **3** (1987) 1109.

[3] T. Provder, Ed., *Particle Size Distribution: Assessment and Characterization*, *ACS Symposium Series* **332** (1987); *Particle Size Distribution II*, ACS Symposium Series **472** (1991).

[4] A.P. Philipse, G.H. Koenderink, *Adv. Colloid Interface Sci.* **100-102** (2003) 613.

in which the distribution is broken up into classes making use of some particle property. Sieving of powders is the classical method for separation based directly on particle size; most other methods rely on the response of particles to external fields or a change in particle interactions. An example of the latter is the fractionation of iron oxide particles by repeated phase separation induced by the addition of salt, which preferentially removes the larger particles[1]. This *fractional distillation* is expected to work for any interparticle attraction, which is size-dependent. The procedure reminds one of fractionating a polymer solution by slow addition of a poor solvent upon which molecules with high molecular weights precipitate first. We will now briefly explain some fractionation methods, which employ external fields.

Magnetic particles can, in principle, be fractionated by an external, inhomogeneous magnetic field $\boldsymbol{B}$. The magnetic force on the particles is[2]

$$\boldsymbol{F} = \left(\boldsymbol{m} \cdot \vec{\nabla}\right)\boldsymbol{B} \qquad [2.3.37]$$

where $\boldsymbol{m}$ is the magnetic moment of the particle, which is proportional to the particle volume. To separate small, paramagnetic colloids, large gradients are needed. They can be produced by magnetizing a steel wool matrix; near curves and edges of the filter large gradients exist, which capture particles from the dispersion[3]. By increasing the magnetization of the matrix, fractions with increasingly smaller particles can be captured. This high-gradient magnetic separation has important applications in the removal of iron oxides from clay dispersions and wastewater. However, its potential for quantitative fractionation is much less developed than for techniques based on sedimentation.

The *disc centrifuge photosedimentometer* (DCP) separates spheres, which sediment radially outward past a detector with a velocity determined by Stokes' law. The technique appears to be robust and sufficiently accurate, for example, to resolve the various components in mixtures of standard polystyrene spheres[3]. For non-spherical colloids, the analysis (as always) is less straightforward than for spheres. We note here that for particles with high aspect ratios, the sedimentation rate is determined mainly by the smaller dimension (c.f. the surface areas in fig. 2.10). For example, for thin rods with diameter $d$ and length $L$, the (orientationally averaged) friction factor is[4]

$$f_0 = 3\pi\eta L / \ln(2L/d) + \alpha \qquad [2.3.38]$$

where $\alpha$ is a number of order unity. Consequently, the sedimentation coefficient of the thin rods is

---

[1] V. Cabuil, R. Massart, J. Bacri, R. Perzynski, and D. Salin, *J. Chem. Res. (S)* (1987) 130.
[2] J. Svoboda, *Magnetic Methods for the Treatment of Minerals*, Elsevier (1987).
[3] T. Provder, 1987, *loc. cit.*
[4] S. Broersma, *J. Chem. Phys.* **32** (1960) 1632.

$$s = \frac{m}{f}\left(1 - \rho V_p\right) \approx \frac{d^2}{12\eta}\left(\rho - \rho_0\right)\ln\left(\frac{2L}{d}\right) \qquad [2.3.39]$$

So here only the distribution in diameters is of importance, which could simplify the DCP analysis. Equation [2.3.39], incidentally, warns us that fractional sedimentation is not useful to decrease the polydispersity in length of rods or width of platelets.

*Sedimentation field-flow fractionation* (sedimentation FFF) fractionates particles in a flow channel with a field acting perpendicular to the stream direction[1]. The (centrifugal) field forces particles to accumulate at one wall of the channel where the viscous drag is large so that downstream displacement of particles is retarded. The distance to the wall depends on the particle size, which leads to size fractionation in the flow direction. The method is quite sensitive and mixtures of well-defined spheres can be analyzed with good resolution. For information on still another fractionation method, hydrodynamic chromatography, we also refer to Provder[2], whose analyses include a comparison of the various particle characterization methods applied to one and the same series of monodisperse PMMA latices.

Once a sufficiently large number of particles have been sampled in each fraction, it may be useful to compare the result with one of the standard mathematical distribution functions, some of which are given below. For a continuous distribution the $n^{th}$ moment is defined as

$$\langle a^n \rangle = \int_0^\infty a^n P(a)\,da \quad ; \quad \int_0^\infty P(a)\,da = 1 \qquad [2.3.40]$$

Here, $P(a)$ is the normalized probability distribution for the radius $a$ and $P(a)da$ is the probability for a radius to be in the interval $a, a + da$. Note that $P(a)$ has the dimensions of reciprocal distance, which is why it is also called the *probability density*. The normal (or Gauss) probability density has the familiar, bell-shaped function and obeys

$$P(a) = \frac{1}{\sigma_a\sqrt{2\pi}}\exp\left[-\left(a - \langle a\rangle\right)^2 / 2\left(\sigma_a\right)^2\right] \qquad [2.3.41]$$

in which $\langle a \rangle$ is the number-averaged radius and $\sigma_a$ is the (*absolute*) *standard deviation* defined by

$$\sigma_a^2 = \left\langle \left(a - \langle a\rangle\right)^2 \right\rangle = \langle a^2\rangle - \langle a\rangle^2 \qquad [2.3.42]$$

which should not be confused with the *relative polydispersity* $s_a$ defined through $s_a^2 = \sigma_a^2 / \langle a\rangle^2$ in appendix 1. Fairly narrow distributions, as for silica and latex spheres, are often fitted reasonably well with a Gauss model as illustrated for silica in

---

[1] J.C. Giddings, F.J.F. Yang, and M.N. Myers, *Anal. Chem.* **46** (1974) 1917.
[2] T. Provder 1987 *loc. cit.*

fig. A1.1. For many other colloidal systems, however, the size distribution is asymmetric. This may be due to various factors, such as a milling process, secondary particle nucleation, the growth mechanism or the tendency of the larger particles to aggregate. For the $\gamma$-$Fe_2O_3$ colloids in a ferrofluid (fig. A1.1), the asymmetric distribution often fits a log-normal probability distribution reasonably well

$$P(a) = \frac{1}{a\sqrt{2\pi \ln z}} \exp\left[ -\frac{\left( \ln \frac{a}{\langle a \rangle} \sqrt{z} \right)^2}{2 \ln z} \right] \qquad [2.3.43]$$

where

$$z = 1 + \frac{\sigma_a^2}{\langle a \rangle^2} = 1 + s_a^2 \qquad [2.3.44]$$

in which, as before, $s_a^2$ is the relative polydispersity. The normalized radius moments of the log-normal distribution are given by[1)]

$$\frac{\langle a^n \rangle}{\langle a \rangle^n} = \left(1 + s_a^2\right)^{n(n-1)/2} \approx 1 + \frac{n(n-1)}{2} s_a^2 \qquad [2.3.45]$$

for $s_a^2 << 1$. So, from a measured or estimated polydispersity, one can compute the higher radius moments and predict the apparent radius obtained by a particular characterization method. As an example of a discrete probability distribution we mention the *Poisson distribution*

$$P(a) = \frac{\langle a \rangle^a}{a!} \exp[-\langle a \rangle] \qquad a = 0,1,... \qquad [2.3.46]$$

This probability function is skewed, but rapidly becomes more symmetrical upon increasing the average $\langle a \rangle$. The Poisson distribution is especially useful when the number of events (here, particle radii) is small. For a large number of random variables, the Poisson distribution is fairly well approximated by normal distribution[2)].

In many cases, a fit of experimental counts to a theoretical probability distribution will be poorer than in fig. A1.1. The data may simply disobey the chosen distribution, for instance, because the distribution is bimodal or the number of counts may be insufficient to draw a clear conclusion anyhow. Luckily it is possible to approximate the moments in [2.3.40] only on the basis of a measurement (or choice) of the relative polydispersity without presupposing any particular size distribution. The approximation is actually the truncated expansion [2.3.45], which is valid for *any* not too broad size distribution as shown in appendix 1.

---

[1)] P.N. Pusey, H.M. Fijnaut, and A. Vrij, *J. Chem. Phys.* **77** (1982) 4270.
[2)] A. Papoulis, *loc. cit.*

## 2.4 Examples of sol preparation

This section provides a few commented preparation methods for various inorganic sols. There are several reasons for such a presentation. Firstly, there is no substitute for learning about practical colloid chemistry than going into the lab to make your own colloids. Secondly, the examples illustrate that quite well-defined colloidal dispersions, used in state-of-the-art colloid research, can be obtained with simple methods. No chemical equipment is needed beyond what is used in a freshman chemistry course – possibly with the exception of the autoclave for the boehmite synthesis. Thirdly, the examples also illustrate that this simplicity in method may be misleading. The outcome of a colloid synthesis is often difficult to explain or to adjust in a predictable manner. Elements of art and surprise remain. The selection below is biased by the author's hands-on experience; many more synthesis examples useful for teaching or research can be found in the general references of sec. 2.5.

### *2.4a Silica sols*

Silica sols are usually prepared in aqueous solutions from waterglass or in ethanol from the precursor tetraethoxysilane (TES). The formation of colloidal silica by acidification of waterglass is extensively documented elsewhere[1,2]. Here we only describe an instructive experiment in which silica supersaturation is generated by a change of solvent, instead of a change of pH[3]. A stock sodium-silica solution ( $Na_2O \cdot SiO_2$ , 27 wt% $SiO_2$ ) is diluted with double distilled water to 0.22 wt% $SiO_2$ . Under vigorous stirring, 0.2 ml of this dilute waterglass solution is rapidly pipetted into 10 ml absolute ethanol. A sudden turbidity increase manifests the formation of small, smooth silica spheres with a diameter around 30 nm and a typical dispersity of 20-30%. The solubility of silica in ethanol is much lower than in water, and it is estimated that in this experiment the supersaturation ratio due to the alcohol addition is on the order of $S \approx O(10)$, which, in view of section 2.2b, should indeed produce very rapid homogeneous nucleation.

