# Instantons in a $U(1)$ Lattice Gauge Theory:
# A Coulomb Dipole Gas

James Glimm[*]

Rockefeller University, New York, NY 10021, USA

Arthur Jaffe[**]

Harvard University, Cambridge, MA 02138, USA

**Abstract.** We study the decomposition $A = A_I + A_{SW}$ of a $U(1)$ lattice gauge field into instanton and spin wave parts. The action also decomposes, $\mathscr{A} = \mathscr{A}_I + \mathscr{A}_{SW} + R$. Here $\mathscr{A}_I$ is a Coulomb dipole gas, $\mathscr{A}_{SW}$ is a zero mass free field, and $R$ is a higher order remainder. We study $\mathscr{A}_I$ in detail, for $d \geq 4$, in the dilute gas case (which corresponds to the low temperature limit of the gauge field theory). We establish the leading behavior of the free energy: $f \sim \varepsilon^{-d} a \zeta$. Here $\varepsilon$ is the lattice spacing, $a$ is a geometrical constant and $\zeta$ is an activity defined in terms of a small number of instanton configurations. Our methods suggest the absence of screening in the dilute dipole gas, $d \geq 4$, in contrast to Debye screening for the dilute monopole gas.

## 1. Introduction

It has been proposed by Gell-Mann and others that a gauge field coupled to a fermion (quark) field may describe the internal structure of protons, neutrons, etc., as quark triplets, etc. In order to account for the strong binding of individual quarks, as well as the observed property of asymptotic freedom of the physical particles, a qualitative understanding of the phase transitions and critical points of the gauge-quark system appears necessary [11]. The important properties of the critical points are inferred from renormalization group and semiclassical methods.

In this paper we study some steps in this program on a rigorous mathematical level. See [8] for an introduction to the mathematics of lattice gauge fields. Following Polyakov [9], we formally derive the instanton interaction in terms of a dilute Coulomb gas for a $U(1)$ lattice gauge theory. For $d \geq 4$, the gas consists of vector charges, and a conservation law restricts the charges to occur only in dipoles. Since the dipole-dipole interaction is not integrable at long distances, standard methods do not apply. Nevertheless, we prove an upper and lower bound on the free

energy of the dipole gas. The essential step is to use a cancellation between the forces produced on a test dipole by oppositely oriented dipoles occurring in a block. Our conclusions suggest that higher order corrections to the dipole forces do not generate a mass (finite correlation length at sufficiently low temperature. This is in contrast to the $d=3$ dilute Coulomb gas of scalar charges for which Debye suggested a finite correlation length, see [1].

We study in detail the action

$$\mathscr{A} = -\beta \sum \cos(\varepsilon^2 gF) . \tag{1.1}$$

We first separate $F = F_I + F_{SW}$ into two parts: The spin wave part $F_{SW}$ characterizes the oscillations withing one cosine well, while the instanton part $F_I$ characterizes the transitions between different wells of the cosine function. The splitting gives rise to a magnetic charge density $Q$ with the geometric interpretation

$$Q = \partial S ; \tag{1.2}$$

namely $Q$ is the boundary $(d-3$ dimensions) of a Dirac surface $S$ $(d-2$ dimensions). We call $Q$ magnetic (rather than electric) since $S$ is the geometric dual of $[F]$ (a discretization of $F$). The divergence of the dual of $F$ is generally called a magnetic charge.

For $d=3$ dimensions, the monopole charge $Q$ occurs as the boundary of the Dirac string $S$. For $d \geq 4$, $Q$ satisfies the conservation law

$$\partial Q = 0 . \tag{1.3}$$

An elementary charge configuration for $d=4$ is illustrated in Figure 1. In that case the vector charges (arrows) bound an elementary plaquette in the dual lattice.
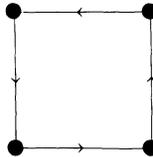


**Fig. 1.** Elementary dipoles, $d=4$

Because of the conservation law (1.3), the elementary charge configurations in $d \geq 4$ are dipolar, in fact two orthogonal dipoles for $d=4$.

With the above splitting of $F$ we obtain for the action the splitting

$$\mathscr{A} = \mathscr{A}_I + \mathscr{A}_{SW} + R . \tag{1.4}$$

Each of the three parts in (1.4) is individually gauge invariant. Here $\mathscr{A}_I$ is a Coulomb interaction for the charges $Q$, while $\mathscr{A}_{SW}$ is the action for a zero mass free field. The remainder $R$ is of order three or more in $Q$ or $F_{SW}$, and formally vanishes at zero lattice spacing. In the detailed analysis which follows, we neglect $R$ and consider the approximate action $\mathscr{A}_I$ and $\mathscr{A}_{SW}$ composed of independent parts. The free field part $\mathscr{A}_{SW}$ can be solved exactly, while the instanton part $\mathscr{A}_I$ requires a detailed analysis. $\mathscr{A}_I$ also vanishes formally at zero lattice spacing.

In this manner we arrive at the analysis of a Coulomb gas of monopoles $(d=3)$ and of dipoles $(d \geq 4)$. The remainder of the paper is devoted to the detailed study of

the statistical mechanics of a dipole gas, see also [12]. The basic difficulties arise from the long range nature of the Coulomb action. We develop an expansion method to study the free energy of the dipoles at low density, i.e. $\beta \gg 1$.

Our detailed analysis involves several steps. First we define dipole clusters $Q_L$ (i.e. average or block charges) to replace the elementary dipole excitation $Q$ of Figure 1. We choose the dipole clusters so that the long range interaction between two clusters is the same as the long range interaction between two elementary dipoles, up to a quadrapole error. Here we have the lengths scales

$$\varepsilon \ll R_L \ll r \; ; \tag{1.5}$$

$\varepsilon$ is the lattice spacing (diameter of the elementary excitation), $R_L = \varepsilon O(\beta^{1/d-2})$ is the diameter of the dipole cluster and $r$ is the inter-cluster separation. The quadrapole-quadrapole interaction is $O(r^{-(d+2)})$, which is integrable at infinity and thus can be treated by standard methods. (The possibility of charge-quadrapole interactions actually requires a second averaging and the replacement of quadrapole by octapole error terms.)

Each dipole cluster has a much reduced self action. The elementary dipoles have self action $O(\beta)$, while the dipole clusters have self action $O(\beta^{-\varepsilon})$. This requires the introduction of an activity

$$\zeta = e^{-O(\beta)} \tag{1.6}$$

which characterizes each averaged dipole cluster. The small activity results in a dilute gas of dipoles in typical configurations, i.e. in an expected density $O(\zeta) \ll 1$.

Using this typical density, we can sum the action up to a statistical length scale $R_{\mathrm{stat}} = \varepsilon \exp(\zeta^{-\delta})$. Thus

$$\varepsilon \ll R_L \ll R_{\mathrm{stat}} \; . \tag{1.7}$$

For $r > R_{\mathrm{stat}}$, we take advantage of the cancellation due to charge symmetry of the dipole gas. For every dipole of positive orientation in Figure 1, there is an elementary dipole with opposite orientation. Thus we can define the total charge $\delta$ of a large number $v$ of dipoles. Configurations which have a large dipole charge are accompanied by a small statistical weight

$$P \leq e^{-\delta^2/2v} \; . \tag{1.8}$$

Here $v = v_+ + v_-$ and $\delta = v_+ - v_-$ is the deviation from dipole neutrality. The entropy factor (1.8) is just the binomial distribution

$$P = 2^{-v} \binom{v}{v_+} \; . \tag{1.9}$$

We use this small factor to produce conditional convergence on the distance scale $r > R_{\mathrm{stat}}$.

