STRONG COUPLING AND MEAN FIELD METHODS IN LATTICE GAUGE THEORIES

Jean-Michel DROUFFE

CERN, Geneva, Switzerland
and
SPhT, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

and

Jean-Bernard ZUBER

SPhT, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

NORTH-HOLLAND PHYSICS PUBLISHING–AMSTERDAM

PHYSICS REPORTS (Review Section of Physics Letters) 102, Nos. 1 & 2 (1983) 1–119. North-Holland, Amsterdam

STRONG COUPLING AND MEAN FIELD METHODS IN LATTICE GAUGE THEORIES

Jean-Michel DROUFFE*

CERN, Geneva, Switzerland

and

Jean-Bernard ZUBER

SPhT, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

Received 8 July 1983

Contents:

* On leave of absence from SPhT, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France.

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 102, Nos. 1 & 2 (1983) 1–119.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 66.00, postage included.

0 370-1573/83/$35.70

1. Generalities

1.1. Introduction

Non-Abelian gauge theories play nowadays a dominant role in particle physics. The mechanism of spontaneous symmetry breaking has enabled Salam and Weinberg to construct a unified theory of electroweak interactions, and grand unified theories are actively studied. On the other hand, asymptotic freedom is the crucial property of the current theory of strong interactions, quantum chromodynamics (QCD). This remarkable property, that only non-Abelian gauge theories enjoy, guarantees that fundamental constituents of hadrons – quarks and gluons – have vanishingly small interactions at short distances. This allows a perturbative treatment in this regime. On the contrary, the long distance behaviour of that theory remains more elusive. Such problems, as the assumed permanent confinement of quarks or the computation of hadron spectrum, are typically strong coupling problems and cannot resort to standard field theoretical methods.

A major breakthrough was accomplished in 1973 when Wilson [1] (and Polyakov [2]) proposed to consider lattice gauge theories. The discretization of space, carefully designed to preserve gauge invariance, offers an ultraviolet regularization and is not expected to affect the long distance behaviour, since the effective coupling at small distance is very weak. A lattice gauge theory may be considered as a model of statistical mechanics and is thus amenable to all techniques used in such models [3, 4] – series expansions, mean field approximations, Monte-Carlo simulations, ... –. This has been done with an Euclidean space-time lattice [1, 5–7] after a Wick rotation, or with a spatial lattice, keeping time continuous, in the Hamiltonian formulation [8].

A model of statistical mechanics approximates a physical system and is used to study its properties as a function of temperature, fields, ... At a non-critical temperature, these properties depend strongly on the choice of the effective Hamiltonian. In contrast with this situation, in lattice gauge theories, one is mainly interested in the approach to the critical point. There, the correlation length becomes much larger than the lattice spacing; at that scale, details of the discretization, i.e. of the lattice action or Hamiltonian, become irrelevant, and the continuous theory is recovered.

Monte-Carlo methods have been thoroughly reviewed by Creutz, Jacobs and Rebbi [9]. The present work reviews two analytical approaches which have been particularly developed over the last ten

years: strong coupling expansions and mean field approximation. This leaves aside many interesting or promising other methods, as variational techniques, Migdal–Kadanoff approximate renormalization group [10], ... By lack of time, of space and of competence, we have also restricted this review to pure gauge and to Higgs-gauge systems. Theories with fermions or finite temperature effects would deserve a separate treatment. For the same reason, we discuss mainly the Euclidean approach, but compare some of the results with those obtained in the Hamiltonian formulation.

Section 2 reviews some particular rigorous or well-established results, to be used in the following. Section 3 deals with strong coupling methods; subsection 3.1 explains the general features of the expansions and the methods to derive them. The series currently known are displayed in the tables of subsection 3.2, and some partial resummation techniques are presented in subsection 3.3. Subsection 3.4 contains a detailed analysis of the various series and their comparison with numerical Monte-Carlo data. Section 4 is devoted to mean field; subsection 4.1 discusses the general framework, and subsection 4.2 applies it to pure lattice gauge theories. Finally, section 5 deals with the case of scalar matter fields coupled to gauge fields. Appendix A gathers some formulae of group theory which may be useful, appendix B elaborates a technical point on the large N limit.

We are quite aware that such a review is highly indigestible. We have tried to make the different parts as independent as possible, to allow a selective reading. The reader who has skipped a part will find cross-references if he or she needs some material presented earlier.

We end this introduction with the usual apologies to those who will find their work improperly quoted or reviewed. We also refer to some general reviews on lattice gauge theories [11–14].

1.2. General conventions and notations

Almost all the way through this paper, we use the Euclidean formulation of lattice gauge theory. A Wick rotation $t \to it$ has been performed, and all calculations are done in the framework of the Euclidean metric. Only at the end of the procedure are the results reinterpreted in the usual physical space. Hopefully, no information is lost in this process. In particular, the physical energy spectrum is unmodified; instead of observing a periodical time evolution, the Wick rotation replaces it by an exponential decrease of field correlations.

A hypercubical regular lattice is commonly used. Relevant notations are

d	dimension of the space.
μ, ν	Greek letters denoting a particular dimension (1 to d).
a	lattice spacing.
$\hat{\mu}, \hat{\nu}$	vectors of length a along the μth direction.
i, j	Latin letters denoting a particular site.
L	lattice size. Depending on the context, free or periodic boundary conditions are used. In some cases, the size L_μ depends on the chosen direction μ.
$\mathcal{N}$	total number of sites ($=L^d$).
$l, ij, i\hat{\mu}$	different notations for a lattice bond joining two neighbouring sites i and $j = i + \hat{\mu}$.
$p, (ijkl), i\hat{\mu}\hat{\nu}$	different notations for the "plaquette", i.e. the elementary square of the lattice, involving the four sites i, $j = i + \hat{\mu}$, $k = i + \hat{\mu} + \hat{\nu}$ and $l = i + \hat{\nu}$; this plaquette is bounded by four lattice bonds.

On the lattice, different kinds of fields are introduced:

ϕ	generic name for all fields. This notation is used in sections describing general

techniques in order to save a lot of writing, especially for the indices. This field ϕ_x^α depends on its location x (which may be a link for gauge fields, or a site for matter fields) and on other internal indices α (spinorial, vectorial, ...).

$D\phi$ corresponding measure, factorized on the field locations as $\Pi_x \, d\mu(\phi_x)$.

$S(\{\phi\})$ action governing the dynamics.

β lattice coupling constant. Also called inverse temperature, or inverse squared (Yang–Mills) coupling constant.

$Z(J)$ generating functional for the Green functions, $Z(J) = \int \exp\{\beta S + J \cdot \phi\} D\phi$.

Z partition function ($= Z(J = 0)$).

$W(J)$ generating functional for connected Green functions ($= \ln Z(J)$).

F free energy per site ($= \ln Z/\mathcal{N}$). (Strictly speaking, this name is incorrect due to the missing $-1/\beta$ factor.).

$\langle\ \rangle$ physical mean values computed with the Boltzmann weight $\exp(\beta S)$.

In the particular case of gauge fields, we write

U_l standard gauge field name, belonging to the fundamental representation of the gauge group G.

U_p shorthand notation for the product of fields U_l along the boundary links of a plaquette p ($U_{(ijkl)} = U_{ij}U_{jk}U_{kl}U_{li}$). This notation is ambiguous (depending on the orientation and on the starting point on the plaquette boundary), but we will use it if this ambiguity is irrelevant (e.g. in $\mathrm{Re\,Tr}\ U_p$).

U_C as preceding, but along a closed curve C.

Some group theory notations are also needed.

G gauge group.

N dimension of the fundamental representation of G. In most applications, G is SU(N).

r, s Latin indices for the representations.

f $r = \mathrm{f}$ is the fundamental representation. This index is often omitted whenever no possible confusion exists.

0 $r = 0$ is the trivial representation.

d_r dimension of the representation r ($d_\mathrm{f} = N$, $d_0 = 1$).

χ_r character (trace of the representation r). $\chi_\mathrm{f}(U) = \mathrm{Tr}\ U$, $\chi_r(\mathbb{1}) = d_r$.

DU invariant Haar measure on G, normalized to 1.

1.3. Gauge fields

Gauge theories are based on the concept of local gauge transformations, i.e. of transformations $g(x) \in \mathrm{G}$ depending on the space point x. Local matter fields $\phi(x)$ transform under the transformation $g(x)$ at the same point. In continuum systems, the requirement of gauge invariance for derivative terms requires the introduction of a local gauge field $A_\mu(x)$ in the Lie algebra which uses both $g(x)$ and its first derivative $\nabla g(x)$ in its transformations.

The direct discretization of this gauge field on the lattice may seem natural, but is rather uninteresting. Indeed, the transcription of the derivative $\nabla g(x)$ in the lattice transformation law leads to the following alternative. Either ∇ is approximated by a nearest neighbour difference; this approximation breaks gauge invariance and restoring this invariance in the continuum limit is problematical; or ∇ is exactly reproduced through a non-local combination (g_i on all sites are used in the transformation), and non-locality may spoil the resulting theory.

In any case, more than one g_i must be used in the transformation of the gauge field. The minimal choice is to use two local gauge transformations on neighbouring sites [1]. Gauge fields U_{ij} take their values in the gauge group G (and not in its Lie algebra) and transform according to

$$U_{ij} \to g_i U_{ij} g_j^{-1} . \tag{1.1}$$

Moreover, the constraint

$$U_{ji} \equiv U_{ij}^{-1} \tag{1.2}$$

is required. On a hypercubical lattice, there are $\mathcal{N}d$ bonds ij on the lattice, and this choice has the same number of degrees of freedom as a direct discretization of the continuum gauge field A_μ. This lattice gauge field which lives on links may be seen as the analogue of the non-local path-ordered phase factor of the continuous theory

$$U_{ij} = P \exp\left\{ \mathrm{i}g \oint_{ij} A_\mu \, \mathrm{d}x^\mu \right\} . \tag{1.3}$$

Before describing the possible dynamics for these gauge fields, we must say a word about the choice of the physical degrees of freedom, i.e. the possible choices of gauge. Given any tree of links on the lattice, it is possible to fix the gauge fields on these links to some assigned value in G. This is done step by step, starting from a given point, by using the transformation law (1.1) to determine recursively the g_j's. Among the possible choices, fixing to $\mathbb{1}$ all the gauge fields located on links parallel to a given direction is known as the axial gauge, whereas fixing them on the comb-like set of links (fig. 1)

$$\begin{aligned} &U_{xd} = \mathbb{1} \qquad \forall x \\ &U_{x,\widehat{d-1}} = \mathbb{1} \qquad \text{for } x_d = 0 \\ &U_{x,\widehat{d-2}} = \mathbb{1} \qquad \text{for } x_d = x_{d-1} = 0 \\ &\qquad \ldots . \end{aligned} \tag{1.4}$$

is sometimes called the maximal gauge.

For Abelian gauge groups, the set $\{U_p\}$ of all plaquette values is sufficient to determine the gauge fields up to a gauge transformation, provided the constraints

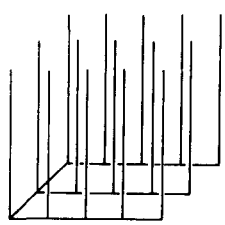


Fig. 1. Comb-like set of links set to 1 in the maximal gauge.

$$\prod_{p\in\partial c} U_p = 1 \tag{1.5}$$

are fulfilled by the product of the six plaquettes bounding every three-dimensional cube c. The U_p's can be taken as new variables (with, in the measure, a δ-function imposing (1.5)). However, this result is difficult to extend to non-Abelian lattice theories [15]. The two-dimensional case, where no constraint (1.5) has to be imposed, is noteworthy: taking plaquette variables makes two-dimensional lattice gauge theories solvable. This will be recalled in section 2.2.

1.4. Pure gauge lattice actions

We want to construct a pure gauge action (without matter fields) for the description of the dynamics of the fields U_{ij}. If $S[U]$ denotes this action, functional of the gauge field configurations, we are interested in the partition function

$$Z = \int \mathrm{D}U\, \mathrm{e}^{\beta S[U]} \tag{1.6}$$

and in expectation values of various operators

$$\langle \mathbb{O} \rangle = Z^{-1} \int \mathrm{D}U\, \mathbb{O}[U]\, \mathrm{e}^{\beta S[U]}. \tag{1.7}$$

In this expression, β is the inverse temperature of the Boltzmann weight $\mathrm{e}^{\beta S}$. Its relation with Yang–Mills coupling constant in the continuous theory will be examined in subsection 2.1. In the construction of an invariant, one is naturally led to consider the product along a closed curve $\mathrm{C} = i_1 i_2 \cdots i_n i_1$

$$U_{\mathrm{C}} = U_{i_1 i_2} U_{i_2 i_3} \cdots U_{i_n i_1}$$

which transforms under the transformation (1.1) as

$$U_{\mathrm{C}} \to g_{i_1}\, U_{\mathrm{C}}\, g_{i_1}^{-1},$$

and to suppress the remaining dependence on g_{i_1} by taking the trace. The simplest contour is the plaquette (since the single back and forth step is ruled out due to the constraint (1.2)). This choice is known as the Wilson action

$$S = \sum_p \frac{1}{N} \mathrm{Re}\, \mathrm{Tr}(U_p). \tag{1.8}$$

Due to the trace and the real part, the result is independent of the starting point and of the orientation chosen to define the quantity U_p. Note also that the real part is automatically taken if one sums over all oriented plaquettes,

$$S = \sum_{\text{or.plaq.}} \frac{1}{2N} \operatorname{Tr}(U_p) .$$

This action has been generalized in order to improve some features of the model in three main directions.

i) Changing the trace to a more general real function on the group. Group theory states that all functions invariant under the transformation (1.1) (function of group classes) are linear combinations of the group characters,

$$S = \sum_p \sum_r \gamma_r \operatorname{Re} \chi_r(U_p)/d_r \tag{1.9}$$

d_r is the dimension of the representation r. The choice is very large; however, two cases are of particular interest.

a) The "fundamental–adjoint" mixed action uses only the fundamental and the adjoint representation,

$$\beta S = \sum \left\{ \frac{\beta_f}{d_f} \operatorname{Re} \chi_f(U_p) + \frac{\beta_a}{d_a} \chi_a(U_p) \right\} . \tag{1.10}$$

Its great interest is to allow us to study easily the role of the centre Z of the group in non-Abelian continuous theories. This centre is thought to be important in confinement properties. Indeed the adjoint representation does not distinguish between group elements which differ by a multiplication by a central element (commuting with all other group elements), while the fundamental representation (which is faithful) does. The two-dimensional phase diagram in the couplings $\beta_f = \beta\gamma_f$ and $\beta_a = \beta\gamma_a$ is thus of interest. Note the interesting boundaries of this phase diagram, which are

$\beta_a = 0$ Wilson action for G (e.g. SU(2))

$\beta_f = 0$ Wilson action for G/Z (e.g. $SU(2)/Z_2 \simeq SO(3)$)

$\beta_a = \infty$ Wilson action for Z (indeed, the configurations are reduced to the only U_p maximizing χ_a, i.e. belonging to Z) (e.g. Z_2)

$\beta_f = \infty$ Trivial theory (since all U_p are $\mathbb{1}$, only pure gauge configurations survive).

b) "geometric" actions using the natural metric induced on Lie groups. They are well-adapted to the vicinity of the continuum limit where the fluctuations around a pure gauge configuration see only the local geometry of the Lie algebra. They may also be useful for some peculiar limits (such as SU(N), $N \to \infty$) where the use of the Wilson action leads to singularities and generates unwanted transitions [16] (see below section 2.5). Manton's action [17] uses this metric directly by writing an action proportional to the squared length separating U_p from unity. For instance, using the U(1) or SU(2) gauge group, the class functions depend only on the angle θ of the rotation U_p, and the metric is $(\mathrm{d}\theta)^2$; therefore, Manton's action is

$$S = \sum_p \theta_p^2 . \tag{1.11}$$

However, the periodicity in the angle variables for the compact groups makes this action multi-valued and singular for conjugate points on the group manifold. This defect is cured for U(1) by

Villain's action [18] which restores the periodicity

$$e^{\beta S} = \prod_p \sum_{n_p} \exp\{-\tfrac{1}{4}\beta(\theta_p - 2\pi n_p)^2\}\,. \tag{1.12}$$