The preparation of so-called *Stöber silica spheres* from the precursor TES in an ethanol-ammonia mixture is well documented[4]. To obtain spheres with a radius of about 60 nm and a typical dispersity of $\sigma \approx$ 10-15%, the procedure is as follows. TES (60 ml, freshly distilled to remove any polymeric species) is injected under the liquid level of a thoroughly stirred mixture of 200 ml ammonia (25%) and 3 litres (preferably distilled) absolute ethanol in a vessel, previously thoroughly cleaned by multiple rinsing with, subsequently, distilled water and absolute ethanol. The TES solution is

---

[1] R.K. Iler *loc.cit.*

[2] K. Andersson, B. Larsson, and E. Lindgren, *Silica Sols and the Use of Sols*, US patent 5603805 (1997).

[3] P.A. Buining, L.M. Liz-Marzan, and A.P. Philipse, *J. Colloid Interface Sci.* **179** (1996) 318.

[4] W. Stöber, A. Fink, and E. Bohn, *J. Colloid Interface Sci.* **26** (1968) 62.

gently stirred in the closed vessel; after about 15-30 min., an increase in turbidity manifests the formation of silica spheres in the alcosol, which grow to their final size over a time scale of hours. Silica growth can be continued by adding small portions of TES to control the final radius[1]. This seeded growth method has the risk of introducing secondary silica nucleation, so samples should be checked with TEM. Secondary particles are usually small enough to be separated from primary spheres by repeated sedimentation. They generally do not disappear by Ostwald ripening because of the very low silica solubility in ethanol.

Stöber silica spheres can be easily silanized by surface modification with 3-methacryloxypropyl-trimethoxysilane (TPM) as follows. TPM (about 1-3 ml per gram of silica) is added to the alcosol, after which the solvent is distilled to reduce the alcosol volume by about 30%. Unreacted TPM is removed afterwards by repeated sedimentation-redispersion cycles. The non-desorbing TPM layer, with a hydrodynamic thickness of a few nm, improves the stability in various organic solvents. Silanes are also very useful to modify silica with fluorescent or phosphorescent dyes, as discussed in refs.[2,3].

With respect to storage of silica sols, the following points should be noted. Aqueous silica sols generally show aging effects; the specific surface area decreases (fig. 2.8) and the pH tends to increase, probably due to sodium hydroxide leaching. Commercial silica sols are usually quite stable, as manifested by a constant (Newton) viscosity. However, over longer periods of time (say one year) the viscosity gradually increases and space-filling gels are often formed. Stöber alcosols may aggregate in the course of time, especially for larger particles with a relatively high ammonia concentration. Removal of ammonia by bubbling nitrogen through the alcosol is one remedy. Another option is distillation together with silanization of the particle surface (see above), leading to TPM-coated silica spheres, which in absolute ethanol have practically unlimited stability.

Hydrophobic silica spheres have been used extensively for the study of hard-sphere colloids in apolar solvents[4]. A suitable surface modification in this respect is the esterification of surface silanol groups of silica spheres under vacuum distillation in a pure octadecyl alcohol melt at about 180°C. A more convenient procedure is to add a Stöber alcosol directly to an excess solution of octadecyl alcohol in triethyl phosphate[5]. After distilling all ethanol and ammonia, the solution is stirred for several days at 140-160°C under a flow of dry nitrogen. The resulting octadecyl-coated

---

[1] A. Philipse, A. Vrij, *J. Chem. Phys.* **87** (1987) 5634.
[2] A. van Blaaderen, A. Vrij, *loc. cit.*
[3] M.P. Lettinga, M. van Zandvoort, C.M. van Kats, and A.P. Philipse, *Langmuir* **16** (2000) 6156.
[4] Some illustrations follow in chapter **5.**
[5] A.M. Nechifor, A.P. Philipse, F. de Jong, J.P.M. van Duynhoven, R.J.M. Egberink, and D.N. Reinhoudt, *Langmuir* **12** (1996) 3844.

particles can be transferred to a stable dispersion in cyclohexane using sedimentation-redispersion cycles[1].

### *2.4b Sulphur sols*

An often quoted example of the formation of a monodisperse sol is La Mer's method[2,3] for preparing sulphur colloids, in which S is gradually formed by the reaction of thiosulfate with acid

$$S_2O_3^{2-} + 2H^+ \rightarrow SO_2 + H_2O + S \qquad [2.4.1]$$

This is a slow reaction, such that growth of sulphur particles occurs on a time scale of hours. La Mer's method is as follows. One (1.00) ml of 1.50 *N* $H_2SO_4$ is added to 995 ml of double distilled water in a one-liter volumetric flask that is thermostatted at 25°C. One (1.00) ml of 1.5 *N* $Na_2S_2O_3$ is added rapidly, after which the flask is quickly made up to 1 liter, mixed thoroughly and returned to the thermostat. Within 1-2 h a weak scattering can be observed from a hand laser (a Tyndall beam), manifesting growing sulphur particles. They continue to grow over a period of about 24 hours after which they settle, presumably because the sulfur colloids become quite large though colloidal instability may also play a role (the Van der Waals attractions must be substantial in this case). Sulphur growth can be stopped by titrating unreacted thiosulfate with an iodine solution in potassium iodide, according to[4]

$$I_3^- + 2S_2O_3^{2-} \rightarrow 3I^- + S_4O_6^{2-} \qquad [2.4.2]$$

Because iodine solutions have an intense yellow to brown colour, even at high dilution, iodine can serve as its own end point indicator. Titration is continued until a barely perceptibly pale yellow sol remains. By applying iodometry after various time intervals on a number of acidified thiosulfate solutions, sols with various particle sizes are obtained.

The sulphur sols are very suitable for a demonstration of the angular dependence of light scattering. When a beam of plane-polarized white light is viewed with the eye in a plane perpendicular to the polarization, spectral colours may be observed at angles, which depend on particle size[5]. This is a clear indication for a narrow size distribution of the sulphur colloids or, to be more precise, of the colloids, which dominate the light-scattering intensity; the presence of small, weakly scattering sub-particles cannot be excluded. Electron microscopy or atomic force microscopy of dried samples of sulphur

---

[1] A.M. Nechifor et al., *loc. cit.*

[2] V.K. La Mer, M.D. Barnes, *J. Colloid Sci* **1** (1946) 71.

[3] A.B. Levit, R.L. Rowell, *J. Colloid Interface Sci.* **50** (1975); for a preparation based on $H_2S$, see: G. Chiu, E.J. Meehan, *J. Colloid Interface Sci.* **62** (1977) 1.

[4] I.A. Vogel, *Textbook of Quantitative Chemical Analysis*, Longman (1989).

[5] V.K. La Mer, M.D. Barnes, *loc. cit.*

sols does not yield images of well-defined spheres (the attempt failed in the author's laboratory), but a variety of morphologies (indeed, containing elemental sulphur), including raspberry-like submicron particles composed of much smaller units. The imaging is hampered by the fact that sulphur colloids easily melt or deform; moreover, crystallizing salts in the drying TEM specimen complicate the picture. So far in the literature, making direct images of them has not supported the presumption that La Mer's method produces monodisperse spheres.

### *2.4c Boehmite and gibbsite sols*

One method to synthesize rod-like colloids employs aqueous aluminium alkoxide solutions to form elongated, crystalline AlOOH (boehmite) particles. The alkoxide is first hydrolyzed[1)] at room temperature in an aqueous HCl-solution, followed by a hydrothermal treatment at about 150°C in an autoclave. By varying the pH and the type and concentration of alkoxide, the length of the boehmite needles can be adjusted in the range of 100-400 nm; the needle thickness is 10-20 nm. The starting aluminium-alkoxides are $Al(OBu^s)_3$ (aluminium tri-*sec*-butoxide, ASB), a volatile, colourless liquid, which hydrolyzes easily due to air moisture and $Al(OPr^i)_3$ (aluminium triisopropoxide, AIP), a white powder, which is less reactive towards moisture. An aqueous HCl solution is made by pouring a concentrated HCl stock into water (never the other way around). The HCl solution must be titrated if its molarity is not precisely known.

To a stirred mixture of 2900 ml of double distilled water and 22 ml of HCl (37%), 59.8 ml ASB is added after which a white precipitate, presumably aluminum hydroxide, is formed. (If the stock ASB is not clear but yellowish, it should be purified by distillation from hydrolysis products). Next, 46.0 g AIP is added, which dissolves within a few hours. The solution is gently stirred at room temperature in a closed vessel for a week. Then the now clear solution is autoclaved for 22 h at 150°C. One option is to heat the solution in partly filled metal pressure vessels with a Teflon inner core, which are slowly rotating inside an oven, as described by Buining et al.[1)]. Commercial equipment for hydrothermal treatment is available. After this treatment, the vessels are allowed to cool to room temperature. Note that hydrolysis of the alkoxides produces alcohols, which increases the pressure in the autoclaved vessels. The Teflon inner cores, see ref.[1)], should not be removed unless they are completely cooled to room temperature, otherwise they will no longer fit in the pressure vessels.

The somewhat turbid, easily flowing boehmite dispersion is dialyzed in cellophane tubes for 1-2 weeks against demineralized water to remove alcohols and salts. The dialyzed dispersion is highly viscous due to the strong, double-layer repulsion in the now nearly salt-free dispersion. Inspection of the dispersion between crossed polarizers reveals permanent birefringence. After dilution, the dispersion exhibits streaming

---

[1)] P.A. Buining, C. Pathmamanoharan, J.B.H. Jansen, and H.N.W. Lekkerkerker, *J. Am. Ceram Soc.* **74** (1991) 1303.

birefringence, which confirms the presence of non-aggregated needles oriented by a flow field. The birefringence is destroyed by addition of some ammonia, which coagulates the boehmite particles. The boehmite dispersion, stored in a plastic bottle, may be stable over a time scale of months to even several years. Storage in glass vessels is not recommended because of possible deposition of soluble silica on the positively charged boehmite. TEM micrographs reveal somewhat irregular rods with an average length probably around 180 nm and a width close to 10 nm; the relative size dispersity is typically 30-40%. Shorter rods with a length of about 100 nm can be prepared by starting with 2850 ml water, 9.7 ml HCl (37%) and 156 ml of ASB. The reproducibility of the dimensions of the boehmite crystals is modest and details of, for example, the hydrothermal treatment may significantly affect the sizes and shapes of the final particles.