## 2. The Definition of Instantons and Spin Waves

We consider the simple cubic lattice $Z_\varepsilon^d$ with lattice spacing $\varepsilon$. The $U(1)$ gauge field is a Lie algebra valued 1-cochain defined on lattice bonds (i.e. simplices connecting

nearest neighbor lattice sites). An element $\gamma$ of the $U(1)$ group can be written

$$\gamma = \exp(i\varepsilon g A), \qquad -\pi \leqq \varepsilon g A \leqq \pi , \tag{2.1}$$

and we define $A$ to be the gauge field. The 2-cochain defined as the coboundary of $A$ is

$$F = \partial^* A = \varepsilon^{-1} \sum_{i \in \partial p} A(i) . \tag{2.2}$$

Thus $F$ is defined on lattice squares (plaquettes $p$). Since four links bound any plaquette,

$$|\varepsilon^2 g F| \leqq 4\pi . \tag{2.3}$$

We have introduced the factor $\varepsilon^{-1}$ in the definition of $\partial^*$ so both $A$ and $F$ formally have limits as $\varepsilon \to 0$. In this limit $A$ is a 1-form, and $F = dA$ is a 2-form. The de Rham cohomology of differential forms on a compact manifold is isomorphic to the simplicial cohomology of a smooth triangulation of the manifold, see [10]. See [7] for the geometry of gauge fields.

The action $\mathscr{A}$ associated with $F$ is defined by

$$\mathscr{A} = -\varepsilon^{d-4} g^{-2} \sum_p \cos(\varepsilon^2 g F) = -\varepsilon^{d-4} g^{-2} \sum_p \cos(\varepsilon^2 g \partial^* A) , \tag{2.4}$$

where the sum ranges over plaquettes. We note that

$$\beta = \varepsilon^{d-4} g^{-2} \tag{2.5}$$

is a dimensionless coupling constant which is inversely related to the strength of the interaction. Thus

$$\mathscr{A} = -\beta \sum_p \cos(\varepsilon^2 g F) . \tag{2.6}$$

We let $[F]$ denote the discretization of $F$. By definition

$$\varepsilon^2 g [F] = 2\pi j, \quad \text{when} \quad j - \tfrac{1}{2} < (2\pi)^{-1} \varepsilon^2 g F \leqq j + \tfrac{1}{2} , \tag{2.7}$$

and where $j$ is an integer. According to (2.3), the possible values of $j$ are 0, $\pm 1$, $\pm 2$.

We note that $\partial^* F = (\partial^*)^2 A = 0$, which we can also write $\delta^* * F = 0$. Here we introduce the duality map $*$. For the purpose of geometric interpretation we let $*$ be the composition of Hodge duality with the metric. Hodge duality defines an isomorphism between $r$-cochains on $Z_\varepsilon^d$ to $(d-r)$-cochains on the dual lattice $*Z_\varepsilon^d$. The metric defines an isomorphism between $(d-r)$ cochains and $(d-r)$ chains (raising and lowering indices). Thus

$$* : r \text{ cochains} \to d - r \text{ chains}$$
$$r \text{ chains} \to d - r \text{ cochains}$$

and satisfies

$$** = (-1)^{r(d-r)} .$$

Furthermore, under the $*$ isomorphism,

$$*\partial^* * = \partial , \qquad *\delta^* * = \delta ,$$

where $\partial$ is the boundary operator, and $\delta$ is its adjoint.

In contrast to the above, $[F]$ does not have vanishing coderivative, and we define the instanton charge density $*Q$ by

$$\partial*[F]=*Q ,\tag{2.8}$$

or with $S=[*F]$, $\partial S=\partial[*F]=Q$, cf. (1, 2). Maxwell's equations have the form

$$\partial*F=0, \qquad \delta*F=-\omega_{d-1}J ,\tag{2.9}$$

which says that $J$ is the source for $F$. Similarly $Q$ is the source for $*[F]$. Of course $\partial Q=0$, or equivalently, $\partial**Q=0$, which is a conservation law for $d>3$. For $d=4$, the Dirac surface has its boundary $Q$ defined on links, i.e. $Q$ is a 1-cochain. Then $\partial Q=0$ is a Kirkoff law: $\sum_i Q(i)=0$, where the sum runs over $2d=8$ oriented links meeting at a given lattice site. Thus $Q$ is concentrated on closed $d-3$ surfaces (loops, for $d=4$) and it follows that the charge distribution consists of dipoles. For $d=4$, the corresponding quantum mechanical soliton ("boxiton") has also been studied [6]. For $d=3$, however, $Q$ is a 0-cochain, concentrated at points, which corresponds to isolated charges.

This distinction of loop charges (or closed $d-3$ surfaces, $d>3$) vs. point charges ($d=3$) is basic to the study of phase transition in $U(1)$ gauge theories, see also [9]. In particular the $d=4$ loop charge explain the expected occurrence of an ordered phase at sufficiently low temperatures $T>0$ for $d\geqq4$, and the absence of this phase for $d=3$. The mechanism, that an appropriate boundary first has dimension greater than zero, and that its entropy is thus dominated at low temperature by its energy, is the same as the mechanism governing phase transitions in the Ising model. Thus we expect that the critical dimension for phase transitions in $U(1)$ gauge theories is $d_c=3$, in contrast to the critical dimension $d_c=1$ for the Ising model. The inequality $d_c\geqq3$ was established in [3].

Now consider the transformation $A(x)\to A(x)+2\pi n_x(\varepsilon g)^{-1}$, with $n_x$ an integer. Under this change of variables,

$$F(p)=(\partial*A)(p)\to F(p)+2\pi(\varepsilon g)^{-1}(\partial*n)(p)\tag{2.10}$$

and

$$[F]\to[F]+2\pi\sum_{\partial p}n_x .$$

Since $\partial*\partial*=0$, $Q$ is unaffected, as is the action $\mathscr{A}$. It follows that $S$, regarded as a $d-2$ surface, can be deformed by such a change of variables, to an arbitrary $d-2$ surface with the same boundary $Q$. Unfortunately, there does not seem to be a natural determination of a unique $*[F]$.

The lattice Laplace operator is

$$\varDelta=-\delta*\partial*-\partial*\delta* ,\tag{2.11}$$

and its dual is

$$\varDelta=-\delta\partial-\partial\delta .$$

Because the topology of $R^d$ is trivial, we can form $\varDelta^{-1}$. The same is true for a finite lattice with Dirichlet boundary conditions or with periodic boundary conditions.

(In the latter case, the topology is trivial except in dimension one.) Consider

$$F_I = [F] + \delta^* \Delta^{-1} * Q \ . \tag{2.12}$$

Then

$$\partial^* F_I = \partial^* F + \partial^* \delta^* \Delta^{-1} * Q$$
$$= *Q + (\partial^* \delta^* + \delta^* \partial^*) \Delta^{-1} * Q = *Q - *Q = 0 \ .$$
$$\tag{2.13}$$

Since the topology is trivial and $f$ is a 2-cochain, we can define $A_I$ by the equation

$$\partial^* A_I = F_I = [F] + \delta^* \Delta^{-1} * Q \ . \tag{2.14}$$

This is the vector potential for instantons. Finally we define the spinwave potential $A_{SW} \equiv A - A_I$ and $F_{SW} = F - F_I$.