This action is generalized to any gauge group using the heat kernel. Indeed, in (1.11), the exponentiated action is Gaussian and can be considered as the response, after a time $t \sim 1/\beta$, to a sudden local heating of an infinite linear rod. However, this rod has the wrong topology of the non-compact group R; it must be replaced by the correct manifold (ring for U(1), S_3 for SU(2), ...). One has to solve the heat equation

$$\Delta f(g, t) = \partial f(g, t)/\partial t, \qquad g \in \mathrm{G}, \qquad t = N/\beta, \qquad f = e^{\beta S} \tag{1.13}$$

where Δ is the Laplace–Beltrami operator, with the initial conditions $f(g, t = 0) = \delta(g = 1)$; the solution gives the heat kernel action [19–25]

$$e^{\beta S} = \prod_p \sum_r d_r \chi_r(U_p) \exp\{-C_r^{(2)}/N\beta\} \tag{1.14}$$

where $C_r^{(2)}$ is the quadratic Casimir invariant for the representation r. For U(1), this is nothing but the preceding Villain's action; the SU(2) heat kernel action reads

$$\exp(\beta S) = \prod_p \sum_{j=0,1/2,\ldots} (2j+1)\frac{\sin(2j+1)\theta_p/2}{\sin\theta_p/2} \exp\{-j(j+1)/2\beta\}\,, \tag{1.15}$$

while only integer values of j are used for the SO(3) group.

ii) Extended plaquette actions use loops larger than the minimal plaquette. The idea is to reduce the lattice artefacts by adding longer range interactions. Hopefully, the rotational invariance effect is restored earlier, and the asymptotic freedom regime extends to lower β. This may be incorporated in a Monte-Carlo approach [26, 27]. Symanzik [28] suggests constructing the lattice action in a systematic way so that the perturbative cut-off dependence is reduced; in the continuum limit behaviour, the first step consists in suppressing the $\mathcal{O}(a^2)$ dependence in the effective continuous Lagrangian [29]. In lattice gauge theories, these approaches add six link loops, such as the "window", the "chair" and the non-planar hexagon (see fig. 2).

iii) Use of other lattices. Their rotation invariance group may be a larger subgroup of the SO(d) continuous rotation group, and again lattice artefacts may be reduced. Few attempts have been made so

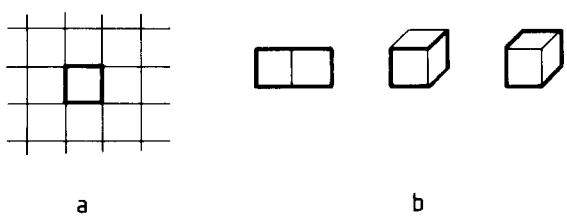


Fig. 2. Loops used in lattice gauge actions. a) the plaquette. b) six link extended loops.

far [30, 31]. Another proposal is to use random lattices [32]. In such systems, the randomness of the lattice sites should restore a mean rotation invariance.

1.5. Order parameter

The various phases of lattice systems are easily characterized if there exists an observable (order or disorder parameter) with a different behaviour in each of them. For instance, the spontaneous magnetization $\langle\sigma\rangle$ in the Ising model is such a parameter; it vanishes identically above the Curie temperature, but takes a finite value under this temperature. This order parameter is local; other observables, such as mass gaps (related to the asymptotic behaviour of the correlation function) are non-local, and may sometimes be used to distinguish the phases.

In lattice gauge theories, there is no local order parameter. Wegner [33] constructed gauge invariant actions as an example of systems without any local order parameter. In particular, the analogue of the spontaneous magnetization, viz. $\langle U_l\rangle$, is not an order parameter. It may be proven (Elitzur's theorem [34], see subsection 2.4) that the average value of any non-gauge invariant quantity vanishes identically, irrespective of the coupling. The Wilson loop [33, 1] is a non-local order parameter which discriminates (in pure gauge systems) between the different phases and which has a nice physical interpretation in terms of confinement properties.

Let us consider the asymptotic behaviour of a large loop C

$$W_r(\mathrm{C}) = \langle \chi_r(U_\mathrm{C})\rangle\,. \tag{1.16}$$

Its physical interpretation is that it measures the variation of the energy when an external static field is introduced along C; stated differently, at some point of C, a quark and an antiquark are created; both propagate along C in opposite directions until they meet and annihilate. Taking for convenience a rectangular loop $R \times T$, we expect a behaviour

$$W(\mathrm{C}) \sim \mathrm{e}^{-TV(R)} \tag{1.17}$$

for large T, where $V(R)$ is the potential energy between two static changes at the distance R, including their self-energy (this formula replaces the periodical quantum behaviour after the Wick rotation leading to the Euclidean metric).

It will be rigorously proven later [35] that this parameter behaves as the exponential of the minimal loop area in the strong coupling region,

$$W(\mathrm{C}) \sim \mathrm{e}^{-KA}\,. \tag{1.18}$$

This leads to a linearly growing potential $V(R) \sim KR$ which prevents the test charges from being separated and implies confinement.

Conversely, we expect, in the low coupling region, a perimeter law. The following argument is, however, perturbative. A rigorous proof exists only for 4-d. U(1) [36] and discrete Abelian gauge groups, while it should fail for SU(2) and other non-Abelian groups in $d \leq 4$ dimensions. Heuristically, to lowest order in g^2 for QED,

$$W(\mathrm{C}) = \left\langle P \exp \mathrm{i}g \oint_\mathrm{C} A\,\mathrm{d}x \right\rangle = \exp\left\{ -\tfrac{1}{2}g^2 \oint_\mathrm{C}\oint_\mathrm{C} \Delta(x-y)\,\mathrm{d}x\,\mathrm{d}y \right\}. \tag{1.19}$$

The propagator Δ favours small separations between x and y and leads to a perimeter behaviour

$$W(C) \sim e^{-\mu P} \tag{1.20}$$

for large loops. In this U(1) case, $V(R)$ is constant for large separations R, leading to charge screening rather than to confinement.

The conclusion is that the string tension, defined as

$$K = \lim_{C \text{ large}} \{-\ln W(C)/A\} \tag{1.21}$$

is possibly a good criterion for discriminating between confined ($K \neq 0$) and non-confined ($K \equiv 0$) phases. However, we already note some important limitations in its use:

a) no matter fields (called quarks in this discussion) are allowed. Otherwise, dynamical quark–antiquark pairs are created when separating the static charges, and screen them: the Wilson parameter behaves as the exponential of the perimeter;

b) the centre of the gauge group must be non-trivial. Otherwise, it is always possible to screen the static charges by gluons (gauge fields) in order to form a gauge invariant object. The corresponding strong coupling diagrams will be displayed in section 3. The Wilson loop has, for instance, a perimeter fall-off with an SO(3) four-dimensional gauge group for all couplings, although a transition is actually observed.

In spite of its defects, the Wilson criterion is presently the most popular order parameter. Using duality arguments, it is possible to construct the dual disorder parameter for Abelian gauge groups. The corresponding quantity generalizes to all groups with a non-trivial centre and is called the 't Hooft loop [37]. We defer its description to subsection 2.3.5, after the review on duality properties.

1.6. Summary of some available numerical results

We do not enter into the technique of Monte-Carlo simulations, nor into a complete survey of the existing data. We refer to another review [9] which is devoted to this aspect, and simply summarize in this subsection the main results obtained for pure gauge theories. The questions of interest are:

– What is the phase structure of the theory and what are the characteristics of each phase?

– What is the nature of the transitions? Indeed, we want to find a second order critical domain represented by the continuous Yang–Mills theory.

– What are the quantitative values of physical observables as we approach these critical domains?

In order to summarize the results on the transitions in a schematic way, we consider three typical gauge groups:

a) Z_2, prototype of discrete Abelian groups, and the simplest possible group. Although the naive continuous limit does not exist for discrete groups, they play an important role in the interpretations in terms of topological excitations, as they are centres of non-Abelian continuous simple groups (SU(2) for Z_2).

b) U(1), compact Abelian group. It corresponds to quantum electrodynamics.

c) SU(2), prototype of continuous non-Abelian compact simple groups, i.e. the groups for which lattice theories have been developed.

Table 1 summarizes the transition pattern for these three categories in various dimensions. The simple case of Z_2 does not call for any particular comment, but for the order of the transition. U(1) is

Table 1
Summary of transition patterns for typical groups

d	Z_2	U(1)	SU(2)
2	confinement for all couplings, no transition		
3	identical to Ising model second order transition near 0.76		
4	(transition at 0.44)	transition of order 2 or higher separating confined phase from Coulomb phase at 0.99	confinement everywhere, but no evidence of a transition, cross-over near 2.2
5 and more	first order transition separating confined and unconfined phases		

more interesting. The low temperature phase (Coulomb phase) appears first in four dimensions and contains a massless photon. There are in particular infinite range correlations in all the weak coupling domain and not only at the transition [36]. This is consistent with the fact that the system represents free QED, a coherent field theory without confinement nor asymptotic freedom.

A comment on discrete subgroups of a given gauge group must be added here. It seems that Z_n with Wilson's action interpolates from Z_2 to U(1) as n increases. Indeed, for $n > 4$, one observes the appearance of a U(1)-like phase for intermediate couplings. This third phase develops between the strong coupling confined phase and the weak coupling unconfined region; the latter shrinks until it vanishes completely at infinite n. Similar effects are observed when non-Abelian groups (such as SU(2)) are simulated by their discrete subgroups; above some critical coupling, the discrete character is smeared by the fluctuations and the system behaves as the continuous one.

Let us turn finally to non-Abelian continuous groups. In four dimensions, due to asymptotic freedom, it is expected that such systems are confined for all couplings and that the continuous limit is reached at the zero coupling limit. This does not exclude, however, the possibility of a transition. With the Wilson SU(2) or SU(3) action, there is some evidence for the absence of such a transition; however, a sharp change in the properties is observed near $\beta = 2.2$ for SU(2) and $\beta = 5.6$ for SU(3). This is seen

– in the string tension (fig. 3 [38]). In the strong coupling region, strong coupling expansions fit the numerical simulation very well and suddenly break at a cross-over point. In the weak coupling phase, the predictions of asymptotic freedom (see subsection 2.1) are well reproduced even far from the continuous limit $\beta = \infty$. A similar behaviour of the mass gap – glueball mass – is also observed;

– in the specific heat, where a bump shows up (see fig. 4 [39]), but without any evidence of a discontinuity, nor divergence, hence excluding a transition;

– in the rotational invariance, restored for some observables in the whole weak coupling phase. Fig. 5 [40] displays equipotential lines between static sources and illustrates this phenomenon that will be discussed in subsection 3.4.3.

Notice that the four-dimensional phase structure depends crucially on the topological properties of the group and on the chosen action. We display, as an example, the fundamental–adjoint mixed action phase diagram for SU(2) (see fig. 6 [41]). Lines are first order transitions. We remark here that centreless groups (such as SO(3) [42]) present a first order transition, in contrast with their universal covering group (such as SU(2)). The interpretation is done using topological excitations described in terms of the discrete centre (here Z_2). Note that the simplicity of the group also plays a role. U(N), which mixes U(1) and SU(N), presents a first order transition in four dimensions, whereas U(1) does not

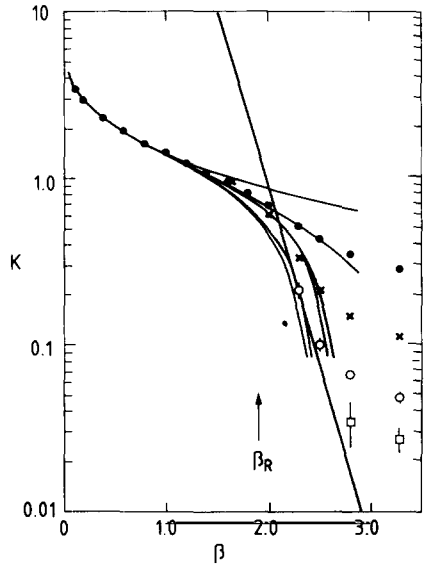


Fig. 3. SU(2) string tension. Monte-Carlo data and their fit by the asymptotic freedom formula. The curves represent the strong coupling estimates, discussed in section 3.4.

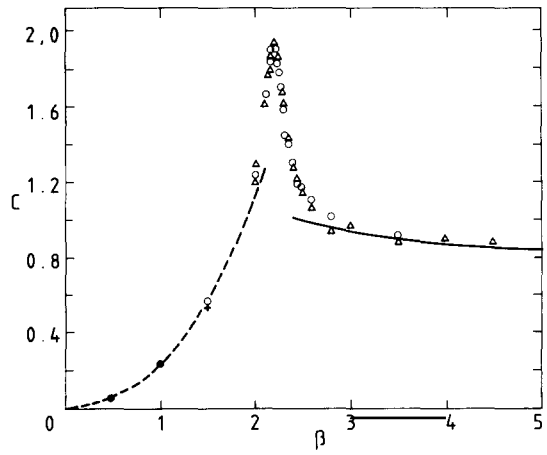


Fig. 4. The specific heat (SU(2) gauge group), showing a bump around the cross-over point $\beta_{co} \simeq 2.2$.

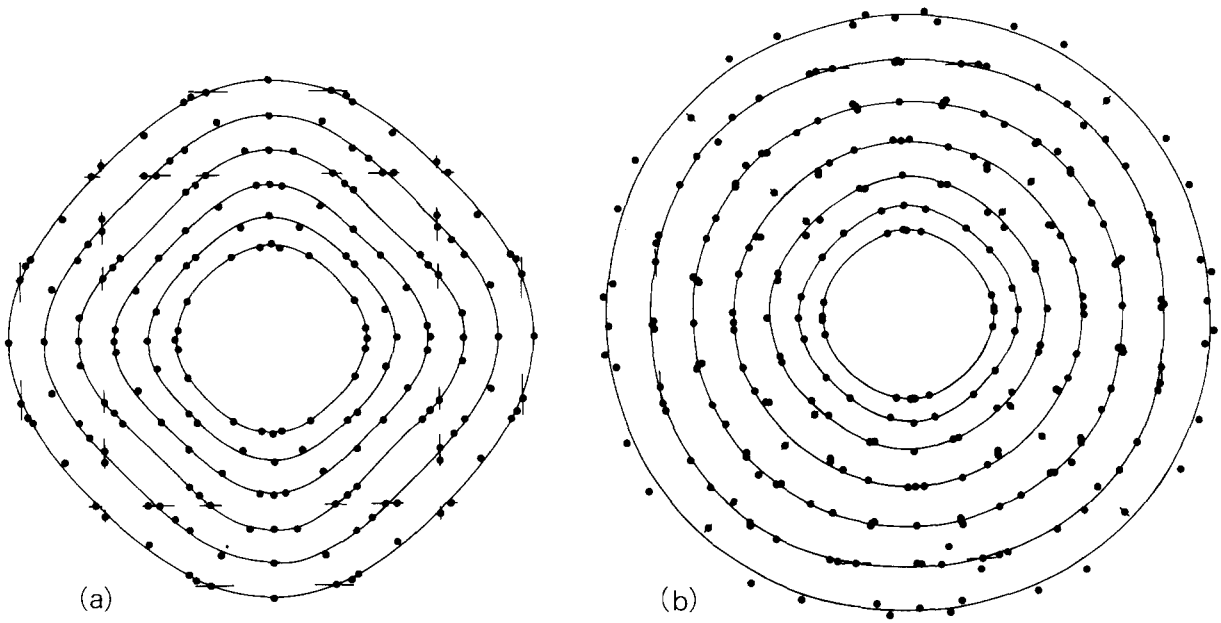


Fig. 5. Restoration of the rotational invariance, seen in the equipotential lines. a) $\beta = 2 < \beta_{co}$, b) $\beta = 2.25 > \beta_{co}$.