The temperature of the hydrothermal treatment greatly influences the particle morphology. At lower temperatures ($T \sim 135°C$), mainly hexagonal gibbsite platelets are observed after 22 h, which apparently recrystallize to boehmite needles at higher temperatures or longer times. These platelets also form when the alkoxide solution (see above), instead of being autoclaved, is stored for several months at room temperature. The fairly monodisperse gibbsite hexagons (typical diameter 150 nm, thickness 13-15 nm) are useful model colloids, in particular, because they can be grafted with polymers to produce organosols of uncharged platelets[1,2]. Boehmite rods have also been coated by silica[3], see also fig. 2.1c.

### *2.4d Ferrofluids*

Ferrofluids are stable colloidal dispersions of single-domain magnetic particles[4], which behave as liquid ferromagnets; the fluid moves towards a magnet and may adopt exotic equilibrium shapes[5]. Most ferrofluids are based on magnetite ($Fe_3O_4$) particles, which oxidize to maghemite ($\gamma$-$Fe_2O_3$), without losing their magnetic properties. The colloids are usually sterically stabilized by a grafted layer of oleic acid and dispersed in non-polar solvents, such as cyclohexane; aqueous sols of magnetite particles are more prone to aggregation when stored over longer periods of time. The traditional method for synthesizing non-aqueous ferrofluids consists of extensive milling of magnetite minerals in an organic solvent in the presence of adsorbing surfactants. Instead of this comminution technique, which may take weeks, a fast condensation route may be used on a laboratory scale. Here, magnetite particles precipitate upon alkalization of a $FeCl_2 / FeCl_3$ solution in what must be an instance of rapid, homogenous nucleation. Particle formation already starts before pH gradients have

---

[1] A.M. Wierenga, T.A.J. Lenstra, and A. Philipse, *Colloids Surf.* **A134** (1998) 359.
[2] F. van der Kooij, E. Kassapidou, and H. Lekkerkerker, *Nature* **406** (2000) 868.
[3] A.P. Philipse, A.M. Nechifor, and C. Pathmamanoharan, *Langmuir* **10** (1994) 4451.
[4] Magnetic pair particle interactions will be discussed in sec. 3.10c.
[5] R. Rosensweig, *Ferrohydrodynamics*, Cambridge University Press (1985).

disappeared by stirring; there is considerable overlap of nucleation and growth, which partly explains the poor control of the particle size distribution in this otherwise convenient synthesis. The following procedure is based on[1,2].

In 380 ml demineralized water, $FeCl_2 \cdot 4H_2O$ (3.29 g, 16.5 mmol) and $FeCl_3 \cdot 6H_2O$ (8.68 g, 32.1 mmol) are simultaneously dissolved. (The hygroscopic properties of anhydrous salts make it more difficult to achieve the correct $Fe^{2+}/Fe^{3+}$ ratio). Under vigorous stirring at room temperature, 25 ml ammonia (25%) is added; a dark precipitate immediately forms. This magnetic precipitate is collected with a permanent magnet and, after decantation of the supernatant, is mixed with 40 ml 2M $HNO_3$, which brings the pH below the isoelectric point of iron oxide, and repeptizes the precipitate. After 5 min. of stirring, the oxidation to maghemite is completed by adding 60 ml of an aqueous 0.35 M $Fe(NO_3)_3$ solution and subsequent refluxing of the stirred solution at its boiling point for 1 h. On a permanent magnet, the maghemite settles as a reddish sediment. After decanting the supernatant and washing the precipitate twice with 100 ml 2 M $HNO_3$ (decant the acid as much as possible), the precipitate is redispersed in 50 ml demineralized water to a stable, black maghemite sol with a typical solid weight concentration of 5-6 g/l. The maghemite particles can now be grafted with oleic acid on a small scale at room temperature. To that end, 2 ml of the aqueous sol is diluted with 50 ml demineralized water, coagulated by adding a few drops of ammonia (25%) and sedimented on a magnet. After decanting the supernatant and washing with 50 ml water, 100 ml water is added to the gently stirred precipitate, followed by the addition of 6-8 ml oleic acid. Within a few minutes, all maghemite colloids migrate into the oil phase where, after separation from the colourless aqueous phase, they are washed three times with 10 ml ethanol to remove water and any excess surfactant. After drying in a nitrogen flow, the oleic acid-coated maghemite particles are easily redispersed in a few milliliters of cyclohexane to form a stable dispersion, which can be manipulated quite effectively with a magnet. At this point, a liquid ferromagnet has been obtained.

TEM micrographs show somewhat irregular maghemite crystallites with an average diameter of typically 10 nm and a relative dispersity around $s \sim 30\%$ (see also figs. 2.2 and A1.1). Thermogravimetry and infrared measurements[2] indicate the presence of covalently bound oleic acid molecules occupying an average surface area of 0.28 $nm^2$.

## 2.5 General references

### 2.5a *Preparation*

Polymer colloids have not been addressed in this chapter. For a suitable entry to the literature on their preparation and characterization, see A. Elaissari, *Colloidal Poly-*

---

[1] A. Bee, R. Massart, *J. Magn. Magn. Mater.* **149** (1995) 6.
[2] G.A. van Ewijk, G.J. Vroege, and A.P. Philipse, *J. Magn. Magn. Mater.* **201** (1999) 31.

*mers, Synthesis and Characterization*, Marcel Dekker (2003).

The literature on inorganic colloid synthesis dates back to the beginnings of colloid science. A very useful entrance, in particular, to the older literature is H.R. Kruyt (Ed.), *Colloid Science I: Irreversible Systems*, Elsevier, (1952). Also see: J.Th.G. Overbeek, *Monodisperse Colloidal Systems, Fascinating and Useful, Adv. Colloid Interface Sci.* **15** (1982) 251-277. (A pointed review on monodisperse colloids and growth mechanisms.)

*Useful texts on nucleation and growth:*

F.F. Abraham, *Homogeneous Nucleation Theory*, Academic Press (1974).

M. Baraton, *Synthesis, Functionalization and Surface Treatment of Nanoparticles*, Americal Scientific Publishers (2003). (Deals with many aspects of particle functionalization and its applications.)

*The Colloid Chemistry of Silica*, H.E. Bergna, Ed., American Chemical Society (1994). (Reviews developments in the study of silica sols and gels since the appearance of Iler's book in 1979.)

M.A. Brook, *Silicon in Organic, Organometallic, and Polymer Chemistry*, Wiley (2000). (Overview of silicon chemistry, including many aspects of silica and functional silanes.)

R.M. Cornell, U. Schwertmann, *Iron Oxides in the* Laboratory, VCH (1991); *The Iron Oxides*, VCH (1996). (The essential text on the topic.)

C.L. De Castro, B.S. Mitchell, in *Synthesis, Functionalization and Surface Treatment of Nanoparticles*, M.I. Baraton, Ed., American Science Publishers (2003).

P.G. Debenedetti, *Metastable liquids: Concepts and Principles*, Princeton University Press (1996). (A good review of nucleation theory.)

J. Delplancke, in *Synthesis, Functionalization and Surface Treatment of Nanoparticles*, M.I. Baraton, Ed., American Science Publishers (2003).

*Colloid Gold: Principles, Methods and Applications* (three volumes), M. Hayat, Ed., Academic Press (1989).

J.P. Jolivet, *Metaloxide Chemistry and Synthesis*, Wiley (2000). (Discusses condensation mechanisms for aqueous cations, and surface chemistry of colloidal oxides.)

J. Livage, M. Henry, and C. Sanchez, *Sol-Gel Chemistry of Transition Metal Oxides* in *Progr. Solid State Chem.* **18** (1988) 259. (Reviews many molecular precursors and their polymerization to inorganic oxides.)

E. Matijevič, *Preparation and Properties of Uniform Size Colloids*, in *Chem. Mater.* **5** (1993) 412. (Reviews monodisperse colloids with a broad spectrum of morphology and composition.)

*Technological Applications of Dispersions*, R.B. McKay, Ed., Marcel Dekker (1994). (Reviews, preparation and properties of colloids used in a variety of applications including paints, paper, ceramics, and plastics.)

A.E. Nielsen, *Kinetics of Precipitation*, Pergamon Press (1964).

S. Odén, *Der Kolloide Schwefel*, Thesis Upsala University (1913). (This still seems to be the latest monograph on sulphur colloids; see also S. Odén, *Kolloid-Z.* **8** (1911) 186.)

M. Ozak, *Preparation and Properties of Well-defined Magnetic Particles*, MRS Bulletin (December 1989) 35.

*Fine Particles Science and Technology*, E. Pelizzetti, Ed., Kluwer (1996). (Provides numerous references to many aspects of colloid synthesis.)

*Particle Growth in Suspensions*, A.L. Smith, Ed., Academic Press (1978). (Proceedings of a Symposium.)

*Fine Particles; Synthesis, Characterization and Mechnisms of Growth*, T. Sugimoto, Ed., Marcel Dekker (2000). (A comprehensive text on the formation routes and mechanisms of inorganic as well as polymeric colloids.)

T. Svedberg, *Die Methoden zur Herstellung Kolloider Lösungen Anorganische Stoffe*, Theodor Steinkopff Verlag (1909). (Probably the first monograph on inorganic colloid synthesis.)

A.G. Walton, *The Formation and Properties of Precipitates*, Interscience Publishers (1967).

H. Weiser, *Inorganic Colloid Chemistry* (two volumes), Wiley (1933). (A rich and still relevant source of detailed preparation methods.)

*Nucleation*, A.C. Zettlemoyer, Ed., Marcel Dekker (1969).

*Separation techniques*

Standard texts on flow in porous media are:

R.E. Collins, *Flow of Fluids through Porous Materials*, Reinhold (1961).

For a treatment of colloidal filtration see also W.B. Russel, *The Dynamics of Colloidal Systems*, University of Wisconsin (1987).

A.E. Scheidegger, *The Physics of Flow through Porous Media*, University of Toronto Press (1974).

**2.5b *Characterization***

An instructive overview of various averages is: J.T. Bailey, W.H. Beattie, and C. Booth, *Average Quantities in Colloid Science*, *J. Chem. Educ.* **39** (1962) 196-202.