For the action we now have

$$\mathscr{A} = -\beta \sum \cos(\varepsilon^2 g \partial^* A)$$
$$= -\beta \sum \cos(\varepsilon^2 g \partial^* A_I) \cos(\varepsilon^2 g \partial^* A_{SW}) + \beta \sum \sin(\varepsilon^2 g \partial^* A_I) \sin(\varepsilon^2 g \partial^* A_{SW}).$$
$$\tag{2.15}$$

Observe that

$$\cos(\varepsilon^2 g \partial^* A_I) = \cos(\varepsilon^2 g \partial^* (-\Delta)^{-1} * Q)$$
$$\sin(\varepsilon^2 g \partial^* A_I) = \sin(\varepsilon^2 g \partial^* (-\Delta)^{-1} * Q) \ . \tag{2.16}$$

so the Dirac surface $S$ or $*[F]$ does not contribute to the action. Also note

$$\langle \delta^* (-\Delta)^{-1} * Q, \partial^* A_{SW} \rangle = \langle (-\Delta)^{-1} * Q, \partial^* \partial^* A_{SW} \rangle = 0 \ .$$

Based on the $\varphi^4$ analysis [4], we expect that the dominant effect on long range order will be given by the decoupled instanton and spin wave fields, i.e. from the terms in (2.15) which are lowest order in $\varepsilon g$ (and formally $\varepsilon, g$ independent as $\varepsilon \to 0$). Thus following Polyakov [9] we introduce the approximate action

$$\mathscr{A}_0 = \text{const} + \tfrac{1}{2} \sum \varepsilon^d (\delta(-\Delta)^{-1} Q)^2 + \tfrac{1}{2} \sum \varepsilon^d (\partial^* A_{SW})^2$$
$$= \text{const} + \mathscr{A}_I + \mathscr{A}_{SW} \ .$$
$$\tag{2.17}$$

After decoupling, we will have long range order only if it occurs in each of the fields $A_I$ and $A_{SW}$. Since the approximate action for $A_{SW}$ is Gaussian, it can be explicitly computed that long range order occurs in $A_{SW}$ for $d \geq 4$, but not for $d \leq 3$ (cf. [3]). Thus for the remainder of the paper we consider only the instanton field generated by the charges $Q$. Since

$$\langle \delta(-\Delta)^{-1} Q, \delta(-\Delta)^{-1} Q \rangle = \langle Q, (-\Delta)^{-1} \partial \delta (-\Delta)^{-1} Q \rangle$$
$$= \langle Q, (-\Delta)^{-2} (\partial \delta + \delta \partial) Q \rangle = \langle Q, (-\Delta)^{-1} Q \rangle , \tag{2.18}$$

we see that the instanton action is

$$\mathscr{A}_I = \tfrac{1}{2} \langle Q, (-\Delta)^{-1} Q \rangle = \tfrac{1}{2} \sum \varepsilon^d q_i (-\Delta)^{-1} q_j$$
$$\sim \tfrac{1}{2} \omega_{d-1} \sum_{i,j} \varepsilon^{2d} q_i r_{ij}^{-d+2} q_j \ . \tag{2.19}$$

Thus the $U(1)$ instantons interact with Coulomb forces in the approximation (2.17). Returning to the exact action (2.15), we expect that the Coulomb forces are the leading long distance parts of the instanton action.

## 3. The Landau Length Scale and the Activity

In the approximation of §2, the pure instanton action is

$$\mathscr{A}_I = \tfrac{1}{2} \sum \varepsilon^d q_i (-\varDelta)^{-1} q_j \,. \tag{3.1}$$

We recognize $q_i$ as the charge density and $\varepsilon^d q_i$ as a charge. Each $q_i$ takes values which are integer multiples of $2\pi\varepsilon^{-3}g^{-1}$, by virtue of (2.7). We let $\beta \equiv \varepsilon^{d-4}g^{-2}$ define the Landau length scale $R_L$ by the formula

$$\beta = (R_L/\varepsilon)^{d-2} \,. \tag{3.2}$$

$R_L$ characterizes the size of individual contributions to (3.1). In other words,

$$R_L = (\varepsilon^{2d-6}g^{-2})1/^{(d-2)} = \varepsilon\beta^{1/(d-2)} \,, \tag{3.3}$$

and the interactions in (3.1) on a length scale less than $R_L$ are strong, while those on a length scale greater than $R_L$ are weak. In particular, if $R_L \gg \varepsilon$, then the *self interaction* is large compared to the *interaction between distinct charges*. The self interaction of a single dipole loop is of the order

$$\varepsilon^{2d}q^2\varepsilon^{-d+2} = O(\varepsilon^{2d-6}g^{-2}\varepsilon^{-d+2}) = O(\beta) \,. \tag{3.4}$$

It is convenient to modify the definition (3.3) in order to make $R_L/\varepsilon$ an integer. We set

$$R_L = \varepsilon[\beta^{1/(d-2)}] \,, \tag{3.3'}$$

where here $[\,\cdot\,]$ denotes the integer part. With this definition, the Landau lattice $Z_{R_L}^d$ is a sublattice of $Z_\varepsilon^d$.

Given a charge configuration $Q$, we will define a decomposition into a typical part $Q_t$ and an atypical part $Q_a$,

$$Q = Q_a + Q_t \,, \tag{3.5}$$

see Chapter 4. The main property of $Q_t$ is that it is a union of connected components of $Q$, and that $Q_t$ has a low density of $d-3$ lattice elements. (Here we consider $Q$ as a $d-3$ chain, and hence as a $d-3$ surface in $R^d$.) Since $\partial$ is a local operator, we have

$$\partial Q_t = 0 \,. \tag{3.6}$$

We now show how to replace the charge distribution $Q_t$ by an averaged charge distribution $Q_{t,L}$ on the Landau lattice. Each elementary dipole of $Q_t$ in the dual $Z_\varepsilon^d$ lattice will be replaced by a large number of weaker dipoles in the dual Landau lattice. In making this replacement, we define the error term $\delta Q_t$ by the equation

$$\delta Q_t = Q_t - Q_{t,L} \,. \tag{3.7}$$

The physical interpretation of $\delta Q_t$ is a quadrapole interaction to compensate for the shorter distance (quadrapolar) forces which differ in $Q_t$ and $Q_{t,L}$. The long range

(dipolar) forces arising from $Q_t$ to $Q_{t,L}$ will agree, since we define them to have equal (average) dipole densities. We also define $Q'$ by

$$Q' = Q_a + Q_{t,L} = Q - \delta Q_t . \tag{3.8}$$

Note

$$Q = Q_a + Q_{t,L} + \delta Q_t . \tag{3.9}$$

Since $\mathscr{A}_I(Q)$ is quadratic in $Q$, for two charge configurations $Q_i, Q_j$ it is convenient to define their mutual action $\mathscr{A}_I(Q_i, Q_j)$ by

$$\mathscr{A}_I(Q_i, Q_j) = \tfrac{1}{2}(\mathscr{A}_I(Q_i + Q_j) - \mathscr{A}_I(Q_i) - \mathscr{A}_I(Q_j)) . \tag{3.10}$$

**Proposition 3.1.** *For any charge distributions* $Q$, $Q_j$,

$$\mathscr{A}_I(Q, Q) = \mathscr{A}_I(Q) , \tag{3.11}$$

$$\mathscr{A}_I\left(Q, \sum_{j=1}^n Q_j\right) = \sum_{j=1}^n \mathscr{A}_I(Q, Q_j) \tag{3.12}$$

*and*

$$\mathscr{A}_I\left(\sum_{j=1}^n Q_j\right) = \sum_{i,j=1}^n \mathscr{A}_I(Q_i, Q_j) . \tag{3.13}$$

*Proof.* These identities follow by expanding the quadratic function $\mathscr{A}_I$, for example

$$\mathscr{A}_I(Q, Q) = \tfrac{1}{2}(\mathscr{A}_I(2Q) - 2\mathscr{A}_I(Q)) = \tfrac{1}{2}(4 - 2)\mathscr{A}_I(Q) .$$

**Corollary.** *For* $Q = Q_a + Q_t = Q_a + Q_{t,L} + \delta Q_t = Q' + \delta Q_t$,

$$\mathscr{A}_I(Q) - \mathscr{A}_I(Q') = \mathscr{A}_I(Q_t) - \mathscr{A}_I(Q_{t,L}) + 2\mathscr{A}_I(Q_a, \delta Q_t) . \tag{3.14}$$

Thus the change in action due to replacing typical $\varepsilon$-dipoles by averaged $R_L$-dipoles is the sum of a quadrapole remainder term, $2\mathscr{A}_I(Q_a, \delta Q_t)$, and the change due to the typical configurations. We estimate (3.14) in Chapter 4.