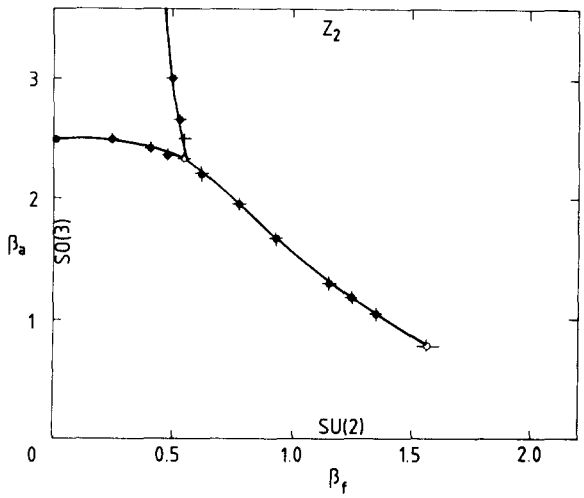


Fig. 6. The fundamental–adjoint mixed SU(2) phase diagram.

and SU(N) does only for $N \geq 4$. In higher dimensions, a first order transition is observed for all the groups.

2. Review of some exact results

2.1. Relation with Yang–Mills perturbative theory

The relation with the continuous theory is obtained by looking at the continuum limit of the action and by parametrizing the lattice gauge fields according to the picture of (1.3). One writes

$$U_{j\mu} = \exp(\mathrm{i}\, g_0\, a\, A_\mu(j))\, U^0_{j\mu}\,. \tag{2.1}$$

Of course, slightly different parametrizations lead to the same result. U^0 is the field configuration near which the lattice gauge field is to be expanded. We take it, for the moment, as a pure gauge configuration ($U^0 = 1$). It is practical to expand the action in two steps. First, using the Baker–Campbell–Hausdorff formula, the plaquette term is written as an exponential

$$U_p = \exp(\mathrm{i}\, g_0\, a^2\, F_{\mu\nu} + \mathrm{o}(a^2))\,, \tag{2.2}$$

with

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + \mathrm{i}\, g_0\, [A_\mu, A_\nu]\,.$$

Then the exponential is expanded and the result is identified with the Yang–Mills action. Additional terms are irrelevant in four dimensions; their dimension is greater than the canonical one. For the generalized plaquette action (1.9), the result is

$$g_0^{-2} = a^{4-d}\beta \sum_r \gamma_r C_r^{(2)}/(N^2-1)\,, \tag{2.3}$$

where $C_r^{(2)}$ is the value of the quadratic Casimir operator in the representation r. For SU(N), $C_{\mathrm{f}}^{(2)} = (N^2-1)/2N$, $C_{\mathrm{a}}^{(2)} = N, \ldots$. In particular, for the four-dimensional SU(N) Wilson action, one gets

$$\beta = 2Ng_0^{-2}. \tag{2.4}$$

In the full theory, the continuum limit is reached in a dynamical way in the approach of a critical point $g_0 \to g_{0_c}$. Indeed at such a point, the lengths of physical interest (correlation length $\xi, \ldots$) become very large as compared to the lattice spacing. Alternatively, one may consider that these physical scales are fixed, and the lattice spacing has to go to zero as $g_0 \to g_{0_c}$. This is nothing but a realization of the renormalization program: the bare coupling has to be adjusted in the ultraviolet limit $a \to 0$ in such a way that physical quantities remain fixed. This process is governed by the renormalization group equation

$$a\, \mathrm{d}\xi/\mathrm{d}a = (a\, \partial/\partial a - \beta(g_0)\, \partial/\partial g_0)\, \xi = 0, \tag{2.5}$$

where

$$\beta(g_0) = -a\, \partial g_0/\partial a|_{\xi}. \tag{2.6}$$

This is solved as

$$\xi = a \exp\left\{ \int^{g_0} \beta(x)^{-1}\, \mathrm{d}x \right\}, \tag{2.7}$$

which determines how a vanishes as g_0 approaches g_{0_c}, zero of the β-function with a negative slope. For non-Abelian four-dimensional gauge theories, there is such an attractive ultraviolet fixed point at the origin $g_0 = 0$, and it is believed to be the only one. More precisely,

$$\beta(x) = -(\beta_0 x^3 + \beta_1 x^5 + \mathcal{O}(x^7)) \qquad (\beta_0 > 0). \tag{2.8}$$

The values β_0 and β_1 are independent of the renormalization scheme. For an SU(N) theory with n_{f} fermions in the representation r, one gets [43]

$$\begin{aligned} \beta_0 &= (4\pi)^{-2} \left[\tfrac{11}{3}N - \tfrac{4}{3}y\right], \\ \beta_1 &= (4\pi)^{-4} \left[\tfrac{34}{3}N^2 - \tfrac{20}{3}Ny - 4C_r^{(2)}y\right], \end{aligned} \tag{2.9}$$

with

$$y = n_{\mathrm{f}} d_r C_r^{(2)}/(N^2-1).$$

One defines the renormalization group invariant mass scale

$$\Lambda_L \equiv a^{-1} (\beta_0 g_0^2)^{-\beta_1/2\beta_0^2} \exp(-1/2\beta_0 g_0^2) \tag{2.10}$$

which is kept fixed as $a \to 0$, $g_0^2 \to 0$. From (2.7), the quantity $\xi \Lambda_L$ has a finite non-zero limit. The actual value of Λ_L is arbitrary; it just fixes the mass scale of the continuous theory.

As an example, let us consider the area law of the Wilson loop (1.18)

$$W(C) \sim e^{-KA} = e^{-\sigma \mathcal{A}} \tag{2.11}$$

where A is the number of plaquettes of the minimal surface, and $\mathcal{A} = Aa^2$ its area. The physical quantity σ is interpreted in the relativistic string model and is related to the Regge slope α' by $\sigma = 1/(2\pi\alpha')$ (experimentally, $\alpha' \simeq 0.90\ \text{GeV}^2$ and hence $\sqrt{\sigma} \simeq 420$ MeV). Therefore the measured lattice quantity K behaves as

$$K \sim \sigma \Lambda_L^{-2} (\beta_0 g_0^2)^{-\beta_1/\beta_0^2} \exp(-1/\beta_0 g_0^2)\,. \tag{2.12}$$

Supplemented by (2.3), this formula permits a direct fit of $\sigma \Lambda_L^{-2}$ from Monte-Carlo data (see fig. 3).

On the other hand, we want to relate lattice predictions to phenomenological results obtained through perturbative calculations. As these calculations have already been done using different renormalization schemes (certainly better adapted to this kind of computation than the lattice regularization), it is necessary to connect Λ_L to the corresponding Λ parameters defined in the continuum theory.

The first calculations were done by evaluating two- and three-point functions at the one-loop level [44]. This calculation is in principle straightforward, but technically involved. Dashen and Gross [45] generalized the background field method to lattice theories; by this trick, the calculations simplify greatly since only the two-point response to the background field at the one-loop level is required. We outline the method, give the main results and refer to the literature [46–49] for detailed calculations.

The basic idea is to shift the field $A_\mu \to A_\mu + A_\mu^0$ by a classical background field A_μ^0. The logarithm of the resulting partition function is a regularized effective action for the background field. Using two different renormalization schemes, the difference between the resulting effective actions is, due to renormalizability, a local gauge invariant quantity. In the infinite cut-off limit, it should have the form

$$\Delta S_{\text{eff}}(A_\mu^0) = \tfrac{1}{4} \int \mathrm{d}x\, (F_{\mu\nu}^0)^2 \left[g_{(1)}^{-2} - g_{(2)}^{-2} + c \right]. \tag{2.13}$$

The condition $\Delta S_{\text{eff}} = 0$ gives the relation between the coupling constants $g_{(1)}$ and $g_{(2)}$ of the two different renormalization schemes. Hence

$$\Lambda_{(1)}/\Lambda_{(2)} = \exp\{c/2\beta_0\}\,. \tag{2.14}$$

It is convenient to use the non-linear parametrization (2.1) instead of the preceding linear shift; however, it can be easily checked that the coefficients of the extra terms vanish if the background field U^0 satisfies the classical equations of motion. The perturbative lattice calculation then proceeds through the usual steps: gauge fixing, corresponding Faddeev–Popov ghosts, Feynman rules, computation of all one-loop diagrams with two external background field lines involved in (2.13).

We now display the result for the generalized pure gauge action (1.9), compared to the minimal subtraction using dimensional regularization

Table 2
Ratios of Λ_{MS}/Λ_L parameters for SU(2) and SU(3) Wilson actions with fermions

$n_f =$	0	1	2	3	4
SU(2)	7.46	9.27	12.09	16.85	25.01
SU(3)	10.85	12.78	15.41	19.09	24.49

$$\ln(\Lambda_{MS}/\Lambda_L) = 0.888 + \left[\left(\sum_r C_r^{(2)2}\gamma_r\right)\Big/\left(\sum_r C_r^{(2)}\gamma_r\right) - N/6\right]\Big/ 8\beta_0 . \tag{2.15}$$

For the Wilson action with fermions, the result is

$$\ln(\Lambda_{MS}/\Lambda_L) = (\ln 4\pi - \gamma)/2 + [1/16N - 0.0849780N + 0.006887y]/\beta_0 , \tag{2.16}$$

leading to the numerical table 2. These results can be easily converted into other schemes (e.g. MOM using ref. [50] results).

2.2. Triviality of two-dimensional theories

In the maximal gauge introduced in 1.3, the two-dimensional partition function factorizes as a product of $\mathcal{N}$ identical integrals

$$\begin{aligned} Z &= \int \prod_l DU_l \exp\Big(\sum_p \beta\chi(U_p)\Big) \\ &= \int \prod_p DU_p \exp\Big(\sum_p \beta\chi(U_p)\Big) = \Big(\int DU \exp(\beta\chi(U))\Big)^{\mathcal{N}} \end{aligned} \tag{2.17}$$

while the expectation value of the Wilson loop reads (fig. 7) (see section 3 and appendix A for notations on characters χ)

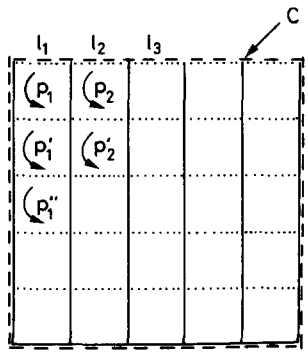


Fig. 7. Integration plaquette variables used in the computation of the Wilson loop in a two-dimensional system.

$$\langle \chi_r(U_C) \rangle = Z^{-1} \int \mathrm{D}U_{l_1} \cdots \chi_r(U_{l_1} \cdots) \, \mathrm{e}^{\beta S}$$

$$= Z^{-1} \int \prod_p \mathrm{D}U_p \, \chi_r(U_{p_1} U_{p_1'} U_{p_1''} \cdots U_{p_2} \cdots) \, \mathrm{e}^{\beta S}$$

$$= d_r \beta_r^A = d_r \exp(A \ln \beta_r) . \tag{2.18}$$

This also results from the rules of strong coupling expansions, as only the leading term survives in two dimensions (see section 3). Notice finally that this triviality of two-dimensional theories is not guaranteed for generalized actions. For example, the Z_2 "window" action which uses the planar six-link loop of fig. 2b reduces to the two-dimensional Ising model, hence gives rise to a phase transition [51].

2.3. Duality

The duality transformation was first established for the two-dimensional Ising model. It has been extended to more complicated systems [33]; in particular, it is applicable to lattice Abelian gauge theories [6, 20, 52].

Physically, duality is interesting in two respects. First, it connects the original model at high (resp. low) temperature to a dual model at low (resp. high) temperature. Secondly, the degrees of freedom of this dual model may be seen as the defects, or disorder variables [53] of the original model. In lattice gauge theories, duality is therefore useful to expose the role of topological excitations, such as vortices or monopoles, in confinement or phase transitions [54].

2.3.1. Equivalence of three-dimensional Ising and Z_2-gauge models

We first illustrate on a simple instance the general formalism of duality transformation, which may seem rather abstract to the unfamiliar reader.

Starting from the three-dimensional Z_2 gauge model, we write the partition function as

$$Z_{\text{gauge}} = 2^{-3\mathcal{N}} \sum_{\{\sigma_l\}} \exp\Big[\beta \sum_p \sigma\sigma\sigma\sigma \Big]$$

$$= 2^{-3\mathcal{N}} \sum_{\{\sigma_l\}} \prod_p \cosh \beta (1 + \sigma\sigma\sigma\sigma \tanh \beta) .$$

The summation over a given σ_l gives either 0 or 1 according to the parity of the power of σ_l in each term of the expanded product. Hence each contributing term of this expansion can be geometrically associated with closed surfaces (may be intersecting) made with those plaquettes carrying a $\tanh \beta$ factor. Hence

$$Z_{\text{gauge}} = (\cosh \beta)^{3\mathcal{N}} \sum_{\substack{\text{closed} \\ \text{surfaces}}} (\tanh \beta)^n$$

where n is the number of plaquettes of the surfaces. A closed surface is also characterized by the different volumes it separates (inner and outer). At the centre of each cube, we introduce a variable μ_i

set to 1 if this cube belongs to the outer volume, -1 otherwise. Each configuration of μ_i (defined, however, up to a global sign) is in one-to-one correspondence with each closed surface. This fact is used to rewrite the expression of the partition function. As a plaquette belongs to the surface only if the two adjacent μ's are opposite, we have

$$n = \sum_{(i,j)} (1 - \mu_i \mu_j)/2$$

where the summation runs over all pairs of neighbours. Hence

$$Z_{\text{gauge}} = \tfrac{1}{2}(\cosh \beta)^{3\mathcal{N}} (\tanh \beta)^{3\mathcal{N}/2} \sum_{\{\mu\}} \exp\Big[-\tfrac{1}{2} \ln \tanh \beta \sum \mu_i \mu_j \Big]$$

which is, up to a multiplicative factor, the partition function of an Ising model for the μ's,

$$F_{\text{gauge}}(\beta) = F_{\text{Ising}}(\beta^*) + \tfrac{3}{2} \ln \sinh 2\beta - \tfrac{1}{2} \ln 2 \tag{2.19}$$

with

$$\beta^* = -\tfrac{1}{2} \ln \tanh \beta$$

or, more symmetrically,

$$\sinh 2\beta \sinh 2\beta^* = 1 . \tag{2.20}$$

The inverse derivation could proceed as follows. The high temperature Ising diagrams are closed curves, which limit some (open) surfaces. Each constitutive plaquette of these surfaces is associated with the bond joining the centres of the two cubes sharing this plaquette. This bond carries a gauge field set to -1. We leave the detailed calculations (leading of course to the same result) to the reader and turn now to the general formulation.

2.3.2. General formulation

We consider fields defined on a d-dimensional a-complex with a k-dimensional interaction. We recall first that an abstract a-complex is [55] a set of elements called cells. A lattice is an a-complex; its cells are sites, links, plaquettes, cubes, ... In general, an a-complex satisfies the following axioms

(i) to each cell is assigned a non-negative integer, called its dimension. The upper bound of the cell dimensions is the dimension of the complex (here d).

(ii) each cell c_k of dimension k (k-cell) is in correspondence with another cell $-c_k$ of the same dimension (cell with opposite orientation).

(iii) to two cells with dimension differing by 1 is associated a signed integer, the incidence number $(c_k : c_{k-1})$ (non-zero if "c_{k-1} is a face of c_k"). One has $(-c_k : c_{k-1}) = (c_k : -c_{k-1}) = -(c_k : c_{k-1})$.

(iv) $\Sigma_{c_k} (c_{k+1} : c_k)(c_k : c_{k-1}) = 0$ (this condition is the basis of homology and cohomology theorems).

A k-chain over the coefficient domain G (Abelian group) is an odd function of the k-cells on G: $c_k^{(i)} \to \phi_i$, with $\phi_i \in G$. This induces a group structure on the k-chains. Any k-chain may then be written as a sum of monomial chains $\Sigma_i \phi_i c_k^{(i)}$. The monomial chain (abusively) denoted $\phi_i c_k^{(i)}$ maps the cell $c_k^{(i)}$

on ϕ_i, and all other cells on the identity element of G. In the lattice example, 2-chains over $Z_2 = \{1, -1\}$ are surfaces made of those plaquettes where $\phi_i = -1$.

