INSTITUTE OF PHYSICS PUBLISHING REPORTS ON PROGRESS IN PHYSICS

Rep. Prog. Phys. **65** (2002) 1577–1632 PII: S0034-4885(02)30922-9

Electrostatic correlations: from plasma to biology

Yan Levin

Instituto de Física, Universidade Federal do Rio Grande do Sul Caixa Postal 15051, CEP 91501-970, Porto Alegre, RS, Brazil

E-mail: levin@if.ufrgs.br

Received 5 July 2002
Published 24 September 2002
Online at stacks.iop.org/RoPP/65/1577

Abstract

Electrostatic correlations play an important role in physics, chemistry and biology. In plasmas they result in thermodynamic instability similar to the liquid–gas phase transition of simple molecular fluids. For charged colloidal suspensions the electrostatic correlations are responsible for screening and colloidal charge renormalization. In aqueous solutions containing multivalent counterions they can lead to charge inversion and flocculation. In biological systems the correlations account for the organization of cytoskeleton and the compaction of genetic material. In spite of their ubiquity, the true importance of electrostatic correlations has come to be fully appreciated only quite recently. In this paper, we will review the thermodynamic consequences of electrostatic correlations in a variety of systems ranging from classical plasmas to molecular biology.

Contents

1. Introduction

Although liquid state theorists have become quite accustomed to looking at the correlation functions for simple and complex fluids, many of the thermodynamic consequences of correlations have not been fully appreciated. To some extent this is the result of the success of mean-field theories for simple fluids. The strategy of applying the mean-field approximation to the many-body problems goes all the way to the pioneering work of van der Waals on liquid–gas phase separation [1]. The success of this theory and its physical transparency has set the stage for future applications of mean-field ideas. These came in the form of the Curie–Weiss theory of magnetism [2] and the Gouy–Chapman [3, 4] theory of diffuse ionic layers.

There are, however, some very familiar systems in which the mean-field contribution to the free energy is identically zero. One such system is the classical two component plasma of positive and negative ions. In order for the thermodynamic limit to exist, the charge neutrality constraint must be imposed. However, for a bulk charge-neutral system the average electrostatic potential is zero, which means that the mean-field contribution to the total free energy vanishes. Thus, the electrostatic free energy of a two component plasma is entirely due to positional correlations between the positive and negative ions. At low temperatures these correlations become so strong that they lead to a phase transition in which the plasma separates into two coexisting high and low density phases [5, 6].

Polar fluids provide another example of a system in which the electrostatic correlations strongly affect the thermodynamics. Perhaps the simplest model of a polar fluid is a system of dipolar hard spheres (DHS). The phase structure of DHS is quite interesting and deserves a separate review [7–9]. Here, we shall confine our attention to the low density disordered fluid phase. Since the average electric field inside a dipolar fluid is zero, it is evident that the mean-field contribution to the free energy also vanishes, and all the non-trivial thermodynamics is, once again, the result of electrostatic correlations.

The thermodynamics of DHS is particularly tricky because of the unscreened long-range interactions. In fact, the very existence of the thermodynamic limit for DHS has been proven only recently [10]. Nevertheless, it has been taken for granted that if the temperature is sufficiently low, the DHS will phase separate into coexisting liquid and gas phases. Indeed, all the theories have been predicting exactly this kind of behaviour [11, 12]. It came, therefore, as a great surprise when the simulations in the early 1990s failed to locate the anticipated liquid–gas critical point [13–15]. Instead, as the temperature was lowered, the simulations found chains of aligned dipoles. Formation of weakly interacting chains, a consequence of strong positional and directional correlations between the dipolar particles, prevented the liquid–gas phase separation from taking place [16–19].

Electrostatic correlations are also crucial in charged colloidal suspensions [20, 21]. In these systems the correlations occur on two different levels. First, there are very strong positional correlations between the highly charged colloidal particles and their counterions. These correlations lead to charge renormalization [22] and to screening of Coulomb interactions between the colloidal particles. In water with monovalent counterions, charge renormalization stabilizes the colloidal suspension against phase separation [23, 24]. In the presence of multivalent counterions, however, the layers of condensed counterions on different colloids can become strongly correlated, leading to a net attraction between the like-charged colloids [25, 26], and to the phase separation [27, 28].

A similar kind of behaviour was also observed in a number of important biological systems. Thus, it was noted that like-charged macromolecules can attract each other in solutions containing multivalent counterions. This attraction manifests itself in *in vitro* formation of toroidal aggregates of concentrated DNA [29, 30], similar to the one found in bacteriophage

heads [31], and in bundle formation of F-actin and tobacco mosaic virus [32]. A number of models have been suggested to explain these curious phenomena. The fundamental ingredients in all these models are the electrostatic correlations [33–35].

Although the phenomena described above are quite complex, we can go a long way towards understanding them by considering some surprisingly simple models and theories. In fact, we shall demonstrate that a lot of the physics of electrostatic correlations is contained within the Debye–Hückel (DH) theory [36] introduced 80 years ago as a way of accounting for the unusual thermodynamic properties of strong electrolytes.

Consideration of only simple physical theories in this review is partially pedagogical, designed for a broad audience not necessarily familiar with the complex machinery of correlation functions and field theories of modern statistical mechanics. For Coulomb systems there is, however, an additional benefit. It is often found that the more sophisticated theories fail when applied to strongly correlated charged fluids. For example, the field theoretic calculations of Netz and Orland [37] find that for charge-asymmetric ($z:1$) electrolytes the reduced critical temperature is a strongly increasing function of charge asymmetry. A completely opposite behaviour is observed in computer simulations—the critical temperature decreases and the critical density increases with the charge asymmetry. In fact, the field theoretic predictions for the critical temperature of asymmetric electrolytes are so far off that they have to be divided by a factor of six just to make them fit on the same graph (figure 1) with the results of simulations and of concurrent theories! The dramatic failure of field theoretic calculations can be attributed to their intrinsically perturbative nature. Similarly, the integral equations, which have proven to be very successful for simple molecular fluids, fail to even converge for strongly asymmetric electrolytes. Furthermore, it has been known for a long time that the hypernetted chain (HNC) equation which is often used to study the Coulomb systems [38], does not posses a true critical region, but only a 'no solution zone' on the border of which compressibility goes to zero with a square root singularity [39, 40]. This behaviour is completely wrong since the compressibility must diverge at the critical point. All these should be contrasted with the physically based DH-like theories, which are in qualitative and often in quantitative agreement with the simulations and experiments (figure 1). For some important many-body systems, however, the integral equations provide the most accurate results. For example, predictions for the electrostatic free energy of the one component plasma (OCP) obtained using the HNC equation are in excellent agreement with the Monte Carlo simulations [44]. The integral equations were also the first to account for the correlation-induced attraction between the like-charged macromolecules [25, 45, 46]. In general, as long as one stays away from the phase transitions, the integral equations provide one of the sharpest tools available to a statistical physicist or chemist. Unfortunately, the approximations involved in constructing the integral equations are not very clear. There exists a great variety of closures to the Ornstein–Zernike equation, each working well for specific kinds of problems. Because of their complexity, we will not talk about the integral equations in this review, referring the interested reader to [47].

2. Electrolyte solutions

Since the work of Faraday early in the nineteenth century the flow of electricity has been associated with the movement of charged particles. The nature of ions (from Greek meaning wanderers), however, was not established. It was Arrhenius who, in 1887, following the experimental work of van't Hoff on osmotic pressure of electrolyte solutions, proposed that when salts and acids are dissolved in water they become ionized [48]. Arrhenius suggested that NaCl dissociates forming cations Na^+ and anions Cl^-. In the spirit of the mean-field theory introduced earlier by van der Waals [1], Arrhenius argued that since the anions and cations are

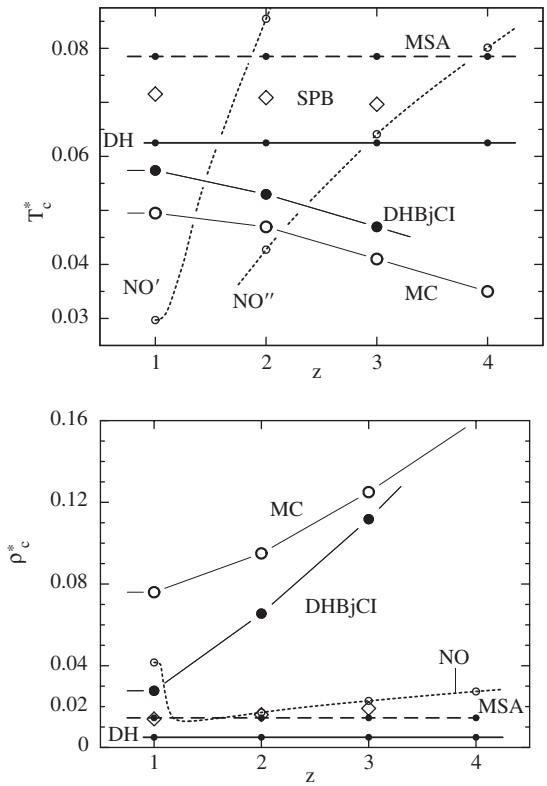

Figure 1. Estimates of the reduced critical temperature, T_c^*, and density, ρ_c^*, for the $(z : 1)$, charge-asymmetric (but equisized) primitive model showing, as labelled, the predictions of pure DH theory (without hard cores), (*b*) the MSA, (*c*) the SPB approximation [41], (*d*) the Netz-Orland field-theoretic treatment [37] which, for T_c^*, have been divided by factors of 6 and 12 in order to bring them on to the plot (see labels NO′ and NO″, respectively), and (*e*) the DHBjCI theory (following the Fisher–Levin approach [5, 6]), point $z = 3$ is a preliminary result. The large open circles for $z = 2$ and 3 represent, the Monte Carlo simulations of Panagiotopoulos and Fisher [42], while the estimate for $T_c^*(4)$ follows from Camp and Patey [43]. (After Banerjee S and Fisher M E unpublished.)

on average uniformly distributed throughout the solution, the average electrostatic potential inside the electrolyte is zero. He then concluded that on average there should not be any interaction between the ions, and the osmotic pressure of, say, a 1 M solution of NaCl should be equivalent to the osmotic pressure of a 2 M solution of a non-electrolyte. All the deviations from this simple rule Arrhenius attributed to the incomplete dissociation of electrolyte, which he then treated as a problem of chemical equilibrium, the thermodynamics of which had already been developed by Gibbs ten years earlier [49]. Soon, however, it became clear that while the theory was working well for weak electrolytes, such as Brønsted acids and bases, it failed for strong electrolytes such as NaCl and HCl, which remained fully ionized even at fairly large concentrations. The disagreements between the theory and the experiment could not be accounted for by the postulate of incomplete dissociation. It appeared that there was a fundamental flaw in the theory advanced by Arrhenius, which relied on the mean-field assumption of non-interacting ions. The situation remained unclear for about 30 years, with various proposals made on how to incorporate the ionic interactions into the framework of Arrhenius' theory [49]. None of these proved very successful at explaining the

experimental measurements, until Debye and Hückel published their, now famous, theory of strong electrolytes [36]. The fundamental insight of Debye and Hückel (DH) was to realize that although ions *on average* are randomly distributed, there exist strong positional correlations between the anions and cations. The depth of Debye's insight can be judged from the fact that he understood the role of electrostatic correlations significantly before the correlation functions became the standard tool of working physicists. Since so many of our results will be based on the fundamental ideas of Debye and Hückel, their theory will provide the starting point for our discussion of thermodynamics of electrostatic correlations.

2.1. The Debye–Hückel theory

Consider the simplest model of an electrolyte solution confined to volume V. The N ions will be idealized as hard spheres of diameter a carrying charge $\pm q$ at their centres. The charge neutrality of the solution requires that $N_+ = N_- = N/2$. The solvent will be modelled as a continuum of dielectric constant ϵ. Although the average potential inside the electrolyte is zero, there are strong positional correlations between the oppositely charged ions. It is convenient to work in spherical coordinates. To calculate the correlational contribution to the Helmholtz free energy, let us fix one ion of charge $+q$ at the origin $r = 0$ and see how the other ions will distribute around it (see figure 2). Inside the region $0 < r \leqslant a$ there are no other charges except for the one fixed at the origin, and the electrostatic potential $\phi(r)$ satisfies the Laplace equation,

$$\nabla^2\phi = 0. \tag{1}$$

For $r > a$ the electrostatic potential satisfies the Poisson equation

$$\nabla^2\phi = -\frac{4\pi}{\epsilon}\rho_q(r), \tag{2}$$

where the charge density can be expressed in terms of the charge–charge correlation functions $g_{++}(r) = g_{--}(r)$ and $g_{+-}(r) = g_{-+}(r)$:

$$\rho_q(r) = q\rho_+ g_{++}(r) - q\rho_- g_{+-}(r). \tag{3}$$

The average densities of positive and negative ions are $\rho_+ = N_+/V$, $\rho_- = N_-/V$; $\rho_+ = \rho_- = \rho/2$.

The correlation functions can be written in terms of the potential of mean force w_{ij}:

$$g_{ij}(r) = \mathrm{e}^{-\beta w_{ij}(r)}, \tag{4}$$

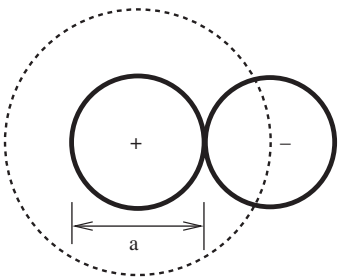

Figure 2. The configuration of closest approach between two oppositely charged ions. The dashed curve delimits the region into which no ions can penetrate, due to the hardcore repulsion.

where $\beta = 1/k_B T$. $w_{ij}(r)$ is the work required to bring ions i and j from infinity to separation r inside the electrolyte solution. In their paper, Debye and Hückel made an implicit approximation of replacing the potential of mean force by the electrostatic potential

$$w_{ij}(r) = q_j \phi_i(r), \tag{5}$$

where q_j is the charge of the jth ion and $\phi_i(r)$ is the electrostatic potential at distance r from the ion i fixed at the origin $r = 0$. With this approximation, equation (2) reduces to the non-linear Poisson–Boltzmann (PB) equation,

$$\nabla^2 \phi = -\frac{4\pi}{\epsilon}\left[q\rho_+ e^{-\beta q\phi} - q\rho_- e^{+\beta q\phi}\right] = \frac{4\pi\rho q}{\epsilon}\sinh(\beta q\phi). \tag{6}$$

Debye and Hückel proceeded to linearize this equation. Technically, linearization is only valid if $\beta q\phi \ll 1$; however, being practically minded, Debye and Hückel linearized first and worried about the consequences later. As was noted later by Onsager, linearization of equation (6) is a necessary step in order to produce a self-consistent theory [50]. The linearized PB equation reduces to the Helmholtz equation,

$$\nabla^2 \phi = \kappa^2 \phi, \tag{7}$$

where the inverse Debye length is

$$\xi_D^{-1} \equiv \kappa = \sqrt{\frac{4\pi q^2 \rho}{k_B T \epsilon}}. \tag{8}$$

The Laplace equation (1) for $r \leqslant a$ and the Helmholtz equation (7) for $r > a$ must be integrated, subject to the boundary condition of continuity of the electrostatic potential and the electric field across the boundary surface $r = a$. For $r \leqslant a$ the electrostatic potential is found to be

$$\phi_<(r) = \frac{q}{\epsilon r} - \frac{q\kappa}{\epsilon(1+\kappa a)}, \tag{9}$$

while for $r > a$,

$$\phi_>(r) = \frac{q\theta(\kappa a)e^{-\kappa r}}{\epsilon r}, \qquad \theta(x) = \frac{e^x}{1+x}. \tag{10}$$

Equation (10) shows that the electrostatic potential produced by the central charge is exponentially screened by the surrounding ionic cloud. Because of the hardcore repulsion the screening, however, appears only at distances larger than $r = a$. This accounts for the presence of the geometric factor $\theta(\kappa a)$ in equation (10). The screening of electrostatic interactions inside the electrolyte solutions and plasmas is responsible for the existence of the thermodynamic limit in these systems with extremely long-range forces.

The electrostatic potential $\phi_<(r)$, equation (9), consists of two terms: the potential produced by the central ion $q/\epsilon r$, and the electrostatic potential induced by the surrounding ionic cloud,

$$\psi = -\frac{q\kappa}{\epsilon(1+\kappa a)}. \tag{11}$$

The electrostatic free energy can now be obtained using the Debye charging process in which all the ions are simultaneously charged from zero to their full charge:

$$F^{el} = Nq\int_0^1 d\lambda\, \psi(\lambda q). \tag{12}$$

The calculation is very similar to the one used to obtain the electrostatic energy stored in a capacitor. While performing the charging it is important to remember that $\kappa(\lambda q) = \lambda\kappa(q)$.

Defining the free energy density as $f = F/V$, the integral in equation (12) can be performed explicitly yielding

$$\beta f^{\mathrm{el}} = \frac{\beta F^{\mathrm{el}}}{V} = -\frac{1}{4\pi a^3}\left[\ln(\kappa a + 1) - \kappa a + \frac{(\kappa a)^2}{2}\right]. \tag{13}$$

For large dilutions equation (13) reduces to the famous Debye limiting law,

$$\beta f^{\mathrm{el}} \approx -\frac{\kappa^3}{12\pi} \sim -\left(\frac{\rho}{T}\right)^{3/2}. \tag{14}$$

Given the free energy, the limiting laws for the osmotic pressure and activity can be easily found [6].

The free energy is not analytic at $\rho = 0$. The singularity at $\rho = 0$ is a consequence of long-range Coulomb interactions, which also manifest themselves in the divergence of the standard virial expansion [51]. The total free energy of the electrolyte F is the sum of electrostatic, equation (13), and entropic contributions. The entropic contribution to the free energy arises from the integration over the momentum degrees of freedom in the partition function, and is equivalent to the free energy of an ideal gas:

$$\beta F^{\mathrm{ent}} = N_+ \ln[\rho_+ \Lambda^3] - N_+ + N_- \ln[\rho_- \Lambda^3] - N_- = N \ln\left[\frac{\rho\Lambda^3}{2}\right] - N, \tag{15}$$

where the de Broglie thermal wavelength is

$$\Lambda = \frac{h}{\sqrt{2\pi m k_{\mathrm{B}} T}}. \tag{16}$$

The osmotic pressure of the electrolyte is

$$P = -\frac{\partial F}{\partial V}\Big|_N, \tag{17}$$

which can also be expressed in terms of the Legendre transform of the negative free energy density $-f$ [6]:

$$P = -f + \mu\rho, \tag{18}$$

where the chemical potential is

$$\mu = \frac{\partial F}{\partial N}\Big|_V = \frac{\partial f}{\partial \rho}. \tag{19}$$

It is a simple matter to see that below the critical temperature T_{c} the total free energy $F = F^{\mathrm{ent}} + F^{\mathrm{el}}$ fails to be a convex function of the electrolyte concentration. This implies the presence of a phase transition. Alternatively, the phase separation can be observed from the appearance of a van der Waals loop in the osmotic pressure equation (18), below the critical temperature T_{c}. The critical parameters are determined from

$$\frac{\partial P}{\partial \rho} = 0, \tag{20}$$

$$\frac{\partial^2 P}{\partial \rho^2} = 0. \tag{21}$$

The coexistence curve can be obtained using the standard Maxwell construction. It is convenient to define the reduced temperature and density as $T^* = k_{\mathrm{B}} T a\epsilon/q^2$ and $\rho^* = \rho a^3$. The critical point of the plasma, within the DH theory, is found to be located at [5, 6]

$$T_{\mathrm{c}}^* = \tfrac{1}{16}, \tag{22}$$

and

$$\rho_{\mathrm{c}}^* = \frac{1}{64\pi}. \tag{23}$$

It is interesting to note that at criticality, $\kappa = 1/a$. This means that in spite of a very low concentration of electrolyte at the critical point, the screening remains very strong. We also observe that the reduced critical temperature for the electrolyte is almost an order of magnitude lower, than for systems in which the particles interact by the short-ranged isotropic potentials. Since the critical point within the DH theory occurs at extremely low density, we are justified in neglecting the excluded volume contribution to the total free energy.

Phase separation of an electrolyte or of a two component plasma is the result of an electrostatic instability arising from the strong positional correlations between the oppositely charged ions. This mechanism is very different from the one driving the phase separation in systems dominated by the short-ranged isotropic forces. In that case the thermodynamic instability is a consequence of the competition between the interparticle attraction and the hardcore repulsion.

The reduced temperature can be written as $T^* = a/\lambda_B$, where $\lambda_B = q^2/k_B T\epsilon$ is the Bjerrum length. For water at room temperature $\lambda_B \approx 7$ Å. This means that one would need ions of size less than 0.4 Å, in order to observe phase separation at room temperature. This is clearly impossible since the minimum hydrated ionic size is about 2–4 Å. Therefore, in order to see phase separation, one is required to look for solutions with λ_B of the order of 40 Å or more. For water, such large values of λ_B correspond to temperatures well below freezing. An alternative is to work with organic solvents which have dielectric constants significantly lower than water. This was the strategy adopted by Pitzer in his studies of ionic criticality [52–54]. Pitzer used liquid salt triethyl-n-hexylammonium triethyl-n-hexylboride ($N_{2226}B_{2226}$) in the diphenyl ether. With this he was able to observe the critical point at room temperature. Pitzer's work has provoked a lot of stimulating controversy because his measurements suggested that the Coulombic criticality belonged to a new universality class [55]. At first sight this might not seem very surprising—after all the Coulomb force is extremely long ranged. On further reflection, the situation is not so clear. Although the bare interaction potential between any two ions is long ranged, inside the electrolyte solution it is screened by the surrounding particles, as is seen from equation (10). The effective interaction potential, therefore, is short ranged, which should place the ionic criticality firmly in the Ising universality class. In fact all the theoretical arguments lead to this conclusion, which seems to be contradicted by Pitzer's experiments. In principle, it is possible that one has to be very close to the critical point before the Ising behaviour becomes apparent. However, even this conclusion is hard to justify theoretically. Estimates of the Ginzburg criterion suggest that the width of the critical region for the Coulombic criticality should be comparable to that of systems with short ranged isotropic interactions [5, 56]. The situation remains unclear.