S.E. Harding, A.J. Rowe, and J.C. Horton, *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, Roy. Soc. of Chem. (1992). (Comprehensive coverage of analytical ultracentrifugation of (bio)polymers; no reference is made to inorganic particles.)

M.B. Huglin, *Light Scattering from Polymer Solutions*, Academic Press (1972). (Includes many practical aspects.)

For a recent overview of characterization methods see: E. Kissa, *Dispersions; Characterization, Testing and Measurement*, Surfactant Series Vol. **84**, Marcel Dekker (1999).

R. Pecora, Dynamic Light Scattering; Applications of Photon Correlation Spectroscopy, Plenum (1983). (Still a basic text.)

A.P. Philipse, *Colloidal Sedimentation and Filtration*, *Current Opinion Colloid Interface Sci.* **2** (1997) 200. (Literature entrance for colloidal spheres and non-spheres.)

E.R. Pike, J.B. Abbiss, *Light Scattering and Photon Correlation Spectroscopy*, Kluwer (1997). (Overview of both experimental and theoretical developments. For more reviews of light scattering see footnotes of sec. 2.3b.)

For extensive information on characterization and fractionation of polydisperse colloids see: *Particle Size Distributions; Assessment and Characterization*, T. Provder, Ed., *Am. Chem. Soc. Symp. Ser.* (1987); *Particle Size Distributions II*, T. Provder, Ed., *Am. Chem. Soc. Symp. Ser.* (1991).

For further types of characterization (electrokinetics, surface charge, steric stabilization, etc.), see the relevant chapters in FICS.

2.72

# APPENDIX 1

## The moment expansion and some applications

Assuming that (electron) microscopy can be employed to the colloids under study, typically 1,000-2,000 counts are needed to obtain a representative size distribution, which no longer changes shape when more particle sizes are sampled. A variety of theoretical distributions for data fitting are available which, together with extensive nomenclature, are treated in detail in the literature[1,2,3], see also 2.3f. Here we focus on some general features of distributions, which do not depend on the applicability of any theoretical fit. We do this on the basis of two data sets in fig. A1.1, namely a size distribution of magnetite particles (from now on called the 'M-distribution,' l.h.s. of this figure) and one of silica spheres (the 'S-distribution,' r.h.s.). The M-distribution is a typical example of an asymmetric distribution with a significant tail of relatively large particles, whereas the symmetric S-distribution agrees very well with a Gaussian. A variety of averages may be defined for these distributions, the type obtained depending on the technique by which the sol is investigated. These averages can be estimated fairly accurately from the relative dispersity $s_a$. Below we will explain this estimate and test it for the distributions in fig. A1.1. We start with the $n^{th}$ moment of a discrete distribution, which is generally defined as

$$\langle a^n \rangle = \frac{1}{N} \sum_{i=1}^{N} a_i^n \qquad [A.1.1]$$

For a continuous distribution this equation may be replaced by [2.3.39]. In [A.1.1] the brackets denote a number average over a total of $N$ particles with radii $a_i$. The first moment is obviously the number-averaged radius $\langle a \rangle$ and the second moment determines the *relative dispersity* $s_a$ defined as

$$s_a^2 = \frac{\langle a^2 \rangle - \langle a \rangle^2}{\langle a \rangle^2} = \frac{\sigma_a^2}{\langle a \rangle^2} \qquad [A.1.2]$$

where $\sigma_a^2 = \langle a^2 \rangle - \langle a \rangle^2$ is the standard deviation, or absolute dispersity, see sec. I.3.7a. Here, $\sigma_i = a_i - \langle a \rangle$ are the fluctuations around the average sphere radius, which by definition cancel each other such that $\langle \sigma \rangle = 0$. From [A.1.1] we obtain

$$\frac{\langle a^n \rangle}{\langle a \rangle^n} = \left\langle \left( 1 + \frac{\sigma}{\langle a \rangle} \right)^n \right\rangle \qquad [A.1.3]$$

To make an expansion in the fluctuations, we make use of the binominal theorem,

---

[1] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, (1965).
[2] M. Alderliesten, *Mean Particle Diameters Part I, Part. Part. Syst. Charact.* **7** (1990) 233-241; *Part II, Part. Part. Syst. Charact.* **8** (1991) 237.
[3] P. Walstra, *Physical Chemistry of Foods*, Marcel Dekker (2002).

which states that for any positive integer $n$

$$(1+y)^n = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} y^k \qquad [A.1.4]$$

From [A.1.3] and [A.1.4] we find the following expansion, which contains a leading term of order $s_a^2$, but no linear $s_a$-term because $\sigma_a = 0$

$$\frac{\langle a^n \rangle}{\langle a \rangle^n} = 1 + \frac{n(n-1)}{2} s_a^2 + \sum_{k=3}^{n} \frac{n!}{(n-k)!k!} \frac{\langle \sigma_a \rangle^k}{\langle a \rangle^k} \qquad [A.1.5]$$

The shape of the size distribution is, at this stage, still arbitrary. If the distribution is symmetrical around the average $\langle a \rangle$, positive and negative fluctuations in [A.1.2] occur with equal probability, so $\langle \sigma_a^k \rangle = 0$ for odd values of k. Consequently, the third moment for any symmetric distribution exactly equals

$$\frac{\langle a^3 \rangle}{\langle a \rangle^3} = 1 + 3s_a^2 \qquad [A.1.6]$$

whatever the width of the distribution. For distributions, which are sufficiently narrow such that $s_a^2 << 1$, we can truncate the expansion [A.1.5] to obtain a result already reported in [2.3.44] for the special case of a log-normal distribution:

$$\frac{\langle a^n \rangle}{\langle a \rangle^n} \approx 1 + \frac{n(n-1)}{2} s_a^2 \qquad s_a^2 << 1 \qquad [A.1.7].$$

This useful approximation (already reported in a light scattering study[1]) relates higher moments of a distribution to the average particle size and dispersity. We have already seen one application in the expression for the apparent Guinier radius in [2.3.8], which is obtained by the application of [A.1.7] to [2.3.7]. In a similar fashion, we find the apparent hydrodynamic radius from dynamic light scattering in [2.3.14]

$$a_h \approx \left(1 + 5s_a^2\right)\langle a \rangle \qquad s_a^2 << 1 \qquad [A.1.8]$$

whereas the specific surface area yields

$$a_s \approx \left(1 + 2s_a^2\right)\langle a \rangle \qquad s_a^2 << 1 \qquad [A.1.9]$$

In the context of sedimentation (sec. 2.3d), we encountered the $z$-averaged particle mass

$$m_z = \frac{\langle m^3 \rangle}{\langle m^2 \rangle} \qquad [A.1.10]$$

which for spheres with identical mass density corresponds to an apparent radius

---

[1] P.N. Pusey, H.M. Fijnaut, and A. Vrij, *J. Chem. Phys.* **77** (1982) 4270.

$$a_z = \left[\frac{\langle a^9\rangle}{\langle a^6\rangle}\right]^{1/3} \approx \left(1+7s_a^2\right)\langle a\rangle \qquad s_a^2 << 1 \qquad [A.1.11]$$

again after expanding up to terms of order $s_a^2$.

Table A.1. Test of [A.1.7] for the data in fig. A1.1.

| Distribution | M | | $M_f$ [1] | | S | |
|---|---|---|---|---|---|---|
| | data[2] | calc.[3] | data[2] | calc.[3] | data[2] | calc.[3] |
| $\langle a\rangle$/nm | 4.3 | - | 3.6 | - | 117.7 | - |
| $s_a$/% | 36.87 | - | 26.04 | - | 9.23 | - |
| $p$ | 2.320 | 2.213 | 1.466 | 1.610 | 1.074 | 1.077 |
| $a_h$/nm | 6.9 | 7.3 | 4.4 | 4.8 | 122.5 | 122.8 |
| $a_G$/nm | 7.4 | 8.1 | 4.5 | 5.2 | 124.0 | 124.2 |
| $a_z$/nm | 7.8 | 8.4 | 4.6 | 5.3 | 124.4 | 124.7 |
| $\bar{a}$/nm | 3.7 | | | | 116.6 | |

[1] $M_f$ is the M distribution in fig. A1.1 after removal of particles with radius $a$ > 5 nm. [2] Directly from particle counts. [3] Calculated using [A.1.7] and the relative dispersity $s_a$ from this table.

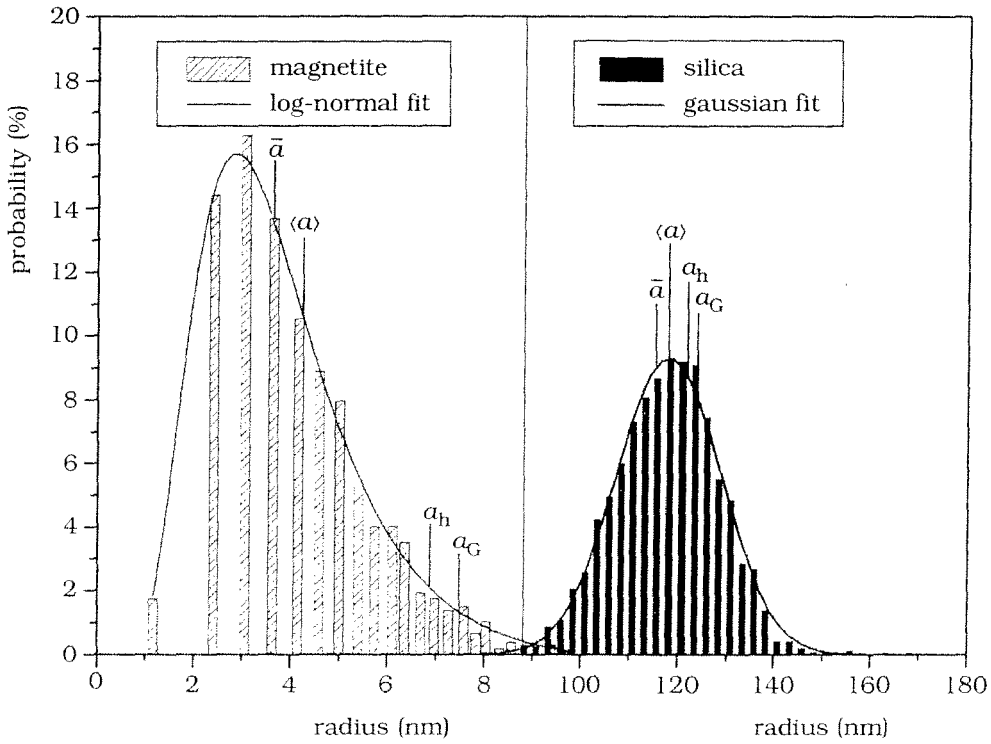



**Figure A1.1.** Size distribution of 1.081 magnetite (*M*) particles obtained from conventional TEM images (courtesy M. Raşa) and 2.892 silica (S) spheres imaged by cryogenic-TEM (courtesy P. Homan). The M-distribution is a fairly broad log-normal one, the S-distribution is a Gaussian. Indicated are the number-average radius $\langle a\rangle$ the *z*-average radius $a_z$, the harmonic mean $\bar{a}$ and apparent radii, which would be found from static $a_G$ and dynamic light scattering $a_h$. See also table 2.1.