Finally, the boundary Δ and coboundary ∇ operators map k-chains into $(k-1)$- and $(k+1)$-chains as follows

$$\begin{matrix}\nabla \\ \Delta\end{matrix} \sum_i \phi_i c_k^{(i)} = \sum_{i,j} \phi_i (c_k^{(i)} : c_{k\pm1}^{(j)}) c_{k\pm1}^{(j)} .$$

Let us introduce a model in this context. The fields ϕ are defined on $(k-1)$-cells and take their values in the Abelian group G. A field configuration is then, in this language, a $(k-1)$-chain over G. Let the action be of the form

$$\beta S = \sum_{k\text{-cells c}} \chi\Big(\prod_{\phi\in\Delta c} \phi\Big) + \sum_{(k-1)\text{-cells}} f(\phi), \tag{2.21}$$

where χ and f are any class functions on the group G. For commodity of notations, Δc denotes the set of cells pertaining to the boundary of the monomial chains associated with c. $k = 1$ corresponds to spin models, $k = 2$ to gauge models, and f to an applied external field. Expanding the exponentiated action on characters, the partition function reads

$$Z = \|G\|^{-n_{k-1}} \sum_{(k-1)\text{-chains}} \Big\{ \prod_{k\text{-cells c}} \sum_r \beta(r) \chi_r\Big(\prod_{\phi\in\Delta c} \phi\Big)\Big\} \Big\{ \prod_{(k-1)\text{-cells}} \sum_r \gamma(r) \chi_r(\phi)\Big\} . \tag{2.22}$$

The group volume $\|G\|$ comes from the normalization of the group measure $\|G\|^{-1} \Sigma_{\phi\in G} 1 = 1$. n_k is the total number of k-cells in the a-complex. The products are expanded (following the strong coupling expansion techniques, described in more detail in section 3). Each term coming from the first product is obtained by assigning a representation r to each k-cell, and therefore corresponds to a k-chain g over the coefficient domain G^* (the set of irreducible representations of G). Since G is Abelian, G^* is an Abelian group for the outer product of representations. This crucial group structure unfortunately does not extend to non-Abelian groups. Similarly, the second product is expanded in terms of $(k-1)$-chains h over G^*. For each term, the summation over the configurations, i.e. the $(k-1)$-chains $\{\phi\}$, is performed using orthogonality relations of characters and provides a selection rule: the contribution is $\|G\|^n$ if $h = \Delta g$, 0 otherwise. Hence the partition function reads

$$Z = \sum_{k\text{-chains g}} \exp\Big\{ \sum_{r\in g} \ln \beta(r) + \sum_{s\in\Delta g} \ln \gamma(s)\Big\} . \tag{2.23}$$

Using the definition of the boundary operator, s is the product of representations r of the k-cells incident to a $(k-1)$-cell; thus the last term is rewritten as a summation over all $(k-1)$-cells c of $\ln \gamma(\Pi_{r\in\nabla c} r)$. The action takes now a form quite similar to the original one, except for the replacement of the boundary operator Δ by the coboundary operator ∇. This difference is removed by going to the dual a-complex.

The dual a-complex is defined as follows. To any cell c_k, associate a $(d-k)$-cell c^*_{d-k}. All incidence numbers remain unchanged; in particular, $n_k = n^*_{d-k}$. On the hypercubical lattice, for nearest neighbour

interactions, this transformation has a nice geometrical visualization. Mark a point at the centre of every hypercube (d-cell); these points will be the nodes of the dual lattice. Two neighbouring d-cells, sharing one $(d-1)$-face, are neighbouring sites on the dual lattice, joined by a link, dual of the common face; and so on.... For instance, the dual of the hypercubical lattice is again a hypercubical lattice shifted by the vector $(a/2, a/2, \ldots)$. In this transformation, boundary and coboundary operators exchange their roles. Hence the action becomes

$$\beta^* S_{\text{dual}} = n^*_{d-k} \ln\|G^*\| + \sum_{(d-k+1)\text{-cells c}} \ln \gamma\Big(\prod_{r\in\Delta c} r\Big) + \sum_{(d-k)\text{-cells}} \ln \beta(r). \tag{2.24}$$

This has exactly the same structure as the initial action. The only changes are the dimension of the cells, the underlying group and the class functions. Note also the interchange of the roles of external field and coupling, and the interchange between strong and weak coupling regions (since $\chi = 0$ (or $f = 0$) lead to all $\beta(r)$ ($\gamma(r)$) vanishing (but $r = 0$), i.e. infinite dual couplings $\ln \beta(r)$ ($\ln \gamma(r)$)). We summarize the conclusions in table 3.

Let us examine in more detail the particular case $f = 0$ of no external field. Note first that $\Delta\Delta$ and $\nabla\nabla$ map any chain on zero. The inverse is not true; one defines the kth homology (cohomology) group as the quotient of the image Δ (∇) of $(k+1)$-chains ($(k-1)$-chains) by the kernel in the set of k-chains of Δ (∇). From now on, we suppose that all $k \neq 0$ homology groups of our a-complex are trivial. This means that any k-chain with null boundary (closed k-chain) is the boundary of a $(k+1)$-chain. For instance, $k = 1$: each closed curve can be considered as the boundary of a surface. This is not generally true, as is shown by the example of a circle on a torus. However, the hypothesis is satisfied for complexes which are lattices in the Euclidean space, as long as one does not impose cyclic boundary conditions.

In the preceding reasoning on the partition function, we are now restricted to only closed k-chains g over G*. This constraint is removed by rewriting them in terms of $(k+1)$-chains. As the correspondence between a closed k-chain g and a $(k+1)$-chain which has g as boundary is not one-to-one (each coefficient of the $(k+1)$-chain can be multiplied by the corresponding coefficient of a given closed $(k+1)$-chain without changing its boundary), this induces a constant multiplicative factor on the partition function. After going into the dual a-complex, we get in this particular case

$$\beta^* S_{\text{dual}} = n_{k-1} \ln\|G\| - \ln x_{k+1} + \sum_{(d-k)\text{-cells c}} \ln \beta\Big(\prod_{r\in\Delta c} r\Big), \tag{2.25}$$

Table 3
Summary of the general duality transformation

	support group	a-complex G	dual a-complex G* (repres. of G)
with external source	field	$\phi \in G$ located on $(k-1)$-cell	$r \in G^*$ located on $(d-k)$-cell
	interaction	$\chi(\phi) = \exp\Big(\sum_r \beta(r)\chi_r(\phi)\Big)$	$\ln \gamma(r)$
	source	$f(\phi) = \exp\Big(\sum_r \gamma(r)\chi_r(\phi)\Big)$	$\ln \beta(r)$
without external source	field	$\phi \in G$ located on $(k-1)$-cell	$r \in G^*$ located on $(d-k-1)$-cell
	interaction	$\chi(\phi) = \exp\Big(\sum_r \beta(r)\chi_r(\phi)\Big)$	$\ln \beta(r)$

where x_{k+1} is the number of closed $(k+1)$-chains. These numbers x_k are easily estimated; a closed k-chain is the boundary of a $(k+1)$-chain defined up to a closed $(k+1)$-chain, and therefore

$$x_k = \|G\|^{n_{k+1}}/x_{k+1}, \qquad \text{with } x_d = \|G\|. \tag{2.26}$$

The dual coupling constants are solutions of the equations

$$\beta(r) = \|G\|^{-1} \sum_{\phi} \beta^*(\phi)\, \chi_r(\phi), \tag{2.27}$$

or

$$\beta^*(\phi) = \|G\|^{-1} \sum_{r} \beta(r)\, \chi_{\phi^{-1}}(r),$$

where the dual characters $\chi_\phi(r)$ are the elements of the inverse matrix $(\chi_r(\phi))^{-1}$ with respect to the indices r and ϕ. This results from orthogonality relations on characters; note that the dual of the dual group G* is G itself.

As noted above, this nice duality operation fails for non-Abelian groups G, because the set of irreducible representations G* is not endowed with a group structure.

Some care must be taken with infinite groups. Starting from a compact group (as G = U(1)), $\|G\|^{-n_{k-1}} \Sigma$ in (2.23) is replaced by a normalized integral on the group. The dual group (as Z) is discrete, non-compact and infinite, and the normalization of the field measure is meaningless; hence the first term n^*_{d-k} in (2.24) must be removed. Similarly, the meaningless term $-\ln x_{k+1}$ in (2.25) must be replaced by a gauge fixing procedure.

2.3.3. Application

All our applications concern hypercubical lattices in the thermodynamic limit. We do not consider other cases here.

With external sources, duality gives results for gauge models ($k = 2$) at $d = 2$ (two-dimensional spin models with an applied external field are dual to gauge models with Higgs field) and $d = 3$ (self-duality relations for three-dimensional gauge models with Higgs field). They are used in section 5.

A trivial application concerns $d = k$ in the absence of an external field. The dual model is just an external constant field, which contains no dynamics. We recover here the triviality of one-dimensional Ising and two-dimensional gauge models.

As a third application, we recover easily the formulae displayed in the example of three-dimensional Z_2-gauge and Ising models, with $d = 3$, $n_1 = n_2 = 3$, $\|G\| = 2$ and k = either 1 or 2. In particular, the transition points of the two models should be related. As the Ising model has a transition at $\beta_I = 0.2217$, the gauge model must also undergo a second order transition at $\beta_c^* = -\frac{1}{2} \ln \tanh \beta_I \approx 0.7613$.

The four-dimensional Z_2 gauge model ($d = 4$, $k = 2$) yields

$$F(\beta) = F(-\tfrac{1}{2} \ln \tanh \beta) + 3 \ln \sinh 2\beta. \tag{2.28}$$

Strong and weak couplings are related by (2.28). Moreover, as the last term on the r.h.s. is regular, any singularity for F at β_c will lead to the existence of another singularity located at $-\frac{1}{2} \ln \tanh \beta_c$.

Therefore, if there is only one transition, it must lie at the self-dual point

$$\beta_c = -\tfrac{1}{2}\ln\tanh\beta_c = \tfrac{1}{2}\ln(1+\sqrt{2})\,. \tag{2.29}$$

Indeed, a first order transition is observed at this point.

Any Abelian compact gauge group is constructed from Z_n and U(1). The Z_n group is self-dual; indeed the irreducible representations of $Z_n = \{\exp(2i\pi k/n);\ k = 1, \ldots, n\}$ are $\exp(2i\pi kr/n)$, labelled by $r = 1, \ldots, n$ and their set is obviously isomorphic to Z_n. Hence, there is duality between three-dimensional Z_n gauge and spin models, and self-duality relations for four-dimensional Z_n gauge systems. The most general action in this case is conveniently rewritten as

$$\exp(\beta S) = \prod \sum_{r=1}^{n} \beta(r) \exp\left(2i\pi r \left(\prod \phi\right)\Big/ n\right) \tag{2.30}$$

and depends only on $[n/2]$ real parameters (where $[m]$ denotes the largest integer contained in m), due to the reality conditions $\beta(r) = \beta(n-r)$ and to the fact that the β's are defined up to an overall factor corresponding to an ineffective additive term in the action. Dual couplings are thus

$$\beta^*(r) = (1/n) \sum_l \beta(l) \cos(2\pi lr/n)\,. \tag{2.31}$$

For instance, application to the four-dimensional Z_3 gauge model (depending on one coupling) yields

$$(e^{-3\beta/2} - 1)\,(e^{-3\beta^*/2} - 1) = 3 \tag{2.32}$$

and the transition must therefore occur at $\beta_c = \frac{2}{3}\ln(1+\sqrt{3})$. For larger n, duality induces only a symmetry in the multidimensional phase diagram, and we refer to the literature [20, 56–58] for detailed results.

The U(1) gauge group is not self-dual. Its irreducible representations are labelled by a signed integer n: $\chi_n(e^{i\theta}) = e^{in\theta}$, and then $U(1)^* \approx Z$, the additive non-compact group of integers. In four dimensions, the Wilson action

$$\beta S = \beta \sum_p \cos\theta_p \tag{2.33}$$

is dual to an action for integer gauge fields n_l

$$\beta^* S_{\text{dual}} = \sum_p \ln I_{n_p}(\beta)\,,$$

with

$$n_{i\hat{\mu}\hat{\nu}} = n_{i\hat{\mu}} + n_{i+\hat{\mu},\hat{\nu}} - n_{i\hat{\nu}} - n_{i+\hat{\nu},\hat{\mu}}\,.$$

A gauge fixing term is implicit in this model. If the Villain action is used instead of (2.33), the resulting dual model is simply the Gaussian integer gauge model

$$\beta^* S_{\text{dual}} = -\sum_p \beta^{-1} n_p^2 + \text{Const}\,.$$

There is an alternative form of this dual action which proves more useful in the discussion of the physics of the phase transition [59–62]. By the Poisson formula, the summation over the integer variables $n_{i\hat{\mu}}$ is traded for a summation over integer variables $J_{i\hat{\mu}}$ satisfying the conservation equations $\Delta_\mu J_{i\hat{\mu}} = 0$ at each site,

$$\begin{aligned}\beta^* S_{\text{dual}} &= -\sum_p \beta^{-1} n_p^2 + \mathrm{i} \sum_{i\mu} 2\pi n_{i\hat{\mu}} J_{i\hat{\mu}} \\ &\simeq -\tfrac{1}{2} \sum_{ij\mu\nu} (2\pi)^2 \beta J_{i\hat{\mu}}\, \Delta^{-1}_{\mu\nu}(i,j)\, J_{j\hat{\nu}} \end{aligned} \tag{2.34}$$

where the inverse of the lattice Laplacian Δ^{-1} depends on the choice of gauge for the n's, but the combination $J\Delta^{-1}J$ does not, thanks to $\Delta_\mu J_{\hat{\mu}} = 0$. The variables $J_{i\hat{\mu}}$ are the topological excitations of the system. Introducing an external electric source in the system reveals that they behave as monopole lines, forming closed loops because of their conservation law. Energy-entropy arguments then show that, at small coupling, there are only a few small monopole loops, whereas, at strong coupling, large loops exist, form a "monopole condensate", and create a linear confining potential between static charges. That the vacuum of the confined phase might be regarded as a magnetic superconductor, with electric flux confined in tubes, was originally proposed by Mandelstam [63] and 't Hooft [64]. The mechanism of monopole condensation in the U(1) model is reminiscent of the physics of the two-dimensional xy-model, and has been confirmed by Monte-Carlo simulations [65]. In four-dimensional Z_n gauge theories with Villain action, there is an exact self-duality, and for $n \geq 4$, the occurrence of two phase transitions. One of them is the analog of the U(1) transition between the Coulomb and confined phases, but there is a third phase at weak coupling, dual of the confined phase, which may be seen as a condensate of electric charges. The second phase transition takes place between the Coulomb phase and this "Higgs phase".

In contrast, in three dimensions, the dual of the U(1) gauge theory is a Gaussian integer spin model, or alternatively, a model of monopoles living on the sites of the lattice. One may show that they disorder the system for all couplings [59, 60, 66], and that the system is always in its confined phase [67].

2.3.4. Dual observables

Up to now, the duality transformation has been applied only to the partition function. It is not very difficult to construct the dual quantity of a given observable. We consider here the plaquette $\langle \chi_r(U_p) \rangle$. Diagrams contributing to this observable have the topology of surfaces with a plaquette boundary p (see section 3 for details). Change in the diagram the representation s carried by the plaquette p to $s \otimes r$ (which is irreducible since G is Abelian); it is clear that the modified diagram contributes to the partition function Z. Thus the observable $\langle \chi_r(U_p) \rangle$ appears as the ratio of a modified partition function in which the coupling constants relative to the plaquette p have been changed from $\beta(s)$ to $\beta(s \otimes r)$, by the original partition function. One applies the duality transformation on this modified partition function. Some straightforward algebra leads to the dual quantity

$$\langle \chi_r(U_p) \rangle = \langle \beta(r \otimes U_p^*) - \beta(U_p^*) \rangle_{\text{dual}}\,. \tag{2.35}$$

With this method, it is easy to make a dictionary between dual observables. For instance, in the three-dimensional Z_2 theory, the Wilson loop is dual to the ratio of Ising partition functions

$$\langle W(C)\rangle = Z_{\text{Ising}}(\beta^* \to -\beta^* \text{ on } S^*)/Z_{\text{Ising}}(\beta^*) = \left\langle \exp\left(-2\sum_S \beta^* \sigma_i \sigma_j\right)\right\rangle \tag{2.36}$$

where, in the numerator, all Ising couplings crossed by some arbitrary surface S^* of boundary C have been reversed. For a large loop, at strong gauge coupling, the left hand side has an area decay law, whereas the right hand side is the exponential of the differences of free energies of two Ising systems, i.e. the interfacial free energy created by the change of couplings

$$e^{-KA} = e^{\Delta F}.$$

In the limit of a large loop extending to infinity, ΔF may also be regarded as the excess free energy when opposite infinitesimal magnetic fields applied on the boundary above and below the surface tend to order the system in two different phases. This interpretation will be useful in the discussion of section 3.4.3.