An alternative to working with electrolyte solutions is to study molten salts, which are classical two component plasmas. In this case the dielectric constant can be taken to be that of vacuum, and ions are no longer hydrated. The reduced critical temperature, $T^* = 1/16$, and the characteristic ionic diameter of about 2 Å, imply that at criticality $\lambda_B \approx 30$, which means that the critical point for a molten salt is located at about 5000 K. It is, indeed, very hard to study critical phenomena at such high temperatures! It seems, therefore, that we are stuck with the low dielectric solvents. An alternative is computer simulations, which are becoming sufficiently accurate to allow measurements of the critical exponents, at least for symmetric 1 : 1 electrolytes. Indeed, the most recent simulations suggest that the Coulombic criticality belongs to the Ising universality class [57].

2.2. The Bjerrum association

The DH theory presented in the previous section was based on the linearization of the PB equation. In view of the strong screening and the rapid decrease of the electrostatic

potential away from the central ion, such a linearization can be justified at intermediate and long distances. It is clear, however, that the linearization strongly diminishes the weight of configurations in which two oppositely charged ions are in close proximity. Linearization underestimates the strength of electrostatic correlations which result in dipole-like structures. At low reduced temperatures characteristic of the critical point, these configurations should be quite important and must be taken into account. One way of doing this, while preserving the linearity of the theory, is to postulate the existence of dipoles with concentration governed by the law of mass action. In the leading-order approximation the dipoles can be treated as ideal non-interacting species [58–60]. The total number of particles $N = \rho V$ is then subdivided into monopoles $N_1 = \rho_1 V$ and dipoles $N_2 = \rho_2 V$. The particle conservation requires that, $N = N_1 + 2N_2$. The free energy of the mixture is $F = F_1^{\rm ent} + F_2^{\rm ent} + F^{\rm el}$, where $F^{\rm el}$ and $F_1^{\rm ent}$ are the entropic and the electrostatic free energies of the monopoles, given by equations (13) and (15), but with $N \to N_1$ and $\rho \to \rho_1$. The entropic free energy of dipoles is

$$\beta F_2^{\rm ent} = N_2 \ln\left[\frac{\rho_2 \Lambda^6}{\zeta_2}\right] - N_2, \tag{24}$$

where the internal partition function of a dipole is

$$\zeta_2(R) = 4\pi \int_a^R r^2 \, {\rm d}r \exp\left(\frac{\beta q^2}{\epsilon r}\right). \tag{25}$$

At low temperatures, the precise value of the cut-off R at which the two ions can be considered to be associated is not very important. Following the original suggestion of Bjerrum [58] we can take this value to be the inflection point of the integral in equation (25), $R_{\rm Bj} = \lambda_{\rm B}/2$. This choice corresponds to the minimum of the integrand in equation (25), which in turn can be interpreted as the probability of finding two oppositely charged ions at the separation r. The minimum then corresponds to a liminal between bound and unbound configurations. A much more careful analysis of the dipolar partition function has been carried out by Falkenhagen and Ebeling based on the re-summed virial expansion [60]. They found that that the low temperature expansion of the Bjerrum equilibrium constant is identical to the equilibrium constant which can be constructed on the basis of the re-summed virial expansion. Since we are interested in the low temperature regime where the critical point is located, the Bjerrum equilibrium constant, $\zeta_2 \equiv \zeta_2(R_{\rm Bj})$, will be sufficient.

It is important to keep in mind that at this level of approximation the electrostatic free energy $F_1^{\rm ent}$ is only a function of the density of free unassociated ions ρ_1, since the dipoles are treated as ideal non-interacting species. The concentration of dipoles is obtained from the law of mass action,

$$\mu_2 = \mu_+ + \mu_-, \tag{26}$$

where the chemical potential of a species s is

$$\mu_s = \left.\frac{\partial F}{\partial N_s}\right|_V. \tag{27}$$

Substituting the expression for the total free energy into the law of mass action leads to

$$\rho_2 = \tfrac{1}{4}\rho_1^2 \zeta_2 \, {\rm e}^{2\beta\mu^{\rm ex}}, \tag{28}$$

where the excess chemical potential is $\mu^{\rm ex} = \partial f^{\rm el}/\partial\rho_1$. The critical point can be located from the study of the convexity of the total free energy as a function of ion concentration ρ. There is, however, a simpler way [6]. We observe that at the Bjerrum level of approximation, dipoles are ideal non-interacting species. This means that they are only present as spectators and do not interact with the monopoles in any way. This implies that only the monopoles can

drive the phase separation. Thus, at the critical point the temperature must still be $T_c^* = 1/16$ and the density of monopoles must still remain $\rho_{1c}^* = 1/64\pi$, as in the case of the pure DH theory. The corresponding density of dipoles at criticality is then given by equation (28), with $T_c^* = 1/16 = 0.0625$ and $\rho_{1c}^* = 1/64\pi = 0.00497$. We find that, at the critical point, the density of dipoles is $\rho_{2c}^* \approx 0.02$. In the vicinity of the critical point there are many more dipoles than monopoles, $\rho_{2c}^*/\rho_{1c}^* \approx 4$. Within the Bjerrum approximation the non-linear correlations, in the form of dipoles, do not affect the critical temperature, but strongly modify the critical density, $\rho_c^* = \rho_{1c}^* + 2\rho_{2c}^* = 0.045$. In spite of the crudeness of approximations, the location of the critical point agrees reasonably well with the Monte Carlo simulations [61–63], $T_c^* = 0.051$ and $\rho_c^* = 0.079$. The coexistence curve, however, is found to have an unrealistic 'banana' shape [6]. To correct this deficiency one must go beyond the 'ideal' dipole approximation and allow for the dipole–ion interaction [5, 6]. Most of the fundamental physics of electrostatic correlations, however, is already captured at the level of the Bjerrum approximation.

3. Two-dimensional plasma and the Kosterlitz–Thouless transition

An electrostatic system which over the years has attracted much attention is the two-dimensional plasma of positive and negative ions interacting through a logarithmic potential. The great interest in two-dimensional plasma is due to the fact that various important physical systems can be mapped directly onto it. Examples include superfluid ^{4}He films, two-dimensional crystalline solids, and XY magnets [64]. Although a continuous symmetry cannot be broken in two dimensions [65], if the Hamiltonian of a system is invariant under an Abelian group, a finite temperature phase transition is possible. This transition occurs as the result of unbinding of the topological defects or 'charges'. The defect-mediated phase transitions belong to the universality class of a two-dimensional plasma.

Thirty years ago Kosterlitz and Thouless (KT) presented a renormalization group study of the two-dimensional plasma [66]. They concluded that at sufficiently low temperatures, the two-dimensional plasma becomes an insulator. All the positive and negative ions pair up forming dipoles. The metal–insulator transition was found to be of an infinite order, characterized by essential singularities in the thermodynamic functions. The KT analysis, however, was restricted to the low ionic densities and it is not clear what happens when the concentration of charged particles is increased. It is tempting to apply to the two-dimensional plasmas, an analysis similar to the one presented earlier for the three-dimensional electrolytes [67].

We shall, therefore, study a fluid of discs with diameter a and charge $\pm q$. The solvent is a uniform medium of dielectric constant ϵ. The bare interaction potential for two ions (i, j) separated by distance r is

$$\varphi(r) = -\frac{q_i q_j}{\epsilon} \ln\left(\frac{r}{a}\right). \tag{29}$$

As in the case of three-dimensional electrolytes, the mean-field contribution to the electrostatic free energy is zero, and all the important physics comes from the electrostatic correlations. We shall account separately for the long-ranged and the short-ranged correlations. The short-ranged correlations lead to the formation of dipolar pairs of density ρ_2, while the long-ranged correlations produce the screening. As in the case of three-dimensional electrolytes, the total density of hard discs is divided between the dipoles and the monopoles, so that $\rho = \rho_1 + 2\rho_2$.

To calculate the electrostatic free energy, we fix one ion and study the distribution of other particles around it. It is important to recall that the Poisson equation in two-dimensions,

$$\nabla^2\phi(r) = -\frac{2\pi}{\epsilon}\rho_q(r), \tag{30}$$

differs from the one in three-dimansions by the normalization factor [67], 2π has replaced the 4π of the three-dimensional Poisson equation.

As before, we shall approximate the potential of mean force by the electrostatic potential and then linearize the Boltzmann factor. Linearization is compensated by the allowance for dipolar formation.

The electrostatic potential for distances $r \leqslant a$ satisfies the Laplace equation $\nabla^2\phi = 0$, while for $r > a$ the potential satisfies the Helmholtz equation $\nabla^2\phi = \kappa^2\phi$, with

$$\kappa = \sqrt{\frac{2\pi q^2 \rho_1}{k_B T \epsilon}}. \tag{31}$$

This equation can be easily integrated yielding the electrostatic potential:

$$\phi_<(r) = -\frac{q}{\epsilon}\ln\left(\frac{r}{a}\right) + \frac{q}{\epsilon}\frac{K_0(\kappa a)}{\kappa a K_1(\kappa a)} \qquad \text{for } \leqslant a, \tag{32}$$

and

$$\phi_>(r) = \frac{q}{\epsilon}\frac{K_0(\kappa r)}{\kappa a K_1(\kappa a)} \qquad \text{for } r > a, \tag{33}$$

where $K_\nu(x)$ are the modified Bessel functions of order ν. For large distances, the electrostatic potential decays exponentially fast. Just as in three dimensions, the electrostatic interactions are screened inside the two-dimensional plasma:

$$\lim_{r\to\infty}\phi_>(r) \approx \frac{q\mathrm{e}^{-\kappa r}}{\epsilon\kappa a K_1(\kappa a)}\sqrt{\frac{\pi}{2\kappa r}}. \tag{34}$$

Equation (32) consists of two terms, the potential produced by the fixed ion and the induced potential due to the ionic atmosphere,

$$\psi = \frac{q}{\epsilon}\frac{K_0(\kappa a)}{\kappa a K_1(\kappa a)}. \tag{35}$$

Given the induced potential, the electrostatic free energy can be obtained using the familiar Debye charging process, equation (12). We find the electrostatic free energy density of a two-dimensional plasma to be

$$\beta f^{\mathrm{el}} = \frac{1}{2\pi a^2}\ln[\kappa a K_1(\kappa a)]. \tag{36}$$

At the Bjerrum level of approximation, the electrostatic free energy depends only on the density of monopoles. The total free energy density is $f = f_1^{\mathrm{ent}} + f_2^{\mathrm{ent}} + f^{\mathrm{el}}$, where

$$\beta f_1^{\mathrm{ent}} = \rho_1 \ln\left[\frac{\rho_1\Lambda^2}{2}\right] - \rho_1 \tag{37}$$

and

$$\beta f_2^{\mathrm{ent}} = \rho_2 \ln\left[\frac{\rho_2\Lambda^4}{\zeta_2}\right] - \rho_2. \tag{38}$$

The internal partition function for a two-dimensional dipole is

$$\zeta_2(R) = 2\pi\int_a^R r\,\mathrm{d}r \exp\left[-\frac{\beta q^2}{\epsilon}\ln\left(\frac{r}{a}\right)\right]. \tag{39}$$

It is convenient to define the reduced temperature and density as $T^* = k_B T\epsilon/q^2$ and $\rho^* = \rho a^2$. We note that for low temperatures, $T^* < 1/2$, the integral in equation (39) converges uniformly

as $R \to \infty$. In this regime it is possible, therefore, to define the internal partition function of a dipole as

$$\zeta_2 \equiv \zeta_2(\infty) = \frac{2\pi a^2 T^*}{1 - 2T^*}. \tag{40}$$

The thermodynamic equilibrium requires that for fixed volume and number of particles the Helmholtz free energy be minimum. This is equivalent to the law of mass action equation (26), which upon the substitution of free energy simplifies to equation (28). In the limit of small concentrations, the excess chemical potential can be expanded in powers of ρ_1 yielding

$$\beta\mu^{\mathrm{ex}} = -\frac{1}{2T^*}\left[\gamma_{\mathrm{E}} + \ln\left(\frac{\kappa a}{2}\right)\right], \tag{41}$$

where γ_{E} is the Euler constant. Substituting equation (41) into equation (28), we find that the concentration of dipoles in the limit $\rho \to 0$ scales as

$$\rho_2 \sim \rho_1^{\theta(T^*)}, \tag{42}$$

where

$$\theta(T^*) = 2 - \frac{1}{2T^*}. \tag{43}$$

For $T^* < 1/4$, the exponent $\theta(T^*) < 0$, and in the limit $\rho_1 \to 0$ the law of mass action cannot be satisfied. This means that in the temperature density plane (T^*, ρ^*), for sufficiently small densities, the line $T^* = 1/4$ corresponds to the critical locus of metal–insulator transitions. Below this line, and for sufficiently small ionic concentrations, no free monopoles can exist. All the ions are paired up into neutral dipolar pairs. The critical line terminates at the tricritical point located at $T^*_{\mathrm{KT}} = 1/4$ and

$$\rho^*_{\mathrm{tri}} = \frac{\mathrm{e}^{-4\gamma_{\mathrm{E}}}}{8\pi} \simeq 0.003\,954. \tag{44}$$

For $T^* < 1/4$ and $\rho^* > \rho^*_{\mathrm{tri}}$ there is a phase separation between an insulating vapour and a conducting liquid phase, figure 3. As the critical line is approached from high temperatures,

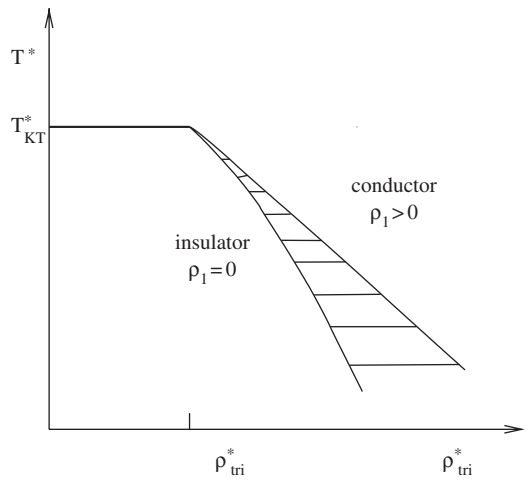

Figure 3. Phase diagram for the two-dimensional plasma within the Debye–Hückel–Bjerrum approximation. We expect the fluctuations to renormalize the KT line, shifting it from its horizontal position and making it density dependent. The topology of the phase diagram should, however, remain the same.

the Debye length diverges as

$$\xi_D \equiv \kappa^{-1} \sim e^{c(\rho)/t^\nu}, \tag{45}$$

where

$$c(\rho) = \frac{1}{4} \ln\left(\frac{\rho_{tri}}{\rho}\right), \tag{46}$$

and

$$t = \frac{T - T_{KT}}{T}. \tag{47}$$

The critical exponent is $\nu = 1$. The KT renormalization group theory [66] predicts the same behaviour for ξ_D except that $\nu = 1/2$. The KT theory, however, leaves unanswered the question of what happens to the metal–insulator transition for higher plasma concentrations. The current theory, on the other hand, shows that the critical line terminates in a tricritical point, after which the metal–insulator transition becomes first order [67]. This topology is also consistent with the findings of Monte Carlo simulations [68, 69]. A more sophisticated theory introduced by Minhagen [70], leads to a very similar phase diagram, except that the tricritical point is replaced by a critical end-point.

We see that the electrostatic correlations are even more important in two-dimensions than in three dimensions. While in three dimensions the electrolyte phase separates into the coexisting liquid and gas phases, both of which contain monopoles and dipoles, in two dimensions the low density vapour phase does not contain any free charges and is an insulator.

4. The one component plasma

The OCP is probably the simplest model of a Coulomb system. It consists of point ions, all of the same sign, inside a rigid neutralizing background of opposite charge. In spite of its simplicity the model is relevant for many physical systems. Some examples are the interiors of stars, liquid metals and magnetically confined electrons. As an approximation, OCP is particularly important since it provides a framework in which ionic correlations can be calculated. Thus, the electrostatic free energy of a homogeneous OCP can be used to account for the counterion correlations in colloidal suspensions and polyelectrolyte solutions. The OCP has been extensively studied over the years. A review of the subject, which still remains very actual, has been presented by Baus and Hansen some twenty years ago [71].

In the spirit of the current work we shall, however, confine our attention to simple analytical theories of the OCP [72, 73]. Our model consists of N ions each carrying charge q inside a uniform neutralizing background of dielectric constant ϵ and volume V. The average ionic density is $\rho = N/V$. The mean electrostatic potential inside the OCP is zero and the free energy is, once again, entirely due to positional correlations between the ions. The advantage of working with the OCP is that it contains only two independent length scales: the average separation between the particles $d = (4\pi\rho/3)^{-1/3}$ and the Bjerrum length $\lambda_B = q^2/k_B T\epsilon$. There is only one dimensionless parameter on which all the thermodynamic quantities depend, $\Gamma = \lambda_B/d$. This is quite distinct from the electrolytes and two component plasmas, for which besides d and λ_B there is a third length scale—the ionic diameter a—and the thermodynamics is parametrized by two dimensionless quantities T^* and ρ^*. For electrolytes the limit $a \to 0$ does not exist, since for point particles the free energy can be lowered indefinitely by collapsing the system into point-like dipolar pairs.

The calculation of free energy of the OCP proceeds along the lines of the DH theory. We fix one ion and study the distribution of other ions around it. The electrostatic potential

satisfies the Poisson equation, equation (2), with the charge density given by

$$\rho_q(r) = q\rho g(r) - q\rho. \tag{48}$$

The correlation function can be expressed in terms of the potential of mean force $w(r)$ as

$$g(r) = \mathrm{e}^{-\beta w(r)}. \tag{49}$$

In the spirit of the DH theory we replace $w(r)$ by $q\phi(r)$. This approximation entails neglect of electrostatic correlations inside the ionic cloud which surrounds the central ion. The approximation should be quite good as long as Γ is not too large. As the next step, linearization of the Boltzmann factor leads to the charge density of particularly simple form,

$$\rho_q(r) = -\frac{\epsilon\kappa^2}{4\pi}\phi(r), \tag{50}$$

where $\kappa^2 = 4\pi\beta q^2\rho/\epsilon$. Substituting this into the Poisson equation we, once again, find the familiar Helmholtz equation (7). This can be easily integrated yielding the electrostatic potential of Yukawa form,

$$\phi(r) = \frac{q\,\mathrm{e}^{-\kappa r}}{\epsilon r}. \tag{51}$$

The charge density must be bounded from below, $\rho_q \geqslant -q\rho$. However, considering equations (50) and (51), it is evident that this condition is violated for sufficiently small separations from the central ion. Something must have gone seriously wrong. It is easy to trace the problem to the linearization of the Boltzmann factor. Clearly at short distances linearization is not justified since the strong electrostatic repulsion between the ions results in a very large electrostatic energy. It appears that in order to understand the physics of the OCP one has to solve the full non-linear PB equation, a task which cannot be performed analytically. Fortunately not everything is lost. A way out was suggested by Nordholm, who noted that the strong repulsion between like-charged ions results in an effective hole surrounding the central ion [72]. Very few ions can penetrate inside the hole since this costs them too much electrostatic energy. Inside this correlation hole, $r \leqslant h$, the electrostatic potential satisfies the Poisson equation (2) with a uniform background charge density, $\rho_q \approx -q\rho$. In the outside region, where the electrostatic interaction is much weaker, the linearization of the Boltzmann factor is justified and the potential satisfies the Helmholtz equation. Therefore, for $r > h$ the electrostatic potential still has a Yukawa form, but with an undetermined prefactor,

$$\rho_q(r) = \frac{A\,\mathrm{e}^{-\kappa r}}{r}. \tag{52}$$

The coefficient A can be obtained from the condition of continuity of $\rho_q(r)$ across $r = h$. We find

$$A = -q\rho h\,\mathrm{e}^{\kappa h}. \tag{53}$$

Charge neutrality requires that

$$q - \frac{4\pi}{3}q\rho h^3 + 4\pi\int_h^\infty r^2\,\mathrm{d}r\,\rho_q(r) = 0. \tag{54}$$

The integral can be performed explicitly and the resulting equation solved to determine the size of the correlation hole:

$$h = \frac{d[\omega(\Gamma) - 1]}{\sqrt{3\Gamma}}, \tag{55}$$

where

$$\omega(\Gamma) = \left[1 + (3\Gamma)^{3/2}\right]^{1/3}. \tag{56}$$

The size of the correlation hole is a monotonically increasing function of the coupling strength. For high temperatures, small couplings, $h \approx \lambda_B$. This is exactly what one might have expected, since on this length scale the electrostatic repulsion becomes comparable to the thermal energy. When the temperature is lowered the kinetic energy diminishes and the particles are strongly scattered by the electrostatic repulsion, causing an increase in the size of the correlation hole. For very low temperatures, large Γ, the size of a correlation hole is equal to the average spacing between the particles, $h \approx d$. This is, once again, the correct limiting behaviour since at low temperatures the ions tend to keep as far away as possible from their neighbours. For $r > h$ the electrostatic potential follows directly from equation (50),

$$\phi_>(r) = \frac{4\pi q\rho h\, \mathrm{e}^{-\kappa(r-h)}}{\epsilon\kappa^2 r}. \tag{57}$$

For $r \leqslant h$ the Poisson equation with a uniform background charge $\rho_q = -q\rho$ must be solved. The solution is easily found to be

$$\phi_<(r) = \frac{q}{\epsilon r} + \frac{2\pi q\rho r^2}{3\epsilon} + \psi. \tag{58}$$

The induced potential ψ can be determined from the condition of continuity of electrostatic potential $\phi_>(h) = \phi_<(h)$, which reduces to

$$\psi = -\frac{k_B T}{2q}\{[1 + (3\Gamma)^{3/2}]^{2/3} - 1\}. \tag{59}$$

The Debye charging process equation (12), yields the electrostatic free energy per particle

$$\frac{\beta F^{\mathrm{el}}}{N} = \frac{1}{4}\left[1 - \omega^2 + \frac{2\pi}{3\sqrt{3}} + \ln\left(\frac{\omega^2+\omega+1}{3}\right) - \frac{2}{\sqrt{3}}\tan^{-1}\left(\frac{2\omega+1}{\sqrt{3}}\right)\right], \tag{60}$$

where $\omega(\Gamma)$ is given by equation (56). In the limit of high temperatures, $\Gamma \to 0$, equation (60) reduces to the Debye limiting law, equation (14). Furthermore, the free energy agrees with the Monte Carlo simulations with an error of less than 10% over a wide range of coupling strengths, $0 < \Gamma < 80$. This suggests that the inclusion of a correlation hole into the DH theory captures most of the essential physics of the OCP.