We now define $Q_{t,L}$. Using (3.6), we write

$$Q_t = \partial S_t \tag{3.15}$$

for some $d - 2$ lattice surface $S_t$. (Henceforth we assume $d \geqq 4$.) This surface $S_t$ is the Dirac surface for the typical charges. We choose $S_t$ to have minimal $d - 2$ area. It follows that if $\{Q_t^{(j)}\}$ is a collection of charge configurations supported in lattice volumes of bounded diameter, then the area of $\{S_t^{(j)}\}$ is also bounded. Define

$$S_a = S - S_t \tag{3.16}$$

so

$$\partial S_t = Q_t , \qquad \partial S_a = Q_a . \tag{3.17}$$

We first assume that $S_t$ is a single $d - 2$ cell in the dual $\varepsilon$ lattice, and we define $S_{t,L}$. In the general case where $S_t$ is the sum of a bounded number of elementary $d - 2$ cells, we define $S_{t,L}$ by linearity.

Let $S_t$ be a single $d-2$ cell in $*Z_\varepsilon^d$, and consider the collection of $(R_L/\varepsilon)^{d-2}$ dipoles (i.e. $d-2$ surfaces) in $*Z_\varepsilon^d$ which both (i) lie in a single $d-2$ cell in $Z_{R_L}^d$ parallel to $S_t$ and (ii) are closest to $S_t$. We define $S_{t,L}$ to be this $d=2$ Landau lattice cell, and we assign a charge strength as follows: Since $S_t$ has strength $2\pi j\varepsilon^{d-2}g^{-1}$, we take the strength of each weak dipole ($\varepsilon$ lattice dipole making up $S_{t,L}$) to be

$$2\pi j\varepsilon^{d-2}g^{-1}(\varepsilon/R_L)^{d-2} = 2\pi^j\varepsilon^{d-2}g^{-1}\beta^{-1} \ . \tag{3.18}$$

Thus the dipole charge is scaled down by $\beta^{-1}$.

We justify this choice by calculating the long range part of the interaction between two pairs of dipole charges, oriented as in Figure 2.



**Fig. 2.** Two dipole pairs on a line, separated by distance $r$, illustrated for $d=4$

Before averaging, the mutual action, up to quadrapole corrections is

$$\omega qq'\left(\frac{d^2}{dr^2}r^{-d+2}\right)\varepsilon^2 \ ,$$

where $q, q'$ are the charges, $\omega$ is a constant and $\varepsilon$ is the charge separation within a dipole pair. After averaging, there are $(R_L/\varepsilon)^{d-3}$ charges on each face of the boundary of $S_L$, and each charge is scaled down by $\beta^{-1}$. The dipole charge separation is $R_L$ in place of $\varepsilon$. Thus for $r \gg R_L$, the mutual action is

$$\omega\beta^{-2}qq'\left(\frac{d^2}{dr^2}r^{-d+2}\right)R_L^2(R_L/\varepsilon)^{2d-6}$$

$$= \omega qq'\left(\frac{d^2}{dr^2}r^{-d+2}\right)\varepsilon^2(R_L/\varepsilon)^{2(d-2)}\beta^{-2}$$

$$= \omega qq'\left(\frac{d^2}{dr^2}r^{-d+2}\right)\varepsilon^2 \ ,$$

up to quadrapole corrections; namely the action is unchanged up to a quadrapole error.

**Proposition 3.3.** *Let $Q_t$ contain at most a bounded number $N$ of charges. Then the self action $\mathscr{A}(Q_{t,L})$ is of magnitude $O(\beta^{-1/(d-2)})$.*

*Proof.* The charge strength per elementary charge in $Q_{t,L}$ is of the order $\varepsilon^{d-3}g^{-1}\beta^{-1}$, since $Q_{t,L} = \partial S_{t,L}$, and the dipoles have strength $\varepsilon^{d-2}g^{-1}\beta^{-1}$ as explained above. Summation over relative separation of the charge pairs yields the factor $\varepsilon^{-d+2}$, while for fixed relative separation there are $(R_L/\varepsilon)^{d-3}$ charge pairs. Thus the self action has magnitude

$$(\varepsilon^{d-3}g^{-1}\beta^{-1})^2(R_L/\varepsilon)^{d-3}\varepsilon^{-d+2} = O(\varepsilon/R_L) = O(\beta^{-1/(d-2)}) \ . \tag{3.19}$$

Note the bound (3.19) uses the fact that $S_{t,L}$ contains a bounded number $N$ of dipoles, since $N^2$ terms occur as a multiplicative factor in the bound.

As a consequence of this argument, each small charge cluster $Q_t$ can be replaced by weaker averaged charge clusters $Q_{t,L}$ without changing the long range interaction. Aside from the necessity of introducing a quadrapole interaction (which we denote by $\mathcal{Q}$), it is also necessary to introduce an activity $\zeta$ to compensate for the large change in self action. For a single $\varepsilon$ lattice dipole, the self action is $c_0\beta$, cf. (3.4). Since the averaged dipole has self action $O(\beta^{-1/(d-2)})$, the activity $\zeta = e^{-O(\beta)}$ is associated with a single charge. Likewise, each small dipole cluster $S_t^{(j)}$ will have an activity $\zeta_j = e^{-O(\beta)}$ to compensate for the large change in self action. We discuss this issue in detail in Chapter 4. In terms of the above equations, the change in self action (yielding the activity) arises from the term $\mathcal{A}_I(Q_t) - \mathcal{A}_I(Q_{t'L})$ in (3.14).

We give a lower bound on $\mathcal{A}_I(Q)$ in terms of the charges in $Q$ which have density $2\pi j \varepsilon^{-3} g^{-1}$, i.e. $\varepsilon$ lattice charges. We can apply this bound to $Q'$ defined by (3.8), where $Q_a$ are the $\varepsilon$ lattice charges.

**Proposition 3.4.** *Let* $Q = Q_a + Q_t$ *where* $Q_a$ *are* $\varepsilon$ *lattice charges. There is a constant* $c_1 > 0$ *such that*

$$\mathcal{A}_I(Q) \geqq c_1 \beta \sum_{Q_a} j^2 , \tag{3.20}$$

*where* $\sum_{Q_a}$ *denotes the sum over all* $d-3$ *cells with nonvanishing charge* $Q_a$, *and where the charge density* $q_i$ *in* $Q_a$ *takes on values* $2\pi\varepsilon^{-3} g^{-1} j$, *for* $|j| \leqq 12$.