In the four-dimensional model, the Wilson loop (order parameter) is similarly dual to a "disorder parameter", the 't Hooft loop $\langle H(C^*)\rangle$, which measures the response of the system to a change of some of its couplings.

2.3.5. 't Hooft loop

The 't Hooft loop operator may actually be defined for any gauge group G with a non-trivial centre Z. We discuss here only the four-dimensional lattice realizations of this idea. Given a closed loop C^* on the dual lattice, one considers an arbitrary surface Σ^* of boundary C^* and the set Σ of plaquettes dual to the plaquettes of Σ^*. The effect of the loop operator is to multiply in the gauge action all plaquettes of Σ by a non-trivial element z of the centre of G

$$\langle H_z(C^*)\rangle = Z^{-1}\prod \int DU_l \exp\left\{\beta \operatorname{Re}\left[\sum_{p\notin\Sigma} \chi(U_p) + \sum_{p\in\Sigma} \chi(zU_p)\right]\right\}. \tag{2.37}$$

For $G = SU(N)$, the centre is Z_N, and the 't Hooft loop may be seen to create a Z_N monopole loop along C^*. In the Z_N theory, it is dual to the electric loop operator, i.e. the Wilson loop; in the three phases described above, one has the three following behaviours

confined phase (small β)	$\langle W\rangle \sim e^{-A}$	$\langle H\rangle \sim e^{-P}$
Coulomb phase (intermediate β)	$\langle W\rangle \sim e^{-P}$	$\langle H\rangle \sim e^{-P}$
Higgs phase (large β)	$\langle W\rangle \sim e^{-P}$	$\langle H\rangle \sim e^{-A}$.

The area fall-off of the 't Hooft loop at small coupling (large β) indicates the suppression of large monopole loops, or alternatively, the existence of a linear confining potential between monopoles. For a further discussion of this disorder operator, its commutation relations with Wilson loops and a detailed study of its asymptotic behaviour, we refer the reader to the literature [37, 58, 68–70]. When the loop extends to the boundary of an entire two-plane through a finite lattice, and periodic boundary conditions are then introduced, the computation of $\langle H\rangle$ amounts to introducing "twisted boundary conditions" [71, 58, 72].

2.4. Elitzur's theorem

This important theorem [34] states that any non-invariant quantity has a vanishing mean value in lattice gauge theories, irrespective of the coupling.

The proof of this assertion is simple; some care, however, is required because of the non-uniformity of the double limit to be taken. Let us consider, in general, a group of transformations acting on fields $\phi \to {}^g\phi$, with an invariant action $S(\phi) = S({}^g\phi)$. The quantity $f(\phi)$ is non-invariant, or, more precisely, has no component transforming according to the trivial representation; in other words

$$\int f({}^g\phi)\, \mathrm{D}g = 0\,. \tag{2.38}$$

For instance, gauge fields, or matter fields, are such quantities. The mean value is computed in a finite system of size $\mathcal{N}$ and with external sources J

$$\langle f(\phi)\rangle_{\mathcal{N},J} = (1/Z_{\mathcal{N},J}) \int \exp\{S(\phi) + J\cdot\phi\}\, f(\phi)\, \mathrm{D}\phi\ ; \tag{2.39}$$

the thermodynamic limit $\mathcal{N}\to\infty$ and the zero external source $J\to 0$ must then be taken in this order

$$\langle f(\phi)\rangle = \lim_{J\to 0}\lim_{\mathcal{N}\to\infty} \langle f(\phi)\rangle_{\mathcal{N},J}\,. \tag{2.40}$$

In the integral on the right hand side of eq. (2.39), we perform the change $\phi \to {}^g\phi$, taking into account the invariance of the measure and of the action. As the result is independent on g, we also integrate over g ($\int \mathrm{D}g = 1$), and hence

$$\langle f(\phi)\rangle_{\mathcal{N},J} = (1/Z_{\mathcal{N},J}) \int\int \exp\{S(\phi) + J\cdot {}^g\phi\}\, f({}^g\phi)\, \mathrm{D}\phi\, \mathrm{D}g\,. \tag{2.41}$$

The set of fields $\{\phi\}$ is now split into a subset $\{\phi''\}$ unchanged by the transformation ($\phi'' = {}^g\phi''$) and the complementary subset $\{\phi'\}$.

The crucial point is that, for sufficiently small sources $\|J\| < \varepsilon$,

$$|\exp\{J'\cdot\phi'\} - 1| \le \eta(\varepsilon) \tag{2.42}$$

with $\eta(\varepsilon)$ vanishing as ε goes to zero, and being independent of both ϕ' and $\mathcal{N}$. The uniformity in ϕ' is trivially satisfied since we deal with compact groups. The uniformity of $\mathcal{N}$ is satisfied if only a finite, $\mathcal{N}$-independent number of degrees of freedom ϕ' are concerned by the transformation. This is not the case in spin systems, since the global transformation rotates all fields, and thus $J\cdot\phi$ is an extensive quantity proportional to $\mathcal{N}$; indeed the theorem is false in this case and spins can take a non-zero mean value, as is well known. Conversely, gauge systems satisfy this condition, because non-invariance of f means that there exists a local gauge transformation g acting on a finite number of degrees of freedom involved in f such that (2.38) holds.

The integral is split into two parts, writing $\exp\{J'\cdot {}^g\phi'\} = 1 + [\exp\{J\cdot {}^g\phi'\} - 1]$. The first part vanishes

in the integration over g, according to (2.38). The second one is bounded in absolute value, using (2.42) and the maximum $\|f\|$ over ϕ of $|f(\phi)|$. Hence

$$|\langle f(\phi)\rangle_{\mathcal{N},J}| \leq \eta(\varepsilon)\,\|f\|\,.$$

Now no problem is encountered when taking the limits in the right order, thanks to the uniformity conditions on $\eta(\varepsilon)$

$$\langle f(\phi)\rangle \equiv 0\,. \tag{2.43}$$

This result applies for pure gauge systems (e.g., $\langle U_l\rangle \equiv 0$) as well as gauge + matter systems.

2.5. *Large N limit*

The large N limit of U(N) or SU(N) gauge theories has received much attention since it has been realized that it leads to a reasonable phenomenological description of hadronic physics [73]. 't Hooft has shown that in continuous field theory, only planar Feynman diagrams survive in the limit. It is legitimate to look at the same limit for lattice gauge theories. In order for the theory to make sense, the coupling has to be rescaled. In the lattice action $(\beta/2N)\,\mathrm{tr}(U+U^\dagger)$, one sets $\beta/2N = N\bar{\beta}$, $\bar{\beta}$ finite. This agrees with the known rescaling of the continuous theory, $\bar{g}_0^2 = Ng_0^2$ (cf. eq. (2.4)), and, in the strong coupling phase, will be shown to lead to a sensible limit. The free energy per site is expected to be proportional to N^2, the number of degrees of freedom of a $N \times N$ unitary matrix.

In a series of works initiated by Gross and Witten [16] and Wadia [74], the structure of U(∞) two-dimensional lattice gauge theories has been unravelled. We recall from section 2.2 that, in two dimensions, plaquette variables are independent and the partition function and other observables may be computed by a single integration

$$Z = \left[\int \mathrm{D}U \exp(\beta\chi(U))\right]^{\mathcal{N}} = \tilde{\beta}_0^{\mathcal{N}}\,. \tag{2.44}$$

The integral $\tilde{\beta}_0$ may be explicitly computed at large N by the saddle point method; although one is dealing with a one-plaquette model, there are still a large number (N^2) of degrees of freedom, and $\ln\tilde{\beta}_0$ may be non-analytic. The non-analyticity – i.e., the phase structure of this two-dimensional model – depends on the explicit form of the action. For Wilson's action, a third order (non-deconfining) phase transition is found at $\bar{\beta} = \frac{1}{2}$. The internal energy $\langle \mathrm{tr}\, U_p\rangle = \partial \ln\tilde{\beta}_0/\partial\beta$ and the string tension $-\ln\bar{\beta}$ have their second derivatives discontinuous at $\bar{\beta} = \frac{1}{2}$. In particular, there are local order parameters in this model: all $\langle \mathrm{tr}\, U^n\rangle$, $n \geq 2$ vanish identically for $\beta \leq \frac{1}{2}$. This phase structure, however, depends on the form of the action; taking the heat kernel or Manton's action leads to regular functions [75, 24].

It is not very clear to which extent these features foreshadow what happens in higher dimensions. Notice, however, that this study of two-dimensional models, in particular the form of the function $\tilde{\beta}_0(\bar{\beta})$, is useful in approximations starting from independent degrees of freedom (strong coupling, mean field).

In higher dimensions, not many exact results are known. It has been recently shown by Eguchi and Kawai [76] that a remarkable equivalence with a theory defined on a single site lattice takes place in the $N \to \infty$ limit. The proof relies on the fact that both models satisfy the same Schwinger–Dyson equations [77], and have the same strong coupling limit. The equivalence, however, breaks down if the $[\mathrm{U}(1)]^N$

subgroup of symmetry of the reduced model is spontaneously broken. This is what seems to happen at a finite value of the coupling [78]. As a consequence, the equivalence holds only at strong coupling and fails below this critical coupling. This failure may be repaired by a quenching procedure of the reduced model. We will not dwell on this reduction in the present review, since it is under current investigation [79–85]. We will only remember from this discussion that, in its strong coupling phase, the U(∞) model is equivalent to a gauge model defined on a hypercube.

3. Strong coupling expansions

3.1. General features

3.1.1. Introduction

As noticed above, the bare coupling constant g_0^2 plays the role of the temperature β^{-1} in statistical models. At large couplings, it is natural to expand the Boltzmann weight in powers of β

$$\exp\{\beta S\} = \prod_p \exp\{\beta \chi(U_p)\} = \prod_p \left[1 + \beta \chi(U_p) + \frac{\beta^2}{2} \chi^2(U_p) + \cdots\right], \tag{3.1}$$

and to integrate term by term over the configuration variables; one may thus compute a series expansion of the partition function, or of various correlation functions. To each term of the expansion is attached a diagram made of those plaquettes retained in the product (3.1). Notice that, in this way, a given plaquette may appear an arbitrary number of times, corresponding to the power of $\chi(U_p)$ in the right hand side of (3.1). In practice, it is more convenient to resum part of this expansion in such a way that a given plaquette appears at most once. This is achieved by the character expansion, and this reduces the number of contributions to a given order in the expansion, making the geometric object simpler, and decoupling the geometric counting of configurations from the group theoretic factor. On the other hand, it is useful to derive rules for computing directly the expansions of physical (intensive) quantities: free energy per site, string tension, correlation lengths, ... This relies on cumulant expansions to be explained in subsection 3.1.6.

Most of these techniques are well known in the context of high temperature expansions of spin models [4]. We recall them briefly, and dwell mainly on features specific to lattice gauge models. We first review some rigorous results on the general properties of these expansions, before turning to the detailed technique.

3.1.2. Convergence of the strong coupling expansion

A well-known, but important property of high temperature (strong coupling) expansions is their finite radius of convergence. In view of the importance of this result, we recall briefly how it is derived and what it implies [35].

The idea is to derive an (overestimated) upper bound on the nth order of the expansion of some arbitrary local observable $\langle X \rangle$. "Local" here means that X involves only a finite set of links $\mathscr{L}$. A particular example is the internal energy $E = \langle S(U_p) \rangle$. This bound is achieved by estimating independently the contribution of a typical diagram with k plaquettes and the number of such diagrams, and still holds in the thermodynamic (infinite volume) limit.

The action per plaquette $\chi(U_p)$ is bounded, since it is a continuous function on a compact set.

Without loss of generality, we may shift it by a constant so that its lower bound is zero. Then, for β small enough and any U_p, we have

$$0 \leq \Omega(U_p) \equiv \exp\{\beta S(U_p)\} - 1 \leq C_1 \beta . \tag{3.2}$$

The constant C_1 only depends on the group and on the form of the action. Now consider a finite lattice Λ. The cluster expansion is obtained by expanding the product over plaquettes in

$$\langle X \rangle = Z_\Lambda^{-1} \int \prod_l \mathrm{D}U_l \prod_{p \subset \Lambda} [1 + \Omega(U_p)] X , \tag{3.3}$$

with

$$Z_\Lambda = \int \prod_l \mathrm{D}U_l \prod_{p \subset \Lambda} [1 + \Omega(U_p)] . \tag{3.4}$$

With each term of the expansion is associated a set of plaquettes D, made of two parts: D_1 is the union of connected components of D sharing some bonds with $\mathscr{L}$, and D_2 its complement in D: $\mathrm{D}_2 = \mathrm{D} \backslash \mathrm{D}_1$; then

$$\langle X \rangle = Z_\Lambda^{-1} \sum_{\mathrm{D} \subset \Lambda} \left\{ \int \prod_{l \subset \mathrm{D}_1 \cup \mathscr{L}} \mathrm{D}U_l \prod_{p \in \mathrm{D}_1} \Omega(U_p) X \right\} \left\{ \int \prod_{l \subset \mathrm{D}_2} \mathrm{D}U_l \prod_{p \in \mathrm{D}_2} \Omega(U_l) \right\} . \tag{3.5}$$

Now, for a given D_1, all D's are obtained by adding to D_1 any diagram unconnected to D_1, that is, any diagram contributing to the partition function $Z_{\Lambda \backslash \overline{\mathrm{D}_1 \cup \mathscr{L}}}$ for the reduced lattice where all plaquettes sharing a link with $\mathrm{D}_1 \cup \mathscr{L}$ have been removed.

$$\langle X \rangle = \sum_{\mathrm{D}_1} \left\{ \int \prod_{l \subset \mathrm{D}_1 \cup \mathscr{L}} \mathrm{D}U_l \prod_{p \in \mathrm{D}_1} \Omega(U_p) X \right\} Z_{\Lambda \backslash \overline{\mathrm{D}_1 \cup \mathscr{L}}} Z_\Lambda^{-1} \tag{3.6}$$

where the sum runs over all D_1, each connected component of which has at least one bond in common with $\mathscr{L}$.

i) Since there are less plaquettes in $\Lambda \backslash \overline{\mathrm{D}_1 \cup \mathscr{L}}$ than in Λ, and since $1 + \Omega \geq 1$, the last ratio in (3.6) is less than one and positive.

ii) Owing to the relation (3.2), the first factor is bounded by

$$\left| \sum_{\mathrm{D}_1} \int \prod_{l \subset \mathrm{D}_1 \cup \mathscr{L}} \mathrm{D}U_l \prod_{p \in \mathrm{D}_1} \Omega(U_p) X \right| \langle C_2 (C_1 \beta)^{|\mathrm{D}_1|} \rangle , \tag{3.7}$$

where $|D_1|$ is the number of plaquettes in D_1.

iii) The number of diagrams D_1 with a given $|\mathrm{D}_1| = k$ satisfies

$$n(k) \leq 2^{|\mathscr{L}|} [16(d-1)]^k . \tag{3.8}$$

Putting all factors together, we find

$$|\langle X \rangle| \leq C_2 \, 2^{|\mathscr{L}|} \sum_k [16(d-1) C_1 \beta]^k . \tag{3.9}$$

As nothing in this expression depends on the finite size of the lattice, we conclude that the strong coupling expansion has a finite radius of convergence, even in the infinite volume limit.

This convergence property has two important consequences, both in the strong coupling regime:

a) exponential clustering of correlation functions

b) area law of the Wilson loop.

a) Let $X^{(t)}$ denote the translate of a local observable X by a lattice vector $t = na$; there exists a constant $m > 0$ such that, for any t,

$$|\langle X_1 X_2^{(t)} \rangle_{\text{conn}}| = |\langle X_1 X_2^{(t)} \rangle - \langle X_1 \rangle \langle X_2 \rangle| < \text{Const. } e^{-mt} \tag{3.10}$$

and

$$am \geq -4 \ln \beta + \text{const.}$$

Indeed, only clusters connecting the supports of X_1 and $X_2^{(t)}$ contribute to the left hand side of (3.10); the dominant ones are tube-like diagrams with $4(n - \text{const.})$ plaquettes, and the contribution of higher orders may be bounded by the same method as above. The actual computation of the mass gap will be reconsidered in the following.

b) The area law of the Wilson loop follows from analogous considerations. Under suitable assumptions about the representation content of the action and of the test charge – the loop – (see [35] and discussion in subsection 3.1.5 below), the only contributing clusters have at least A plaquettes (A being the minimal area of the loop).