Some caution must be taken when using the OCP to model real physical systems. For $\Gamma > 3$, the isothermal compressibility and pressure of the OCP become negative [71]. This is a consequence of treating the background as a rigid entity and neglecting its pressure. How this can be corrected, in practice, depends on the kind of problem that one wants to study. If one wants to use the OCP to model dense ionized matter, the suitable background is the degenerate electron gas. When the free energy of background is added to the OCP the pressure and the compressibility remain non-negative for all values of Γ. An alternative approach was suggested by Weeks, who defined the OCP as the classical 'dense-point limit' of a two component plasma [74]. In this limit, the background is treated as an infinitely dense cloud of point particles each carrying an infinitesimal charge e, so that $e\rho_{\mathrm{back}}$ remains constant. Presence of such background does not affect the electrostatics of the OCP, but regularizes its pressure and isothermal compressibility, making them non-negative. An interesting by-product of this analysis is the conclusion that the freezing of the OCP occurs without any change in density, i.e. the volume per particle in the fluid and the solid phases is the same [74]. The freezing transition happens at $\Gamma \approx 180$ and the resulting solid phase has the BCC structure [75, 71, 76].

4.1. Confined one component plasmas

In 1971 Crandall and Williams suggested that electrons trapped on the surface of liquid helium, ^{4}He, can crystallize, forming a two-dimensional Wigner crystal [77]. Eight years later this

order–disorder transition was observed experimentally by Grimes and Adams [78]. In this system, electrons obey classical mechanics, since the Fermi energy is much smaller than $k_B T$. Similar crystallization can occur in the inversion layer near the surface of a semiconductor; however, in this case the quantum effects are important and the electrons form a degenerate quantum gas [79].

The trapped electrons above the liquid ^{4}He can be modelled as a confined quasi-two-dimensional plasma of particles interacting by a $1/r$ potential. This model is also appropriate for the study of correlations between the condensed counterions on the surface of colloidal particles.

The average spacing between the confined electrons is $d = (\pi\sigma)^{-1/2}$, where σ is the average surface density, $\sigma = N/A$. The dimensionless quantity parameterizing the strength of electrostatic interactions is $\Gamma = q^2/\epsilon k_B T d$. For an infinitesimally thin layer separating two mediums of dielectric constants ϵ_1 and ϵ_2, the important parameter is the average dielectric constant $\epsilon = (\epsilon_1 + \epsilon_2)/2$. It has been observed in computer simulations [80] that the two-dimensional OCP crystallizes into a triangular Wigner crystal for $\Gamma > 130$. This value is also in close agreement with the experiments of Grimes and Adams.

We can gain much insight into the thermodynamics of two-dimensional OCP using the, now familiar, DH theory. Our model consists of a plasma of point particles of charge q and of a neutralizing background, confined to an interface located at $z = 0$ between the two dielectric half-spaces. For $z < 0$ the dielectric constant is ϵ_1 and for $z > 0$ the dielectric constant is ϵ_2. Since the half-spaces do not contain any free charges, the electrostatic potential everywhere satisfies the Laplace equation $\nabla^2\phi = 0$. The electrostatic free energy is obtained by fixing one particle and calculating the induced potential resulting from the redistributions of other ions in the $z = 0$ plane. It is convenient to adopt the cylindrical coordinate system, (ϱ, φ, z), so that the fixed ion is located at $\varrho = 0$, $z = 0$. Using the azimuthal symmetry and the fact that the electrostatic potential vanishes at infinity, the solution to the Laplace equation can be written as [81]

$$\phi_1(\varrho, z) = \int_0^\infty A_1(k) J_0(k\varrho)\, \mathrm{e}^{kz}\, \mathrm{d}k \qquad \text{for } z < 0, \tag{61}$$

and

$$\phi_2(\varrho, z) = \int_0^\infty A_2(k) J_0(k\varrho)\, \mathrm{e}^{-kz}\, \mathrm{d}k \qquad \text{for } z > 0, \tag{62}$$

where $J_0(x)$ is the Bessel function of order zero.

The functions $A_1(k)$ and $A_2(k)$ can be determined from the boundary conditions, which are: continuity of the electrostatic potential,

$$\phi_2(\varrho, 0) = \phi_1(\varrho, 0), \tag{63}$$

and discontinuity of the displacement field across the $z = 0$ plane. The discontinuity results from the inhomogeneous distribution of interfacial charge induced by the fixed ion:

$$[\epsilon_2 \boldsymbol{E}_2(\varrho, z) - \epsilon_1 \boldsymbol{E}_1(\varrho, z)] \cdot \hat{n} = 4\pi\sigma_q(\varrho). \tag{64}$$

From charge neutrality the *average* interfacial charge is zero so that $\sigma_q(\varrho)$ is the result of ionic correlations,

$$\sigma_q(\varrho) = \frac{q\delta(\varrho)}{2\pi\varrho} - q\sigma + q\sigma\, \mathrm{e}^{-\beta q\phi(\varrho,0)}. \tag{65}$$

The first term of equation (65) is the surface charge density of the fixed ion, the second term is due to the uniform negative background, while the last term is the surface charge density of ions confined to the interface. We have, once again, approximated the potential of mean

force by the electrostatic potential. In the spirit of the DH theory we shall now linearize the Boltzmann factor. The surface charge density becomes

$$\sigma_q(\varrho) = \frac{q\delta(\varrho)}{2\pi\varrho} - \frac{\epsilon\phi(\varrho,0)}{2\pi\lambda_{\rm GC}}, \tag{66}$$

where

$$\lambda_{\rm GC} = \frac{k_{\rm B}T\epsilon}{2\pi q^2\sigma} \tag{67}$$

is the Gouy–Chapman length.

The continuity of electrostatic potential requires that $A_1(k) = A_2(k)$. Substituting equations (61) and (62) into equation (64) and using equation (66), we find the electrostatic potential over the full range $-\infty < z < \infty$ to be

$$\phi(\varrho,z) = \frac{q}{\epsilon}\int_0^\infty \frac{k}{k+\lambda_{\rm GC}^{-1}} J_0(k\varrho)\,{\rm e}^{-k|z|}\,{\rm d}k. \tag{68}$$

For $z = 0$ the integral can be performed explicitly yielding

$$\phi(\varrho,0) = \frac{q\tau_0(\varrho/\lambda_{\rm GC})}{\epsilon\varrho}, \tag{69}$$

where the functions $\tau_\nu(x)$ are defined as [82]

$$\tau_\nu(x) = 1 - \frac{\pi x^{1-\nu}}{2}[H_\nu(x) - N_\nu(x)], \tag{70}$$

with $H_\nu(x)$ and $N_\nu(x)$ being the Struve and Bessel functions of order ν, respectively. For large values of x, $\tau_0 \approx 1/x^2$, so that, asymptotically,

$$\phi(\varrho,0) \approx \frac{q\lambda_{\rm GC}^2}{\epsilon\varrho^3}. \tag{71}$$

We conclude that in the case of a confined plasma there is no exponential screening; instead, the electrostatic potential is purely algebraic and has the form of a dipole–dipole interaction [71].

There is a well-known argument in condensed matter physics going back all the way to Bloch [83], Peierls [84] and Landau [85] in the 1930s, which states that a continuous symmetry cannot be broken in two dimensions. This means that there cannot exist a true two-dimensional crystalline order, since it requires breaking translational symmetry. The argument was made rigorous by Mermin, who proved it for particles interacting through short-ranged potentials [65]. It is quite simple to see how this conclusion arises. Suppose there is a two-dimensional crystal, one can then calculate the mean-square displacement δ^2 of one particle from its equilibrium position due to thermal fluctuations. It is found that $\delta^2 \sim T \ln L$, where L is the characteristic crystal size. For $L \to \infty$, the mean-square displacement diverges for any finite temperature, implying that in the thermodynamic limit a two-dimensional crystal is unstable to thermal fluctuations. Although there is no true long-range order in two dimensions for systems with short-range forces, there exists a pseudo-long-range order characterized by an algebraically decaying correlation function. It is not clear, however, to what extent this conclusion applies to the two-dimensional OCP, whose particles interact by a long-ranged $1/r$ potential. Certainly, in this case, Mermin's proof is no longer valid. However, since the effective interaction potential inside a two-dimensional OCP decays as $1/r^3$, which is short-ranged in two dimensions, it suggests that there should not be any long-range order. Whether there is a true long-range order or a pseudo-long-range order for a two-dimensional OCP remains uncertain. Simulations find that for $\Gamma \approx 130$ there is a crystallization transition. It is, however, difficult to say whether the crystalline state has a true long-range order or a pseudo-long-range order [80]. It is also unclear if the transition is of first order or continuous, belonging

to the KT universality class [66, 64]. Existence of the thermodynamic limit for confined two-dimensional plasmas can also be attributed to the effective renormalization of the interaction potential from a non-integrable $1/r$ to an integrable (in two dimensions) $1/r^3$ form.

The Helmholtz free energy of a two-dimensional plasma can be obtained directly from equation (69). We need to know the induced potential felt by the central ion due to other particles. In the limit $\varrho \to 0$, the electrostatic potential reduces to

$$\phi(\varrho, 0) \approx \frac{q}{\epsilon\varrho} + \frac{q}{\epsilon\lambda_{\mathrm{GC}}} \ln\left(\frac{\varrho}{2\lambda_{\mathrm{GC}}}\right). \tag{72}$$

The first term of this expression is the potential produced by the central ion, while the second term is the induced potential felt by the fixed ion. We note that the induced potential is actually divergent in the limit $\varrho \to 0$. This is the consequence of the failure of linearization of the PB equation. This deficiency can be corrected in the same way as was done for the three-dimensional OCP, by introducing a correlation hole of radius h around each particle. Unfortunately, in the present geometry this leads to calculations which are no longer tractable analytically. From our study of the three-dimensional OCP we can, however, make some reasonable approximations. In the limit of high temperatures, small Γ, the size of the correlation hole should be such that the electrostatic and the thermal energies become approximately equal,

$$\frac{q^2}{\epsilon h} \approx k_{\mathrm{B}}T. \tag{73}$$

This means that $h \approx \lambda_{\mathrm{B}}$. We can use this value as the short-distance cut-off in the calculation of free energy. The induced potential then becomes

$$\psi \approx \frac{q}{\epsilon\lambda_{\mathrm{GC}}} \ln\left(\frac{\lambda_{\mathrm{B}}}{2\lambda_{\mathrm{GC}}}\right). \tag{74}$$

The free energy is obtained through the usual Debye charging process, equation (12). Recalling that $\lambda_{\mathrm{B}}(\lambda q) = \lambda^2\lambda_{\mathrm{B}}(q)$ and $\lambda_{\mathrm{GC}}(\lambda q) = \lambda_{\mathrm{GC}}(q)/\lambda^2$, where λ is the charging parameter, in the limit of high temperatures $\Gamma \to 0$, the reduced free energy per particle is found to be

$$\frac{\beta F^{\mathrm{el}}}{N} \approx \Gamma^2 \ln(\Gamma). \tag{75}$$

Equation (75) is precisely the leading order term of the resumed virial expansion obtained by Totsuji [86, 87].

For low temperatures, the OCP crystallizes into a triangular lattice. The Madelung energy of this lattice is

$$\frac{\beta U}{N} = -1.106\,103\Gamma. \tag{76}$$

This equation provides a surprisingly good fit not only for the free energy of solids, but also for the free energy of fluids at sufficiently high values of Γ. Comparing to the results of the Monte Carlo simulations [80] we find that for $\Gamma = 5$ the error accrued from using equation (76) to calculate the total electrostatic free energy is about 30%. For $\Gamma = 20$ this error drops to 11% and for $\Gamma = 50$ it goes down to 6%. Recalling that the crystallization transition occurs at $\Gamma \approx 130$, we see that the equation (76) works well into the fluid phase. It is reasonable, therefore, to approximate the electrostatic free energy of a fluid for $\Gamma > 5$ by

$$\frac{\beta F^{\mathrm{el}}}{N} = -1.106\,103\Gamma. \tag{77}$$

The reason why the electrostatic free energy of a fluid is so well approximated by the free energy of the crystal, is a consequence of strong electrostatic correlations.

5. Asymmetric systems

Up to now we have considered only symmetric plasmas and electrolytes. In practice, however, it is unlikely that both cations and anions have exactly the same size and magnitude of charge. It is, therefore, important to explore the thermodynamics of a general $Z:1$ electrolyte in which cations have charge Zq and diameter a_c, while anions have charge $-q$ and diameter a_a. Unfortunately, as soon as the asymmetry is introduced, the internal inconsistency enters into the PB equation [50]. Recall that the cation–anion correlation function can be expressed in terms of the potential of mean force w_{+-}:

$$g_{+-}(r) = \mathrm{e}^{-\beta w_{+-}(r)}. \tag{78}$$

The $w_{+-}(r)$ is the work needed to bring the cation and anion from infinity to separation r inside an electrolyte. Clearly this work is invariant under the permutation of particle labels $w_{+-}(r) = w_{-+}(r)$. This means that

$$g_{+-}(r) = g_{-+}(r). \tag{79}$$

The PB equation, which serves as the basis, for the DH theory, approximates the potential of mean force by $w_{+-}(r) = q_-\phi_+(r)$. The self consistency condition, equation (79), then requires that

$$q_+\phi_-(r) = q_-\phi_+(r). \tag{80}$$

Because of the non-linear nature of the PB equation this condition cannot be satisfied except for symmetric electrolytes. The linearization prescription intrinsic to the DH theory allows equation (80) to hold for ions of different valence, but with the *same* ionic diameter, $a_c = a_a$.

We see that as soon as the symmetry between the cations and anions is broken, the physics and the mathematics of the problem becomes significantly more complex. In the limit of very large asymmetries, $Z \to \infty$ and $a_c \gg a_a$ a new simplification, however, enters into the game.

5.1. Colloidal suspensions

A typical colloidal suspension often studied experimentally consists of polystyrene sulphonate spheres of diameter 10 nm–1 μm and 10^3–10^4 ionizable surface groups. Because of the large surface charge, the colloidal particles tend to repel each other, forming crystals, even at fairly low volume fraction of less than 10%. Using the periodic structure of the lattice, the thermodynamics of a colloidal crystal can be studied fairly straightforwardly. Each colloidal particle can be thought to be confined to a Wigner–Seitz (WS) polyhedral cell. A further approximation replaces the polyhedral WS cell by a sphere [22].

5.2. Colloidal lattices

We shall model the colloidal particles as hard spheres of radius a carrying Z ionizable groups of charge $-q$ distributed uniformly on the surface. The counterions will be idealized as point particles of charge $+q$. The suspension of $N_p = \rho_p V$ polyions and $N_c = ZN_p = \rho_c V$ counterions is confined to a volume V. As usual, the solvent will be treated as a uniform continuum of dielectric constant ϵ. For sufficiently large polyion concentrations colloidal suspension crystallizes. Using the lattice symmetry, we restrict our attention to *one* colloidal particle and its counterions inside a spherical WS cell of radius R such that

$$\rho_p = \frac{1}{(4\pi/3)R^3}. \tag{81}$$

Using statistical mechanics it is possible to show that the osmotic pressure inside a cell is proportional to the concentration of counterions at the cell boundary [88]:

$$\beta P = \rho_c(R). \tag{82}$$

The thermodynamics of a crystalline colloidal suspension now reduces to the calculation of the distribution of counterions inside a WS cell. This can be done using a simple mean-field picture. The electrostatic potential inside a WS cell satisfies the Poisson equation (2) with the counterion charge density approximated by the normalized spherically symmetric Boltzmann distribution:

$$\rho_q(r) = Zq N_p \frac{e^{-\beta q\phi(r)}}{4\pi \int_a^R r^2\, dr\, e^{-\beta q\phi(r)}}. \tag{83}$$

The non-linear PB equation can be solved numerically to yield the electrostatic potential and the distribution of counterions inside the cell. In practice, it is more convenient to work with the electric field

$$\boldsymbol{E}(\boldsymbol{r}) = -\nabla\phi(\boldsymbol{r}). \tag{84}$$

The Poisson equation can then be rewritten as

$$\nabla \cdot \boldsymbol{E}(\boldsymbol{r}) = \frac{4\pi}{\epsilon}[\rho_q(\boldsymbol{r}) + q_p(\boldsymbol{r})], \tag{85}$$

where $q_p(\boldsymbol{r})$ is the polyion charge density,

$$q_p(\boldsymbol{r}) = -\frac{Zq}{4\pi a^2}\delta(|\boldsymbol{r}| - a)]. \tag{86}$$

To simplify the calculations we have uniformly smeared the charge of the polyion over its surface. Integrating both sides of equation (85) and taking advantage of the divergence theorem, the electric field at distance r from the polyion is

$$E(r) = -\frac{1}{\epsilon r^2}[Zq - \alpha(r)], \tag{87}$$

where

$$\alpha(r) = \int_{|\boldsymbol{r}'|<r} d^3\boldsymbol{r}'\, \rho_q(\boldsymbol{r}') \tag{88}$$

is the counterion charge inside a sphere of radius r centred on the colloidal particle. Using the gauge in which $\phi(a) = 0$ the electrostatic potential is

$$\phi(r) = -\int_a^r dr E(r) \tag{89}$$

and the PB equation reduces to an integral equation for the electric field. Note that equation (87) naturally incorporates the boundary conditions

$$E(a) = -\frac{Zq}{\epsilon a^2} \tag{90}$$

and

$$E(R) = 0. \tag{91}$$

Equation (87) can be solved iteratively to yield the counterion density profile from which all other thermodynamic functions are straightforwardly determined.

For aqueous colloidal lattices with monovalent counterions, the PB equation is in excellent agreement with the experiments and simulation. The PB equation, however, does not account for the correlations between the counterions and breaks down for low dielectric solvents or for aqueous suspensions with multivalent ions. Fortunately, in the case of colloidal lattices, it is fairly straightforward to account for these effects using the density functional theory (DFT).

5.3. Density functional theory for colloidal lattices

We shall now construct the Helmholtz free energy functional for a WS cell of a colloidal lattice. The Helmholtz free energy is a functional of the average local counterion concentration

$$\rho_c(\boldsymbol{r}) = \left\langle \sum_{i=1}^{N_c} \delta(\boldsymbol{r} - \boldsymbol{r}_i) \right\rangle, \tag{92}$$

where the brackets denote a Boltzmann average over all the particle positions. The free energy consists of electrostatic and entropic contributions. The entropic contribution is simply that of an inhomogeneous ideal gas:

$$\beta F^{\text{ent}}[\rho_c(\boldsymbol{r})] = \int \mathrm{d}^3\boldsymbol{r}\, \rho_c(\boldsymbol{r})\{\ln[\rho_c(\boldsymbol{r})\Lambda^3] - 1\}. \tag{93}$$

The electrostatic contribution is the result of Coulomb interactions between the counterions and the polyion, as well as the self-energy of the ionic cloud:

$$F^{\text{el}}[\rho_c(\boldsymbol{r})] = q \int \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}^3\boldsymbol{r}'\, \frac{q_p(\boldsymbol{r}')\rho_c(\boldsymbol{r})}{\epsilon|\boldsymbol{r} - \boldsymbol{r}'|} + \frac{q^2}{2} \int \mathrm{d}^3\boldsymbol{r}\, \mathrm{d}^3\boldsymbol{r}'\, \frac{\rho_c(\boldsymbol{r}')\rho_c(\boldsymbol{r})}{\epsilon|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{94}$$

Equation (94) is the mean-field approximation for the total electrostatic free energy. It does not account for the electrostatic correlations between the counterions surrounding the colloidal particle. Clearly, if there is a counterion present at position $\boldsymbol{r}$, there is a reduced probability of finding another counterion in its vicinity. This information is not included in equation (94). One can attempt to account for the electrostatic correlations using the local density approximation (LDA),

$$\beta F^{\text{cor}}[\rho_c(\boldsymbol{r})] = \int \mathrm{d}^3\boldsymbol{r}\, \rho_c(\boldsymbol{r})\, f_{\text{cor}}[\rho_c(\boldsymbol{r})]. \tag{95}$$

The correlational free energy can be approximated by the free energy of a three-dimensional OCP equation (60),

$$f_{\text{cor}}[\rho_c(\boldsymbol{r})] \approx f_{\text{ocp}}[\rho_c(\boldsymbol{r})] = \frac{1}{4}\left[1 - \omega^2 + \frac{2\pi}{3\sqrt{3}} + \ln\left(\frac{\omega^2 + \omega + 1}{3}\right) - \frac{2}{\sqrt{3}}\tan^{-1}\left(\frac{2\omega + 1}{\sqrt{3}}\right)\right], \tag{96}$$

with

$$\omega[\rho_c(\boldsymbol{r})] = \left[1 + (3\Gamma[\rho_c(\boldsymbol{r})])^{3/2}\right]^{1/3}, \tag{97}$$

where the local coupling strength is,

$$\Gamma[\rho_c(\boldsymbol{r})] = \lambda_B \left[\frac{4\pi\rho_c(\boldsymbol{r})}{3}\right]^{1/3}. \tag{98}$$

Although equation (96) was derived for a plasma in a uniform neutralizing background, it can also be used to approximate the correlational electrostatic free energy of counterions confined in a WS cell, the role of a neutralizing background being played by the confining electric field produced by the colloidal particle.

The equilibrium charge distribution is determined from the minimization of the total Helmholtz free energy:

$$F = F^{\text{ent}} + F^{\text{el}} + F^{\text{cor}} \tag{99}$$

subject to the constraint of particle conservation:

$$\int \mathrm{d}^3\boldsymbol{r}\, \rho_c(\boldsymbol{r}) = N_c. \tag{100}$$

This is equivalent to minimization of the grand potential

$$\Omega = F - \mu_c N_c, \tag{101}$$

where the chemical potential of counterions μ_c is the Lagrange multiplier. Performing the calculation, leads to the equilibrium counterion density profile

$$\rho_c(r) = N_c \frac{e^{-\beta q\phi(r) - \beta\mu^{ex}(r)}}{4\pi \int_a^R r^2\, dr\, e^{-\beta q\phi(r) - \beta\mu^{ex}(r)}}, \tag{102}$$

where the excess chemical potential is

$$\mu^{ex}(\boldsymbol{r}) = \frac{\delta F^{cor}}{\delta\rho_c(\boldsymbol{r})}. \tag{103}$$

We see that in the absence of correlations the density profile reduces to the Boltzmann distribution equation (83). The counterion density, which enters into the expression for the excess chemical potential, can be expressed in terms of the electric field:

$$\rho_c(\boldsymbol{r}) = \frac{\epsilon}{4\pi q}\nabla\cdot\boldsymbol{E}, \tag{104}$$

which due to spherical symmetry simplifies to

$$\rho_c(r) = \frac{\epsilon}{4\pi q r^2}\frac{\partial(r^2 E)}{\partial r}. \tag{105}$$

Substituting equation (102) into the Poisson equation (85) and using equations (89) and (105) leads to an integro-differential equation for the electric field, equation (87).