*Proof.* Let $B = (-\Delta)^{-1} Q$ and observe that

$$\mathcal{A}_I(Q) = \tfrac{1}{2}\langle \delta B, \delta B\rangle .$$

For a given charge $q$ in $Q$, let $\sum^{(q)}$ denote the sum over all $d-2$ cells of the dual lattice for which $q$ lies on the boundary. By definition

$$\delta = \varepsilon^{-1} \sum{}^{(q)} .$$

Note that

$$\delta B = \partial\delta(-\Delta)^{-1} S = S + \partial\delta\Delta^{-1} S$$

and hence

$$\partial\delta B = \partial S = Q . \tag{3.21}$$

In other words, $\delta B$ satisfies a conservation law, $\delta\delta B = 0$, and in particular for a charge $q \in Q_a$,

$$\sum{}^{(q)} \delta B = 2\pi\varepsilon^{-2} g^{-1} j . \tag{3.22}$$

However, each $d-2$ cell occurs at most $2(d-2)$ times in the iterated sum $\sum_{Q_a}\sum^{(q)}$, and so omitting terms from $Q_{t,L}$,

$$\mathcal{A}_I(Q) = \tfrac{1}{2}\langle \delta B, \delta B\rangle \geqq \varepsilon^d 2^{-2}(d-2)\sum_{Q_a}\sum{}^{(q)} (\delta B)^2 . \tag{3.23}$$

Each sum $\sum^{(q)}$ contains $6 = 2(d-(d-3))$ terms, and so by the Schwarz inequality

$$|\sum{}^{(q)}\delta B|^2 \leqq 36 \sum{}^{(q)} (\delta B)^2 .$$

Thus by (3.22)–(3.23),

$$\mathscr{A}_I Q') \geqq [\pi^2/36(d-2)]\varepsilon^{d-4}g^{-2}\sum_{Q_a}j^2 = c_1\beta\sum_{Q_a}j^2 \ ,$$

where $c_1 = \pi^2/36(d-2)$.

We end this section by remarking that a further change in the charge configuration eliminates the long distance part of the quadrapolar interaction $\mathscr{Q}$. In fact, replacing each array of $(R_L/\varepsilon)^{d-2}$ piecewise constant dipoles in $S_{t,L}$ by a linear combination of $2^d$ translates, we eliminate the quadrupole interaction up to an octapole correction $\mathscr{O}$. So that all error terms arising from linear contributions to charge minus average charge can be estimated with octapole remainders, we perform this one additional averaging in the definition of $Q_{t,L}$.

## 4. The Dilute Dipole Gas for $d \geqq 4$: Upper Bound on $Z$

We study $\mathscr{A}_I$ defined in (2.19) for a dipole (instanton) gas and establish bounds on the partition function

$$Z = \sum_Q e^{\mathscr{A}_I(Q)} \ . \tag{4.1}$$

We obtain the leading contribution to the free energy

$$f_V = V^{-1}\ln Z, \quad \beta \to \infty \ . \tag{4.2}$$

To any finite configuration of charges $Q$, we associate the action

$$\zeta_Q = e^{-\mathscr{A}_I(Q)} \tag{4.3}$$

defined by (2.19). As in (3.4), we see that

$$\mathscr{A}_I(Q) = c_Q\beta \ , \tag{4.4}$$

for a constant $c_Q$ which can be expressed explicitly in terms of $(-\varDelta)^{-1}$ via (2.19). The constant $c_Q$ is invariant under scale transformations, and thus (if we use scale transformations to identify charge configurations $Q$ on different lattice spacings) the constant $c_Q$ is independent of $\varepsilon$.

Our main assertion is that it is sufficient to examine a finite number of $Q$'s to determine $f_V$ to leading order. Let $\mathscr{Q}$ be the set of all finite charge configurations, and let

$$\mathscr{Q}_M = \{Q \in \mathscr{Q} : \text{diam } Q \leqq M\varepsilon\}$$

$$\zeta_0 = \sup_Q\{\zeta_Q : Q \in \mathscr{Q}_M\} \ . \tag{4.5}$$

Since $\zeta_Q$ is translation invariant, the sup above is taken over the finite set $\mathscr{Q}_M/Z^d$. Here $M$ is a constant to be chosen later. Let $a_0$ denote the multiplicity of the maximum, i.e. the number of $Q$'s for which $\zeta_Q = \zeta_0$. In counting $a_0$, we identify two charge configurations if they are related one to another by a lattice translation. We state our main result.

**Theorem 4.1.** *There is a $\gamma > 0$ and a choice of M in (4.5) such that*

$$f_V = a_0 \varepsilon^{-d} \zeta_0 (1 + O(\beta^{-\gamma})) , \tag{4.6}$$

*for $V > V(\beta)$ sufficiently large and for $\beta$ sufficiently large.*

*Remark.* The choice $V > V(\beta)$ eliminates surface boundary effects in leading order. Fabrizio De Angelis, de Falco and Guerra [2] have shown that $\lim_V f_V$ exists for the lattice gauge theory, and Griffiths [5] show this for $d = 3$ magnetic dipoles. For the instanton gas considered here, we believe that $\lim f_V = f$ also exists and that $f = \varepsilon^{-d} \sum \zeta_Q + O(\zeta_0^{1+\theta})$, for $\theta < 1$. In this section we prove a lower bound on $f_V$ of the form (4.6).

First let us explain the significance of Proposition 3.4. Let $c_0$ be the coefficient $c_Q$ in (4.4) for $Q$ consisting of a single $\varepsilon$ lattice charge loop (charge with density of order $\varepsilon^{-3} g^{-1}$). Let $Q = Q_a + Q_t$ be a decomposition into $\varepsilon$ lattice charges $Q_a$ and disjoint charges $Q_t$. If $Q_a$ contains $L \geq c_0/c_1$ charges, then $Q$ does not contribute to the maximum in (4.5). Thus only a bounded number of charges will contribute to the maximum in (4.5), independent of $M$.

We now specify precisely the typical configurations and the application of the construction of Section 3. Decompose the lattice into hypercubic blocks $\Delta$ with side length $2R_L$. Let $Q$ be a charge configuration, and let $N(Q, \Delta)$ be the total number of charges in $\Delta$ and in $\bar{\Delta}$, the union of $\Delta$ and the $3^d - 1$ blocks adjacent to $\Delta$. If

$$N(Q, \Delta) \leq 4 \cdot 3^d c_0/c_1 \equiv N , \tag{4.7}$$

then $Q \upharpoonright \Delta$ is called typical; otherwise $Q \upharpoonright \Delta$ is called atypical. The typical charge configurations are more complicated than the bounded charge configurations (4.5). In fact a typical configuration may contain several widely separated clusters.

**Proposition 4.2.** *Let $\delta > 0$ be given and let $Q = \sum Q_i$ be a typical configuration, composed of clusters $Q_i$ having minimal separation $D\varepsilon$. Then for $D \geq D(N, \delta)$ sufficiently large*

$$|\mathscr{A}(Q_i, Q_j)| \leq \delta \beta N^{-2} , \tag{4.8}$$

*for $i \neq j$.*

*Proof.* From the definition (3.10), we have

$$\mathscr{A}_I(Q_i, Q_j) \sim \tfrac{1}{2} \omega_{d-1} \sum_{\substack{q_i \in Q_i \\ q_j \in Q_j}} \varepsilon^{2d} q_i r_{ij}^{-d+2} q_j .$$

Summing over charges $2\pi j \varepsilon^{-3} g^{-1}$ and $r_{ij} \geq D\varepsilon$ gives

$$|\mathscr{A}_I(Q_i, Q_j)| \leq c_{Q_i, Q_j} \beta D^{-d+2} , \tag{4.9}$$

where $c_{Q_i, Q_j}$ is a constant. In fact $c_{Q_i, Q_j} \leq O(N^2)$. Thus (4.8) holds for $D(N, \delta)$ sufficiently large.