We now put [A.1.7] to the test for the distributions in fig. A1.1 with results summarized in table A.1. It turns out that [A.1.7] quantitatively applies for the S distribution even though the silica spheres have a dispersity of 9 percent. For the magnetite particles, with a significant dispersity of 37 percent, approximation [A.1.7] still provides a reasonable estimate. Table A.1 also shows the results of a hypothetical (but practically not unfeasible) fractionation by which large magnetite particles with a radius $a > 5$ nm (286 out of a total of 1,081 particles) are removed from the M distribution, to illustrate the effect of the tail of large particles. Removal of this tail substantially reduces the discrepancy between the static ($a_G$) and dynamic ($a_h$) light scattering radius, and the number averaged radius $\langle a \rangle$. Note in table A1 that [A.1.2] correctly predicts the trend in radii going from the $M$ to the $M_f$ distribution. Some further remarks on the basis of this table are the following. A frequently used, dimensionless dispersity measure (also for polymers) is the ratio $p$ between the weight-averaged molecular mass and the number-averaged molecular mass. For spheres with identical mass density, this ratio equals

$$p = \frac{\langle a^6 \rangle}{\langle a^3 \rangle^2} \qquad [A.1.12]$$

where the $a^6$ term stems from the square of the particle mass. Substitution of the approximation [A.1.7] shows that to leading order in $s_a^2$, $p$ equals

$$p \approx 1 + 9 s_a^2 \qquad s_a^2 << 1 \qquad [A.1.13]$$

Clearly, the ratio $p$ is not a very sensitive measure in comparison with the relative dispersity $s_a$ itself, as also becomes clear from the numbers in table A.1.1. Another observation from this table is the marked increase relative to $\langle a \rangle$ due to polydispersity of the Guinier radius $a_G$ and the hydrodynamic radius $a_h$ obtained from static and dynamic light scattering, respectively. An increase by a factor of two is to be expected for a polydispersity near $s_a \approx 40\%$.

In table A.1, the harmonic mean

$$\bar{a} = \langle a^{-1} \rangle^{-1} \qquad [A.1.14]$$

is also included as an example of an average, which emphasizes the small-particle tail of a distribution. This tail will be relevant for Ostwald ripening and the amount of dissolved material due to the Gibbs-Kelvin effect, which depends on the reciprocal particle radius, see [2.2.51].

Universiteit Utrecht 

# Soft Condensed Matter

## Problem Sets

Utrecht University

September 2014

## Soft Condensed Matter 2005 Problems Chapter 1

**1)** In order to get a feeling of where the formula for the Brownian time (Eq. 1.5 ) comes from, we can derive it for a simpler case already encountered in the field of the mechanics of point objects. We want to calculate how long it takes if drops a ball with radius *R* and buoyant mass *m* in a liquid. Assume that the friction factor *f* is proportional to the speed *v* and is given by $6\pi\eta R$.

a) What is the terminal velocity of the ball if it is give that the $\eta = 1 * 10^{-2}$ Pa.s and that the buoyant mass of the ball is 1 kg and its radius 1 m ($g$ = 10 m/s$^2$).
b) Derive an equation for the relaxation time mentioned by solving the second order differential equation that describes this situation.
c) What is the relaxation time for the situation mentioned in a)?
d) What is the terminal velocity and the relaxation time for a colloidal sphere with a radius of 1 μm, assume a particle density difference of 1 g/cm$^3$ with the solvent?
e) What distance does the particle travel if it is assumed to diffuse Brownian during this time?
f) What is the gravitational height of the same colloidal sphere?

**2)** Let us assume the human lifetime is extended to 10.000 year, what size of particle would one still consider colloidal? What will ultimately limit the size of a colloid?

**3)** The rotational diffusion constant of a spherical particle with radius 'a' in a medium with viscosity η is given by:

$$D_0^{\,r} = \frac{k_B T}{8\pi\eta a^3}$$

What particle size would you choose if you wish to study their rotational diffusion experimentally, for instance in an aqueous (η = 1 * 10$^{-2}$ Pa.s [Pa·s]) system?

**4)** **a)** Show that for colloidal particles dispersed in a liquid, the equilibrium number of particles, *N*, at a height h above a reference level, $h_0$, is given by:

$$N = N_0 \exp[-(m-m')g(h-h_0)/k_B T]$$

where $N_0$ is the number of particles at height $h_0$ and m' is the mass of fluid displaced by a particle of mass m.

**b)** Svedberg (1928) gives the following Table for the sedimentation equilibrium of a gold sol under gravity:

| Height (μm) | Number of particles | Height (μm) | Number of particles |
|---|---|---|---|
| 0 | 889 | 600 | 217 |
| 100 | 692 | 700 | 185 |
| 200 | 572 | 800 | 152 |
| 300 | 426 | 900 | 125 |
| 400 | 357 | 1000 | 108 |
| 500 | 253 | 1100 | 78 |

Assume the particles have radius 21 nm and density 19.3 $g \cdot cm^{-3}$ and the temperature is 20 ˚C. Estimate the Boltzmann constant, *k*, from the equation derived in (a) and then calculate Avogadro's number, $N_A$, assuming $R$=8.31 $J \cdot K^{-1} \cdot mol^{-1}$.

**c)** Repeat the calculation with a radius of 22 nm and note how sensitive the answer is to this variable.

## 2. Problems

1) The van der Waals gas.
a) Calculate the critical density, temperature, and pressure.
b) Show that $p_c/\rho_c kT_c = 3/8$ independent of $a$ and $b$. This is called the law of corresponding states.

2) Verify Eq. (2.24).

3) Second virial coefficient for charged colloids.
a) Calculate $B_2$ for the hard-core Yukawa interaction, assuming $\varepsilon \ll kT$ .
b) Examine the limiting cases $\kappa d \gg 1$ and $\kappa d \ll 1$.

4) The virial route to thermodynamics. Derive Eq. (2.48) starting from the canonical partition sum. *Hint*: transform to the dimensionless variables $\mathbf{s}_i = \mathbf{r}_i/V^{1/3}$ .

5) Potential of mean force.
a) Sketch the mean force and the potential of mean force for a pair of hard spheres at a volume fraction of 0.49.
b) Explain how hard particles could feel mean forces that are repulsive and attractive.

6) Radial distribution function up to first order in density
a) Sketch the radial distribution functions of low-density fluids of particles interacting with the Lennard-Jones, hard-core Yukawa, and hard-sphere potentials.
b) For the hard-sphere fluid show that the structure factor up to first order in volume fraction $\eta$ is given by

$$S(\mathbf{q}) = 1 + \frac{24\eta}{(qd)^3}(qd\cos qd - \sin qd).$$

Also sketch this function.

7) The square well fluid.
Consider a fluid of spheres interacting with the so-called 'square well potential':

$$\phi(r) = \begin{cases} \infty & (r \le d) \\ -\varepsilon & (d < r < d+\delta), \\ 0 & (r \ge d+\delta) \end{cases}$$

Here, $d$ is the diameter of the spheres and $\varepsilon > 0$ is the well depth.
a. Calculate the second virial coefficient of the square well fluid.
b. Do you think the square well fluid has a liquid-gas phase separation? Give arguments for your answer.
c. For sufficiently dilute fluids the radial distribution function can be approximated as $g(r) = \exp(-\phi(r)/kT)$. Use this to show that the internal energy per unit volume can be approximated as

$$\frac{U}{V} = \frac{3}{2}\rho kT - \frac{2\pi}{3}\rho^2 \varepsilon e^{\varepsilon/kT}\left((d+\delta)^3 - d^3\right).$$

d. In one graph sketch the radial distribution functions of the square well fluid and the hard sphere fluid at volume fractions of 0.01 and 0.40.

## 3. Problems

1) How does equation (3.11) change if the incident light has perpendicular polarization? And parallel?

2) Calculate the scattering cross section of a Rayleigh scatterer.

3) Derive (3.14).

4) *The form factor of a clay particle*
Clay particles have the shape of a uniform, but very small, thin disk. Propose a method to measure the radius *R* of the disks with light scattering. Derive a formula with which to extract *R* from the measured data.