3.1.3. Character expansion

We use notations adapted to continuous Lie groups (see section 1.2) but the case of discrete groups is trivial to reinstate. The Haar measure on G satisfies

$$\mathrm{D}U = \mathrm{D}U^{-1} = \mathrm{D}(UV) \qquad \forall V \in \mathrm{G}, \tag{3.11}$$

and we have normalized it to unity ($\int \mathrm{D}U = 1$). The pure gauge action introduced in (1.9) uses class functions, i.e. functions satisfying

$$f(U) = f(VUV^{-1}) \qquad \forall V \in \mathrm{G}. \tag{3.12}$$

Irreducible characters (traces of irreducible representations) form a complete basis for these class functions. For the irreducible representation r, the $d_r \times d_r$ matrix elements are written $D^r_{\alpha\beta}(U)$, and the corresponding character is $\chi_r(U) = \Sigma_\alpha D^r_{\alpha\alpha}(U)$. From the orthogonality and completeness relations

$$\int \mathrm{D}U \, D^r_{\alpha\beta}(U) \, D^{s\,*}_{\gamma\delta}(U) = \delta_{rs}\delta_{\alpha\gamma}\delta_{\beta\delta}/d_r \tag{3.13}$$

$$\sum_{r,\alpha,\beta} d_r \, D^r_{\alpha\beta}(U) \, D^{r\,*}_{\alpha\beta}(V) = \delta(U, V), \tag{3.14}$$

it results that the irreducible characters form an orthonormal basis

$$\int \mathrm{D}U\, \chi_r(U)\, \chi_s^*(U) = \delta_{rs}\,, \tag{3.15}$$

$$\sum_r d_r\, \chi_r(UV^{-1}) = \delta(U, V)\,. \tag{3.16}$$

In particular, we have the useful formula

$$\int \mathrm{D}U\, \chi_r(U)\, \chi_s(U^{-1}V) = \delta_{rs} d_r^{-1} \chi_r(V)\,. \tag{3.17}$$

Any class function f may be decomposed in its Fourier components as

$$f(U) = f(VUV^{-1}) = \sum_r \chi_r(U)\, f_r\,,$$

with

$$f_r \equiv \int \mathrm{D}U\, \chi_r^*(U)\, f(U)\,. \tag{3.18}$$

In particular, the Boltzmann weight reads

$$\exp(\beta\chi) = \sum_r \tilde{\beta}_r\, \chi_r(U)\,. \tag{3.19}$$

Take as an example the group SU(2). A matrix U is parametrized as

$$U = \cos(\theta/2) + \mathrm{i}\boldsymbol{\sigma}\cdot\hat{n}\,\sin(\theta/2) \qquad (0 \le \theta < 4\pi) \tag{3.20}$$

in terms of Pauli matrices; the normalized Haar measure is

$$\mathrm{D}U = \sin^2(\theta/2)\frac{\mathrm{d}\theta}{2\pi}\frac{\mathrm{d}^2\hat{n}}{4\pi} \tag{3.21}$$

and the characters read

$$\chi_j(U) = \sin(j+\tfrac{1}{2})\theta/\sin(\theta/2), \qquad j = 0, \tfrac{1}{2}, 1, \ldots. \tag{3.22}$$

Now the expansions of the exponentiated actions $\exp\{\frac{1}{2}\beta\,\chi_{1/2}(U)\}$ and $\exp\{\frac{1}{3}\beta\,\chi_1(U)\}$ (SO(3) action, or part of the SU(2) fundamental–adjoint mixed action) read respectively in terms of modified Bessel functions

$$\exp\{\tfrac{1}{2}\beta\,\chi_{1/2}(U)\} = \exp\{\beta\cos(\theta/2)\} = \sum_j 2(2j+1)\frac{I_{2j+1}(\beta)}{\beta}\chi_j(U) \tag{3.23}$$

and

$$\exp\{\tfrac{1}{3}\beta\chi_1(U)\} = \exp\{\tfrac{1}{3}\beta(1+2\cos\theta)\} = \sum_{\text{integer } j} (I_j(\tfrac{2}{3}\beta) - I_{j+1}(\tfrac{2}{3}\beta))\exp(\tfrac{1}{3}\beta)\chi_j(U)\,. \tag{3.24}$$

Table 4 summarizes basic results on characters and expansion coefficients for some groups of particular importance and for Wilson's action. For more details on notations and derivations, the reader is referred to appendix A.

The character coefficients for the heat kernel action are trivial to identify from the definition of eq. (1.14)

$$\tilde{\beta}_r = \exp\{-C_r^{(2)}/N\beta\} \tag{3.25}$$

in terms of the quadratic Casimir operator in the representation r, $C_r^{(2)}$. For SU(2), $C_j^{(2)} = j(j+1)$, $j = 0, \frac{1}{2}, 1, \ldots$; for U(1), $C_l^{(2)} = l^2$, and for the representation (λ, μ) of SU(3), $C_{(\lambda,\mu)} = \lambda + \mu + (\lambda^2 + \lambda\mu + \mu^2)/3$. Casimir invariants for low-lying representations of SU(N) may be found in [86].

Let us list some relevant properties of the coefficients $\tilde{\beta}_r$ in (3.19):

i) for a real function as the exponentiated action, conjugate representations r and $\bar{r}$, which both appear in the sum if they are inequivalent, contribute equally

$$\tilde{\beta}_r = \tilde{\beta}_{\bar{r}}\,. \tag{3.26}$$

ii) Since

$$\tilde{\beta}_r = \int \chi_r^*(U)\exp\{\beta\chi(U)\} \tag{3.27}$$

Table 4
Character expansion coefficients for Wilson action

	$\tilde{\beta}_0$	(r) and $\chi(U)$	d_r	$\tilde{\beta}_r = \beta_r d_r \tilde{\beta}_0$
Z(2)	$\cosh\beta$	U	1	$\sinh\beta$
U(1)	$I_0(\beta)$	$n(\geq 0 \text{ or } <0)$, $\exp(in\phi)$	1	$I_n(\beta)$
SU(2)	$2I_1(\beta)/\beta$	j half-integer, $\dfrac{\sin(j+1/2)\theta}{\sin(\theta/2)}$	$2j+1$	$I_{2j}(\beta) - I_{2j+2}(\beta) = \dfrac{2(2j+1)}{\beta} I_{2j+1}(\beta)$
SO(3)	$e^{\beta/3}(I_0(2\beta/3) - I_1(2\beta/3))$	j integer, $\dfrac{\sin(j+1/2)\theta}{\sin(\theta/2)}$	$2j+1$	$e^{\beta/3}(I_j(2\beta/3) - I_{j+1}(2\beta/3))$
U(N)	$\det I_{i-j}(\beta/N)$	$l_1 \geq \cdots \geq l_N$ $(\geq 0 \text{ or } <0)$	$\prod_{i<j} \dfrac{l_i - l_j + j - i}{j - i}$	$\det I_{l_j - j + i}(\beta/N)$
	$1 \leq i, j \leq N$	$U = \operatorname{diag}(\exp(i\alpha_j))$		
		$\chi = \dfrac{\det(\exp i(l_j + N - j)\alpha_i)}{\det(\exp i(N-j)\alpha_i)}$		
SU(N)	$\sum_{n=-\infty}^{\infty} \det I_{i-j+n}(\beta/N)$	same as U(N), but $l_N = 0$	same	$\sum_{n=-\infty}^{\infty} \det I_{l_j - j + i + n}(\beta/N)$
U(∞)	$\exp(\beta/2N)^2 = \exp(N^2\beta_f^2)$	low-lying representations		$\dfrac{\sigma_{(r)}}{\lvert r\rvert!}\dfrac{\sigma_{(s)}}{\lvert s\rvert!}(N\beta_f)^n \exp(N^2\beta_f^2)$
or	(for $\beta_f \leq 1/2$)	are labelled by two Young		$(n \ll N)$
SU(∞)		tableaux		
		$\{r\} = l_1 \geq \cdots \geq l_r \geq 0$		
		$\{s\} = k_1 \geq \cdots \geq k_s \geq 0$		
		$\sum l_j + \sum k_j = n < N$		

is an integral over a compact domain, it is an entire function of β, and the ratios

$$\beta_r = d_r^{-1}\,\tilde{\beta}_r/\tilde{\beta}_0 \tag{3.28}$$

to be used in the following, are meromorphic functions of β, with no poles on the real axis on which $\tilde{\beta}_0$ cannot vanish. This remark will be important in the analysis of the series, where complex poles may cause trouble.

Also notice that these analyticity properties hold for any given group, but that taking the limit $N \to \infty$ of e.g. SU(N) or U(N) may generate singularities [16].

iii) An important property of this character expansion, relevant for strong coupling expansions, is that only a finite number of terms in (3.19) contribute to a finite order in β. In general,

$$\tilde{\beta}_r = \int \mathrm{D}U\,\chi_r^*(U)\exp\{\beta\chi(U)\} = \frac{\sigma_r}{r!}\,\beta^{\nu_r}(1+\mathcal{O}(\beta))\,, \tag{3.29}$$

Table 5
Expressions of β and β_r in terms of $\beta_1 \equiv t$

U(1)
$$\beta_n = \frac{t^n}{n!}(1+\mathrm{O}(t^2))$$
$$\beta_2 = \frac{1}{2}t^2+\frac{1}{6}t^4+\frac{5}{48}t^6+\frac{7}{90}t^8+\frac{541}{8640}t^{10}+\cdots$$
$$\beta_3 = \frac{1}{6}t^3+\frac{1}{8}t^5+\frac{13}{120}t^7+\frac{431}{4321}t^9+\cdots$$
$$\beta_4 = \frac{1}{24}t^4+\frac{1}{20}t^6+\frac{1}{18}t^8+\frac{367}{6048}t^{10}+\cdots$$
$$\cdots\cdots\cdots$$
$$\frac{\beta}{2} = t+\frac{1}{2}t^3+\frac{5}{12}t^5+\frac{19}{48}t^7+\frac{143}{360}t^9+\frac{1769}{4320}t^{11}+\frac{103543}{241920}t^{13}+\frac{2614699}{5806080}t^{15}+\frac{124086301}{261273600}t^{17}+\cdots$$

SU(2)
$$\beta_j = \frac{1}{(2j+1)!}(2t)^{2j}(1+\mathrm{O}(t^2))$$
$$\beta_1 = \frac{2}{3}t^2+\frac{2}{9}t^4+\frac{16}{135}t^6+\frac{8}{135}t^{10}+\frac{128}{8505}t^{12}+\cdots$$
$$\beta_{3/2} = \frac{1}{3}t^3+\frac{4}{15}t^5+\frac{2}{9}t^7+\frac{496}{2835}t^9+\cdots$$
$$\cdots\cdots\cdots$$
$$\frac{\beta}{4} = t+\frac{2}{3}t^3+\frac{2}{3}t^5+\frac{32}{45}t^7+\frac{308}{405}t^9+\frac{2264}{2835}t^{11}+\frac{34708}{42525}t^{13}+\frac{103216}{127575}t^{15}+\frac{55168}{70875}t^{17}+\cdots$$

SU(3)
$$\beta_{(\lambda,\mu)} = \frac{1}{d_{(\lambda,\mu)}}\frac{(3t)^{\lambda+\mu}}{(\lambda+\mu)!}\,T_{\lambda\mu}^{(\lambda+\mu)}(1+\mathrm{O}(t)) \qquad \text{(see appendix A)}$$
$$\beta_{(2,0)} = \beta_6 = \frac{3}{4}t^2-\frac{9}{16}t^4+\frac{297}{80}t^5-\frac{405}{32}t^6+\frac{6237}{160}t^7-\frac{70227}{640}t^8+\cdots$$
$$\beta_{(3,0)} = \beta_{10} = \frac{9}{20}t^3-\frac{81}{80}t^5+\frac{567}{80}t^6-\frac{2187}{80}t^7+\frac{1037367}{11200}t^8+\cdots$$
$$\beta_{(1,1)} = \beta_8 = \frac{9}{8}t^2-\frac{9}{8}t^3+\frac{81}{32}t^4-\frac{81}{16}t^5+\frac{1863}{160}t^6-\frac{2187}{80}t^7+\frac{43011}{640}t^8+\cdots$$
$$\beta_{(2,1)} = \beta_{15} = \frac{9}{10}t^3-\frac{63}{40}t^4+\frac{81}{20}t^5-\frac{81}{10}t^6+\frac{81}{5}t^7-\frac{89181}{3200}t^8+\cdots$$
$$\cdots\cdots\cdots$$
$$\frac{\beta}{8} = t-\frac{3}{2}t^2+\frac{9}{2}t^3-\frac{45}{4}t^4+\frac{243}{8}t^5-\frac{6237}{80}t^6+\frac{3159}{16}t^7-\frac{61479}{128}t^8+\frac{89667}{80}t^9-\frac{22076307}{8960}t^{10}+\frac{636417}{128}t^{11}-\frac{220994163}{25600}t^{12}+\frac{73555371}{7168}t^{13}-\cdots$$

where ν_r is the smallest integer so that $\sigma_r \equiv \int DU \chi_r^*(U) \chi^\nu(U) \neq 0$. In the previous cases of SU(2), with the action in the fundamental (resp. adjoint) representation, we have $\nu_j = 2j$ (resp. j), meaning that we need $2j$ (resp. j) times the spin 1/2 (resp. 1) representation to build up spin j. The successive terms of the expansion (3.29) may be worked out using the explicit expressions of $\tilde{\beta}_r$ in terms of modified Bessel functions, etc..... It is often useful to recast these expansions as series in $\beta_f \equiv t$, the fundamental character coefficient. This is done on table 5 for the groups U(1), SU(2), SU(3).

iv) The coefficients β_r defined in (3.28) are positive and vary between 0 and 1, as β grows from 0 to infinity. Indeed they vanish at $\beta = 0$ according to (3.29), and as $\beta \to \infty$, if the dominant saddle-point of the integral is at $U = 1$,

$$\tilde{\beta}_r = \text{const. } d_r \frac{\exp\{\beta \chi(\mathbb{1})\}}{\beta^{X/2}} (1 + \mathcal{O}(\beta^{-1})) \tag{3.30}$$

where X is the dimension of the Lie algebra (number of infinitesimal generators) of the gauge group.

Once the expansion (3.19) has been done for each plaquette, we can write the Boltzmann weight for a given configuration of links as

$$\exp(\beta S) = \prod_p \tilde{\beta}_0 \Big[1 + \sum_{r \neq 0} d_r \beta_r \chi_r(U_p) \Big] \tag{3.31}$$

in terms of $\tilde{\beta}_0$ and β_r. Changing the convention for the orientation of plaquettes changes U_p into $U_p^\dagger$, but does not affect the expansion, thanks to (3.26). The product in the right hand side may then be expanded. To each term is associated a diagram made of the plaquettes with a non-trivial representation. In this way, every plaquette appears at most once in each diagram, but decorated with a representation label. Clearly the integration over link variables of (3.28), (possibly in the presence of source terms for the computation of some correlation functions) leads to selection rules on the possible diagrams. We shall discuss them in the case of the partition function and of the Wilson loop. Let us first classify the geometric objects we are dealing with.

3.1.4. Some remarks on combinatorial topology

We examine now the topological properties of the relevant sets of plaquettes. Let $\mathscr{L}$ be such a set of plaquettes, connected or not, each plaquette appearing at most once. Let n_2 be the number of plaquettes, n_1 the number of distinct links bordering these plaquettes, and n_0 the number of endpoints of these links. A connected subset $\mathscr{L}'$ is homeomorphic to a simple surface (with possible boundaries) if each link of $\mathscr{L}'$ belongs to at most two plaquettes. Links belonging to a single plaquette form the boundary. Any set of plaquettes is therefore homeomorphic to a set of simple connected surfaces with boundaries (regular components). The boundaries are a set of simple arcs, each of which

– either borders a surface (true boundary),

– or is a singular line bordering n (≥ 3) surfaces (n-fold branch line).