Unfortunately, for large colloidal charges this equation has no stable solutions [89–91]. There is an unbounded increase in the concentration of counterions in the vicinity of the colloidal surface resulting from the failure of the LDA [92]. To overcome this difficulty Groot [91] has suggested the use of a weighted density approximation (WDA), which has proved quite successful in its application to other problems of condensed matter physics [93,94]. Within this approach the correlational free energy is given by

$$F^{cor}[\rho_c(\boldsymbol{r})] = \int d^3\boldsymbol{r}\, \rho_c(\boldsymbol{r}) f_{cor}[\bar{\rho}_c(\boldsymbol{r})], \tag{106}$$

where $\bar{\rho}_c(\boldsymbol{r})$ is the *average* local density,

$$\bar{\rho}_c(\boldsymbol{r}) = \int d^3\boldsymbol{r}'\, w(|\boldsymbol{r}-\boldsymbol{r}'|)\rho_c(\boldsymbol{r}). \tag{107}$$

The weight function $w(r)$ can be determined from the thermodynamic requirement that

$$\frac{\delta^2\beta F}{\delta\rho_c(\boldsymbol{r})\delta\rho_c(\boldsymbol{r}')} = \frac{\delta(\boldsymbol{r}-\boldsymbol{r}')}{\rho_c(\boldsymbol{r})} - C_2(|\boldsymbol{r}-\boldsymbol{r}'|), \tag{108}$$

where $C_2(|\boldsymbol{r}-\boldsymbol{r}'|)$ is the direct correlation function. In particular, this equation must hold in the limit of a homogeneous OCP, the direct correlation function for which can be obtained using the theory developed in section 4 [91],

$$C_2(r) = -\frac{\lambda_B}{h} \qquad \text{if } r \leqslant h, \tag{109}$$

and

$$C_2(r) = -\frac{\lambda_B}{r} \qquad \text{if } r > h. \tag{110}$$

Performing the calculation, we find that the weight function is well approximated by [91]

$$w(r) = \frac{3}{2\pi h^2}\left(\frac{1}{r} - \frac{1}{h}\right)\Theta(h-r), \tag{111}$$

where $\Theta(x)$ is the Heaviside step function, $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x < 0$. The WDA is significantly more computationally demanding than the LDA. Its advantage, however, is the numerical stability for all values of the colloidal charge. The results based on the WDA are in excellent agreement with the Monte Carlo simulations. The WDA gives us a good handle on the thermodynamics of colloidal lattices. At lower concentrations, when the crystalline structure has melted, the situation unfortunately is no longer so clear cut. In this case a simple picture based on the WS cell is not sufficient and new methods must be developed [23, 95]. Unfortunately, the standard techniques of the liquid state theory based on integral equations are powerless in the case of highly asymmetric colloidal systems. The field theoretic methods also fail when applied to this difficult problem, figure 1. Furthermore, even the experimental situation is far from clear. Ise and co-workers claim to have seen stable clusters of colloidal particles in highly de-ionized colloidal suspensions. Tata *et al* even report an observation of a full equilibrium vapour–liquid-like phase separation [96]. These experiments, however, have been challenged by Palberg and Würth, who demonstrated that the phase separation observed by Tata *et al* was the result of non-equilibrium salt gradients produced by the ion exchange resin [97,98]. In the colloidal science community the possibility of a liquid–vapour phase separation in highly de-ionized colloidal suspensions has met with a large amount of scepticism. The usual argument against the phase transition is based on the Derjaguin–Landau–Verwey–Overbeek (DLVO) colloidal pair potential [99, 100].

It is easy to understand the nature of the DLVO potential based on the DH theory. If the size of colloidal particles is shrunk to zero, $a \to 0$, then due to screening by counterions, the interaction energy between two 'point' colloids would be of a Yukawa form:

$$V_0(r) = (Zq)^2 \frac{\mathrm{e}^{-\kappa r}}{\epsilon r}, \tag{112}$$

where the inverse Debye length is

$$\xi_{\mathrm{D}}^{-1} \equiv \kappa = \sqrt{\frac{4\pi Z q^2 \rho_{\mathrm{p}}}{k_{\mathrm{B}} T \epsilon}}. \tag{113}$$

Now, consider the electrostatic potential outside the fixed colloidal particle of radius a and charge $-Zq$, equation (10)

$$\phi_{>}(r) = -\frac{Zq\theta(\kappa a)\,\mathrm{e}^{-\kappa r}}{\epsilon r}, \qquad \theta(x) = \frac{\mathrm{e}^{x}}{1+x}. \tag{114}$$

Evidently the factor $\theta(\kappa a)$ accounts for the fact that screening starts only outside the cavity, $r > a$. We also can think of equation (114) as the potential of a point particle with an effective charge $Q_{\mathrm{p}} = Zq\theta(\kappa a)$. An advantage of this alternative point of view is that the interaction energy for two 'point' particles is simply given by equation (112) with $Zq \to Q_{\mathrm{p}}$. This leads directly to the famous DLVO potential

$$V_{\mathrm{DLVO}}(r) = (Zq)^2 \theta^2(\kappa a) \frac{\mathrm{e}^{-\kappa r}}{\epsilon r}. \tag{115}$$

This potential is purely repulsive [101], which naively suggests that a charged colloidal suspension is stable against a liquid–gas phase separation. Sogami and Ise, therefore, have argued that the DLVO potential must be incorrect, since it cannot account for the inhomogeneities observed experimentally [102]. In its stead, they proposed a different interaction potential derived on the basis of the Gibbs free energy. The potential found by Sogami and Ise contains a minimum [102], which implies that at short enough separations the two like-charged colloidal particles attract! What is most surprising is that the attraction appears even for monovalent counterions, i.e in the absence of strong correlations between the

colloidal double layers. Furthermore, water is an incompressible fluid so that it is difficult to see how a change of paradigm from Helmholtz to the Gibbs free energy can lead to such a profound modification of the interaction potential. Inconsistency in the results based on the Helmholtz and the Gibbs free energies has been carefully re-examined by Overbeek, who has traced the discrepancy to a flaw in Sogami and Ise's calculations [103].

It is important to stress that the repulsive two-body interactions do not, in general, preclude the possibility of a liquid–gas phase separation in a multicomponent fluid. In fact van Roij and Hansen found, within the linearized density functional theory, that it is possible for a colloidal suspension with polyions interacting by the repulsive DLVO potential to phase separate into coexisting liquid and gas phases [20]. Before entering into the discussion of colloidal fluids it is, however, important to introduce a new fundamental concept—the colloidal charge renormalization.

5.4. Charge renormalization

Although the non-linear PB equation cannot be solved analytically for a spherical geometry, the numerical solution indicates that the electrostatic potential far from colloidal particles saturates as a function of the bare colloidal charge [22]. This suggests that the thermodynamics of highly charged colloidal systems can be based on a linearized PB equation but with the bare colloidal charge replaced by an effective renormalized charge. The original concept of colloidal charge renormalization is due to Alexander *et al*, but is well pre-dated in the polyelectrolyte literature, where the phenomenon is known as the Manning counterion condensation [104–106].

To understand better colloidal charge renormalization, let us first consider a uniformly charged plane at fixed potential ψ_s inside a salt solution of concentration c. The electrostatic potential at distance x from the plane satisfies the PB equation,

$$\frac{d^2\phi(x)}{dx^2} = \frac{8\pi cq}{\epsilon}\sinh(\beta q\phi). \tag{116}$$

Since at the moment we are considering aqueous suspensions containing only monovalent ions, the electrostatic correlations are insignificant and the mean-field PB approximation is sufficient. Multiplying both sides of equation (116) by $d\phi/dx$ allows us to perform the first integration. Since the potential vanishes in the limit $x \to \infty$, we find

$$\frac{1}{2}[\phi'(x)]^2 = \frac{8\pi c}{\epsilon\beta}\left[\cosh(\beta q\phi) - 1\right]. \tag{117}$$

The second integration yields [107]

$$\phi(x) = \frac{2k_BT}{q}\ln\frac{1 + e^{-\kappa x}\tanh(\beta q\psi_s/4)}{1 - e^{-\kappa x}\tanh(\beta q\psi_s/4)}, \tag{118}$$

where the inverse Debye length is,

$$\kappa = \sqrt{8\pi c\lambda_B}. \tag{119}$$

In the limit of large surface potentials this expression simplifies to

$$\phi(x) = \frac{2k_BT}{q}\ln\frac{1 + e^{-\kappa x}}{1 - e^{-\kappa x}}. \tag{120}$$

For separations from the plane larger than the Debye length, equation (120) becomes

$$\phi(x) = \frac{4k_BT}{q}e^{-\kappa x}. \tag{121}$$

An important observation is that for large surface potentials, $\beta q\psi_s/4 \gg 1$, the electrostatics away from the plane is completely insensitive to the surface charge density.

Now, let us consider a highly charged colloidal particle of valence Z and radius a inside a symmetric 1 : 1 electrolyte of concentration c. The electrostatic potential at a distance r from the centre of a colloidal particle satisfies the PB equation (6). For distances $r > a + \xi_D$ the electrostatic potential is small and the PB equation can be safely linearized leading to the Helmholtz equation (7). This can be easily integrated yielding the electrostatic potential,

$$\phi(r) = A\frac{e^{-\kappa r}}{r}, \tag{122}$$

where A is the integration constant. To find its value, lets restrict our attention to suspensions in which the $\xi_D \ll a$. In practice, this is not a very strong restriction. For salt solutions at physiological concentrations $\xi_D \approx 8$ Å while the characteristic colloidal size is on the order of 1000 Å. Even for solutions with very low salt content, in the millimolar range, the Debye length is of the order of 100 Å. Under these conditions all the curvature effects associated with the spherical geometry of colloidal particles are effectively screened at separations $a + \xi_D < r < 2a$, and the electrostatic potential is well approximated by that of a uniformly charged plane, equation (121). Comparing equations (121) and equation (122) the value of the integration constant follows directly, and the electrostatic potential at distance $r > a + \xi_D$ from the centre of colloidal particle is

$$\phi(r) = \frac{4k_B T a\, e^{-\kappa(r-a)}}{qr}. \tag{123}$$

This is the asymptotic solution of the full non-linear PB equation for $\kappa a \gg 1$. Comparing this to the solution of linearized PB, equation (114), it is evident that the two are identical as long as *the bare colloidal charge is replaced by the renormalized charge*. For highly charged particles, equation (123) shows that the renormalized charge saturates at [108]

$$Z_{ren}^{sat} = \frac{4a(1+\kappa a)}{\lambda_B}. \tag{124}$$

While the previous analysis was carried out for one colloidal particle inside an electrolyte solution, the concept of charge renormalization is quite general and can be applied to colloidal suspensions under various conditions [109–111]. The difficulty of defining the effective charge for suspensions at non-zero concentrations resides in the complexity of accounting for the consequences of colloidal interactions. The standard practice is to study one colloidal particle inside a spherical WS cell whose radius is determined by the volume fraction of colloids [22]. While this procedure is fully justified for colloidal lattices, its foundation is less certain for fluidized suspensions. To find the renormalized charge, one numerically solves the full non-linear PB equation and matches the electrostatic potential to the solution of the linearized equation at the cell boundary. Alternatively, the osmotic pressures inside the WS cell calculated using the non-linear and linear equations are matched in order to define the effective charge. One should remember, however, that while at the level of the non-linear PB equation the osmotic pressure is directly proportional to the concentration of ions at the cell boundary, equation (82), this is not the case for the linearized PB equation. The various procedures lead to similar values of the renormalized charge. In the case of salt-free suspensions, the effective charge is found to saturate at [22]

$$Z_{ren}^{sat} \approx \frac{\chi a}{\lambda_B}, \tag{125}$$

where χ is an approximately linearly increasing function of colloidal concentration for suspensions with volume fraction larger than 1%. For suspensions with colloidal volume fraction between 1% and 10% the value of χ varies from around 9 to 15 [112, 113].

5.5. Colloidal fluid

In this section we will apply the insights gained from the study of one and two component plasmas to the exploration of the stability of charged colloidal suspensions against gas–liquid phase separations. We note that the large size asymmetry between colloids and counterions leads to very different equilibration timescales. On the timescale of polyion motion, the counterions are always equilibrated. This suggests that the calculation of free energy should be done in two stages [114]. First, we shall trace out the counterion degrees of freedom, leading to effective many-body interactions between the colloidal particles. Then we will use these effective interactions to calculate the colloid–colloid contribution to the total free energy. The procedure is similar to the one used in the McMillan–Mayer theory of solutions [115].

We shall first calculate the contribution to the total free energy arising from the polyion–counterion interactions [23, 95]. Consider a suspension in thermal equilibrium. While the colloidal particles are more or less uniformly distributed throughout the solution, the positions of counterions are strongly correlated with the positions of polyions. As a leading order approximation we can, therefore, take the polyion–polyion correlation function to be

$$g_{\mathrm{pp}} = 1 \tag{126}$$

and the polyion–counterion correlation function to be

$$g_{\mathrm{pc}} = \mathrm{e}^{-\beta q \phi(r)}. \tag{127}$$

Choosing the coordinate system in such a way that it is centred on top of one of the colloidal particles, the electrostatic potential at distance $r < a$ satisfies the Laplace equation, while for distances $r > a$ it satisfies the Poisson equation, equation (2). Based on equations (126) and (127) the charge density in the region $r > a$ can be approximated by,

$$\rho_q(r) = -Zq\rho_{\mathrm{p}} + q\rho_{\mathrm{c}}\,\mathrm{e}^{-\beta q\phi(r)}. \tag{128}$$

In the spirit of the DH theory we shall linearize the exponential [36, 6]. The distribution of charge around the colloid reduces to

$$\rho_q(r) = -\beta q^2 \rho_{\mathrm{c}} \phi(r). \tag{129}$$

For $r > a$ the electrostatic potential, therefore, satisfies the Helmholtz equation (7) with κ given by equation (113). The solution to this equation is

$$\phi_>(r) = -\frac{Zq\theta(\kappa a)\,\mathrm{e}^{-\kappa r}}{\epsilon r}, \tag{130}$$

while the solution to the Laplace equation for $r \leqslant a$ is

$$\phi_<(r) = -\frac{Zq}{\epsilon a(1+\kappa a)}. \tag{131}$$

The electrostatic energy due to polyion–counterion interaction is

$$u_{\mathrm{p}} = \frac{1}{2}\int \mathrm{d}^3\boldsymbol{r}\,[\rho_q(\boldsymbol{r}) + q_{\mathrm{p}}(\boldsymbol{r})]\phi(r), \tag{132}$$

where $\rho_q(\boldsymbol{r})$ is the charge density of counterions given by equation (129), and $q_{\mathrm{p}}(\boldsymbol{r})$ is the charge density of a polyion:

$$q_{\mathrm{p}}(\boldsymbol{r}) = -\frac{Zq}{4\pi a^2}\delta(|\boldsymbol{r}| - a). \tag{133}$$

Performing the integration we find

$$u_{\mathrm{p}} = \frac{Z^2q^2}{2\epsilon(1+\kappa a)}\left[\frac{1}{a} - \frac{\kappa}{2(1+\kappa a)}\right]. \tag{134}$$

The electrostatic *free energy* of a polyion inside the suspension is obtained using the Debye charging process [116]:

$$\mathcal{F}_{\rm p} = \int_0^1 {\rm d}\lambda \frac{2u_{\rm p}(\lambda q)}{\lambda} = \frac{Z^2 q^2}{2\epsilon a(1+\kappa a)}. \tag{135}$$

Note that this free energy is the sum of the polyion self energy,

$$\mathcal{F}_{\rm p}^{\rm self} = \frac{Z^2 q^2}{2\epsilon a}, \tag{136}$$

and the solvation energy,

$$\mathcal{F}_{\rm p}^{\rm solv} = -\frac{Z^2 q^2 \kappa a}{2\epsilon a(1+\kappa a)}, \tag{137}$$

which the polyion gains from being inside the 'ionic sea'. The electrostatic free energy due to interaction between all the polyions and counterions is

$$F^{\rm pc} = -\frac{Z^2 q^2 N_{\rm p} \kappa a}{2\epsilon a(1+\kappa a)}. \tag{138}$$

We have effectively integrated out the counterion degrees of freedom. This, however, leaves us with the effective many-body potentials of interaction between the colloidal particles. For dilute suspensions, the pairwise interaction potential should be the dominant one. The two-body interaction potential can be obtained from the solution of the Helmholtz equation for two colloidal particles [117, 118]. At large separations this leads directly to the DLVO interaction potential, equation (115). This potential has been extensively tested experimentally and found to work very well for bulk colloidal suspensions [119]. Since the DLVO potential is short ranged, the contribution to the total free energy arising from the colloid–colloid interaction can be calculated in the spirit of the traditional van der Waals theory, through the second virial term. A more sophisticated calculation of the colloid–colloid free energy relies on the Gibbs–Bogoliubov variational bound,

$$F^{\rm pp} \leqslant F_0 + \langle V_{\rm DLVO} \rangle_0, \tag{139}$$

where the reference system is taken to be the fluid of hard spheres, whose diameter plays the role of a variational parameter. The free energy resulting from the polyion–polyion interaction, $F^{\rm pp}$, can be approximated by the lowest variational bound of equation (139). The calculation is somewhat involved, so we refer the interested reader to the original papers [120, 121, 20, 122].

The entropic mixing free energy of colloids and their counterions is simply that of an ideal gas:

$$\beta F^{\rm ent} = Z N_{\rm p}\big[\ln\left(Z\rho_{\rm p}\Lambda_{\rm c}^3\right) - 1\big] + N_{\rm p}\big[\ln\left(\rho_{\rm p}\Lambda_{\rm p}^3\right) - 1\big], \tag{140}$$

where $\Lambda_{\rm c}$ and $\Lambda_{\rm p}$ are the de Broglie thermal wavelengths of counterions and polyions, respectively.

The total free energy of colloidal suspension is the free energy needed to solvate colloids in the sea of other polyions and counterions $F^{\rm pc} + F^{\rm pp}$, and the free energy of mixing, $F^{\rm ent}$,

$$F = F^{\rm pc} + F^{\rm pp} + F^{\rm ent}. \tag{141}$$

The osmotic pressure is

$$P = -\frac{\partial F}{\partial V}\Big|_{N_{\rm p}}. \tag{142}$$

It is found that for suspensions with

$$\mathcal{C} \equiv \frac{Z\lambda_{\rm B}}{a} > 15.2 \tag{143}$$

the pressure is not a convex function of the colloidal concentration, implying existence of a thermodynamic instability. At criticallity the colloidal volume fraction is around 1%. The crucial question is whether this result is reliable. In order to calculate the electrostatic free energies, we were forced to linearize the Boltzmann factor. While this is a reasonable approximation at large separations away from the polyions, linearization is clearly invalid in the vicinity of the colloidal surface. There, the strong electrostatic interactions result in an accumulation of counterions and the effective polyion charge renormalization. Therefore, the linear theory can be used *only if* the bare colloidal charge is replaced by the effective renormalized charge, $Z \to Z_{\rm eff}$, in all the expressions. It was found, however, that the bare charge does not increase without limit but saturates at the value given by equation (125). Substituting $Z \to Z_{\rm eff}$, into the definition of $\mathcal{C}$, equation (143), we see that $\mathcal{C} < 15$ for all the values of the bare charge Z in the critical region. The critical threshold, therefore, cannot be reached, meaning that a de-ionized *aqueous* suspension with *monovalent* counterions is stable against a liquid–gas phase separation for all colloidal charges and sizes. This conclusion has also been confirmed by more detailed calculations and simulations [23, 24, 123, 27, 124].

The result that the non-linear terms omitted within the DH approximation stabilize de-ionized colloidal suspensions against a liquid–vapour phase separation has also been obtained by von Grünberg *et al* [125–127] and Tamashiro and Schiessel [128] based on the analysis of the full non-linear PB equation inside a WS cell. The numerical integration of the non-linear PB shows that the osmotic pressure is a monotonically increasing function of colloidal concentration. This means that at the level of the WS approximation the suspension is thermodynamically stable. von Grünberg *et al* and Tamashiro and Schiessel, however, demonstrate that the *linearized* PB *equation* leads to the negative compressibility and the osmotic pressures for highly charged colloidal particles. This erroneously suggests the presence of a thermodynamic instability. Clearly the instability is an artefact of the linearization. Furthermore, our calculations show that any linear theory, which does not take into account the colloidal charge renormalization, is likely to lead to an incorrect prediction of a liquid–vapour phase separation [20, 129] in de-ionized aqueous suspensions with monovalent counterions.

It is curious that the 'linear' correlations between the colloids and the counterions, responsible for the screening of electrostatic interactions, are also the ones driving the suspension towards the phase separation. On the other hand, the 'non-linear' correlations responsible for the counterion condensation and the colloidal charge renormalization, stabilize the suspension against a phase transition.

6. Polyelectrolyte solutions

Polyelectrolytes are polymers with ionizable groups which have a tendency to dissociate in polar solvents [130, 131]. The good water solubility of polyelectrolytes is due to large favourable gain in the solvation free energy resulting from hydration of charged monomers and counterions. Unlike polyampholytes [132–135], whose monomers can either be cationic (positive) or anionic (negative), all charged monomers of a polyelectrolyte carry charge of the same sign. Depending on the sign of this charge, polyions are either cationic or anionic. Over the last few decades polyelectrolytes have found many industrial applications ranging from water treatment and oil recovery to detergents and superabsorbants. The biological importance of polyelectrolytes, however, has been realized much earlier. After all the most important biomolecule, DNA, is an anionic polyelectrolyte [106].

Unlike the simple polymeric fluids, the thermodynamics of which is fairly well understood, polyelectrolyte solutions still remain to a large extent enigmatic. The difficulty in studying

polyelectrolytes resides in the combination of polyion flexibility [136–138, 130, 139–141], the long-ranged nature of the Coulomb force, as well as a large charge and size asymmetry between the polyions and the counterions. There are, however, some polyelectrolytes whose polyions are rigid molecules. This allows us to bypass the complications associated with the statistics of polyion conformations. We have already come across this kind of system in our exploration of charged colloidal suspensions. Many biologically relevant polyelectrolytes are also fairly rigid. Persistence length (the distance over which the polymer can be considered to be rod-like) for double stranded DNA is of the order of 500 Å. Actin filaments, which are the building blocks of a cytoskeleton, have persistence length even larger, of the order of microns. This should be compared with the Debye length at physiological concentrations of 150 mM of NaCl, which is about 8 Å. Clearly under these conditions the flexibility of the DNA or the actin filaments can be considered an irrelevant perturbation.