With the choice of $D = D(N, \delta)$ as required by Proposition 4.2, we divide $Q \upharpoonright \bar{\Delta}$ into clusters separated by at least $D\varepsilon$. Each cluster $Q_j$ (as $Q \upharpoonright \bar{\Delta}$ itself) has at most $N$ charges, so each cluster has diameter at most $ND\varepsilon$. Choose $M = ND$ in (4.5). Thus if $Q$ is typical in $\Delta$,

$$Q \upharpoonright \bar{\Delta} = \sum Q_j , \qquad Q_j \in \mathscr{Q}_M , \tag{4.10}$$

and (4.8) holds. In case the cluster $Q_j$ lies in two or more of the $3^d$ blocks of $\bar{\varDelta}$, we assign $Q_j$ to the block containing its "lower corner" in some lexicographic order of lattice sites in $Z^d$. For $\beta$ sufficiently large $(R_L \gg \varepsilon)$ each $Q_j$ associated with $\varDelta$ satisfies

$$\partial Q_j = 0 \ . \tag{4.11}$$

Let $J(\varDelta)$ index clusters associated with $\varDelta$. For typical blocks $\varDelta$ let

$$Q_\varDelta = \sum_{j \in J(\varDelta)} Q_j \ . \tag{4.12}$$

Then $Q_\varDelta$ is the configuration to which we apply the construction of Chapter 3, giving $Q_{\varDelta,L}$, which agrees with $Q_\varDelta$ up to the octapole remainder. Let

$$Q_t = \sum_\varDelta Q_\varDelta \ , \qquad Q_{t,L} = \sum_\varDelta Q_{\varDelta,L} \ ,$$
$$\partial Q_\varDelta = Q_\varDelta - Q_{\varDelta,L} \ . \tag{4.13}$$

**Lemma 4.3.** *For $\beta$ sufficiently large,*

$$\mathscr{A}_I(Q_\varDelta) - \mathscr{A}_I(Q_{\varDelta,L}) \geq \sum_{j \in J_\varDelta} \mathscr{A}_I(Q_j)(1 - \delta - \beta^{-1}) \ . \tag{4.14}$$

*Moreover we may take $\delta = 0$ if $J_\varDelta$ contains a single element.*

*Proof.* Using Proposition 3.1,

$$\mathscr{A}_I(Q_\varDelta) = \sum_{i,j \in J_\varDelta} \mathscr{A}_I(Q_i, Q_j) = \sum_{i \in J_\varDelta} \mathscr{A}_I(Q_i) + \sum_{\substack{i \neq j \\ i,j \in J_\varDelta}} \mathscr{A}_I(Q_i, Q_j) \ .$$

By Proposition 4.2, we bound the intercluster action by

$$\sum_{i \neq j} \delta \beta N^{-2} \leqq \delta \beta \ .$$

Since $Q_j$ belongs to the finite set up to translations, of typical charge clusters,

$$\beta \leqq \text{const} \sum_{j \in J_\varDelta} \mathscr{A}_I(Q_j) \ , \tag{4.15}$$

using the constants of (4.4). Thus with a new $\delta$,

$$\mathscr{A}_I(Q_\varDelta) \geqq (1 - \delta) \sum_{j \in J_\varDelta} \mathscr{A}_I(Q_j) \ . \tag{4.16}$$

Clearly we may take $\delta = 0$ if $J_\varDelta$ contains a single element.

We bound $\mathscr{A}_I(Q_{\varDelta,L})$ using Proposition 3.3 Thus

$$|\mathscr{A}_I(Q_{\varDelta,L})| \leqq O(1)\beta^{-1/(d-2)} \ ,$$

and by (4.15)

$$|\mathscr{A}_I(Q_{\varDelta,L})| \leqq O(1)\beta^{-1/(d-2)}\beta^{-1} \sum_{j \in J_\varDelta} \mathscr{A}_I(Q_j)$$

$$\leqq \beta^{-1} \sum_{j \in J_\varDelta} \mathscr{A}_I(Q_j) \tag{4.17}$$

for $\beta$ large. The desired bound follows from (4.16–4.17).

**Lemma 4.4.** *There exists $\gamma > 0$ such that for $\beta$ sufficiently large,*

$$\mathscr{A}_I(Q_t) - \mathscr{A}_I(Q_{t,L}) \geq \sum_\Delta \sum_{j \in J_\Delta} \mathscr{A}_I(Q_j)(1 - \delta_\Delta - \beta^{-\gamma}) \tag{4.18}$$

*where $\delta_\Delta = 0$ if only one cluster is associated with $\Delta$.*

*Proof.* We expand using Proposition 3.1,

$$\begin{aligned}
\mathscr{A}_I(Q_t) &- \mathscr{A}_I(Q_{t,L}) \\
&= \sum_\Delta (\mathscr{A}_I(Q_\Delta) - \mathscr{A}_I(Q_{\Delta,L})) + \sum_{\Delta \neq \Delta'} (\mathscr{A}_I(Q_\Delta, Q_{\Delta'}) - \mathscr{A}_I(Q_{\Delta,L}, Q_{\Delta',L})) .
\end{aligned} \tag{4.19}$$

The first terms in (4.19) are bounded below by Lemma 4.3. The sum over $\Delta \neq \Delta'$ is split into two parts: the sum over adjacent pairs and the sum over nonadjacent pairs. The sum over adjacent pairs is bounded using Proposition 4.2 and the argument in the proof of Lemma 4.3. For nonadjacent pairs we write the terms as

$$\begin{aligned}
\sum_{\Delta, \Delta'} &(\mathscr{A}_I(Q_\Delta - Q_{\Delta,L}, Q_{\Delta'}) + \mathscr{A}_I(Q_{\Delta,L}, Q_{\Delta'} - Q_{\Delta',L})) \\
&= \sum_{\Delta, \Delta'} \mathscr{A}_I(\delta Q_\Delta, Q_{\Delta'} + Q_{\Delta',L}) .
\end{aligned} \tag{4.20}$$

The sum (4.20) contributes to the octapole remainder. Due to the subtractions in the definition of $\delta Q_\Delta$,

$$|\mathscr{A}_I(\delta Q_\Delta, Q_{\Delta'} + Q_{\Delta',L})| \leq O(1)\beta D^{-d-1} ,$$

where

$$D = \varepsilon^{-1} \operatorname{dist}(\Delta, \Delta') \geq (R_L/\varepsilon) = \beta^{1/(d-2)} .$$

This bound follows as in the proof of Proposition 4.2, cf. (4.9). The additional powers of $D$ in the denominator arise from the multiple differences which occur in the definition of $\delta Q_\Delta$.

The summation over $\Delta$ and $\Delta'$ is achieved by taking $\operatorname{dist}(\Delta, \Delta') = 2nR_L$ where $\boldsymbol{n}$ is a vector with integer components, and summing $\boldsymbol{n}$ over $Z^d$. Thus, using (4.15) and $\sum D^{-d-1} \leq O(R_L/\varepsilon)^{-d-1} \leq O(\beta^{-1-\gamma})$,

$$\sum_{\Delta, \Delta'} |\mathscr{A}_I(\delta Q_\Delta, Q_{\Delta'} + Q_{\Delta',L})| \leq O(\beta^{-\gamma}) \sum_\Delta \sum_{j \in J_\Delta} \mathscr{A}_I(Q_j) .$$

**Lemma 4.5.** *There exists $\gamma > 0$ such that for $\delta > 0$ and $\beta$ sufficiently large*

$$\mathscr{A}_I(Q) \geq (c_1 - \delta)\beta \sum_{Q_a} j^2 + \sum_\Delta \sum_{j \in J_\Delta} \mathscr{A}_I(Q_j)(1 - \delta_\Delta - \beta^{-\gamma}) . \tag{4.21}$$

*Proof.* We write

$$\mathscr{A}_I(Q) = \mathscr{A}_I(Q') + \mathscr{A}_I(Q_t) - \mathscr{A}_I(Q_{t,L}) + 2\mathscr{A}_I(Q_a, \delta Q_t) . \tag{4.22}$$