All connected surfaces with boundaries may be classified according to

i) their orientability. Two neighbouring plaquettes are coherently oriented if they induce opposite orientations on their common link. A surface is orientable if there exists an orientation of all the plaquettes such that all neighbouring plaquettes are coherently oriented.

ii) the number b of holes, i.e. of simple closed contours making the boundary.

iii) the genus of the surface

$$g = 2 - (n_2 - n_1 + n_0 + b).$$

This number is related to Euler's characteristics and is a topological invariant.

Any orientable surface of genus g with b holes is homeomorphic to a sphere with $b + g$ holes, g of them being joined in pairs with first kind (orientability-conserving) handles. The sphere corresponds to $b = g = 0$, the torus to $b = 0$, $g = 2$, and the disk to $b = 1$, $g = 0$.

Any non-orientable surface of genus g with b holes is homeomorphic to a sphere with $b + g$ holes, the g last ones being closed using a Moebius band. The Moebius band corresponds to $b = 1$, $g = 1$, the projective plane to $b = 0$, $g = 1$ and the Klein bottle to $b = 0$, $g = 2$.

3.1.5. Classification of diagrams. Group theoretic factors

The partition function Z is obtained by integrating over the link variables U_l the expression (3.28). The following classification applies to Z, to its logarithm and its derivatives (internal energy, specific heat, ...).

From the orthogonality relation (3.15) with $r \neq 0$, $s = 0$ giving zero, it is clear that no true boundary occurs. All diagrams are closed, although they may of course have singular lines. Consider first the integration over an inner link of a regular component. A typical integration has the form

$$\int \mathrm{D}U_l \chi_r\Big(U_l\Big[\prod_{l'} U_{l'}\Big]\Big)\, \chi_s\Big(U_l^{-1}\Big[\prod_{l''} U_{l''}\Big]\Big) = \delta_{rs}\, d_r^{-1}\, \chi_r\Big(\Big[\prod_{l'} U_{l'}\Big]\Big[\prod_{l''} U_{l''}\Big]\Big). \tag{3.32}$$

Hence orthogonality forces all adjacent plaquettes with a coherent orientation to carry the same representation. The previous integral has to be used repeatedly, together with

$$\int \mathrm{D}U\, \chi_r(UU_1U^{-1}U_2) = d_r^{-1}\, \chi_r(U_1)\, \chi_r(U_2),$$

$$\int \mathrm{D}U\, \chi_r(UU_1UU_2) = d_r^{-1}\, \delta_{r\bar{r}}\, \chi_r(U_1U_2^{-1}). \tag{3.33}$$

The latter appears only for a non-orientable surface: we see that such a surface contributes only for self-conjugate representations $r \sim \bar{r}$.* If there are n_1' inner links with n_0' inner end-points (i.e. not belonging to the boundary), there are $n_1' - n_0' = n_1 - n_0$ integrals to carry out, and n_0' trivial ones. This yields a contribution

$$\beta_r^{n_2}\, d_r^{n_2 - n_1 + n_0} \prod_{\text{boundaries}} \chi_r(U_{\text{boundary}}) \tag{3.34}$$

where the product is to be replaced by one if there is no boundary. The power of d_r is $2 - g - b$ for a regular component. For example, diagrams with the topology of the sphere (resp. a torus) give contributions $\Sigma_{r \neq 0} \beta_r^{n_2} d_r^2$ (resp. $\Sigma\, \beta_r^{n_2}$). For more complicated diagrams, one is left with group in-

*For self-conjugate unitary representations, $D^r = VD^{r*}V^{\dagger}$ with V a unitary symmetric matrix (may be not unity!).

tegrations along the singular lines. For example, a diagram made of p disks of $m_1, \ldots, m_p$ plaquettes, soldered along a single closed contour (p-fold singular line) contributes

$$\Theta_{m_1,\ldots,m_p} = \sum_{r_1,\ldots,r_p \neq 0} \beta_{r_1}^{m_1} \cdots d_{r_1} \cdots N_{r_1,\ldots,r_p} \tag{3.35}$$

where

$$N_{r_1,\ldots,r_p} = \int \mathrm{D}U \chi_{r_1}(U) \cdots \chi_{r_p}(U) \tag{3.36}$$

is the number of times the trivial representation is contained in $r_1 \otimes \cdots \otimes r_p$.

In section 3.2, we list all such group theoretical factors associated with the various topologies involved in diagrams with sixteen or less plaquettes.

Let us turn to the computation of a Wilson loop. For a large loop of minimal area A (minimal number of plaquettes needed to fill the loop) and under some conditions outlined hereafter, the leading contribution to $W_r(\mathrm{C})$ is proportional to β_r^A. This area fall-off is one of the main virtues of the lattice formulation of gauge theories. We already mentioned that this means a (linear) confinement in the strong coupling regime (see remarks on the convergence and the reliability of strong coupling results in section 3.1.2). This result actually relies on two assumptions, namely

i) β_r does not vanish, or, more generally, there exist representations $s_1, s_2, \ldots$ such that $r \subset s_1 \otimes s_2 \otimes \ldots$, with $\beta_{s_1}, \beta_{s_2}, \ldots \neq 0$.

ii) there exists no set of representations such that r appears in the decomposition of the tensorial product $\otimes_i [s_i \otimes \bar{s}_i]$, $\beta_{s_i} \neq 0$.

If i) is not fulfilled, the Wilson loop in representation r is obviously a bad order parameter. A trivial example is provided by SU(2) with an action invariant under a transformation of the centre $U_p \to -U_p$, and $r = \frac{1}{2}$. If condition ii) is not fulfilled, e.g., if s exists such that $r \subset s \otimes \bar{s}$, it is possible to form tube-like diagrams around the loop C (see fig. 8), thus giving a perimeter fall-off to the Wilson loop. Physically it means that we have screening rather than confinement. The test charge r is screened by dynamical pairs s and $\bar{s}$ to form a group singlet. An example of such a case for SU(2) is given by $r = 1$, screened either by $1/2 \otimes 1/2$ or by $1 \otimes 1$. Comparison of the two types of contribution β^{L^2} versus $\beta^{16(L-1)}$ shows that a cross-over from area law to perimeter law occurs at $L \simeq 15$. If the gauge group has a non-trivial centre, a sufficient condition is that χ_r transforms non-trivially under transformations belonging to the centre. In particular, this is always the case for the fundamental representation of SU(N).

To go beyond the leading term β_r^A, we retain plaquettes not belonging to the minimal surface, or carrying a representation other than r in the expansion of $\exp\{\beta S\}$. Relevant diagrams now have C as a

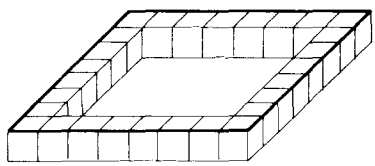


Fig. 8. Diagram contributing to the perimeter fall-off of the Wilson loop.

boundary and may be ordered according to the number of extra plaquettes, in addition to the minimal number A. As in the case of the partition function above, different group integrals correspond to different topologies. These coefficients for diagrams with twelve plaquettes or less are listed in section 3.2.1.

It must be clear now that these considerations may be extended to various observables: small closed loops, plaquette–plaquette correlation functions, 't Hooft loop. The plaquette–plaquette correlation function, for example, which is of particular physical interest since it yields the glueball mass, is built up from tube-like diagrams joining the two source plaquettes.

3.1.6. Cumulants, connected diagrams and extensivity

As noticed above, the direct evaluation of the expansion in powers of β is not very useful in the actual computation of the series. We review, however, this method which leads to some important results. The technique is related to the general theory of diagrammatic expansions. The statements will be illustrated in the three particular cases of the high temperature Ising model (I), perturbative field theory (F) and, of course, lattice pure gauge theory (G).

The action is split into a free part S_0 ((F): quadratic part; (I) and (G): 0) and an interaction part S_{int}. The generating functional is rewritten as

$$\begin{aligned} Z(J) &= \int \exp\{S(\phi) + J \cdot \phi\}\, \mathrm{D}\phi \\ &= \exp\{S_{\text{int}}(\partial/\partial J)\} \int \exp\{S_0(\phi) + J \cdot \phi\}\, \mathrm{D}\phi \\ &= \exp\{S_{\text{int}}(\partial/\partial J)\} \exp W_0(J) \\ &= [1 + S_{\text{int}}(\partial/\partial J) + \tfrac{1}{2} S_{\text{int}}^2(\partial/\partial J) + \cdots] \exp\{W_0^{(1)} J + W_0^{(2)} J^2/2 + \cdots\} \end{aligned} \tag{3.37}$$

where $W_0(J)$ is the connected generating functional of the free theory ((F): propagator $\int\int J(x)\, \Delta(x-y)\, J(y)\, \mathrm{d}x\, \mathrm{d}y$; (I) (resp. (G)): $\Sigma_x\, u(J_x)$, sum over the sites (resp. links) of the Fourier transform of the field measure). Each monomial of the bracket contains derivatives with respect to the sources J, from which arises a sum of terms when applied to the exponential $\exp W_0$. Each resulting term is interpreted as a diagram made of "vertices" joined to "sites".

– S_{int} is a sum of monomials $(1/k!) V^{(k)}_{i_1 i_2 \ldots i_k} \phi_{i_1} \ldots$, where $i_1, \ldots i_j$ denote all indices (location as well as internal indices). A "k-vertex" (drawn as a black dot from which k lines originate) is associated with every such monomial and contributes a factor $V^{(k)}_{i_1 i_2 \ldots i_k}$ ((F): usual vertices of Feynman diagrams; (I) S_{int} being quadratic, only 2-vertices exist and are usually represented by a line joining 2 neighbouring sites instead of a black dot; (G): S_{int} is quartic and the vertex is usually represented by a plaquette, although we occasionally draw it as a black dot).

– a "p-site" (drawn as an open dot from which p lines originate) is associated with a pth derivative of W_0: its contribution is the cumulant $W_0^{(p)}$. ((F): W_0 being quadratic, there are only 2-sites which are represented as a Feynman propagator line; (I) (resp. (G)): sites are located at the nodes (resp. bonds) of the lattice and contribute a factor $u^{(p)}(J_x)$ for the location x.)

The rule for computing the generating functional $Z(J)$ is the following: draw all possible diagrams with sites and vertices joined together by internal lines (external lines pertaining to a site are also allowed and contribute each a factor J); divide each contribution by the order of the symmetry group of the diagram. $Z(J)$ is the sum of all the contributions.

Due to the simplifications occurring in field theory (only 2-sites) and in statistical spin models (only 2-vertices), this expansion is used in practice. In lattice gauge theory, the method has to be refined because the number of diagrams increases drastically with the order, making things quite untractable. For instance, the reader may list all the diagrams of 4th order in β in the Z_2 gauge partition function of a one plaquette world (i.e., derive by this method the expansion of β_0 in terms of β); the result contains 16 topologically different diagrams and one finds among them spheres, torii, projective planes, Klein bottles, . . . ! Improvement of the method may consist in resumming sub-classes of diagrams. We will later present some of them ($1/d$ expansion). The character expansion can be regarded as such a reordering of the β series, grouping all vertices or sites pertaining to the same location.

The first refinement of the method is the restriction to connected diagrams. In this discussion of the cumulant expansion, connectivity refers to the abstract graph made of "sites" and "vertices", not to the geometrical object made (e.g. in lattice gauge theory) of links and plaquettes; in particular, two different "sites" can share the same lattice location, while belonging to different connected parts. It is clear from the diagrammatic rules that a disconnected diagram contributes exactly as the product of the contributions of its connected parts. Hence the connected generating functional $W(J) \equiv \ln Z(J)$ is the sum of contributions of all connected diagrams computed according to the preceding rule. In particular, in a lattice with periodic boundary conditions, translation invariance implies that the same diagram occurs $\mathcal{N}$ times, as long as its spatial extension does not allow it to close using the boundary conditions. The first occurrence of such an exception is the diagram consisting of all L^2 plaquettes in a given 2-plane, which generates only L^{d-2} different copies of itself instead of $\mathcal{N}$ by translation. Hence

$$Z = e^{\mathcal{N}F} + \mathcal{O}(L^{d-2}\beta^{L^2}) . \tag{3.38}$$

In contrast, in the character expansion, there is no simple definition of connectivity such that $\ln Z$ is the sum of connected diagrams. This is due to the fact that 2 plaquettes cannot share the same lattice location and therefore spurious contributions to Z generated in the exponentiation of the sum of connected diagrams have to be subtracted out. However, the property of eq. (3.38) proved using the cumulant formalism (namely that to any finite order in β, the expansion of Z exponentiates) is still true in the character expansion.

Münster [87] has used a cluster version of the character expansion and a moment-cumulant transformation to prove this result directly. His proof extends to large rectangular Wilson loops. It shows that, to any order smaller than the side of the loop

$$\ln W(\mathrm{C}) = -KA + \mu P + C \tag{3.39}$$

where P is the perimeter of the loop. The perimeter and constant terms reflect edge and corner effects around the loop, and can be disposed of by extending the loop to the boundary of a finite lattice and by taking periodic boundary conditions. In the latter case, we have simply

$$\ln W(\mathrm{C}) = -KA .$$

We use these results in the next subsection to derive rules for the free energy and the string tension in the character expansion formalism.

The cumulant expansion may also be applied to the large N limit of U(N) theories. This is relegated to appendix B.

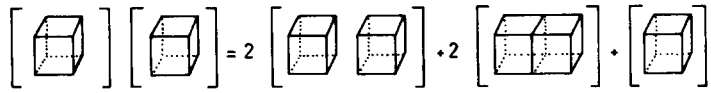
3.1.7. Geometric factors. Configuration number

It follows from the previous section that, to any finite order and for periodic boundary conditions, the series expansion for Z (resp. $W(C)$) exponentiates in the form $e^{\mathcal{N}F}$ (resp. e^{-KA}). From this follows a simple rule to derive the expansion of F (resp. K). In the diagrammatic expansion of Z (resp. of $W(C)$), it will be identified with the coefficient of the term linear in $\mathcal{N}$ (resp. in A, for $\mathcal{N}=0$). The number of inequivalent positions of a given diagram on the lattice is called its configuration number (c.n.), its linear part in $\mathcal{N}$ (resp. in A) is the reduced configuration number (r.c.n.). Configuration numbers of connected diagrams (for which c.n. $=\mathcal{N}$ r.c.n.) are easily computed; the only delicate point is a possible non-obvious symmetry which reduces the c.n. by a factor. For example, the torus diagram (16.2 of table 6) has a r.c.n. $\frac{1}{2}\times\frac{1}{2}d(d-1)\times\frac{1}{2}(d-2)(d-3)$, where the factors $\frac{1}{2}$ reflect the three symmetries between its axes $(1\leftrightarrow 2)$, $(3\leftrightarrow 4)$ and $(1,2)\leftrightarrow(3,4)$.

We recall that, in the character expansion, disconnected diagrams (i.e. made of pieces having no plaquette in common) contribute to F and/or K. Their r.c.n. are determined in a recursive way. Let D_1 and D_2 be two diagrams, $[D_1]$, $[D_2]$ their c.n. One may write the identity

$$[D_1]\,[D_2]=\sum_{D=D_1\cup D_2} n_D[D] \tag{3.40}$$

where the sum runs over all inequivalent diagrams D (connected or not) which may be split into $D_1\cup D_2$ in n_D different ways. If the r.c.n. of the connected diagrams in the right hand side of (3.40) have already been computed, keeping the term linear in $\mathcal{N}$ (resp. in A) gives an equation for the r.c.n. of the disconnected contributions (the left hand side has no linear contribution to $\mathcal{N}$ or A). For example,



$$\left[\mathcal{N}\frac{d(d-1)(d-2)}{6}\right]^2 = 2\left[\quad\right] + 2\,\mathcal{N}\frac{d(d-1)(d-2)(2d-5)}{2} + \mathcal{N}\frac{d(d-1)(d-2)}{6}\,.$$

Therefore the r.c.n. x of the contribution of two disconnected cubes to F is

$$x=-d(d-1)(d-2)(12d-29)/12\,.$$

Another example, quoted from the computation of the string tension, is displayed in fig. 9. In higher orders, it may happen that a given r.c.n. appears in several such equations, thus providing a useful cross-check of the counting procedure.