The thermodynamics of rod-like polyelectrolytes can be explored using the same theoretical tools used for spherical colloidal suspensions. Rod-like molecules can undergo nematic and smectic phase transitions. For ordered periodic structures, the WS cell formalism can be employed to obtain most of the relevant thermodynamics [142, 143]. At low volume fraction, when a polyelectrolyte solution is disordered, this strategy is no longer valid and a different methodology must be used. This can be constructed along the same lines taken for colloidal suspensions [144,145]. The fundamental role of electrostatic correlations between the polyions and the counterions appears as Debye screening of polyion–polyion interactions and the renormalization of a polyion charge. In polyelectrolyte literature the polyion–counterion association leading to polyion charge renormalization is known as the Manning condensation [104, 106, 105]. Here, we will show that the Manning condensation is very similar to the charge renormalization found in colloidal suspensions.

First, we shall briefly review Manning's original argument [104]. Manning was interested in deriving the limiting (low density) laws for polyelectrolyte solutions, similar to the ones found by Debye and Hückel for simple electrolytes, see equation (14). The salient feature of the DH limiting laws is that they do not depend on the specifics of the electrolyte, i.e. size or hydration. For example, the osmotic pressure at low ionic concentrations is found to be a function only of the ionic charge, temperature and concentration. The question then arises if such a limiting law is also possible for polyelectrolyte solutions. The answer to this question is far from obvious. The strong electrostatic interaction between the polyions and counterions favours the accumulation of counterions in the vicinity of polyions. It is, therefore, possible that even at very large dilutions the physics of a polyelectrolyte solution remains that of a strongly interacting system for which no limiting law should be anticipated [106].

6.1. Manning condensation

Consider a simple model of a polyelectrolyte solution. The rod-like polyions of concentration ρ_p, idealized as rigid cylinders of length L and radius a, carrying Z ionized groups—each of charge $-q$ uniformly spaced along the major axis of the cylinder—inside a uniform dielectric solvent of constant ϵ. The counterions of concentration $\rho_c = Z\rho_p$ will be treated as point particles of charge q. For simplicity we will restrict our attention to the situation in which there is no added salt.

In the low density limit we can neglect the discreteness of the polyion charge distribution and assume a uniform line-charge density,

$$\lambda_0 = -\frac{Zq}{L} \equiv -\frac{q}{b}, \tag{144}$$

where b is the separation between the successive charged monomers along the polyelectrolyte chain. The bare interaction potential between a long charged cylinder and a counterion is

$$\phi = -\frac{2q\lambda_0}{\epsilon} \ln\left(\frac{r}{r_0}\right), \tag{145}$$

where r_0 is the arbitrarily chosen point of zero potential. A polyion–counterion two-body partition function is

$$\zeta_1 = L\int_a^R \mathrm{e}^{-\beta q\phi(r)}\,\mathrm{d}^2 r = \pi L r_0 \frac{(R/r_0)^{2-2\xi} - (a/r_0)^{2-2\xi}}{1-\xi}, \tag{146}$$

where R is the cut-off distance at which a counterion can still be considered to be bound to the polyion. The Manning parameter is defined as

$$\xi = \frac{|q\lambda_0|}{\epsilon k_\mathrm{B} T} = \frac{q^2}{\epsilon k_\mathrm{B} T b}. \tag{147}$$

The integral in equation (146) remains finite for all values of ξ. Manning noticed, however, that if the limiting laws exist, the thermodynamic functions should be independent of the polyion diameter. However, if $a = 0$ the integral in equation (146) diverges as $\xi \to 1^-$. Manning interpreted this divergence as an indication of the counterion condensation. For values of $\xi > 1$ he supposed that n counterions condense onto the polyion, reducing proportionately its effective charge density from λ_0 to

$$\lambda_n = \lambda_0 \frac{Z-n}{Z}. \tag{148}$$

To find the number of condensed counterions n, Manning postulated that for $\xi > 1$ the effective reduced line-charge density

$$\xi_\mathrm{eff} \equiv \frac{|q\lambda_n|}{\epsilon k_\mathrm{B} T} \tag{149}$$

saturates at one:

$$\xi_\mathrm{eff} = 1. \tag{150}$$

If the 'renormalized' ξ_eff is used in equation (146), instead of the 'bare' ξ when $\xi > 1$, the polyion–counterion partition function remains finite. Equations (149) and (150) determine the number of condensed counterions to be

$$\begin{aligned} n^* &= Z\left(1 - \frac{1}{\xi}\right) \qquad \text{for } \xi > 1 \\ n^* &= 0 \qquad\qquad\quad \text{for } \xi \leqslant 1 \end{aligned} \tag{151}$$

Once n^* is determined, the rest of the thermodynamic functions can be calculated quite easily. The nice thing about Manning's argument is that it is so simple. On the other hand it contains some points which might leave a more mathematically inclined reader quite disturbed. Manning relied on the existence of the limiting laws to establish the limiting law in the first place! This is a circular logic which is not always guaranteed to work. It is interesting to apply the same argument to the case of a two-dimensional plasma of particles interacting through a logarithmic potential [146], section 3. Consider an anion–cation two-body partition function, equation (39). Following Manning, let us look at the limit $a \to 0$. In this case, we find that the integral in equation (39) diverges for the temperatures

$$T < T_\mathrm{d} \equiv \frac{q^2}{2k_\mathrm{B}\epsilon}. \tag{152}$$

Thus, we might incorrectly conclude that the metal–insulator transition also happens at T_d. In reality, we know that the KT transition occurs at half this temperature:

$$T = T_{KT} \equiv \frac{q^2}{4k_B\epsilon}. \tag{153}$$

For any finite value of the particle diameter, all the thermodynamic functions are analytic at T_d and the singularities only appear at $T = T_{KT}$. Therefore, it seems far from obvious if Manning's argument is valid for real polyelectrolytes with monomers and counterions of finite diameter. To explore this in more detail we can appeal to the Debye–Hückel–Bjerrum theory, previously constructed for low dielectric electrolytes, section 2.2.

6.2. Counterion association

We will restrict our attention to the salt-free infinite dilution limit. The calculations, however, can be extended to solutions of finite polyelectrolyte concentration as well as to polyelectrolytes with salt or even amphiphiles [145, 147]. As in the case of colloidal fluids, we would like to trace out the degrees of freedom associated with the counterions. This leads to the effective many-body interactions between the polyions. In the limit of infinite dilution the contribution from these interactions to the total free energy can be neglected.

Consider a dilute polyelectrolyte solution in thermal equilibrium. If we take a snapshot, we will see polyions distributed more or less uniformly throughout the solution, with no specific orientation. The counterions, on the other hand, will be clustered in the vicinity of polyions. This picture suggests that there are only weak positional correlations between the polyions and strong positional correlations between the polyions and the counterions [148]. The average distribution of charge around a polyion can, therefore, be approximated by equation (128).

Suppose we choose a coordinate system so that it is centred on one of the polyions, with the z-axis along the polyion's major axis. The electrostatic potential for $r < a$ satisfies the Laplace equation, while for $r \geqslant a$ it satisfies the Poisson equation with the charge distribution approximated by equation (128). As in the case of colloids, we would like to linearize the exponential Boltzmann factor. This, however, is prohibited by the fact that electrostatic interactions are very strong in the vicinity of the polyion surface. Thus, in order to linearize the exponential, the short distance polyion–counterion correlations must be explicitly taken into account. In our earlier study of colloidal suspensions this was done by introducing an effective renormalized charge. Here, we will take a somewhat different approach. The main consequence of short distance electrostatic correlations is the polyion–counterion association. This is very similar to the concept of Bjerrum association, which has proved so successful for symmetric 1 : 1 electrolytes. A polyelectrolyte solution can be thought of as being composed of free unassociated counterions of density ρ_f, as well as of clusters of density ρ_n, consisting of one polyion and some number $0 \leqslant n \leqslant Z$ of associated counterions. The polyions with no associated counterions are treated as 0-clusters. Particle conservation requires that

$$\sum_{n=0}^{Z} \rho_n = \rho_p. \tag{154}$$

and

$$\rho_f + \sum_{n=0}^{Z} n\rho_n = Z\rho_p. \tag{155}$$

The distribution of cluster sizes $\{\rho_n\}$ can be determined from the equilibrium condition that the Helmholtz free energy be minimum. Once the clusters are explicitly introduced into the theory,

the exponential factor in equation (128) can be linearized, since at large distances $\beta q\phi(r) < 1$ and at short distances the 'non-linearities' are accounted for through the cluster formation. After linearization, the PB equation reduces to the Helmholtz equation with κ given by,

$$\kappa = \sqrt{\frac{4\pi q^2 \rho_{\mathrm{f}}}{k_{\mathrm{B}} T \epsilon}}. \tag{156}$$

The linear equation can be easily solved yielding the electrostatic potential outside an n-cluster

$$\phi(r) = \frac{2\lambda_n}{\epsilon} \frac{K_0(\kappa r)}{\kappa a K_1(\kappa a)}, \tag{157}$$

where λ_n is the linear cluster charge density of an n-cluster, equation (148), and $K_\nu(x)$ is the modified Bessel function of order ν. For distances $r < a$ the electrostatic potential is found to be

$$\phi(r) = -\frac{2\lambda_n}{\epsilon} \ln\left(\frac{r}{a}\right) + \frac{2\lambda_n}{\epsilon} \frac{K_0(\kappa a)}{\kappa a K_1(\kappa a)}. \tag{158}$$

The electrostatic energy due to n-cluster–counterion interaction can be obtained from equation (132) and the electrostatic *free* energy follows from the Debye charging process as in equation (135). In the limit of infinite dilution, the electrostatic free energy of an n-cluster is

$$\beta \mathcal{F}_n = -\frac{(Z-n)^2}{Z} \xi \left[\frac{2\gamma_{\mathrm{E}} - 1}{2} + \ln\left(\frac{\kappa a}{2}\right)\right] + \mathrm{O}(\rho_{\mathrm{f}}). \tag{159}$$

The electrostatic free energy density, $f = F/V$, due to all cluster–counterion interactions is

$$\beta f^{\mathrm{pc}} = -\frac{\xi}{Z} \sum_{n=0}^{Z} (Z-n)^2 \rho_n \left[\frac{2\gamma_{\mathrm{E}} - 1}{2} + \ln\left(\frac{\kappa a}{2}\right) + \mathrm{O}(\rho_{\mathrm{f}})\right]. \tag{160}$$

The cluster–cluster and the counterion–counterion contributions are of higher order in density and, in the limit of infinite dilution, can be neglected. The only contributions which must still be taken into account are the entropic free energy of mixing and the free energy necessary to construct the isolated clusters. Both of these can be concisely written as

$$\beta f^{\mathrm{ent-cl}} = \rho_{\mathrm{f}} \left[\ln\left(\rho_{\mathrm{f}} \Lambda_{\mathrm{c}}^3\right) - 1\right] + \sum_n \rho_n \left[\ln\left(\frac{\Lambda_n^{3(n+1)} \rho_n}{\zeta_n}\right) - 1\right], \tag{161}$$

where the de Broglie thermal wavelength of an n-cluster is

$$\Lambda_n = \frac{h}{\sqrt{2\pi m_n k_{\mathrm{B}} T}}, \tag{162}$$

and m_n is the cluster geometric mean mass,

$$m_n = (m_{\mathrm{p}} m_{\mathrm{c}}^n)^{1/(n+1)}. \tag{163}$$

The internal partition function of an n-cluster is

$$\zeta_n = \frac{1}{n!} \int \prod_{i=1}^{n} \frac{\mathrm{d}^3 \boldsymbol{r}_i}{\Lambda_n^3} \, \mathrm{e}^{-\beta U}, \tag{164}$$

were U is the usual Coulomb potential. A suitable cut-off must be chosen in order to define what constitutes a cluster. Evaluation of the integral in equation (164) represents a formidable task. Fortunately, as we shall see, for polyelectrolyte solutions at infinite dilution, specific knowledge of ζ_n proves unnecessary. The total free energy density is then

$$f(\{\rho_n\}) = f^{\mathrm{pc}} + f^{\mathrm{ent-cl}}. \tag{165}$$

To find the equilibrium cluster distribution, this free energy must be minimized subject to constraints of particle conservation, equations (154) and (155). The minimization is equivalent to the law of mass action,

$$\mu_n = \mu_0 + n\mu_{\rm f}, \tag{166}$$

where the chemical potential of n-clusters is

$$\mu_n = \frac{\partial f}{\partial \rho_n}, \tag{167}$$

and the chemical potential of free ions is

$$\mu_{\rm f} = \frac{\partial f}{\partial \rho_{\rm f}}. \tag{168}$$

Substitution of the total free energy into equation (166) leads to the n-cluster distribution

$$\rho_n = \zeta_n \rho_0 \rho_{\rm f}^n \, {\rm e}^{\beta(\mu_0^{\rm ex} + n\mu_{\rm f}^{\rm ex} - \mu_n^{\rm ex})}, \tag{169}$$

which is actually a set of Z coupled algebraic equations. At large dilutions, the excess chemical potential of n-clusters is

$$\beta\mu_n^{\rm ex} = -\frac{(Z-n)^2}{Z}\xi\left[\frac{2\gamma_{\rm E}-1}{2} + \ln\left(\frac{\kappa a}{2}\right)\right] + {\rm O}(\rho_{\rm f}), \tag{170}$$

and of free ions is

$$\beta\mu_{\rm f}^{\rm ex} = -\xi\sum_{n=0}^{Z}\frac{(Z-n)^2\rho_n}{2Z\rho_{\rm f}} + {\rm O}(\rho_{\rm f}). \tag{171}$$

The internal partition function of an n-cluster is independent of density. Recalling that $\kappa \sim \sqrt{\rho_{\rm f}}$, equation (169) simplifies to

$$\rho_n \sim \rho_0 \rho_{\rm f}^{g(n)}, \tag{172}$$

where the exponent is

$$g(n) = \frac{\xi}{2Z}n^2 - \xi n + n. \tag{173}$$

In the limit of infinite dilution, $\rho_{\rm f} \to 0$, only the cluster of size

$$n_{\rm M} = Z\left(1 - \frac{1}{\xi}\right), \tag{174}$$

which minimizes $g(n)$ survives. In this limit the cluster size distribution takes a particularly simple form,

$$\rho_n = \rho_{\rm p}\delta_{n\,n_{\rm M}}. \tag{175}$$

This is precisely the cluster size distribution postulated by Manning based on his heuristic argument. The osmotic pressure of a polyelectrolyte solution can be obtained through the Legendre transform of the negative free energy density [146],

$$P = -f(\rho_{\rm f}, \{\rho_n\}) + \mu_{\rm f}\rho_{\rm f} + \sum_n \mu_n\rho_n, \tag{176}$$

which leads directly to the Manning limiting law for pressure [104]:

$$\beta P = \left(1 - \frac{1}{2\xi}\right)Z\rho_{\rm p} \qquad \text{for } \xi \leqslant 1, \tag{177}$$

$$\beta P = \frac{Z\rho_{\rm p}}{2\xi} \qquad \text{for } \xi > 1. \tag{178}$$

The discontinuity in the slope of pressure as a function of temperature has provoked a lot of speculation that the Manning condensation is a real thermodynamic phase transition.

Kholodenko and Beyerlein even went so far as to identify Manning condensation with the KT transition [149]. That this is incorrect follows already from our discussion in section 6.1. For the two-dimensional plasma with logarithmic interactions, the KT transition occurs at half the equivalent Manning temperature. The two phenomena, therefore, have nothing in common [146]. Furthermore, while the KT is a real thermodynamic phase transition characterized by the diverging Debye length, for rod-like polyelectrolytes the discontinuity in the slope of pressure appears only in the double limit $\rho_p \to 0$, $L \to \infty$. If the order of limits is interchanged, there is no singularity and no counterion condensation. In addition, directly at the condensation threshold $\xi = 1$, the Debye length remains finite [145].

For polyelectrolyte solutions containing polyions of *finite* size and at *non-zero* concentration, there is also a polyion–counterion association [145]. In this case, however, the distribution of cluster sizes is no longer a delta-function, but rather a bell-shaped curve centred on the value n^*. We find that n^* depends on the concentration of polyelectrolyte and is somewhat larger than the limiting Manning value n_M. The pressure remains an analytic function of ξ, showing that for real polyelectrolyte solutions, the counterion condensation is actually a crossover phenomenon, very similar to the micellar formation in amphiphilic systems [150].

7. Multivalent counterions

Up to now we have been concentrating our attention on aqueous solutions with monovalent counterions. It was already mentioned that in this case the correlations between the condensed counterions can be neglected. To understand why, let us compare the characteristic electrostatic energy of a counterion–counterion interaction to the characteristic thermal energy $k_B T$,

$$\Gamma = \frac{\alpha^2 q^2}{\epsilon d k_B T}, \tag{179}$$

where α is the counterion valence and d is the average separation between the n condensed counterions on the surface of a colloidal particle of radius a. Since $n\pi d^2 = 4\pi a^2$,

$$d = \frac{2a}{\sqrt{n}}, \tag{180}$$

and the coupling strength becomes

$$\Gamma = \frac{\alpha^2 \lambda_B \sqrt{n}}{2a}. \tag{181}$$

Now, let us consider highly charged latex particles with $Z = 7000$ and $a = 1000$ Å, in water at room temperature. From equation (125), taking $\chi = 15$, $Z_{\text{eff}}^{\text{sat}} = 2100$, which means that 4900 monovalent ($\alpha = 1$) counterions are condensed onto the particle. The coupling strength of the counterion–counterion interaction is then $\Gamma \approx 0.25$, which clearly shows that the electrostatic interactions between the condensed counterions are very weak. We can make this observation even more general. The high surface charge concentration σ_m encountered in nature is of the order of one elementary charge per 100 Å^2. Let us suppose that the suspension consists of highly charged colloidal particles with surface charge density σ_m. Clearly this means that there will be a lot of counterion condensation. For a salt-free colloidal suspension containing multivalent counterions, the number of condensed counterions will be approximately

$$n^* \approx \frac{Z}{\alpha}. \tag{182}$$

The radius of a colloidal particle can be expressed as

$$a = \sqrt{\frac{Z}{4\pi\sigma_{\rm m}}}. \tag{183}$$

Substituting equations (182) and (183) into equation (181) we find that the maximum counterion–counterion coupling strength is,

$$\Gamma_{\rm max} \approx \alpha^{3/2}\lambda_{\rm B}\sqrt{\pi\sigma_{\rm m}}. \tag{184}$$

For monovalent counterions $\Gamma_{\rm max} \approx 1.3$, for divalent counterions $\Gamma_{\rm max} \approx 3.6$, and for trivalent counterions $\Gamma_{\rm max} \approx 6.8$. Although $\Gamma_{\rm max}$ is an overestimate, it clearly shows that for highly charged colloidal particles, correlations between the condensed multivalent counterions cannot be ignored.

7.1. Overcharging

One consequence of strong electrostatic correlations is the phenomenon known as 'overcharging' [151–160]. Overcharging occurs as the result of highly favourable gain in electrostatic free energy due to strong positional correlations between the condensed counterions.

To understand better how the overcharging of colloidal particles comes about let us consider a simple case of one colloidal particle with a uniform surface charge $-Zq$ and radius a, at zero temperature [161, 162]. The question that we would like to answer is how many α-valent counterions should be placed on top of the colloidal particle in order to minimize the electrostatic energy of the resultant polyion–counterion complex? Naively we might suppose that the number of condensed counterions should be such as to neutralize completely the colloidal charge. This, indeed, would be the case if the charge of counterions was uniformly smeared over the surface of the colloid. In reality, the counterions are discrete entities and can gain favourable energy by maximizing their separation from one another. Let us calculate the electrostatic energy of the polyion–counterion complex:

$$E_n = \frac{Z^2q^2}{2\epsilon a} - \frac{Z\alpha nq^2}{\epsilon a} + F_n^{\alpha\alpha}. \tag{185}$$

The first term is the self energy of a polyion, the second term is the electrostatic energy of interaction between the polyion and n condensed α-ions, and the last term is the electrostatic energy of repulsion between the condensed counterions. Now, consider a one component plasma of n α-ions on the surface of a sphere of radius a but with a *uniform, neutralizing background charge* $-\alpha nq$. The electrostatic energy of this OCP can be expressed as the sum of contributions arising from the counterion–counterion interaction, counterion–background interaction, and the self energy of the background,

$$F_n^{\rm OCP} = F_n^{\alpha\alpha} - \frac{\alpha^2n^2q^2}{\epsilon a} + \frac{\alpha^2n^2q^2}{2\epsilon a}. \tag{186}$$

Substituting this expression into equation (185), the electrostatic energy of the polyion–counterion complex simplifies to

$$E_n = \frac{(Z-\alpha n)^2q^2}{2\epsilon a} + F_n^{\rm OCP}. \tag{187}$$

For low temperatures, the condensed counterions try to maximize their separation from one another. In the planar geometry the ground state corresponds to a triangular Wigner crystal. A similar arrangement of counterions will also be found on the surface of a spherical colloidal

particle, up to some topological defects. The electrostatic energy of a planar OCP has been discussed in section 4.1. For a spherical OCP the electrostatic energy at zero temperature is

$$F_n^{\mathrm{OCP}} = -M\frac{\alpha^2 q^2 n^{3/2}}{2\epsilon a}. \tag{188}$$

where M is the Madelung constant. Because of the topological difference between the plane and the surface of a sphere, we expect that the Madelung constant will not be exactly the same in the two cases. The difference, however, should not be very large as was confirmed in recent Monte Carlo simulations [162]. For concreteness, we shall use $M = 1.106$, the value for the planar OCP.