5) *Scattering by a Gaussian polymer coil*
Debye pointed out in 1947 that when calculating the scattering by a polymer molecule it is convenient to consider the basic scattering particle to be the statistical chain segment. Interference between waves scattered from different chain segments is then incorporated into the structure factor. The chain segment is usually so small compared to the wavelength of light that it can be considered a point scatterer (*i.e.* its form factor equals unity.) If the polymer solution is dilute then only interference between segments in the same polymer molecule is important. The structure factor (3.27) depends only on the difference vector $\mathbf{r}_{jk}$ between two such segments. For a Gaussian polymer coil the probability density function for this vector is given by (see the Chapter on polymers)

$$p(\mathbf{r}_{jk}) = \left( \frac{3}{2\pi b^2 |j-k|} \right)^{3/2} \exp\left( -\frac{3 r_{jk}{}^2}{2b^2 |j-k|} \right), \tag{P1}$$

where *b* is the segment length.

a. Show that the form factor for a Gaussian coil with *N* segments can be written as

$$P(\mathbf{q}) = \frac{1}{N^2} \sum_j \sum_k e^{-\frac{1}{6} q^2 b^2 |j-k|}. \tag{P2}$$

*Hint:* Contrary to most problems involving spherical symmetry the integral is best done in normal Cartesian coordinates. The integrals can be performed by completing the squares in the exponent and making use of the integral

$$\int_{-\infty}^{+\infty} e^{-ax^2} dx = \sqrt{\pi/a}\ .$$

b. Next, evaluate (P2) by approximating the sums with integrals. Show that

$$P(\mathbf{q}) = \frac{2}{y^2}\left( e^{-y} - 1 + y \right), \tag{P3}$$

where

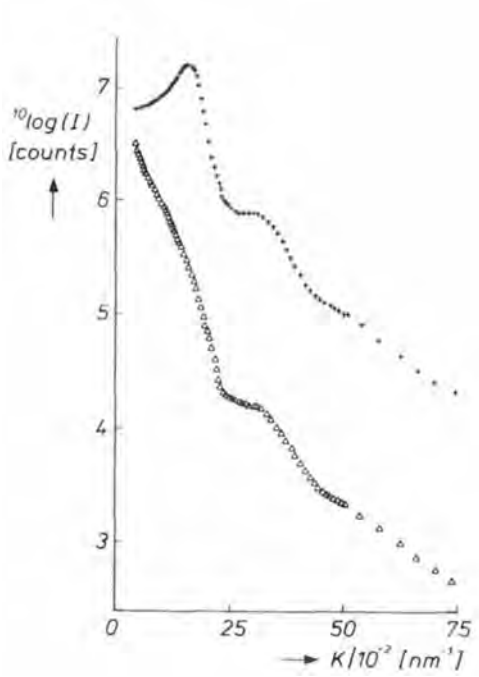
$$y = \tfrac{1}{6} N q^2 b^2 . \tag{P4}$$

c. Now derive a lowest order in *q* approximation for the form factor and confirm that the radius of gyration of a Gaussian chain is given by

$$R_g^2 = \tfrac{1}{6} N b^2 . \tag{P5}$$

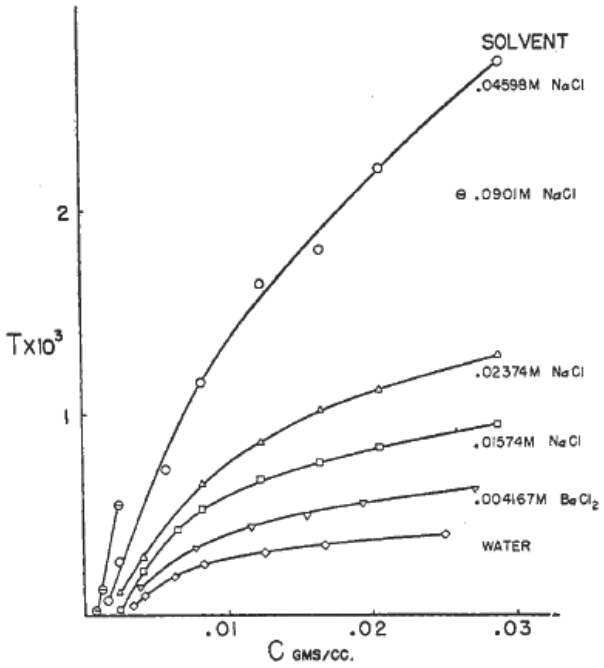
6) Determine the systematic vanishings for a body-centered cubic lattice (*bcc*). How would you distinguish bcc from fcc?

7) The graph shows small-angle X-ray scattering data on two suspensions of colloidal silica spheres in the solvent cyclohexane. The upper curve is a sample with a concentration of 0.714 g/cm$^3$, the lower curve 0.01 g/cm$^3$. The vertical axis shows the scattered intensity $I$ and the horizontal axis shows the scattering vector $K$ (multiplied by 100).




a. Estimate the size of the spheres used in this experiment.
b. Explain why the upper curve has a peak in the low-$K$ range, while the lower curve does not.
c. Why do the curves look almost the same in the high-$K$ range?
d. Estimate the average distance between the particles in the suspension.
e. How would the measured data change if a silica concentration of 1.0 g/cm$^3$ is used?

8) The graph below shows measurements by Peter Debye of the amount of light (546 nm) scattered by solutions of the cationic surfactant dodecylamine hydrochloride in water. $T$ is a number proportional to the scattered intensity at low scattering angle, and $c$ is the surfactant concentration (in grams per cubic centimeter). The measurements were repeated with salts added at the concentrations shown on the right of the graph.




(a) Explain why the scattering goes to zero at a concentration $c_0$ that differs from zero. Also explain why $c_0$ decreases as more NaCl is added.

In the syllabus equation (4.26) was derived for the scattered intensity. If we ignore all the constants in this equation it can be written as

$$I_s(q) \propto \rho V_p^2 P(q) S(q). \qquad \text{(I)}$$

Here $\rho$ is the number of scattering particles per cubic centimeter and $V_p$ the volume of one particle. $P(q)$ is the form factor, $S(q)$ the structure factor, and $q$ the scattering vector. For the experiment shown here we are allowed to make two approximations:

I. The particles are much smaller than the wavelength of light, so we can take the limit $q \to 0$.

II. In the low concentration range shown in the graph, particle interactions can be approximated up to the second virial coefficient $B_2$.

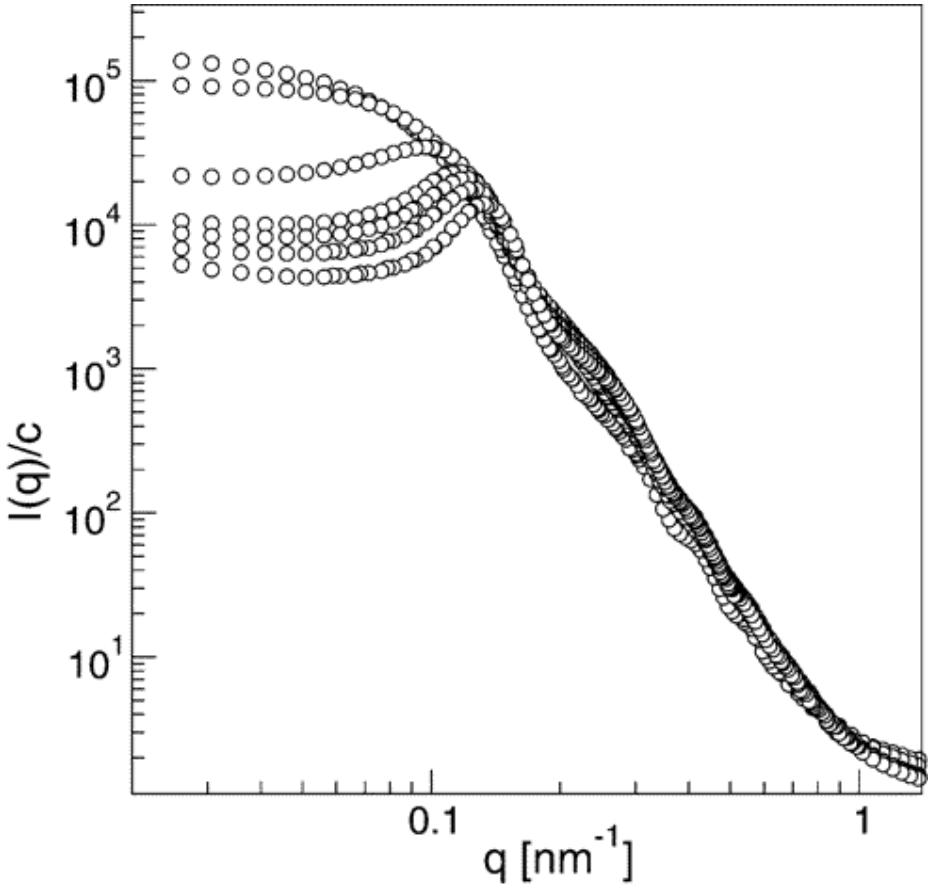
(b) Show that these approximations lead to a scattered intensity given by

$$I_s \propto \frac{\rho V_p^2}{1+2B_2\rho}. \qquad \text{(II)}$$

Argue that this equation qualitatively describes the form of the measured data.

(c) Use the data and equation II to estimate by what factor $V_p$ increases when 0.04598 M NaCl is added to a surfactant solution in water.

9) The figure below shows small angle neutron scattering data on suspensions of sterically stabilized polystyrene colloids dispersed in water [Zackrisson et al., Langmuir **21**, 10835 (2005)]. The quantity plotted is the scattered intensity $I(q)$ at scattering vector $q$ divided by the weight concentration $c$ of colloids. The seven sets of data correspond to colloid concentrations of 0.40, 0.35, 0.33, 0.30, 0.21, 0.048, and 0.0042 g/mL.




a. Which set of data corresponds to which colloid concentration? Explain your answer.

b. Estimate the average distance between particles at the highest concentration.

c. From the low-$q$ scattering data, prepare a plot of the inverse of the osmotic compressibility $\frac{1}{kT}\left(\frac{\partial \Pi}{\partial \rho}\right)$ versus the concentration $c$. Interpret the result.

d. Estimate the second virial coefficient $B_2/V_0$ of this colloidal system, given that the particle density is $d$=1.076 g/mL. Interpret the result.
*Hint:* The number density $\rho$ and weight concentration $c$ are simply related through $\rho = c/V_0 d$ . $V_0$ is the volume of a particle.

10) *Brownian motion*

a. Find an expression for the probability density function $P(\Delta r,t)$ describing the displacement of a Brownian particle in a time $t$.

b. Show that the mean square displacement is given by $\left\langle \Delta r^2 \right\rangle = 6Dt$ .