The action $\mathscr{A}_I(Q')$ is bounded below using Proposition 3.4, with $Q' = Q_a + Q_{t,L}$. Thus

$$\mathscr{A}_I(Q') \geq c_1 \beta \sum_{Q_a} j^2 .$$

The difference $\mathscr{A}_I(Q_t) - \mathscr{A}_I(Q_{t,L})$ is bounded below by Lemma 4.4. The remaining term $\mathscr{A}_I(Q_a, \delta Q_t)$ contributes to the octapole remainder, and it is bounded by

$$|\mathscr{A}_I(Q_a, \delta Q_t)| \leqq O(\delta\beta) \sum_{Q_a} j^2 \ .$$

*Proof of Theorem 4.1* (upper bound on $Z$). By Lemma 4.5,

$$e^{-\mathscr{A}_I(Q)} \leqq e^{-(c_1-\delta)\beta \sum_{Q_a} j^2} \prod_{\Delta} \prod_{J_\Delta} e^{-\mathscr{A}_I(Q_j)(1-\delta_\Delta-\beta-\gamma)} \ .$$

We substitute $\zeta_j = e^{-\mathscr{A}_I(Q_j)}$. In order to simplify the sum over configurations, and to make the $Q_a, Q_t$ sums factorize, we introduce some spurious terms. We allow overlapping $Q_a, Q_t$ configurations, and we regard two or more elementary clusters in the same or adjacent blocks as separate events with activity

$$(\zeta_{j_1} \dots \zeta_{j_l})^{1-\delta_\Delta} = \zeta_{j_1 \dots j_l} \ . \tag{4.23}$$

We then allow arbitrary multiplicities of these $\zeta_{j_1}, \zeta_{j_1 j_2} \dots$ events in any block. This allows us to bound the summation over charge configurations, $\sum_{\{Q\}}$, by independent sums $\sum_{\{Q_a\}} \sum_{\{Q_t\}}$ over atypical and typical charge configurations. Thus

$$Z = \sum_{\{Q\}} e^{-\mathscr{A}_I(Q)} \leqq \left( \sum_{\{Q_a\}} e^{-\frac{1}{2}c_1\beta\Sigma Q_a j^2} \right) \left( \sum_{\{Q_t\}} \prod_{\Delta} \prod_{J_\Delta} \zeta_{j_1} \dots \zeta_{j_1 \dots j_2} \right)$$

$$\leqq \left( \sum_{\{Q_a\}} e^{-\frac{1}{2}c_1\beta\Sigma Q_a j^2} \right) \left( 1 + \sum_j \left( \frac{R_L}{\varepsilon} \right)^d \zeta_j + \sum_{j_1 j_2} \left( \frac{R_L}{\varepsilon} \right)^{2d} \zeta_{j_1 j_2} + \dots \right)^{V R_L^{-d}} \ .$$

Here we remark there are a finite number of terms for the typical summation in a given block $\Delta$. The factors $(R_L/\varepsilon)^d$, $(R_L/\varepsilon)^{2d}$, etc. arise from possible locations of the clusters in $\Delta$. Since $(R_L/\varepsilon)^d = \beta^{d/d-2}$, and $\zeta_j \leqq \zeta_0$, $\zeta_{j_1 j_2} \leqq \zeta_0^{2(1-\delta_\Delta)}, \dots$,

$$Z \leqq \left( \sum_{\{Q_a\}} e^{-\frac{1}{2}c_1\beta\Sigma Q_a j^2} \right) \left( 1 + \left( \frac{R_L}{\varepsilon} \right)^d a_0\zeta_0 (1 + O(\beta^{-\gamma})) \right)^{V R_L^{-d}}$$

$$\leqq \left( \sum_{\{Q_a\}} e^{-\frac{1}{2}c_1\beta\Sigma Q_a j^2} \right) e^{a_0\zeta_0(1+O(\beta^{-\gamma}))V\varepsilon^{-d}} \ .$$

Here we have used the fact that $\delta_\Delta$ does not appear for a single cluster, to identify the leading term in the exponent as $a_0\zeta_0$.

The sum over atypical configurations can be carried out explicitly. Since $Q_a$ has at least $4 \cdot 3^d c_0/c_1$ charges on the $\varepsilon$ lattice, the activity of each atypical configuration in $\Delta$ is at least $\zeta_0^2$. Thus the atypical terms can be included in the error $O(\beta^{-\delta})$.

## 5. The Lower Bound on $Z$

In this section we prove the lower bound on $Z = \sum_{\{Q\}} e^{-\mathscr{A}_I}$, which completes the proof of Theorem 4.1. Since each term in the sum over configurations is nonnegative, we achieve a lower bound by omitting certain configurations $Q$. The remaining

configurations, called typical, differ from the typical configurations of Section 4 (and in any case are a set of measure zero in the infinite volume measure).

We introduce an infinite sequence $R_1 = \varepsilon$, $R_2 = R_L$,

$$R_3 = \zeta_0^{-2\delta} R_L$$
$$R_4 = \exp(\zeta_0^{-\delta}) R_L \tag{5.1}$$
$$R_n = 2^n R_4, \quad n \geqq 4$$

of length scales. The first restriction defining a typical configuration $Q$ is that $Q$ must be a sum of translates of the $a_0$ distinct $Q_j \in \mathcal{Q}_M$ with activity $\zeta_0 = \zeta_{Q_j}$, as in (4.5). Furthermore we allow only one such elementary cluster in each block of length $R_3$; within this block, the cluster must have a distance at least $\zeta_0^{-\delta} R_L = \zeta_0^{\delta} R_3$ from the boundary of the block.

Let

$$\mathscr{A}_I(Q_i, R, R') = \sum \mathscr{A}_I(Q_i, Q_j) \tag{5.2}$$

where the sum runs over all elementary clusters $Q_j \neq Q_i$ in $Q$ at a distance between $R$ and $R'$ from the cluster $Q_i$. For typical configurations,

$$\mathscr{A}_I(Q_i, 0, \zeta_0^{\delta} R_3) = 0 \tag{5.3}$$

because the minumum intercluster distance is $2\zeta_0^{\delta} R_3$.

**Lemma 5.1.** *For $\beta$ sufficiently large and for a typical configuration $Q$,*

$$|\mathscr{A}_I(Q_i, 0, R_4)| \leqq \tfrac{1}{2}\zeta_0^{\delta} . \tag{5.4}$$

*Proof.* At separation $r$,

$$|\mathscr{A}_I(Q_i, Q_j)| \leqq O(r/R_L)^{-d} \beta^{-1/(d-2)} .$$

At minimum separation $r = 2\zeta_0^{\delta} R_3 = 2\zeta_0^{-\delta} R_L$

$$|\mathscr{A}_I(Q_i, Q_j)| \leqq 2^{-d} \zeta_0^{\delta d} \beta^{-1/(d-2)} .$$

For $Q_j$'s separated by one or more intermediate block of size $R_3$, the discrete sum can be estimated by an integral, giving

$$|\mathscr{A}_I(Q_i, 0, R_4)| \leqq O(\zeta_0^{\delta d}) \beta^{-1/(d-2)} + O(R_3/R_L)^{-d} \beta^{-1/(d-2)} \ln(R_4/R_L)$$
$$\leqq \tfrac{1}{2}\zeta_0^{\delta} .$$

Here we have substituted $R_3^{-d}$ as an upper bound on the density of charge clusters.