The effective charge of the polyion–counterion complex in units of $-q$ is

$$Z_{\mathrm{eff}} = Z - \alpha n^*. \tag{189}$$

where n^* is the number of condensed α-ions which minimize the electrostatic energy:

$$\left.\frac{\mathrm{d}E_n}{\mathrm{d}n}\right|_{n^*} = 0. \tag{190}$$

The effective charge is found to be

$$Z_{\mathrm{eff}} = -\frac{1+\sqrt{1+4\gamma^2 Z}}{2\gamma^2}, \tag{191}$$

where

$$\gamma = \frac{4}{3M\sqrt{\alpha}}. \tag{192}$$

We see that the effective charge of the polyion–counterion complex is inverted compared to the bare charge Z of the colloidal particle, so that the complex is overcharged. For highly charged colloids, the effective charge scales as the square root of the bare charge [163, 162]:

$$Z_{\mathrm{eff}} \approx -\frac{\sqrt{Z}}{\gamma}. \tag{193}$$

The analysis above was conducted for one colloidal particle at zero temperature. For solutions at finite concentration and temperature we face the same difficulties already encountered in our earlier discussion of charge renormalization in colloidal suspensions and polyelectrolyte solutions. In fact the problems are even more severe, since for multivalent counterions the PB equation fails completely. A logical step is to appeal to the density functional theory. The difficulty with the DFT lies in constructing a suitable density functional which can take into account the electrostatic correlations. For $Z:\alpha$ suspensions without salt, such a functional was presented in section 5.3. It was found using the DFT and the WS cell formalism [91], that the effective charge Z_{eff}, as the function of the bare charge Z, after reaching the maximum decreases, asymptotically going to zero as $Z \to \infty$. This behaviour is in striking contrast to the saturation of the effective charge predicted by the PB theory [22]. Unfortunately, it is difficult to construct a suitable density functional for suspensions containing salt. One alternative is to use the integral equations. The numerical complexity of these theories, however, tends to obscure the essential physics of the problem. Below, we shall present a simple phenomenological model of overcharging. Our goal is not quantitative accuracy, but rather physical insight into the mechanisms leading to the overcharging in the polyelectrolyte solutions.

7.2. Overcharging in electrolyte solutions

Consider a dilute colloidal suspension containing a monovalent salt at concentration c, and an α-valent salt at concentration c_α. In aqueous solution the monovalent salt dissociates producing 1 : 1 electrolyte, while the α-valent salt dissociates into α : 1 electrolyte. The inverse Debye length is

$$\xi_{\mathrm{D}}^{-1} = \kappa = \sqrt{8\pi\lambda_{\mathrm{B}} I}, \tag{194}$$

where the ionic strength is

$$I = \tfrac{1}{2}\left(\alpha^2 c_\alpha + \alpha c_\alpha + 2c\right). \tag{195}$$

For simplicity we shall restrict our attention to suspensions with monovalent salt near physiological concentrations—150 mM of NaCl. Under these conditions the Debye length is around 8 Å, and the interactions between the colloidal particles can be completely neglected. Furthermore, since the electrostatic attraction between the highly charged α-ions and colloids is so much stronger than the interaction between the monovalent counterions and colloids, the effective charge of the polyion–counterion complex is completely determined by the number of condensed α-ions.

We will define the counterions as free (not-associated) if they are farther than some distance δ from the colloidal surface. The 'agglomerate' is then defined as the polyion with its δ-sheath of surrounding counterions. We note, however, that not all of the α-ions in the agglomerate are actually associated with the polyion. The reason for this is that many of these ions have sufficient kinetic (thermal) energy to leave the vicinity of the colloidal particle. Only the counterions which have a total (kinetic plus potential) energy less than zero can be considered bound to the polyion. Unfortunately, it is not easy to come up with a practical implementation of this energetic criterion, except within a molecular dynamics simulation. On the other hand, a simple geometric criterion based on the polyion–counterion separation is easily implemented. Care, however, must be taken not to count all of the α-ions inside the agglomerate as belonging to the polyion–counterion complex. Below, we shall see how this can be accomplished.

We shall take δ to be of the order of a few angstroms, corresponding to the radius of a hydrated ion, $\delta \approx 2$ Å. Since the agglomerate is in contact with the bulk, its size is determined by the minimum of the grand potential function,

$$\Omega(n) = F(n) - n\mu_\alpha, \tag{196}$$

where $F(n)$ is the Helmholtz free energy of the agglomerate and μ_α is the chemical potential of α-ions inside the sheath. In thermal equilibrium, the chemical potential of α-ions inside the sheath equals the chemical potential of α-ions in the bulk of the suspension. For low bulk concentrations, μ_α can be approximated by the chemical potential of an ideal gas,

$$\mu_\alpha = \ln(c_\alpha \Lambda^3). \tag{197}$$

The Helmholtz free energy of an agglomerate is then

$$F_n = E_n + F_n^{\mathrm{solv}} + F_n^{\mathrm{ent}}. \tag{198}$$

The electrostatic free energy of an isolated polyion–counterion agglomerate E_n is given by equation (187); the solvation energy that the agglomerate gains when placed in an ionic environment is

$$F_n^{\mathrm{solv}} = -\frac{(Z-\alpha n)^2 q^2 \kappa a}{2\epsilon a(1+\kappa a)} \tag{199}$$

(see equation (137)) and the entropic energy of counterions inside the sheath is

$$F_n^{\rm ent} = k_{\rm B}T[n\ln(\rho_n\Lambda^3) - n], \tag{200}$$

where

$$\rho_n = \frac{n}{4\pi a^2\delta}. \tag{201}$$

For high valence counterions (strong-coupling limit) the free energy of the OCP can be approximated by that of a Wigner crystal, see section 4. If more accuracy is needed, one can use the extrapolation formulae based on Monte Carlo simulations [80]. Here, however, we shall content ourselves with the simplest approximation. The grand potential of an agglomerate is,

$$\beta\Omega(Z,n) = \frac{(Z-\alpha n)^2\lambda_{\rm B}}{2a(1+\kappa a)} - M\frac{\alpha^2\lambda_{\rm B}n^{3/2}}{2a} + n\ln\left(\frac{\rho_n}{c_\alpha}\right) - n. \tag{202}$$

The number of α-ions, n^*, inside an agglomerate is determined from the minimum of the grand potential:

$$\left.\frac{\partial\Omega(Z,n)}{\partial n}\right|_{n^*} = 0. \tag{203}$$

From the previous discussion recall that it is incorrect to associate the effective charge of the polyion–counterion complex with the value of n^*, i.e. $Z_{\rm eff} \neq Z - \alpha n^*$. The reason for this is that not all of the α-ions inside the δ-sheath are actually bound to the polyion. The real number of condensed counterions is $n^* - n_{\rm o}^*$, where the overestimate $n_{\rm o}^*$ can be obtained by considering the number of α-ions within the distance δ from the surface of a 'neutral' polyion, $Z = 0$:

$$\left.\frac{\partial\Omega(0,n)}{\partial n}\right|_{n_{\rm o}^*} = 0. \tag{204}$$

The effective charge of the polyion–α-ion complex is then

$$Z_{\rm eff} = Z - \alpha(n^* - n_{\rm o}^*). \tag{205}$$

If a small electric field is applied to the suspension, it is the $Z_{\rm eff}$ which will determine the electrophoretic mobility of colloidal particles [160, 110, 164, 111].

In figures 4 and 5 we present the effective colloidal charge as a function of concentration of divalent and trivalent counterions, for a suspension containing colloidal particles with $Z = 4000$ and $a = 1000$ Å. It is curious to note the appearance of a minimum in the effective charge as a function of trivalent ion concentration. In figure 6 we present a plot of the effective charge as a function of bare charge for colloids with $a = 1000$ Å, in a suspension containing monovalent salt at concentration $c = 0.15$ M and trivalent counterions at $c_3 = 0.01$ M. We note that unlike suspensions containing monovalent counterions, the effective charge does not saturate, instead it reaches the maximum value and then falls off sharply. For colloids with $Z \approx 11\,500$, the bare colloidal charge is completely neutralized by the α-ion condensation. In figure 7 we show the dependence of $Z_{\rm eff}$ on the amount of monovalent salt. For small concentrations of α-ions, the effective charge of the complex is found to increase with the concentration of monovalent salt, asymptotically approaching the bare value. However, when the concentration of multivalent ions reaches the critical value, there is a qualitative change of behaviour. At this point the effective charge is no longer a monotonically increasing function of the monovalent salt concentration. Instead, after reaching the maximum, $Z_{\rm eff}$ begins to decline, eventually going through the isolectric point ($Z_{\rm eff} = 0$) and charge inversion, figure 8.

We have considered only the simplest form of overcharging involving colloids and multivalent microions. There are a number of interesting variations. Recent developments in the field of gene therapy require construction of safe and efficient trans-cellular gene delivery

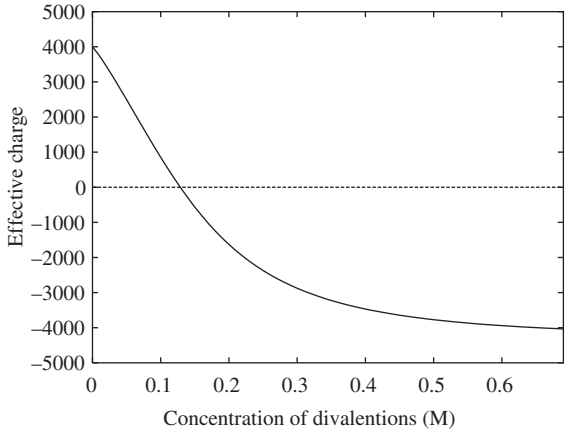

Figure 4. The effective (renormalized) charge of colloidal particles of $Z = 4000$, $a = 1000$ Å inside a suspension containing a monovalent salt at physiological concentration of 0.15 M, as a function of concentration of *divalent* counterions.

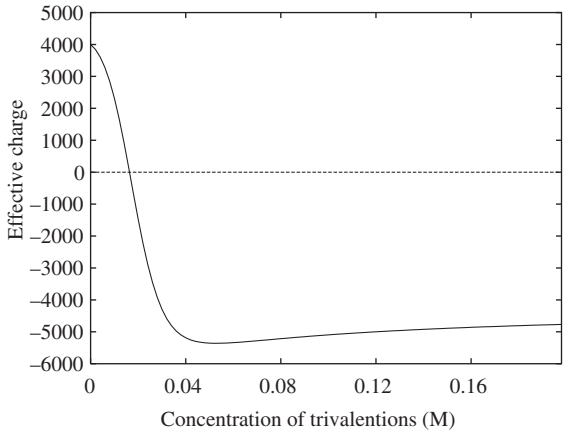

Figure 5. The effective (renormalized) charge of colloidal particles of $Z = 4000$, $a = 1000$ Å inside a suspension containing monovalent salt at a physiological concentration of 0.15 M, as a function of concentration of *trivalent* counterions.

systems [165–167]. Both the DNA and the phospholipids, which are the main constitutive component of a cellular membrane, are negatively charged. There is, therefore, a strong electrostatic repulsion between the DNA and the cellular membrane. This repulsion inhibits the transfection of naked DNA into the cells. Furthermore, *in vivo* the unprotected DNA is rapidly degraded by the nucleases present in plasma [168].

Much of the effort of gene therapy has been concentrated on viral transfection. A retrovirus (virus which incorporates its genetic material into the host genome) or adenovirus (which does not) has its genetic material removed and substituted by the gene that needs to be replicated. The modified viruses are then made to infect the cells, thus effectuating transfer of the genetic material. There are, unfortunately, a number of serious complications involved in this procedure. These range from a strong immunological response of an organism against the infecting virus, to potentially deadly consequences arising from the recombinant viral

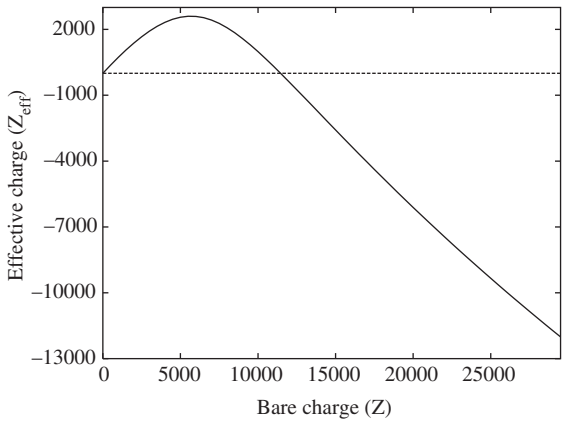

Figure 6. The effective (renormalized) charge of a colloidal particle with $a = 1000$ Å as a function of the bare charge Z, inside a suspension containing monovalent salt a at physiological concentration of $c = 0.15$ M and the trivalent ions at $c_3 = 0.01$ M.

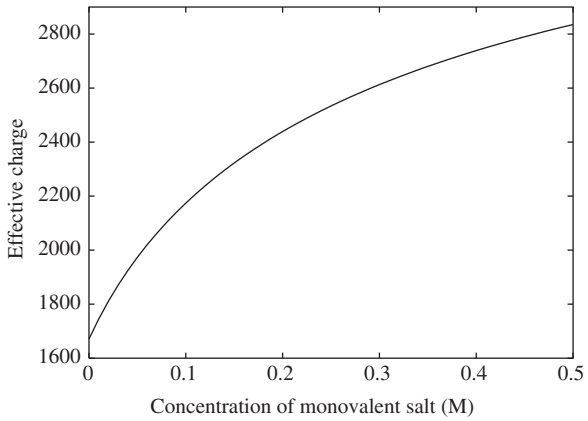

Figure 7. The effective (renormalized) charge of colloidal particles with $Z = 4000$ and $a = 1000$ Å inside a suspension containing trivalent counterions with $c_3 = 0.01$ M, as a function of concentration of monovalent salt c.

structures [168]. All these factors have led to attempts to develop non-viral transfection methods. One of the promising approaches relies on formation of the DNA–cationic lipid complexes, or lipoplexes for short. The hydrophobic interaction between the lipid tails, in addition to the cationic charge of their head groups, favours their agglomeration in the vicinity of polyions. For sufficiently long hydrocarbon chains, the gain in the hydrophobic energy resulting from the lipid condensation onto the DNA is sufficient to lead to the charge reversal of a lipoplex [169]. This happens even though the charged head group of the lipid is monovalent. The overcharging in this case is the consequence of hydrophobicity of lipid molecules and is not the result of electrostatic correlations.

A different method for gene transfection uses multivalent counterions such as Ca^{++}. The efficiency of Ca^{++} as a transfection agent might be due to its ability to neutralize or even invert the helical DNA charge without making it collapse. In this respect, it is quite different from the transition metal ions such as Mn^{++} and Cd^{++} which neutralize the DNA charge, but also lead to its condensation [170]. Formation of a neutral or overcharged DNA–Ca^{++} complex

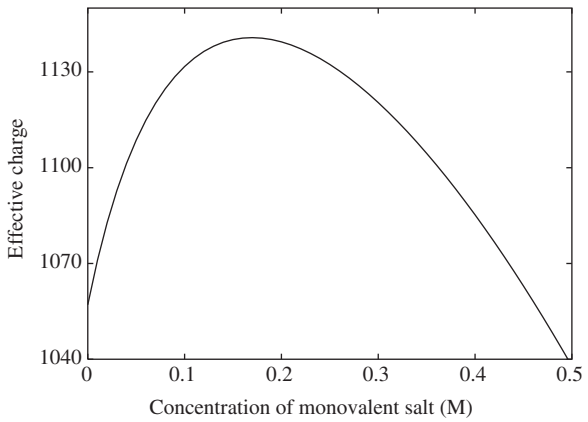

Figure 8. The effective (renormalized) charge of colloidal particles with $Z = 4000$ and $a = 1000$ Å inside the suspension containing trivalent counterions with $c_3 = 0.0135$ M, as a function of concentration of monovalent salt c.

allows the DNA to come into close proximity of the cellular membrane. Transfection might then be able to proceed through a reptation-like motion of the DNA through a cellular pore.

A very curious form of overcharging is found to occur in cellular compaction of the DNA. Virtually all of the DNA in nuclei of eukaryotic cells exists as a highly organized nucleoprotein fibre called chromatin. At the lowest level of the chromatin hierarchical structure is a nucleosome. Nucleosome resembles a thread wound around a cylindrical spool. The thread is the DNA molecule, while the spool is an octameric protein composed of eight smaller proteins called histones. Each octamer has 220 basic residues of which 148 are on the surface of the protein and are exposed to the solvent [171]. The rest are inside the protein core and are unlikely to be ionized. The maximum charge of an octamer is, therefore, $Q_{\rm oc}^{\rm max} = +220e$, but is likely to be significantly lower than this value under normal physiological conditions. Each octamer is encircled by 1.8 turns (147 base pairs) of the DNA thread, carrying a net charge of $-294e$. A nucleosome is, therefore, strongly overcharged by the associated DNA [157, 172–174].

It is interesting to speculate how nature is using this overcharging in the overall organization of the chromatin. The degree of compaction achieved inside the nucleus is quite astonishing. The total length of the DNA in the nucleus of a human cell is about 3.3 billion base pairs. If extended it would be more than 1 m long. Yet, it is compacted into a nucleus of diameter of 10 μm! Furthermore, all this compaction is done in such a way that the DNA is easily transcribed and replicated. This is, indeed, an astonishing feat of chemical engineering!

8. Like-charge attraction

8.1. Confined suspensions

One of the most curious phenomena which has produced much healthy debate in the condensed matter community is the appearance, under some circumstances, of attraction between like-charged macromolecules. The first observation that something might be missing in the traditional DLVO theory of colloidal stability came from the experimental observations of Ise *et al* of void structures inside the highly de-ionized suspensions [102]. The finding of voids and clusters has lead Sogami and Ise to propose a modification of the DLVO pair potential. In its stead they suggested a new potential obtained from the considerations of the Gibbs free

energy of suspension. The Sogami–Ise potential has created a lot of stimulating controversy, which is still in progress. The problem has been re-analysed by Overbeek [103], who has argued that the Sogami–Ise theory contains a basic thermodynamic inconsistency, which when resolved leads to the usual DLVO potential. This, however, has failed to settle the issue and a number of papers are still being published on the thermodynamics of suspensions whose particles interact by the Sogami–Ise potential.

The recently developed digital video microscopy (DVM) has the potential of putting an end to this long standing debate. DVM provides the possibility of explicitly measuring the interaction potential between two macromolecules in suspension. For dilute bulk suspensions, the interaction potential between two spherical colloidal particles was found to be completely consistent with the DLVO theory [119]. A surprising result appeared, however, when a highly de-ionized suspension was confined between glass plates [119, 175–177]. It was discovered that when particles were close to a wall, the pair interaction potential developed a strong attractive component. The attraction was quite long ranged, comparable in its extent to the diameter of colloidal particle.

It is well known that the glass–water interface has a significant negative charge due to dissociation of silanol groups. This charge density is comparable to that of colloidal particles. Furthermore, the absence of multivalent ions signifies that electrostatic correlations play only a marginal role. Under these conditions the PB theory should work quite well. Indeed, the first numerical solution of the PB equation for two macromolecules inside a cylindrical pore has found an attraction between the two polyions [178]. The triumph of the PB theory was, however, short lived. Soon afterwards the mathematical proofs were published demonstrating that the PB equation is incapable of producing attraction between like-charged particles in a confined geometry [179–182]. The numerical calculation was flawed, a consequence of the intrinsic difficulty of solving numerically non-linear partial differential equations with complicated boundary conditions. Indeed, it is extremely difficult to understand how a boundary can possibly change the interaction potential from repulsive to attractive. This is not to say that the presence of a boundary does not have a profound effect on the interactions. Consider one colloidal particle inside an electrolyte solution at a fixed distance from a wall or an interface. In general, the interface is characterized by a dielectric discontinuity. For the moment, however, let us ignore this complication and suppose that the two sides have exactly the same dielectric constant. The interface, then, separates two half-spaces, one containing electrolyte and another electrolyte-free. The fact that all the charges are confined to one half-space strongly modifies the distribution of ions around the colloid. While the ionic cloud is spherically symmetric for colloidal particles far from the interface (in the bulk), it develops a strong asymmetry, which results in a net dipole moment of the colloid–electrolyte system. This dipole produces an electric field

$$E \sim \frac{1}{r^3}, \tag{206}$$

which, because of the interface, cannot be screened. If there are two macromolecules separated by a distance r along the interface, the electric field produced by one macromolecule interacts with the dipole moment $\boldsymbol{p}$ of the charge distribution induced by the second macromolecule. This leads to the effective interaction potential which is repulsive and falls off as

$$w(r) \approx \boldsymbol{p} \cdot \boldsymbol{E} \sim \frac{1}{r^3}, \tag{207}$$

along the direction of an interface [183–187, 21]. This is exactly the same kind of effective interaction found for two-dimensional OCP, as seen in section 4.1.

It is clear that the existence of an interface or a wall strongly modifies the interaction potential between two macromolecules. Instead of an exponential screening found in the bulk,

the interactions along the interface fall off algebraically as $1/r^3$. The potential, however, is still repulsive, and it is difficult to see how anything can modify this conclusion at the level of electrostatics.

An interesting suggestion, which seems to account for the apparent attraction between the like-charged colloids near a wall, has been recently advanced by Squires and Brenner [188]. These authors attributed the attraction to non-equilibrium hydrodynamic flows which were not properly accounted for in experiment.

While the hydrodynamics seems to be able to explain the apparent attraction between the colloids near a wall, it is not sufficient to explain the results of experiments in which colloids are sandwiched between two glass plates, since, in this geometry, the hydrodynamic attraction mediated by one wall is suppressed by the second wall [189]. Furthermore, hydrodynamics does not help to understand the long-lived metastable crystalline structures observed by Grier *et al* [177, 190] when a low density suspension is compressed against a glass plate. If the interactions between particles are effectively repulsive, once the constraint is removed the crystals should melt within seconds. Instead, some crystalline regions are found to survive for as long as an hour. What is even more surprising is that the crystallites are actually three-dimensional, extending far beyond the region were the pairwise surface-mediated attraction is found. This phenomenon is very similar to the voids observed by Ise *et al*.

The nature of confinement-induced attraction between the like-charged particles remains an open question. However, we must stress again that concentration on pair potentials when studying *thermodynamic stability* of colloidal suspensions is a serious oversimplification. The DLVO theory was proposed as an indicator of *dynamical stability* against flocculation, driven by short-ranged van der Waals forces. If the equilibrium structure of a colloidal suspension is in question, inter-colloid pair potential is not sufficient and the full free energy must be considered. We have already seen in section 5.5 that a large gain in solvation free energy obtained from the polyion–counterion interactions, strongly favours the phase separation of suspension into the coexisting high and low density phases. This tendency is opposed by the counterion condensation, which renormalizes the effective colloidal charge. The theory presented in section 5.5 suggests that colloidal suspensions should phase separate when $\mathcal{C} > 15.2$. The counterion condensation, however, prevents $\mathcal{C}$ from reaching this threshold. Nevertheless for highly charged colloidal particles, $\mathcal{C}$ can come very close to the critical value, equation (125). This suggests that a de-ionized suspension of highly charged particles might actually be very close to criticality. This regime will be characterized by strong density fluctuations, which might appear as coexisting domains of voids and crystallites.