11) *The Random Walk*

Consider the following idealization of a random walk in one dimension. A particle starts in the origin and makes a step of size $l_x$ randomly in the positive or negative $x$ direction, so that:

$$x_{i+1} = x_i + l_x \quad \text{probability } \tfrac{1}{2}$$
$$x_{i+1} = x_i - l_x \quad \text{probability } \tfrac{1}{2}$$

It is clear that, after a large number $N$ of steps, the expectation value of the $x$ coordinate of the particle is zero.

a. Now show that

$$\left\langle x_N^2 \right\rangle = N l_x^2 .$$

b. Assuming that the particle travels ballistically with a speed $v$, show that

$$\left\langle x^2 \right\rangle = v l_x t .$$

c. Next, consider a random walk in three dimensions. Furthermore, allow the random walk step size $l$ to be selected from a probability distribution $P(l)$. Show that

$$\left\langle r^2 \right\rangle = vt \frac{\left\langle l^2 \right\rangle}{\left\langle l \right\rangle}. \tag{0.1}$$

d. For diffusion of molecules in a dilute gas the pdf for the step size $P(l)$ can be found by considering the collision probability between molecules. Consider an ensemble of molecules that have just undergone a collision. After traveling a distance $l$ there are $N(l)$ molecules left that have gone without colliding a second time. The number of these molecules undergoing another collision in the next $dl$ meters is then proportional to $dl$ and $N(l)$. Verify that this leads to

$$P(l) = \frac{e^{-l/l_{mf}}}{l_{mf}},$$

where $l_{mf} = \left\langle l \right\rangle$ is called the *mean-free path* of the molecules.

e. Finally, show that the diffusion coefficient of these molecules is given by the following formula, which is well known in the kinetic theory of gases:

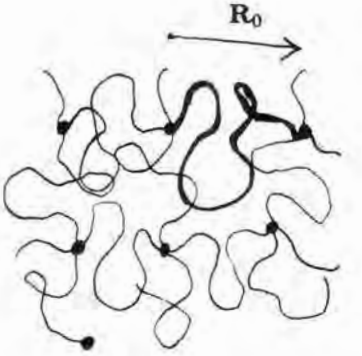
$$D = \tfrac{1}{3} v l_{mf} . \tag{0.2}$$

12) Suppose that a dilute colloidal dispersion contains equal numbers of two types of spherical particles. One population has a radius of 20 nm, the other 40 nm. Describe the form of the intensity autocorrelation function that one would measure with DLS.

# 5. Polymers: problems

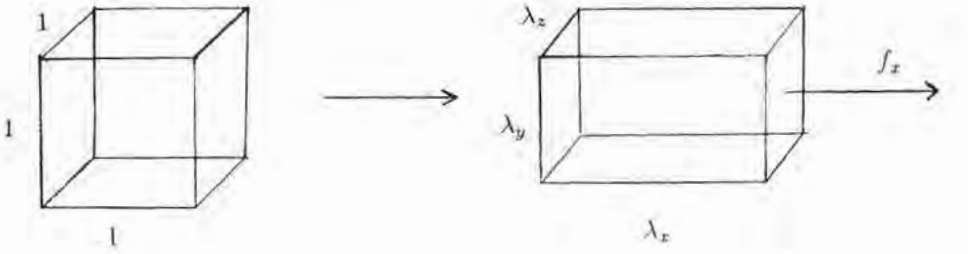
1. In this problem we will have a closer look at some properties of the solution to the diffusion equation.
   a. Check that probability distribution (2.7) is a solution to equation (2.4).
   b. Show that this probability distribution is normalized.
   c. What is the formula corresponding to (2.7) describing a diffusing particle?
   d. Make a sketch of the formula in c. for different moments in time, for instance as a function of $x$ for $y = z = 0$. What happens if we take $t \to 0$ ? What is the physical significance of this?
   e. Calculate the standard deviation in $\mathbf{R}^2$ and compare this to $\langle \mathbf{R}^2 \rangle$.

2. This problem tries to describe the elasticity of polymer molecules in a rubber. Rubber consists of molten polymers connected to each other at certain points by socalled *cross links*:




A schematic representation of a rubber.

The concentration of cross links is $\nu$, so there are also on the order of $\nu$ subchains per unit volume. $\mathbf{R}_0$ is the end-to-end distance of a subchain. In a polymer melt (sub)chains practically behave like undisturbed ideal chains, so that we immediately have some formulas to our disposal. When a rubber cube is stretched, there is conservation of volume to a good approximation (after all a polymer melt is a kind of liquid, which are generally incompressible):




Stretching of a unit cube of rubber.

In this picture the socalled stretching factors are $\lambda_x, \lambda_y$ and $\lambda_z$. Conservation of volume requires $\lambda_x \lambda_y \lambda_z = 1$. Symmetry considerations further give: $\lambda_y = \lambda_z$. We presume that the rubber network deforms in the same way as the total cube:

$$\mathbf{R}_0 = (R_{0x}, R_{0y}, R_{0z}) \rightarrow \mathbf{R} = (\lambda_x R_{0x}, \lambda_y R_{0y}, \lambda_z R_{0z})$$

Now we have sufficient elements to calculate the elastic force.

a. What is the change in free energy of one single ideal chain under the deformation described above?

b. Argue that the free energy change of the total cube is given by:

$$\Delta A = \frac{k_B T \nu}{2}(\lambda_x^2 + \lambda_y^2 + \lambda_z^2 - 3)$$

c. Let $\lambda_x = \lambda$ and use conservation of volume and symmetry to further simplify this formula.

d. Now show that the required force for stretching is given by:

$$f_x = k_B T \nu \left(\lambda - \frac{1}{\lambda^2}\right)$$

Discuss this formula and the way it depends on the variables involved.

3. From a theoretical point of view it is interesting to see how polymers behave in $d$ dimensions. For $d = 2$ one can think of a polymer at a surface. However, also unphysical dimensions $d \geq 4$ are theoretically interesting, since polymer coils turn out to behave always like ideal coils. The most important point is the dependence of the magnitude of the coil on $N$. A calculation in $d$ dimensions using Flory's lattice model, gives insight into this.

a. Argue that equation (3.5) may be generalized to

$$\frac{A(R)}{k_B T} = -\ln W(R) = cst - (d-1)\ln R + \frac{d}{2Nb^2}R^2 + \frac{N^2 \nu_c}{2R^d}$$

b. Take $\nu_c = 0$ (ideal chain) first and calculate $R = R_0^*$ where the free energy has a minimum. Compare this with the square root of the mean square $\sqrt{\langle \mathbf{R}^2 \rangle} = bN^{1/2}$ (note that for this part dimension $d = 1$ is not covered by this method).

c. Verify that for a non–ideal polymer the minmization leads to

$$\left(\frac{R^*}{R_0^*}\right)^{d+2} - \left(\frac{R^*}{R_0^*}\right)^{d} \approx \frac{N^2 \nu_c}{(R_0^*)^d} \approx \frac{\nu_c}{b^d} N^{2-d/2} \approx N^{2-d/2}$$

We can leave out numerical factors like $d/(d-1)$ and take $\nu_c/b^d = 1$, since here we are only interested in the dependence on $N$.

d. We can now further analyze this equation. For long chains (large $N$) it is nearly always possible to neglect one of the terms in the equation with respect to the other two. Use this fact to determine:

| $R^* \approx bN^{\nu}$ | |
|---|---|
| $d \leq 4$ | $\nu = \frac{3}{2+d}$ |
| $d \geq 4$ | $\nu = \frac{1}{2}$ |

e. Would it have been clear from the start that $\nu = 1$ for $d = 1$? For 4 dimensions and higher a polymer chain with excluded volume also behaves ideally: think of the reason for this.

Computer simulations give:

| $d$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\nu$ | 1 | 0.75 | 0.588 | 0.50 |

Compare these results with those of the simple Flory model.

4. At the end of section 3.3 a remark was made about the relation between the free energy of a chain and the second virial coefficient. The second virial coefficient may be better known in the context of the pressure. In this problem the connection is clarified and a relation is made to the free energy of a concentrated polymer solution within the socalled *Flory–Huggins* approximation. The interaction between polymer segments within a single polymer chain is given by the last term of (3.5) using transformation (3.10):

$$\frac{A_{ww}(R)}{k_B T} = \frac{N^2 v}{2R^3} = \frac{v}{2} Nc$$

where segment concentration $c(= N/R^3) = N/V$. From the free energy you can go to the pressure by means of the thermodynamic relation:

$$P = -\left(\frac{\partial A}{\partial V}\right)_N$$

a. Use this to show that the interaction term in the free energy leads to a pressure term quadratic in the segment concentration (of which the coefficient is directly related to the second virial coefficient).

b. For a more concentrated system of $n_p$ polymers of $N$ segments and $n_s$ solvent molecules (occupying $\Omega = n_p N + n_s$ lattice sites and with a volume $V = \Omega v_c$, where $v_c$ is the fixed volume per lattice site) we have volume fraction $\phi = n_p N/\Omega$ and the (Helmholtz) free energy of mixing $A_m$ (within the Flory–Huggins approximation):

$$A_m(\Omega, \phi) = \Omega k_B T f_m(\phi) = \Omega k_B T \left[ \frac{1}{N} \phi \ln \phi + (1 - \phi) \ln(1 - \phi) + \chi \phi (1 - \phi) \right]$$

The free energy of mixing $A_m$ is the difference in free energy between the mixed and unmixed system. The expression for the pressure now also requires a constant number of polymers:

$$P = -\left(\frac{\partial A}{\partial V}\right)_{n_p, N}$$

Could you argue that we obtain the *osmotic* pressure $\Pi$ of a polymer solution against pure solvent if we take $A = A_m$ within this expression?

c. Now calculate $\Pi$ from $A_m$. Try to find the following intermediate results:

$$\Pi = \frac{k_B T}{v_c} \left[ -f_m(\phi) + \phi \frac{df_m}{d\phi} \right]$$

$$= \frac{k_B T}{v_c} \left[ -\ln(1 - \phi) + \frac{\phi}{N} - \phi - \chi \phi^2 \right]$$

d. Use this to make a low density expansion. Could you explain the term linear in $\phi$? What is the second virial coefficient and how does it compare to the one found in a.?