For distances $R > R_4$, the bound on $\mathscr{A}_I(Q_i, 0, R)$ depends on cancellations between oppositely oriented dipoles. To achieve this cancellation, we impose further restrictions on the typical configurations. The map $q \to -q$ is a symmetry of $\mathscr{A}_I$ which reverses the orientation of the elementary dipole clusters. Thus the number $a_0$ of these clusters is even, and we can label them as $Q_{\pm j}, j = 1, \ldots, a_0/2$, with $Q_{-j} = -Q_j$ the orientation reversed image of $Q_j$. We divide $R^d$ into $d$-cells (block spins) of length $R_n$, for each length scale $R_n$, $n \geqq 5$. In any block spin of length $R_n$, let $v_j$ be the number of times the elementary dipole cluster $Q_j$ occurs. Our final

restriction defining a typical configuration is that

$$|v_j - v_{-j}| \leqq (R_n/R_3)^{3d/4} \tag{5.5}$$

for each $j = 1, \ldots, a_0/2$, each $R_n$-block and each $n = 5, 6, \ldots$.

**Lemma 5.2.** *For $\beta$ sufficiently large and for typical configurations,*

$$|\mathscr{A}_I(Q_i, 0, \infty)| \leqq \zeta_0^\delta . \tag{5.6}$$

*Proof.* Write

$$\mathscr{A}_I(Q_i, 0, \infty) = \mathscr{A}_I(Q_i, 0, R_4) + \sum_{n=4}^{\infty} \mathscr{A}_I(Q_i, R_n, R_{n+1}) .$$

We divide the spherical shell contributing to $\mathscr{A}_I(Q_i, R_n, R_{n+1})$ into $O(R_n/R_{4[n/5]})^d$ blocks of length $R_{4[n/5]}$. In each block

$$|v_j - v_{-j}| \leqq (R_{4[n/5]}/R_3)^{3d/4}$$

is the deviation from neutrality. The dipole interaction strength for a dipole with separation $R$ is bounded by $O(\beta^{-1/(d-2)})(R/R_L)^{-d}$. Thus

$$\begin{aligned}
|\mathscr{A}_I(Q_i, R_n, R_{n+1})| &\geqq \sum_j O(\beta^{-1/(d-2)})(R_n/R_L)^{-d}|v_j - v_{-j}|(R_n/R_{4[n/5]})^d \\
&\leqq O(\beta^{-1/(d-2)})R_L^d R_3^{-3d/4} R_{4[n/5]}^{-d/4} \\
&\leqq O(\beta^{-1/(d-2)})2^{-[n/5]d} \zeta_0^{3d/2} e^{-\zeta_0^\delta/4} .
\end{aligned}$$

Summing over $n$, and choosing $\beta$ large,

$$\sum_{n=4}^{\infty} |\mathscr{A}_I(Q_I, R_n, R_{n+1})| \leqq \tfrac{1}{2} \zeta_0^d$$

so that Lemma 5.1 completes the proof.

*Proof of Theorem 4.1* (completed). We now let $\sum'$ denote the sum over typical configurations. In a bounded region $\Lambda$ of volume $V$, a typical configuration $Q$ is specified by the number $N = N(Q)$ of dipole clusters, the choice of $\pm$ orientation for each cluster, the $(a_0/2)^N$ choices of $|j|$ and the

$$\binom{V/R_3^d}{N}(R_3/\varepsilon)^{dN}(1 - O(\zeta_0^\delta))^N \tag{5.7}$$

possible positions of the clusters. Here the factor $(1 - O(\zeta_0^\delta))^N$ results from the excluded boundary region for the location of each cluster in its block of length $R_3$. Thus

$$Z = \sum_Q e^{-\mathscr{A}_I} \geqq \sum_Q{}' e^{-\mathscr{A}_I} \leqq \sum_N \sum_{\{Q:N(Q)=N\}} \sum_{\pm} e^{-\zeta_0^\delta N}(a_0/2)^N \zeta_0^N . \tag{5.8}$$

Substitution of (5.7) and $\sum_{\pm} = 2^N$ in (5.8) would give

$$\begin{aligned}
Z &\geqq (1 + a_0(R_3/\varepsilon)^d e^{-\zeta_0} \zeta_0(1 - O(\zeta_0^\delta)))^{V/R_3^d} \\
&\geqq [1 + a_0 \varepsilon^{-d} \zeta_0(1 - O(\beta^{-\delta}))]^V \tag{5.9}
\end{aligned}$$

and complete the proof. Thus it remains to analyze the restriction placed on $\sum'$ by the restriction to typical configurations.

For a particular $R_n$-length block, the number of choices of $\pm$ which violate (5.5) is bounded by

$$e^{-O(|v_j - v_{-j}|^2/|v_j + v_{-j}|)} 2^{|v_j + v_{-j}|} \leq e^{-O(R_n/R_3)^{6d/4}/(R_n/R_3)^d} 2^{|v_j + v_{-j}|}$$

$$\leq e^{-O(2^{n/2} \zeta_0^\delta e^{\frac{1}{4}\zeta\bar\sigma^{-\delta}})} 2^{|v_j + v_{-j}|} .$$

Thus the complementary set occurs for at least

$$[1 - e^{-O(n/2 \zeta_0^\delta e^{\frac{1}{4}\zeta\bar\sigma^{-\delta}})}] 2^{|v_j + v_{-j}|}$$

$$\cong \left[1 - \left(\frac{R_3}{R_n}\right)^d e^{-O(2^{n/2} \zeta_0^\delta e^{\frac{1}{4}\zeta\bar\sigma^{-\delta}})}\right]^{(R_n/R_3)^d} 2^{|v_j + v_{-j}|}$$

choices. Multiplying these values, we see that (5.5) is satisfied on all $R_n$-length blocks at once for

$$\left[1 - \left(\frac{R_3}{R_n}\right)^d e^{-O(2^{n/2} \zeta_0^\delta e^{\frac{1}{4}\zeta\bar\sigma^{-\delta}})}\right]^{V/R_3^d} 2^N$$

choices of $\pm$. Multiplying over $n$ shows that

$$\sum_{\pm} \geq \left[1 - O(1) \sum_n \left(\frac{R_3}{R_n}\right)^d e^{-O(2^{n/2} \zeta_0^\delta e^{\frac{1}{4}\zeta\bar\sigma^{-\delta}})}\right]^{V/R_3^d} 2^N .$$

Substitution in (5.8) now yields (5.9) and completes the proof.

## References

1. Brydges, D.: A rigorous approach to Debye screening in dilute classical Coulomb systems. Commun. math. Phys., to appear
2. Fabrizio De Angelis, G., de Falco, D., Guerra, F.: Lattice gauge models in the strong coupling regime. Preprint
3. Glimm, J., Jaffe, A.: Quark trapping for $U(1)$ lattice gauge fields. Phys. Letters **66**B, 67—69 (1977)
4. Glimm, J., Jaffe, A. Spencer, T.: Convergent expansion about mean field theory, Parts I and II. Ann. Phys. **101**, 610—630; **101**, 631—669 (1976)
5. Griffiths, R. B.: Free energy of interacting magnetic dipoles. Phys. Rev. **176**, 655—659 (1968)
6. Kogut, J., Susskind, L.: Hamiltonian formulation of Wilson's lattice gauge theories. Phys. Rev. D**11**, 395—408 (1975)
7. Mayer, M.: Gauge fields as quantized connection forms. In: Conference on differential geometric methods in mathematical physics (eds. K. Bleuler, A. Reetz). Lecture notes in mathematics, Vol. 570. Berlin-Heidelberg-New York: Springer 1977
8. Osterwalder, K., Seiler, E.: Gauge field theories on the lattice. Ann. Phys. to appear
9. Polyakov, A. M.: Quark confinement and topology of gauge theories. Nucl. Phys. B **120**, 429—458 (1977)
10. Singer, I., Thorpe, J. A.: Lecture notes on elementary topology and geometry. Berlin-Heidelberg-New York: Springer 1976
11. Wilson, K.: Confinement of quarks. Phys. Rev. D**10**, 2245—2259 (1974)
12. Høye, J. S., Stell, G.: Ionic solution theory for non-ideal solvents—potential of mean force between ions. Preprint