8.2. Correlation-induced attraction

DNA in aqueous solution is highly ionized due to dissociation of phosphate groups. This ionization results in one of the highest charge densities found in nature, one electron charge every 1.7 Å. In spite of this huge charge concentration, over a metre of DNA is packed into a nucleus of few micrometres. This efficient compaction is accomplished with the help of cationic proteins. The bacteriophages (viruses that infect bacteria) also use multivalent cations to package their DNA. Thus, the T7 bacteriophage head is 10^{-4} times smaller than the unpacked form of its DNA [191]. Furthermore, it is found that if the multivalent polyamines, known to exist in the host bacteria, are added to an *in vitro* solution containing DNA, the chains condense, forming toroids very similar in size and shape to the ones found *in vivo* [29, 30]. To produce condensation, multivalent counterions must somehow induce attraction between the different parts of the DNA [192–196]. The toroidal geometry is the result of the high intrinsic rigidity of the DNA molecule, which has a persistence length of $\xi_p = 500$ Å. If the compaction is done

in such a way that the local radius of curvature, r_c, exceeds $\xi_p/2\pi$ the cost in elastic energy will be prohibitively high. The requirement that $r_c > \xi_p/2\pi$, therefore, results in toroidal or spool-like condensates [197, 198].

In eukaryotic cells the cytosol is traversed by a complicated network of microfilaments which are made of a protein called F-actin [32, 199]. In spite of its high negative charge density F-actin, in the presence of multivalent counterions, agglomerates forming a network of bundles [200]. Addition of monovalent salt screens the electrostatic interactions and re-dissolves the bundles [201]. What is the action of multivalent counterions which induces attraction between the like-charged macromolecules [90, 202, 203, 33, 204, 34, 35, 170, 205–212]? To understand this we shall look at some very simple models [34].

Consider, first, two parallel polyions separated by a distance d inside a dilute solution containing α-valent ions. The polyions will be idealized as rigid lines of charge of length $L = Zb$. Each line has Z monomers of charge $-q$ spaced uniformly along the chain. The solvent is a uniform medium of dielectric constant ϵ. It is convenient to define the reduced polyion charge density as $\xi = q^2/\epsilon k_B T b$ [104]. A simple Manning argument then suggests that for $\xi > 1/\alpha$

$$n_c = \frac{Z}{\alpha}\left(1 - \frac{1}{\alpha\xi}\right) \tag{208}$$

α-ions condense onto each polyion. This is the lower bound on condensation, since the Manning argument does not take into account the correlations between the condensed counterions. Nevertheless, even this simple estimate suggests that 88% of the DNA's charge should be neutralized by the divalent counterions.

The associated counterions are free to move along the length of the DNA. We shall suppose that the only effect of condensation is the local renormalization of the monomeric charge from $-q$ to $(-1+\alpha)q$. Let us define the occupation variables σ_{ij}, with $i = 1, 2, \ldots, Z$ and $j = 1, 2$ in such a way that $\sigma_{ij} = 1$, if a counterion is condensed at the ith monomer of the jth polyion, and $\sigma_{ij} = 0$ otherwise. The occupation variables obey the constraint

$$\sum_{i=1}^{Z} \sigma_{i1} = \sum_{i=1}^{Z} \sigma_{i2} = n, \tag{209}$$

where n is the number of condensed α-ions. The interaction energy between the two polyions is

$$H = \frac{1}{2\epsilon} \sum_{i,i'=1}^{Z} \sum_{j,j'=1}^{2} \frac{q^2(1-\alpha\sigma_{ij})(1-\alpha\sigma_{i'j'})}{r(i, j; i', j')}, \tag{210}$$

where the sum is restricted to $(i, j) \neq (i', j')$, and

$$r(i, j; i', j') = b\sqrt{|i - i'|^2 + (1 - \delta_{jj'})x^2} \tag{211}$$

is the distance between the monomers located at (i, j) and (i', j'). $\delta_{jj'}$ is the Kronecker delta, and $x = d/b$. The partition function is

$$Q = \sum_{\{\sigma_{ij}\}}{}' \exp(-\beta H), \tag{212}$$

where the prime indicates that the trace is done under the constraint of equation (209). The force between the two polyions is

$$F = \frac{1}{b\beta} \frac{\partial \ln Q}{\partial x}. \tag{213}$$

This model is so simple that for polyions with not too high values of Z, the partition function can be solved explicitly [34]. For larger Z the model can be easily simulated. In figure 9 we present the force as a function of separation for two polyions with $Z = 20$ and n condensed divalent counterions. We see that in spite of the net like-charge, the two polyion–counterion complexes can attract each other at sufficiently small separations. Furthermore, we find that a critical number $n_1 = Z/2\alpha$ of condensed α-ions is necessary for the attraction to appear and that for monovalent counterions the interaction is always repulsive.

What is the nature of this attraction? To understand this let us consider the limit of small separations between the polyions. The configurations which dominate the partition function for $x \to 0$ are the ones in which condensed counterions on one polyion face the bare monomeric charges of the second polyion. In this limit the positions of condensed counterions on the two polyions become strongly correlated. The electrostatic energy of such configurations is

$$E \approx \frac{2n(1-\alpha)q^2}{\epsilon bx} + \frac{(Z-2n)q^2}{\epsilon bx}, \tag{214}$$

where n is the number of condensed counterions. The first term of equation (214) is the electrostatic energy due to attractive interaction between n condensed counterions and the bare monomers, while the second term is the repulsive interaction between $Z - 2n$ uncompensated monomeric charges. We see that, *if there are $n > Z/2\alpha$ condensed counterions*, the first term of equation (214) dominates and the force becomes attractive for sufficiently short separations. For $n < Z/2\alpha$ the force is repulsive for all temperatures. This is the general mechanism of attraction between the two polyion–counterion complexes at any finite temperature. To minimize the electrostatic *free* energy, the positions of condensed counterions on the two polyions become correlated. For sufficiently short separations and number of condensed counterions exceeding the critical threshold n_1, the correlation-induced attraction dominates the monopolar repulsion between the two complexes.

The strength of the counterion correlations increases with a decrease in temperature. The state of maximum correlation is at $T = 0$. This, however, does not imply that the attraction between two polyions is maximum at zero temperature [213]. The reason for this is that the ground state configuration, which corresponds to the lowest electrostatic energy, is not, in general, the configuration which maximizes the attractive force. For low temperatures and

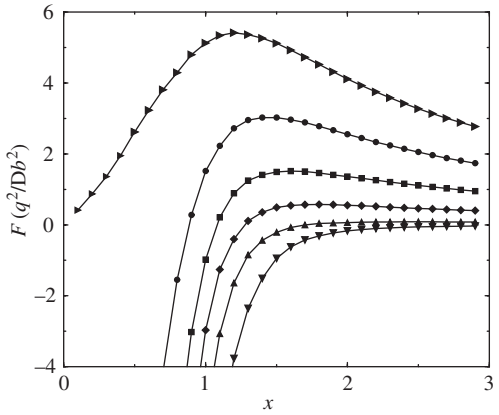

Figure 9. Force versus distance between polyions for $Z = 20$, $\alpha = 2$, $\xi = 2.283$ (corresponding to polymethacrylate) and $n = 5, \ldots, 10$ (from top to bottom) in the Monte Carlo simulation [34]. Positive force signifies repulsion between the complexes, while the negative force implies the existence of attraction.

short separations there are configurations which have force more attractive than the force in the ground state. For example, consider a polyion with $Z = 6$ and $n = 3$ condensed divalent counterions. The ground state corresponds to a staggered arrangement of counterions on the two polyions. The counterions on one polyion form the pattern $+-+-+-$ while on the second polyion they form a complimentary pattern $-+-+-+$. This leads to attraction between the complexes. However, it is easy to see that the configuration in which the counterions on the two polyions form the patterns $+++---$ and $---+++$, has larger electrostatic energy, but leads to stronger attraction. At finite temperature the total force, being a weighted mean of forces associated with all the configurations can, therefore, become more attractive than the force at $T = 0$. This curious behaviour, however, is confined to very small separations between the polyions. At larger distances the modulus of the attractive force is a monotonically decreasing function of temperature.

If the number of condensed counterions exceeds

$$n_{\mathrm{u}} = \frac{Z(\alpha + 1)}{2\alpha}, \tag{215}$$

the interaction between the two lines of charge becomes, once again, repulsive. The n_{u} corresponds to an overcharged configuration in which each polyion has the effective charge

$$Z^{\mathrm{u}}_{\mathrm{eff}} = Z - \alpha n_{\mathrm{u}} = -\frac{(\alpha - 1)}{2} Z. \tag{216}$$

It is curious that exactly this kind of re-entrant behaviour is observed for the DNA condensates [214–217]. In the absence of multivalent counterions, the DNA inside the solution has an extended configuration. When the concentration of multivalent salt is slowly raised, there comes a point at which the gain in electrostatic energy due to polyion–α-ion association overcomes the entropic loss due to α-ion confinement in the vicinity of the polyions. The condensed monovalent counterions are then released into solution and are replaced by the polyvalent α-ions. After the critical number of α-ions is associated to the polyion, the interaction between the separate segments of the DNA becomes attractive. This plus the intrinsic rigidity of the molecule drives its condensation into toroidal bundles. As the concentration of multivalent ions in solution increases further, the DNA–α-ion complex becomes overcharged. When the effective charge of the polyion reaches $Z^{\mathrm{u}}_{\mathrm{eff}}$, the interaction between the segments of the DNA becomes, once again, repulsive and the bundles re-dissociate. Clearly this simple model is incapable of accounting for all the intricacies of the DNA condensation; nevertheless, it sheds a lot of light on the role of electrostatic correlations in this interesting and important phenomenon.

The calculations above were presented for a very idealized model of interacting lines of charge. It is quite simple to modify the theory to account for finite polyion diameter. This modification, however, does not significantly affect the predictions of the theory. Attraction appears at small separations between the polyion surfaces—about 7 Å—after the critical number of α-ions is condensed onto the polyions [218]. We find that for macromolecules of finite diameter less counterions are needed to induce attraction than for the two lines of equivalent charge density [218]. Furthermore, the charge–charge correlations along the polyion are of very short range [218, 219], showing the absence of any long-range order between the condensed counterions, contrary to earlier speculations [202, 203, 35].

A very interesting experimental feature of rigid polyelectrolyte solutions containing multivalent counterions is that the correlation-induced attraction does not lead to phase separation [220]. Instead rod-like polyions associate into bundles of well-defined thickness. The precise bundle morphology depends on the persistence length of the polyions; all the bundles, however, tend to have a well-defined cross-sectional diameter. What can account

for this curious phenomenon? Correlation-induced attraction should favour an unconstrained growth of bundles [35]. In this respect the situation is very similar to that of an ionic crystal whose electrostatic energy is negative and is unbounded from below. What then cuts off the bundle size? The question is still not fully settled, but the indications are that the size of a bundle is controlled by the kinetics of its formation [221]. To understand this better, let us, once again, consider our simple model of two lines of charge with $n_1 < n < Z/\alpha$ condensed α-ions [213]. The polyions are located on two parallel planes separated by a distance d. They are free to rotate in their respective planes. The relative angle between the two lines is θ, with $\theta = 0$ corresponding to two parallel polyions, figure 10. It is convenient to define the adimensional free energy as

$$\mathcal{F} = -\frac{1}{\xi} \ln Q, \tag{217}$$

where Q is the partition function for the two polyions at fixed separation and angular orientation and ξ is the Manning parameter. In figure 11 the reduced free energy is plotted as a function

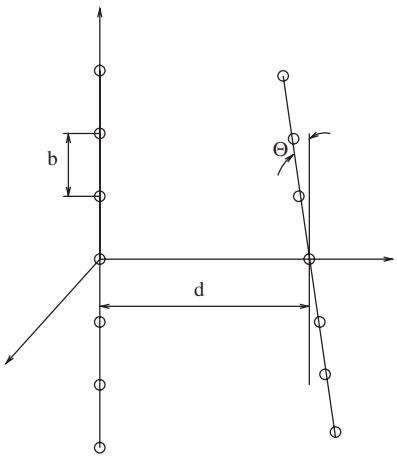

Figure 10. Two rod-like polyions of $Z = 7$.

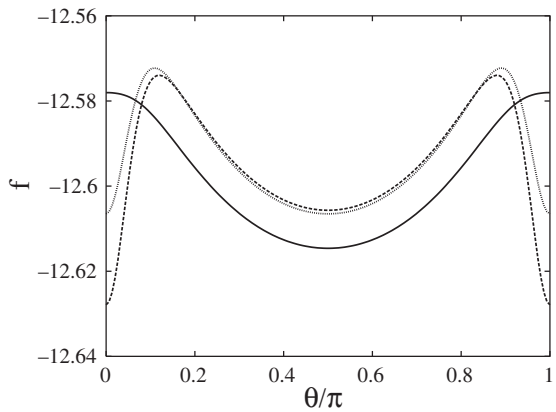

Figure 11. Reduced free energy $\mathcal{F}$ as a function of the angle θ for $Z = 9$, $n = 4$, $\alpha = 2$, $\xi = 2$. The curves shown are for: $x = 2$ (——), $x = 1.3676$ (· · · · · ·), and $x = 1.3$ (- - - -). Notice that the free energies for $\theta = 0$ and $\theta = \pi/2$ are equal in the second case [213].

of θ for various separations between the polyions. The free energy has two extrema at $\theta = 0$ and $\theta = \pi/2$, corresponding to the parallel and the perpendicular orientation between the polyions. In general, in situations when the attraction appears, the perpendicular configuration has the lowest free energy at large separations between the polyions, while at small distances the parallel configuration minimizes the free energy of the model. However, in some cases, a re-entrant behaviour is observed. The perpendicular configuration is the global minimum for both large and small distances, while the parallel configuration minimizes the free energy at intermediate distances.

At large separations the like-charged polyion–counterion complexes repel and are perpendicular to one another. As the distance between the polyions decreases, the global minimum passes from $\theta = \pi/2$ to $\theta = 0$, with $\theta = \pi/2$ becoming metastable. In order for polyions to form a bundle they must align, which means they have to overcome an activation barrier which separates the metastable minimum at $\theta = \pi/2$ from the global minimum at $\theta = 0$. It has been suggested that the height of the barrier grows with the number of polyions which are already in the bundle [221]. There comes a point when the barrier becomes so high that the thermal fluctuations are unable to overcome it and no new polyion can join the bundle. This puts an end to the bundle growth.

8.3. Counterion polarization

The correlation-induced attraction between the polyions discussed in the previous section is short ranged, decaying exponentially with the separation between the polyions. The line of charge model is, however, too simple to account for the counterion polarization [222]. Consider for example a rigid cylinder-like polyion or a spherical colloidal particle with a layer of condensed counterions. Suppose that there is no overcharging so that only some fraction of the polyion charge is neutralized by the condensed counterions. The electric field produced by one complex polarizes the counterions of the other complex, figure 12. This mechanism certainly provides an attractive component to the interaction potential [223]. However, can this induced attraction overcome the monopolar repulsion from the uncompensated colloidal charge $Z_{\rm eff}$? To understand this let us consider two colloidal particles or globular proteins separated by a distance R inside an electrolyte solution with Debye length $\xi_{\rm D}$. Each polyion has radius a, charge Z and n condensed α-ions. The leading order interaction between the two macromolecules at separation $R > \xi_{\rm D}$ is of DLVO form:

$$V_{\rm DLVO}(R) = Z_{\rm eff}^2 q^2 \theta^2 (\kappa a) \frac{{\rm e}^{-\kappa R}}{\epsilon R}, \tag{218}$$

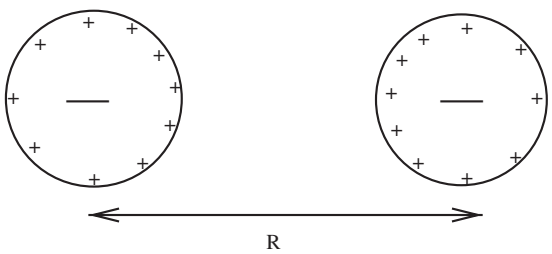

Figure 12. Two polyions with condensed counterions. Note that the net charge $Z_{\rm eff}$ of one complex polarizes the counterions of the other complex.

where $Z_{\rm eff} = Z - \alpha n$ The major corrections come from two effects. First, the presence of polyions produces holes in the ionic atmosphere. The charge of a hole is minus the charge of the counterions excluded from the region of space occupied by the macromolecule [117,118]. For example, the charge of the hole produced by the polyion $\mathcal{P}_2$ in the ionic atmosphere polarized by the polyion $\mathcal{P}_1$ is

$$Q_{h,2}(R) \approx -\frac{4\pi a^3 \rho_q(R)}{3}, \tag{219}$$

where the charge density at position R is given by the familiar DH expression,

$$\rho_q(R) = -\frac{\epsilon\kappa^2}{4\pi}\phi(R) = Z_{\rm eff} q \kappa^2 \theta(\kappa a)\frac{{\rm e}^{-\kappa R}}{4\pi R}. \tag{220}$$

Use of the linearized DH theory is justified by the fact that in equation (220) the renormalized effective charge enters. The electrostatic energy of the polyion–hole interaction is obtained using the charging process in which $q \to \lambda q$ and

$$V_{\rm h}(R) = Q_{\rm h,2}(R)\int_0^1 {\rm d}\lambda\, \phi(R;\lambda q) = \frac{1}{6}\kappa^2 a^3 Z_{\rm eff}^2 q^2 \theta^2(\kappa a)\frac{{\rm e}^{-2\kappa R}}{\epsilon R^2}, \tag{221}$$

at fixed κ. The second major correction to the DLVO potential is the result of polarization of the condensed layer of counterions surrounding the polyion. Thus, the electric field produced by the complex $\mathcal{P}_1$ induces a dipole moment in the complex $\mathcal{P}_2$, and vice versa. The local surface charge density of $\mathcal{P}_2$ is

$$\sigma(\boldsymbol{r}) = -\sigma_- + \sigma_+ {\rm e}^{-\beta\alpha q\phi(\boldsymbol{r})}, \tag{222}$$

where

$$\sigma_- = \frac{Zq}{4\pi a^2}, \tag{223}$$

$$\sigma_+ = \frac{\alpha n q}{4\pi a^2}, \tag{224}$$

and $\phi(\boldsymbol{r})$ is the local electrostatic potential felt by the condensed counterions at position $\boldsymbol{r}$. In the weak coupling limit, $\Gamma < 1$, the exponential in equation (222) can be linearized yielding the surface charge distribution

$$\sigma(\boldsymbol{r}) = \Delta\sigma - \frac{\epsilon\phi(\boldsymbol{r})}{2\pi\lambda_{\rm GC}}, \tag{225}$$

where $\Delta\sigma = \sigma_+ - \sigma_-$ and the Gouy–Chapman length is

$$\lambda_{\rm GC} = \frac{\epsilon}{2\pi\sigma_+\beta\alpha q}. \tag{226}$$

It is a straightforward calculation in electrostatics to show that the induced dipole moment of the complex $\mathcal{P}_2$ is

$$\boldsymbol{p} = \frac{2a^4\epsilon\boldsymbol{E}}{3\lambda_{\rm GC} + 2a}. \tag{227}$$

The electric field produced by $\mathcal{P}_1$ at distance R from its centre is

$$\boldsymbol{E}(\boldsymbol{R}) = Z_{\rm eff} q\theta(\kappa a)\frac{{\rm e}^{-\kappa R}(1+\kappa R)\boldsymbol{R}}{\epsilon R^3}. \tag{228}$$

The electrostatic energy of the induced dipole of $\mathcal{P}_2$ inside the electric field produced by $\mathcal{P}_1$ is obtained using the charging process

$$V_{\rm p} = \int_0^1 {\rm d}\lambda\, \boldsymbol{p}(\lambda q)\cdot\boldsymbol{E}(\lambda q), \tag{229}$$

where it is important to remember that $\lambda_{GC}(\lambda q) = \lambda^{-2}\lambda_{GC}(q)$. The same argument can be equally well applied to the interaction of $\mathcal{P}_2$ with the hole produced by $\mathcal{P}_1$ and to the polarization of $\mathcal{P}_1$ by the electric field produced by $\mathcal{P}_2$. Summing up all these contributions leads to

$$W(R) = Z_{\rm eff}^2 q^2 \theta^2(\kappa a)\frac{e^{-\kappa R}}{\epsilon R} - Z_{\rm eff}^2 q^2 \kappa^2 a^3 \theta^2(\kappa a) h\left(\frac{a}{\lambda_{GC}}\right)\frac{e^{-2\kappa R}}{\epsilon R^2}, \qquad (230)$$

where the scaling function $h(x)$ is,

$$h(x) = \frac{2}{3} - \frac{3}{2x}\ln\left(1 + \frac{2x}{3}\right). \qquad (231)$$

The correction to the DLVO potential is repulsive (hole dominated) for $a/\lambda_{GC} < 1.716\ldots$ and is attractive (dipole dominated) for $a/\lambda_{GC} > 1.716\ldots$. We note, however, that it is always 'doubly' screened and is, therefore, smaller than the leading order DLVO term for separations $R > \xi_D$ [224]. At shorter distances the approximations used to arrive at equation (230) fail, and the correlations between the condensed counterions must be explicitly taken into account [223]. For multivalent counterions, the electrostatic correlations result in short-ranged attraction, similar to the one discussed in the previous section.

9. Conclusions

We have explored the role of electrostatic correlations in systems ranging from classical plasmas to molecular biology. We saw how the positional correlations between the ions of an electrolyte can result in a thermodynamic instability. We also saw how the strong correlations between the polyions and the counterions lead to colloidal charge renormalization which stabilizes de-ionized suspensions against a phase separation. For two-dimensional plasmas, electrostatic correlations are responsible for the metal–insulator transition. The critical behaviour of superfluid ^{4}He films, the roughening transition of crystal interfaces [225], the melting of two-dimensional solids, and the criticality of XY magnets [64] are all governed by the 'electrostatic' interactions between the topological defects (charges). Nature has learned to take full advantage of the electrostatic correlations to efficiently package huge amounts of genetic material into tiny regions of space.

Throughout this review we have come to rely on some simple models in order to understand complex physical phenomena. While these models are often sufficient to grasp the underlying physics, it is quite easy to push the models too far. This is particularly the case when one deals with specific structural properties of biomolecules [226–228]. As soon as the length scales of the order of a few angstroms become important, approximation of water as a uniform dielectric medium is no longer sufficient [229, 230]. Under these conditions reliance on simple models, which treat macromolecules and solvents as dielectrics, is probably no more than wishful thinking. A careful path must be treaded between simplification and over-simplification.