5. In this problem we will consider mixtures of 2 types of molten polymers A ($N_A$ segments) and B ($N_B$ segments), socalled *blends*. The second component is now also a polymer, so that the corresponding term in the entropy of mixing within the Flory-Huggins approximation (4.2) must here be divided by the number of segments within this polymer, $N_B$:

$$f_m(\phi) = \frac{1}{N_A}\phi \ln\phi + \frac{1}{N_B}(1-\phi)\ln(1-\phi) + \chi\phi(1-\phi)$$

with a generalized $\chi \equiv \frac{z(\varepsilon_{AA}+\varepsilon_{BB}-2\varepsilon_{AB})}{2k_BT}$. We assume that both polymers have the same number of segments ($N_A = N_B = N$).

a. Within the above expression for $f_m(\phi)$ compare (qualitatively) the relative importance of the entropy and the energy and try to predict whether the polymers will easily mix.
b. Show that for $N_A = N_B = N$ the expression for $f_m(\phi)$ is symmetric.
c. Determine the positions of the minima and maxima of $f_m(\phi)$. (Hint: show that this implies $\ln(\frac{\phi}{1-\phi}) = -N\chi(1-2\phi)$; assume for one of the solutions to this equation $\phi \ll 1$ and check if this approximation is justified afterwards).
d. Make a sketch of $f_m(\phi)$. Estimate the values of the coexisting volume fractions.
e. Calculate the critical point of this system. What does this tell you about the miscibility of polymers?
f. Could you think of a method to mix these polymers anyway?

6. Describe and explain (within 1 to 2 pages) the different regions and boundaries in the figure displayed on page 31.

# 6. DLVO Potential & Techniques to Measure Interaction Forces

**1)** Show that if the interaction potential between molecules are of the form $C/r^n$ for all distances, the energy of that molecule will depend on the shape of the container if $n \leq 3$. (Why does the same argument not hold for the free energy?)

**2)** Show that for two weakly interacting dipoles the Boltzmann averaged first order approximation to the energy is of the form $1/r^6$.

**3a)** Show that the interaction energy of an atom *A* with a permanent dipole with moment $m_1$ with an atom *B* with a polarizability $\alpha_2$ is given by:

$$w(r,\theta) = -\frac{1}{2}\frac{m_1^2 \alpha_2}{(4\pi\varepsilon\varepsilon_0)^2 r^6}\left[1+3\cos^2\theta\right]$$

Where $\theta$ is the angle between the dipole and the distance vector *r* between the atoms.

**3b)** Average this energy rotationally to arrive at the $1/r^6$ term for the induction or Debye forces.

**4)** Calculate the interaction energy between two infinite half-spaces using Hamaker's approach of pair-wise summation of $1/r^6$ interactions (note that this energy needs to be given per unit surface area).

**5a)** Calculate the Van der Waals interaction energy between two identical spheres with radius *a* a distance *h* apart using pair-wise summation. Note that in this equation the interaction depends only on *h/a* (as it should for scale invariant $1/r^6$ potentials).

**5b)** What is the limiting formula for distances $h/a << 1$?

**5c)** Show that this limiting form is compatible with the Derjaguin approximation between two spheres.

**5d)** Check how good this limiting formula is by calculating the interaction at a distance $h=0.01a$.

**6a)** Colloidal particles often aggregate in non-polar liquids (hydrocarbons, oils) because of the attractive Van der Waals forces between them. This is often a nuisance but can be prevented by coating the particles with a surfactant or polymer layer whose refractive index matches that of the liquids. Explain this phenomenon.

**6b)** At an ACS conference Dr. X from Colloids Corp. describes a colloidal dispersion of silica spheres (diameter 0.5 μm, smooth surface) in oil, where by coating the spheres with a 'matching layer' of surfactant, the depth of the potential well was reduced by a factor 10 as ascertained by light scattering measurements. When asked about the thickness of the layer, Dr. X replied that this is proprietary information. What was the thickness of the layer?

**7a)** Show that the non-linear PB equation for a flat plate and 1-1 electrolyte can be written in dimensionless units ($X=x\kappa$, and $\Phi=e\psi/kT$) as:

$$\frac{1}{2}\frac{d}{dX}\left(\frac{d\Phi}{dX}\right)^2 = \frac{d\cosh\Phi}{dX}$$

**7b)** Solve this equation and show that its solution is:

$$\Phi = 2\ln\left(\frac{1+t_0 e^{-X}}{1-t_0 e^{-X}}\right) \quad \text{with} \quad t_0 = \tanh\left(\frac{\Phi_0}{4}\right)$$

**7c)** Show that for $\Phi << 1$ this equation reduces to the DH limiting case (Eq. 32).

**7d)** Another interesting limiting case is obtained for large $X$. Show that in this case the potential can be approximated with:

$$\Phi = 4t_0 e^{-X}$$

**7e)** Interpret the physical meaning of the (limit of) the prefactor.

**8)** Use the equation derived at 7d) in combination with the superposition approximation (using the force method) *and* the Derjaguin approximation for spheres to obtain the following free energy of interaction between two identical spheres (for distances $\kappa^{-1} << x << a$) at a constant surface potential (note the different functional form as compared with Eq. 39):

$$V_{dlo}(h) = 4kT\frac{a}{l_{Bj}}t_0^2 e^{-\kappa h}$$

with the Bjerrum length as defined in Chapter 1 and $h$ the distance between the spheres.

**9)** If we combine the Van der Waals forces for short distances between spheres (means also in the Derjaguin approximation) with the Equation derived in problem 8 we get for this limit of the DLVO potential:

$$V(h) = 4kT\frac{a}{l_{Bj}}t_0^2 e^{-\kappa h} - \frac{A_{eff}\,a}{12h}$$

We can now analyze this equation for conditions of stability between the spheres.

**9a)** First show for what conditions this equation has a maximum.
Let's now analyze a few concrete examples. Example 1: $A_{eff}$ = 20 x $10^{-20}$ J (~50 $kT$) and $\kappa^{-1}$ = 10 nm ($10^{-3}$ M 1-1 electrolyte in water, ε~80).

**9b)** For what surface potentials will there be a maximum in the stability curve?
Example 2: $A_{eff}$ = 20 x $10^{-20}$ J (~50 $kT$) and $\Phi_0 = 1$.

**9b)** For what range of 1-1 salt concentration will there be a maximum in the DLVO potential?

**9c)** Are the values given for 9b still within the approximations made in deriving the DLVO potential used?

# 7. Problems

1. a) A spherical droplet with radius $R$ is divided in $N$ droplets of radius $r$. Derive a formula for the total, and the specific surface area of the $N$ droplets. Calculate the increase in specific surface area by dividing a micron-size water droplet into droplets of 10 nm.
   b) A colloidal sphere of radius $R$ is composed of molecules with specific volume $v_m$. A fraction $F$ of the molecules is exposed at the sphere surface. Find $F$ as a function of $R$, and apply your result to colloids with radius $R$ = 1 mm, 1 $\mu$m, 10 nm and 1 nm.

2. a) Show that $\Delta G^* = A^*\gamma/3$ [2.2.19] is valid for spheres as well as cubes. Make an educated guess how $\Delta G^*$ looks like, if the faces $i$ of the cube have different surface tensions $\gamma_i$.
   b) Derive the formation Gibbs energy $\Delta G$ of a sphero cylinder (radius $a$, length $L+2a$) in a solution with supersaturation $S$.
   Explain why the cylinder only grows spontaneously in length above a critical diameter.

3. a) A sphere with radius $R$ is suspended in a solution with molecule concentration $c(r)$ at a distance $r$ from the sphere center. Derive the diffusion-controlled frequency $J$ of collisions between molecules and sphere. Clearly explain the various steps and assumptions.
   b) We now wish to model the diffusion controlled growth of a disc with radius $R$ on a flat surface on which molecules diffuse around; now $c(r)$ is a surface concentration. Try to derive the equivalent of $J$ in 3a) for this two-dimensional case, and discuss your findings.

4. a) Use your result from 3a) to calculate the homogeneous nucleation rate of spherical particles in a supersaturated solution.
   b) Estimate the homogeneous nucleation rate for silica spheres in an aqueous solution (pH = 7) for a 'soluble silica' concentration of $c_m$ = 200 ppm, 500 ppm and $10^3$ ppm (see also pages 2.23 and 2.24).

5. a) Derive eq. [2.2.58] for the heterogeneous nucleation of a spherical cap on a flat substrate. Plot the function $f(\theta)$ as a function of the contact angle $\theta$.
   b) Suppose the spherical cap in Fig. 2.9 is a silica nucleus on a surface in contact with water. Estimate the required contact angle $\theta$ such that the nucleation flux of silica is enhanced by a factor of $10^3$, in comparison to the homogeneous flux in 4b) for the case of $c_m$ = 200 ppm.

6. a) A bimodal distribution of colloidal silica spheres is growing at neutral pH by diffusion controlled silica precipitation. Show that the relative width of the distribution decreases in time.

b) We now suddenly increase the pH to pH ~ 12, such that silica spheres gradually dissolve. Discuss the fate of the relative width in time.

7. a) Consider a solution containing a small silica sphere with radius $a_s$ and a much larger sphere with radius $a_l >> a_s$. Describe what will happen according to the Gibbs-Kelvin equation.
b) Formulate the diffusion flux for molecules diffusing from the small to the large sphere. Derive an expression for the derivative $da_s/dt$; explain any assumptions.
c) Estimate the dissolution rate of silica sphere in water at neutral pH near a flat glass surface.

8. a) For a continuous size distribution the moments are defined by:

$$< a^n >= \int_0^\infty a^n P(a) da$$

The Gauss (or normal) probability density has the form:

$$P(a) = A \exp\left[ -(a - < a >)^2 / 2(\sigma_a)^2 \right]$$

where

$$\sigma_a{}^2 = < a^2 > - < a >^2 .$$

Determine the constant $A$.

b) Start from the above definition of the $n^{th}$ moment and show how to obtain the approximation:

$$\frac{< a^n >}{< a >^n} \simeq 1 + \frac{n(n-1)}{2} s_a{}^2 ; \qquad s_a{}^2 = \frac{\sigma_a{}^2}{< a >^2} << 1$$

c) We determine <$a$> using transmission electron microscopy (TEM). How would you check whether you have sufficient 'counts' for a reliable <$a$>?
d) For a suspension of silica spheres it turns out that <$a$> = 100 nm with $s_a$ = 6%. Predict what would be the radius obtained from a) the specific surface area and b) static light scattering. Do the same for spheres with <$a$> = 100 nm and $s_a$ = 40%.