Acknowledgments

I am grateful to J J Arenzon, M C Barbosa, A Diehl, M E Fisher, J E Flores-Mena, P Kuhn, X-J Li, J Stilck, and M N Tamashiro who have collaborated closely with me on topics covered in this review. Special thanks are due to Michael Fisher and Shubho Banerjee who have kindly provided me with figure 1 which compares the predictions of various theories of asymmetric electrolytes. Last but not least, I would like to thank Eric Westhof without whose encouragement this review would not have been written.

References

[1] van der Waals J D 1873 *PhD Thesis* Leiden
[2] Weiss P 1907 *J. Phys. Radium* (Paris) **6** 667
[3] Gouy G L 1910 *J. de Phys.* **9** 457
[4] Chapman D L 1913 *Phil. Mag.* **25** 475
[5] Fisher M E and Levin Y 1993 *Phys. Rev. Lett.* **71** 3826
[6] Levin Y and Fisher M E 1996 *Physica* A **225** 164
[7] Groh B and Dietrich S 1994 *Phys. Rev.* E **50** 3814
[8] Groh B and Dietrich S 1996 *Phys. Rev.* E **54** 1687
[9] Groh B and Dietrich S 1997 *Phys. Rev. Lett.* **79** 749
[10] Banerjee S, Griffiths R B and Widom M 1998 *J. Stat. Phys.* **93** 109
[11] de Gennes P G and Pincus P 1970 *Phys. Kondens. Mater.* **11** 189
[12] Rushbrooke G S, Stell G and Høye J S 1973 *Molec. Phys.* **26** 1199
[13] Weis J J and Levesque D 1993 *Phys. Rev. Lett.* **71** 2729
[14] Caillol J-M 1993 *J. Chem. Phys.* **98** 9835
[15] van Leeuwen M E and Smit B 1993 *Phys. Rev. Lett.* **71** 3991
[16] Levin Y 1999 *Phys. Rev. Lett.* **83** 1159
[17] Sear R P 1996 *Phys. Rev. Lett.* **76** 2310
[18] van Roij R 1996 *Phys. Rev. Lett.* **76** 3348
[19] Tavares J M, Telo da Gama M M and Osipov M A 1997 *Phys. Rev.* E **56** R6252
[20] van Roij R and Hansen J P 1997 *Phys. Rev. Lett.* **79** 3082
[21] Hansen J P and Löwen H 2000 *Annual Rev. Phys. Chem.* **51** 209
[22] Alexander S *et al* 1984 *J. Chem. Phys.* **80** 5776
[23] Levin Y, Barbosa M C and Tamashiro M N 1998 *Europhys. Lett.* **41** 123
[24] Diehl A, Barbosa M C and Levin Y 2001 *Europhys. Lett.* **53** 86
[25] Patey G N 1980 *J. Chem. Phys.* **72** 5763
[26] Guldbrand L, Jonsson B, Wennerstrom H and Linse P 1984 *J. Chem. Phys.* **80** 2221
[27] Linse P and Lobaskin V 1999 *Phys. Rev. Lett.* **83** 4208
[28] Lobaskin V 2000 *J. Mol. Liq.* **84** 131
[29] Bloomfield V A 1991 *Biopolymers* **31** 1471
[30] Bloomfield V A 1997 *Biopolymer* **44** 269
[31] Klimenko S M 1967 *J. Mol. Biol.* **23** 523
[32] Tang J X 1996 *Ber. Bunsenges. Phys. Chem.* **100** 796
[33] Ha B-Y and Liu A J 1997 *Phys. Rev. Lett.* **79** 1289
[34] Arenzon J J, Stilck J F and Levin Y 1999 *Eur. Phys. J.* B **12** 79
[35] Shklovskii B I 1999 *Phys. Rev. Lett.* **82** 3628
[36] Debye P W and Hückel E 1923 *Phys. Z.* **24** 185
[37] Netz R R and Orland H 1999 *Europhys. Lett.* **45** 726
[38] Belloni L 1986 *Phys. Rev. Lett.* **57** 2026
[39] Fisher M E and Fishman S 1981 *Phys. Rev. Lett.* **47** 421
[40] Belloni L 1993 *J. Chem. Phys.* **98** 8080
[41] Sabir A K, Bhuiyan L B and Outhwaite C W 1998 *Mol. Phys.* **93** 405
[42] Panagiotopoulos A Z and Fisher M E 1902 *Phys. Rev. Lett.* **88** 045101
[43] Camp P J and Patey G N 1999 *J. Chem. Phys.* **111** 9000
[44] Ng K-C 1974 *J. Chem. Phys.* **61** 2680
[45] Kjellander R and Marcelja S 1984 *Chem. Phys. Lett.* **112** 49
[46] Kjellander R and Marcelja S 1986 *J. Phys. Chem.* **90** 1230
[47] Hansen J P and McDonald I R 1976 *Theory of Simple Liquids* (New York: Academic)
[48] Arrhenius S 1887 *Z. Phys. Chem.* **1** 631
[49] Glasstone S 1946 *Textbook of Physical Chemistry* (New York: van Nostrand)
[50] Onsager L 1933 *Chem. Revs.* **13** 1933
[51] Fisher M E, Li X-J and Levin Y 1995 *J. Stat. Phys.* **79** 1
[52] Pitzer K S 1995 *J. Phys. Chem.* **99** 13070
[53] Narayanan T and Pitzer K S 1995 *J. Chem. Phys.* **102** 8118
[54] Narayanan T and Pitzer K S 1994 *Phys. Rev. Lett.* **73** 3002
[55] Wiegand S *et al* 1998 *J. Chem. Phys.* **109** 9038
[56] Fisher M E and Lee B P 1996 *Phys. Rev. Lett.* **77** 3561

[57] Luijten E, Fisher M E and Panagiotopoulos A Z cond-mat/0112388, unpublished
[58] Bjerrum N 1926 *Kgl. Dan. Vidensk. Selsk. Mat.-fys. Medd.* **7** 1
[59] Ebeling W 1968 *Z. Phys. Chem.* (Leipzig) **238** 400
[60] Falkenhagen H and Ebeling W 1971 *Ionic Interactions* ed S Petrucci (New York: Academic)
[61] Caillol J M, Levewque D and Weis J J 1997 *J. Chem. Phys.* **107** 1565
[62] Orkoulas G and Panagiotopoulos A Z 1999 *J. Chem. Phys.* **110** 1581
[63] Yan Q and de Pablo J J 1999 *J. Chem. Phys.* **111** 9509
[64] Nelson D 1983 *Phase Transitions and Critical Phenomena* ed C Domb and J L Lebowitz (New York: Academic) p 1
[65] Mermin N D 1968 *Phys. Rev.* **171** 272
[66] Kosterlitz J and Thouless D 1973 *J. Phys.* C **6** 1181
[67] Levin Y, Li X-J and Fisher M E 1994 *Phys. Rev. Lett.* **73** 2716
[68] Caillol J-M 1994 *J. Chem. Phys.* **100** 2161
[69] Orkoulas G and Panagiotoupolos A Z 1996 *J. Chem. Phys.* **104** 7205
[70] Minnhagen P 1987 *Rev. Mod. Phys.* **59** 1001
[71] Baus M and Hansen J P 1980 *Phys. Rep.* **59** 1
[72] Nordholm S 1984 *Chem. Phys. Lett.* **105** 302
[73] Tamashiro M N, Levin Y and Barbosa M C 1999 *Physica* A **268** 24
[74] Weeks J D 1981 *Phys. Rev.* B **24** 1530
[75] Pollock E L and Hansen J P 1973 *Phys. Rev.* A **8** 3110
[76] Rogers F J, Young D A, DeWitt H E and Ross M 1983 *Phys. Rev.* A **28** 2990
[77] Crandall R S and Williams R 1971 *Phys. Lett.* A **34** 404
[78] Grimes C C and Adams G 1979 *Phys. Rev. Lett.* **42** 795
[79] Chaplik A V 1971 *Sov. Phys. JETP* **35** 395
[80] Gann R C, Chakravarty S and Chester G V 1979 *Phys. Rev.* B **20** 326
[81] Flores-Mena J E, Barbosa M C and Levin Y 2001 *Phys. Rev.* E **63** 066104
[82] Velazquez E S and Blum L 1997 *Physica* A **244** 453
[83] Bloch F 1930 *Z. Phys.* **61** 206
[84] Peierls R E 1935 *Ann. Inst. Henri Poincare* **5** 177
[85] Landau L D 1937 *Phys. Z. Sowjetunion* **II** 26
[86] Totsuji H 1975 *J. Phys. Soc. Japan* **39** 253
[87] Totsuji H 1976 *J. Phys. Soc. Japan* **40** 857
[88] Wennerström H, Jönsson B and Linse P 1982 *J. Chem. Phys.* **76** 4665
[89] Penfold R, Nordholm S, Jönsson B and Woodward C E 1990 *J. Chem. Phys.* **92** 1915
[90] Stevens M J and Robbins M O 1990 *Europhys. Lett.* **12** 81
[91] Groot R D 1991 *J. Chem. Phys.* **95** 9191
[92] Barbosa M C, Deserno M and Holm C 2000 *Europhys. Lett.* **52** 80
[93] Tarazona P 1985 *Phys. Rev.* A **31** 2672
[94] Curtin W A and Ashcroft N W 1985 *Phys. Rev.* A **32** 2909
[95] Tamashiro M N, Levin Y and Barbosa M C 1998 *Physica* A **258** 341
[96] Tata B V R, Rajalakshmi M and Arora A 1992 *Phys. Rev. Lett.* **69** 3778
[97] Palberg T and Wurth M 1994 *Phys. Rev. Lett.* **72** 786
[98] Tata B V R and Arora A K 1994 *Phys. Rev. Lett.* **72** 787
[99] Derjaguin B V and Landau L 1941 *Acta Physicochimica* (USSR) **14**
[100] Verwey E J W and Overbeek J T G 1948 *Theory of the Stability of Lyophobic Colloids* (Amsterdam: Elsevier)
[101] Robbins M O, Kremer K and Grest G S 1998 *J. Chem. Phys.* **88** 3286
[102] Sogami I and Ise N 1984 *J. Chem. Phys.* **81** 6320
[103] Overbeek J. T G 1987 *J. Chem. Phys.* **87** 4406
[104] Manning G S 1969 *J. Chem. Phys.* **51** 924
[105] Oosawa F 1971 *Polyelectrolytes* (New York: Marcel Dekker)
[106] Manning G S 1978 *Q. Rev. Biophys.* II **2** 179
[107] Russel W B, Saville D A and Schowalter W R 1989 *Colloidal Dispersions* ed G K Batchelor (Cambridge: Cambridge University Press)
[108] Trizac E, Bocquet L and Aubouy M cond-mat/0201510, unpublished
[109] Quesada-Pérez M, Callejas-Fernández J and Hidalgo-Álvarez R 2000 *Phys. Rev.* E **61** 574
[110] Fernández-Nieves A, Fernández-Barbero A and de las Nieves F J 2000 *Langmuir* **16** 4090
[111] Fernández-Nieves A, Fernández-Barbero A and de las Nieves F J 2001 *Phys. Rev.* E **63** 041404

[112] Stevens M J, Falk M L and Robbins M O 1995 *J. Chem. Phys.* **104** 5209
[113] Belloni L 1998 *Colloids and Surfaces* A **140** 227
[114] Löwen H, Hansen J P and Madden P A 1993 *J. Chem. Phys.* **98** 3275
[115] McMillan W G and Mayer J E 1945 *J. Chem. Phys.* **13** 276
[116] Marcus R A 1955 *J. Chem. Phys.* **23** 1057
[117] Li X-J, Levin Y and Fisher M E 1994 *Europhys. Lett.* **26** 683
[118] Fisher M E, Levin Y and Li X-J 1994 *J. Chem. Phys.* **101** 2273
[119] Crocker J M and Grier D G 1994 *Phys. Rev. Lett.* **73** 352
[120] Mansoori G A and Canfield F B 1969 *J. Chem. Phys.* **51** 4958
[121] Firey B and Ashcroft N W 1977 *Phys. Rev.* A **15** 2072
[122] Warren P B 2000 *J. Chem. Phys.* **112** 4683
[123] Linse P and Lobaskin V 2000 *J. Chem. Phys.* **112** 3917
[124] Linse P 2000 *J. Chem. Phys.* **113** 4359
[125] von Grünberg H H, van Roij R and Klein G 2001 *Europhys. Lett.* **55** 580
[126] Deserno M and von Grünberg H H cond-mat/0202029, unpublished
[127] Klein R and von Grünberg H H 2001 *Pure Appl. Chem.* **73** 1705
[128] Tamashiro M N and Schiessel H, unpublished
[129] Chan D Y C, Linse P and Petris S N 2001 *Langmuir* **17** 4202
[130] Barrat J-L and Joanny J-F 1996 *Adv. Chem. Phys.* **94** 1
[131] Netz R and Andelman D cond-mat/0203364, unpublished
[132] Higgs P G and Joanny J-F 1991 *J. Chem. Phys.* **94** 1543
[133] Kantor Y and Kardar M 1995 *Phys. Rev.* E **51** 1299
[134] Dobrynin A V and Rubinstein M 1995 *J. Phys.* II (France) **5** 677
[135] Levin Y and Barbosa M C 1996 *Europhys. Lett.* **31** 513
[136] de Gennes P G, Pincus P and Velasco R M 1976 *J. Phys.* (France) **37** 1461
[137] Khokhlov A R 1980 *J. Phys.* A **13** 979
[138] Stevens M J and Kremer K 1995 *J. Chem. Phys.* **103** 1669
[139] Dobrynin A V, Rubinstein M and Obukhov S P 1996 *Macromolecules* **9** 2974
[140] Micka U, Holm C and Kremer K 1999 *Langmuir* **15** 4033
[141] Lee N and Thirumalai D 2001 *Macromolecules* **34** 3446
[142] Fuoss R M, Katchalsky A and Lifson S 1951 *Proc. Natl. Acad. Sci. USA* **37** 579
[143] Deserno M and Holm C 2001 *Electrostatic Effects in Soft Matter and Biophysics* vol 46, ed C Holm *et al* (Dordrecht: Kluwer)
[144] Levin Y 1996 *Europhys. Lett.* **34** 405
[145] Kuhn P, Levin Y and Barbosa M C 1998 *Macromolecules* **31** 8347
[146] Levin Y 1998 *Physica* A **257** 408
[147] Kuhn P, Levin Y and Barbosa M C 1998 *Chem. Phys. Lett.* **298** 5
[148] Löwen H 1994 *J. Chem. Phys.* **100** 6738
[149] Kholodenko A L and Beyerlein A L 1995 *Phys. Rev. Lett.* **74** 4679
[150] Israelachvili J N, Mitchell D J and Ninham B W 1976 *J. Chem. Society-Faraday Trans.* II **72** 1525
[151] Torrie G M and Valleau J P 1980 *J. Chem. Phys.* **73** 5807
[152] Gonzales-Tovar E, Lozada-Cassou M and Henderson D 1985 *J. Chem. Phys.* **83** 361
[153] Lozada-Cassou M, Saavedra-Barrera R and Henderson D 1982 *J. Chem. Phys.* **77** 5150
[154] Messina R, Tovar E G, Lozada-Cassou M and Holm C cond-mat/0111335, unpublished
[155] Nguyen T T and Shklovskii B I cond-mat/0005304, unpublished
[156] Nguyen T T, Grosberg A Y and Shklovskii B I 2000 *Phys. Rev. Lett.* **85** 1568
[157] Grosberg A Y, Nguyen T T and Shklovskii B I 2002 *Rev. Mod. Phys.* **74** 329
[158] Nguyen T T, Grosberg A Y and Shklovskii B I 2000 *Phys. Rev. Lett.* **85** 1568
[159] Nguyen T T, Grosberg A Y and Shklovskii B I 2000 *J. Chem. Phys.* **113** 1110
[160] Tanaka M and Grosberg A Y 2001 *J. Chem. Phys.* **115** 567
[161] Messina R, Holm C and Kremer K 2000 *Phys. Rev. Lett.* **85** 872
[162] Messina R, Holm C and Kremer K 2001 *Phys. Rev.* E **64** 021405
[163] Shklovskii B I 1999 *Phys. Rev.* E **60** 5802
[164] Golestanian R 2000 *Europhys. Lett.* **52** 47
[165] Friedmann T 1997 *Sci. Am.* **276** 80
[166] Felgner P L 1987 *Sci. Am.* **276** 86
[167] Felgner P L and Ringold G M 1989 *Nature* **337** 387
[168] Hope M J, Mui B, Ansell S and Ahkong Q F 1998 *Mol. Membrane Biol.* **15** 1

[169] Kuhn P S, Levin Y and Barbosa M C 1999 *Physica* A **274** 8
[170] Kornyshev A A and Leikin S 1999 *Phys. Rev. Lett.* **82** 4138
[171] Khrapunov S N, Dragan A I, Sivolob A V and Zagariya A M *Biochimica et Biophysica Acta-Gene Structure and Expression* **1351** 213
[172] Andelman D and Joanny J-F 2000 *Cr. Acad. Sci. IV-Phys.* **1** 1153
[173] Park S Y, Bruinsma R F and Gelbart W M 1999 *Europhys. Lett.* **46** 493
[174] Mateescu E M, Jeppesen C and Pincus P 1999 *Europhys. Lett.* **46** 493
[175] Kepler G M and Fraden S 1994 *Phys. Rev. Lett.* **73** 356
[176] Grier D G and Crocker J C 2000 *Phys. Rev.* E **61** 980
[177] Grier D G 2000 *J. Phys.: Condens. Matter* **12** A85
[178] Bowen W R and Sharif A O 1998 *Nature* **393** 663
[179] Neu J 1999 *Phys. Rev. Lett.* **82** 1072
[180] Sader J and Chan D Y C 1999 *J. Colloid Interfacial Sci.* **213** 268
[181] Trizac E 2000 *Phys. Rev.* E **62** R1465
[182] Trizac E and Raimbault J L 1999 *Phys. Rev.* E **60** 6530
[183] Jancovici B 1982 *J. Stat. Phys.* **28** 43
[184] Jancovici B 1982 *J. Stat. Phys.* **29** 263
[185] Goulding D and Hansen J P 1998 *Mol. Phys.* **95** 649
[186] Goulding D and Hansen J P 1999 *Europhys. Lett.* **46** 407
[187] Allen R, Hansen J P and Melchionna S 2001 *Phys. Chem. Chem. Phys.* **3** 4177
[188] Squires T and Brener M 2001 *Phys. Rev. Lett.* **86** 5266
[189] Grier D G Private communication, unpublished
[190] Larsen A E and Grier D G 1996 *Phys. Rev. Lett.* **76** 3862
[191] Gosule L C and Schellman J A 1976 *Nature* **259** 333
[192] Strey H H, Podgornik R, Rau D C and Parsegian V A 1998 *Curr. Opin. Struc. Biol.* **8** 309
[193] Podgornik R and Parsegian V A 1998 *Phys. Rev. Lett.* **80** 1560
[194] Golestanian R, Kardar M and Liverpool T B 1999 *Phys. Rev. Lett.* **82** 4456
[195] Solis F and de la Cruz M 2000 *J. Chem. Phys.* **112** 2030
[196] Ariel G and Andelman D cond-mat/0206361, unpublished
[197] Odijk T 1998 *Biophys. J.* **75** 1223
[198] Park S Y, Harries D and Gelbart W M 1998 *Biophys. J.* **75** 714
[199] Tang J X and Janmey P A 1996 *J. Biol. Chem.* **271** 8556
[200] Borukhov I, Bruinsma R F, Gelbart W M and Liu A J 2002 *J. Chem. Phys.* **117** 462
[201] Tang J X, Szymanski P T, Janmey P A and Tao T 1997 *Eur. J. Biochem.* **247** 432
[202] Rouzina I and Bloomfield V 1996 *J. Chem. Phys.* **100** 9977
[203] Grønbech-Jensen N, Mashl R J, Bruinsma R F and Gelbart W M 1997 *Phys. Rev. Lett.* **78** 2477
[204] Levin Y, Arenzon J J and Stilck J F 1999 *Phys. Rev. Lett.* **83** 2680
[205] Kornyshev A A and Leikin S 2000 *Phys. Rev.* E **62** 2576
[206] Allahyarov E, D'Amico I and Lowen H 1998 *Phys. Rev. Lett.* **81** 1334
[207] Kardar M and Golestanian R 1999 *Rev. Mod. Phys.* **71** 1233
[208] Grønbech-Jensen N, Beardmore K M and Pincus P 1998 *Physica* A **261** 74
[209] Lau A W C, Levine D and Pincus P 2000 *Phys. Rev. Lett.* **84** 4116
[210] Solis F and de la Cruz M 1999 *Phys. Rev.* E **60** 4496
[211] Schmidt A 1999 *J. Chem. Phys.* **110** 113
[212] Schmidt A B 2001 *Physica* A **293** 21
[213] Stilck J F, Levin Y and Arenzon J J 2002 *J. Stat. Phys.* **106** 287
[214] Pelta J *et al* 1996 *Biophys. J.* **71** 48
[215] Pelta J *et al* 1996 *J. Biol. Chem.* **271** 5656
[216] Saminathan M *et al* 1999 *Biochemistry* **38** 3821
[217] Nguyen T T, Rouzina I and Shklovskii B I 2000 *J. Chem. Phys.* **112** 2562
[218] Diehl A, Carmona H A and Levin Y 2001 *Phys. Rev.* E **64** 011804
[219] Deserno M, Arnold A and Holm C cond-mat/0206126, unpublished
[220] Ha B-Y and Liu A J cond-mat/0003162, unpublished
[221] Ha B-Y and Liu A J 1999 *Europhys. Lett.* **46** 624
[222] Belloni L 2000 *J. Phys.: Condens. Matter* **12** R549
[223] Levin Y 1999 *Physica* A **265** 432
[224] Ninham B and Parsegian V A 1971 *J. Theoret. Biol.* **31** 405
[225] Chui S T and Weeks J D 1976 *Phys. Rev.* B **14** 4978

[226] Gilson M K *et al* 1985 *J. Mol. Biol.* **184** 503
[227] Penfold R, Warwicker J and Jonsson B 1998 *J. Phys. Chem.* B **102** 8599
[228] Spassov V and Bashford D 1998 *Prot. Sci.* **7** 554
[229] Auffinger P and Westhof E 1998 *Curr. Opin. Struc. Biol.* **8** 227
[230] Auffinger P, LouiseMay S and Westhof E 1996 *Faraday Discuss.* **103** 151