3.1.8. Computation of the glueball masses

The computation of the glueball masses (inverse correlation lengths) is more delicate. In section 3.1.2, it has been shown that, at strong coupling, connected correlation functions are bounded by a decreasing exponential. At large distance, one actually expects a behaviour

$$G(r)=\langle \mathcal{O}_1(0)\,\mathcal{O}_2(r)\rangle_{\text{conn}}$$

$$\underset{r\to\infty}{\sim} A_1(r)\,e^{-m_1(\hat{r})r}+A_2(r)\,e^{-m_2(\hat{r})r}+\cdots \tag{3.41}$$

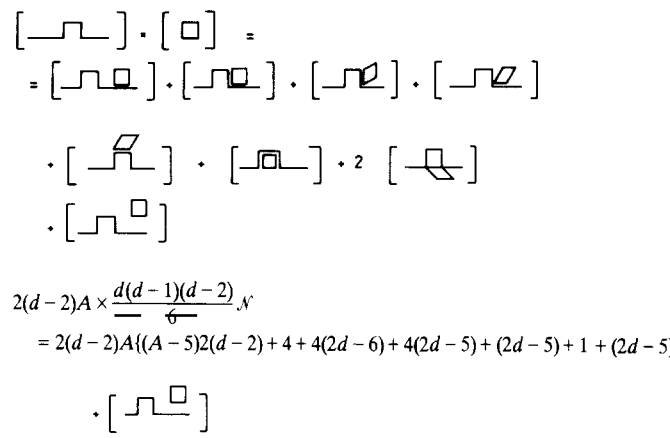


Fig. 9. Determination of the c.n. of a disconnected contribution to the string tension. Diagrams have been represented by their cross-section.

with several mass gaps $m_1(\hat{r}) < m_2(\hat{r}) < \cdots$, depending on the direction of observation $\hat{r}$. In a lattice gauge theory, the simplest choice for $\mathcal{O}_1$ and $\mathcal{O}_2$ is to take two distant plaquettes in the fundamental representation, in various relative orientations

$$G(r) = \langle \operatorname{tr} U_{p_1}(0) \operatorname{tr} U_{p_2}(r) \rangle - \langle \operatorname{tr} U_{p_1} \rangle \langle \operatorname{tr} U_{p_2} \rangle . \tag{3.42}$$

Only the smallest mass contributes to the leading asymptotic behaviour, but it may be that these masses are degenerate in the strong coupling limit; then the identification of the different masses $m_1, m_2, \ldots$ on the strong coupling expansion may be problematical.

This is a well-known problem in the context of spin models. For example, the first two orders of the connected correlation function of the two-dimensional Ising model at high temperature do not sum up as a single exponential (see fig. 10a):

$$\langle \sigma_0 \sigma_{\tau = na} \rangle_{\text{conn}} = t^n [1 + n(n+1)t^2 + \mathcal{O}(t^4)] \neq \mathrm{e}^{-nam} .$$

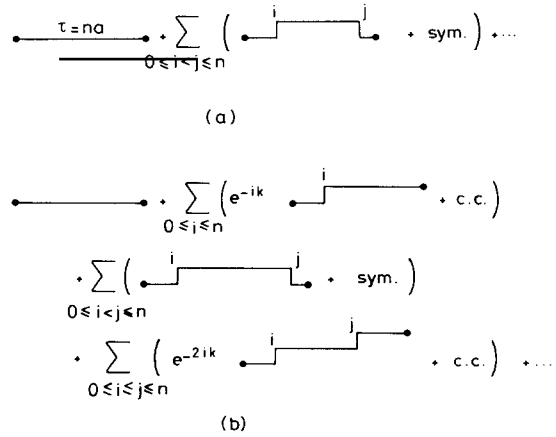


Fig. 10. Leading contributions to the spin–spin correlation function of the two-dimensional Ising model. a) $\langle \sigma_0 \sigma_{rr} \rangle$, b) $\Sigma_r \mathrm{e}^{-ikr} \langle \sigma_0 \sigma_{rr} \rangle$.

In such a case, where one looks at the asymptotic behaviour along an axis of the lattice, one may use the transfer matrix formalism [88, 89, 35]: e^{-m_1}, e^{-m_2} in (3.41) correspond to various eigenvalues of the transfer matrix, and, to disentangle them, it is suggested to project the correlation functions on eigenstates of the transfer matrix. As the latter commutes with "spatial" discrete translations and rotations (in the $(d-1)$-space orthogonal to the "time" axis $\mu = (1, 0, \ldots)$), one may consider linear combinations of definite spatial momentum p, and which transform under an irreducible representation of the $(d-1)$-cubic group [90]. Moreover, for gauge groups like SU(3) in which the orientation of plaquettes matters, there is a discrete symmetry (C-invariance) $U_l \to U_l^*$, $\mathrm{tr}\, U_p \to (\mathrm{tr}\, U_p)^* = \mathrm{tr}\, U_p^\dagger$, and states may be classified according to a C-parity $+1$ or -1. For example, the rotation invariant, C-even combination, which should give the lowest mass gap of four-dimensional lattice gauge theories, reads

$$\sum_{\boldsymbol{x}} \sum_{\substack{\text{6 spatial} \\ \text{directions} \\ \text{of } p_1 \text{ and } p_2}} \exp(-\mathrm{i}\boldsymbol{p}\boldsymbol{x}) \langle \mathrm{Re}\, \mathrm{tr}(U_{p_1}(\boldsymbol{0}, 0))\, \mathrm{Re}\, \mathrm{tr}(U_{p_2}(\boldsymbol{x}, \tau)) \rangle_{\mathrm{conn}} = G(\boldsymbol{p}, \tau)\,.$$

It implies a sum over the six plaquettes forming the cube centred at 0 (resp. at $\boldsymbol{x}$), over $\boldsymbol{x}$, and over internal orientations of the plaquettes p_1 and p_2, with a fixed time separation. At large τ, one has

$$G(\boldsymbol{p}, \tau) \sim \exp\{-E_0(\boldsymbol{p})\tau\} \tag{3.43}$$

and the mass gap $m_0 = E_0(\boldsymbol{0})$ [91] and the dispersion relation $E_0(\boldsymbol{p})$ [92] may be computed by strong coupling expansions. It is known that such a method solves the above-mentioned problem for the d-dimensional Ising model. For example, if $d = 2$ (see fig. 10b),

$$\begin{aligned} G(\boldsymbol{p}, \tau) &= \sum_{\boldsymbol{x}} \exp(-\mathrm{i}\boldsymbol{p}\boldsymbol{x})\, \langle \sigma_{0,0} \sigma_{x,\tau} \rangle_{\mathrm{conn}} \\ &= [1 + 2(n+1)t \cos p + 2\{(n+1)^2 \cos^2 p - (n+1)\sin^2 p\}t^2 + \mathcal{O}(t^3)]t^n \\ &= t^n \exp\{2(n+1)\,(t \cos p - t^2 \sin^2 p + \mathcal{O}(t^3))\} \end{aligned} \tag{3.44}$$

whence

$$am_0 = -\ln t - 2t + \mathcal{O}(t^3)\,, \tag{3.45}$$

$$a\,E_0(p) = -\ln t - 2(t \cos p - t^2 \sin^2 p) + \mathcal{O}(t^3)\,,$$

in agreement with the exact results [93–95]

$$am_0 = -\ln[t(1+t)/(1-t)] \tag{3.46}$$

$$a\,E_0(p) = \mathrm{Arg}\cosh(\cosh m_0 + 1 - \cos p)\,.$$

Moreover, with appropriate skew-periodic boundary conditions [94], the term $(n+1)$ in the exponent of (3.44) is changed into n, and a simple prescription to identify $E_0(p)$ is to take the term linear in n in the expansion.

The rotation invariant plaquette–plaquette correlation function has also a strong coupling series that exponentiates, because the next contributions are of a much higher order. Typically,

$$G(\mathbf{0}, \tau = an) = t^{4n}\,(1 + \mathcal{O}(t)) + t^{8n}\,(1 + \mathcal{O}(t)) + \cdots \tag{3.47}$$

where the term t^{8n} comes from higher representations on the tube joining p_1 to p_2. Similarly, other linear combinations of plaquette–plaquette correlation functions $\langle \mathrm{tr}\, U_{p_1}\, \mathrm{tr}\, U_{p_2}\rangle_{\mathrm{conn}}$ or of more complicated loops may be constructed, transforming under other representations of the cubic group and odd or even under C-parity, and give the mass gaps in other channels [96, 97].

One may also look at the exponential decay of correlations in directions which differ from the lattice axes. Then the previous method is more delicate to set up, and an alternative way to extract the corresponding mass gaps from strong coupling series has been proposed [98–100, 96]. One constructs the $\frac{1}{2}d(d-1)\times\frac{1}{2}d(d-1)$ matrix of correlations between all possible plaquette directions

$$G_{\mu\nu,\rho\sigma}(\boldsymbol{r}) = \langle \mathrm{tr}\, U_{\mu\nu}(\mathbf{0})\, \mathrm{tr}\, U_{\rho\sigma}(\boldsymbol{r})\rangle_{\mathrm{conn}} \tag{3.48}$$

where Re or Im is understood in front of tr U, depending on whether $C = +1$ or -1. The Fourier transform of $G(\boldsymbol{r})$ is denoted $\tilde{G}(\boldsymbol{k})$, with $\boldsymbol{k}$ now a d-vector

$$\tilde{G}_{\mu\nu,\rho\sigma}(\boldsymbol{k}) = \sum_{\boldsymbol{r}} \exp(-\mathrm{i}\boldsymbol{k}\boldsymbol{r})\, G_{\mu\nu,\rho\sigma}(\boldsymbol{r})\,. \tag{3.49}$$

The large r behaviour of $G(r)$ is dominated by singularities of $\tilde{G}(\boldsymbol{k})$ at $\boldsymbol{\kappa}$, i.e. $\det \tilde{G}^{-1}(\boldsymbol{\kappa}) = 0$. The corresponding eigenvector of G behaves as $\exp(\mathrm{i}\boldsymbol{\kappa}\cdot\boldsymbol{r}) = \exp(-mr)$, with $m = -\mathrm{i}\boldsymbol{\kappa}\cdot\hat{r}$.

This method is well adapted to strong coupling calculations. To leading order, $\tilde{G}_{\mu\nu,\rho\sigma}$ may be seen as describing the random motion of a plaquette (fig. 11), and

$$\tilde{G}^0(\boldsymbol{k}) = (1 - t^4\, M(\boldsymbol{k}))^{-1} \tag{3.50}$$

where the matrix M describes a unit step of the random walk:

$$M_{\mu\nu,\rho\sigma} = M_{\rho\sigma,\mu\nu} = C M_{\nu\mu,\rho\sigma} \qquad C = \pm 1$$

$$M_{\mu\nu,\mu\nu} = 2 \sum_{\rho\neq\mu,\nu} \cos k_\rho$$

$$M_{\mu\nu,\mu\rho} = \begin{cases} 4\cos(k_\nu/2)\cos(k_\rho/2) & \text{if } C = +1 \\ 4\sin(k_\nu/2)\sin(k_\rho/2) & \text{if } C = -1\,. \end{cases}$$

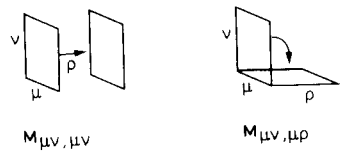


Fig. 11. Random walk of a plaquette.

To that order, the relevant poles have

$$e^{-m} = e^{i\kappa\hat{r}} = A(\hat{r})\, t^{4\sum \hat{r}_\mu} ,$$

hence

$$m = -4\left(\sum \hat{r}_\mu\right) \ln t - \ln A(\hat{r}) \tag{3.51}$$

where the constant A measures the entropy of the shortest path. In four dimensions for example, along an axis, $A = 1$ and

$$m = -4 \ln t + \cdots ,$$

whereas along the diagonal ($r = (1/2, \ldots, 1/2)$), $A = 6^2$

$$m = -8 \ln t - 2 \ln 6 + \cdots .$$

Higher order corrections may be incorporated into this scheme. They correspond to "self-energy" insertions Π in the propagator G, i.e. to inclusions of walls inside the tube or to decorations or deformations of the tube; $\tilde{G}$ takes the form

$$\tilde{G}(\boldsymbol{k}) = (1 - t^4 M - \Pi)^{-1} . \tag{3.52}$$

In that way, the calculation up to order β^4 has been carried out for the on-axis and diagonal glueball masses; on-axis results agree with those obtained through the method presented earlier in this section.

3.1.9. Computation steps and cross-checks

In practice, a strong coupling computation will consist of several successive steps: listing all topologically distinct diagrams, connected or not, computing their group theoretical factors (as a function of the β_r's) and their reduced configuration number, and re-expressing the result as series in the desired variable (β, or $t = \beta_t, \ldots$).

The first step, enumeration of diagrams, is the most tedious and delicate. It can be done "by hand", with obvious limitations (there are 26 distinct diagrams with 16 plaquettes or less for the free energy (see table 6), about 300 with 14 plaquettes or less for the string tension). This enumeration "by hand" suffers also from a difficulty of notations: how to draw diagrams in four (or more) dimensions in a clear and unambiguous way. As an alternative to the standard drawings, used for example in table 6, one may use a description by contour lines, valid for an arbitrary regular component. For definiteness, take a regular closed surface; each link is common to two and only two plaquettes. Starting from the centre $x^{(0)}$ of some plaquette μ, ν, follow the path $x_\mu = x_\mu^{(0)}$ on the surface. It goes across plaquettes of directions $\nu, \rho, \ldots$, until it closes back at $x^{(0)}$. This is represented by a closed oriented curve labelled by μ, crossed by closed curves indexed by $\nu, \rho, \ldots$. The orientation convention is that crossing the ν-curve oriented from right to left corresponds to a shift in the direction $+\nu$. This procedure is iterated for each centre of each plaquette, until every plaquette has been traversed twice. The resulting network of curves is a faithful description of the regular surface (see for example fig. 12); the original surface can be

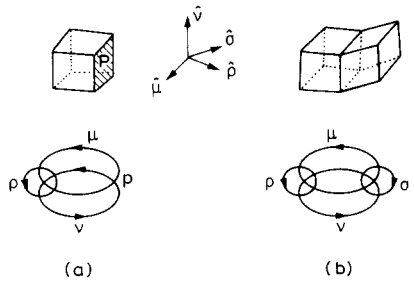


Fig. 12. Representation of diagrams by the system of contour lines. The number of intersections counts the number of plaquettes. Going from a) to b) corresponds to pulling plaquette p into the σ-direction, which is represented by an additional closed curve, labelled by σ and encircling p.

reconstructed (up to a displacement on the lattice) from this picture. Therefore this representation may be used for our book-keeping. Similarly, all diagrams with the topology of a disk contributing to the string tension may be represented in that way. Notice that this representation might also be used to generate distinct diagrams, if we were able to write all the constraints and selection rules on these patterns.

It may be thought that it is relatively easy to generate all diagrams by computer, using for instance successive accretions of three-dimensional cubes. However, in the simplest algorithms, the same diagram is often generated a large number of times; to avoid double counting requires an elaborate pattern recognition program and reduces the efficiency in a dramatic way. More elaborate algorithms can be designed [101], but are difficult to set up because the number of special cases grows rapidly. Using such a program K. Wilson [102] has been able to generate rather long series (22nd order) for the free energy of some four-dimensional pure gauge theories.

The computation of the group theoretical factors and of the geometrical weights (reduced configuration number) has been explained in the two previous subsections. Finally, the manipulations on the resulting series (substitution of one variable for another one, ...) are the ideal field of application of symbolic computer programs. In this review, we shall mainly present expansions in the original inverse coupling β, best suited for a direct comparison with Monte-Carlo data, or in the first character parameter $\beta_f \equiv t$, more appropriate for comparing series pertaining to different groups.

At the end of such a laborious computation, what are the possible cross-checks? If the calculation has been carried out in a general way, valid for all groups and dimensions, it must pass a few tests:

– the three-dimensional Z_2 gauge theory is, by duality, equivalent to the three-dimensional Ising model. Hence, strong coupling expansions of observables in the former must coincide with low temperature series of dual observables in the latter. For example, free energies are related by eq. (2.19). The gauge string tension is identical to the Ising surface tension, while plaquette–plaquette and spin–spin correlation lengths can be identified. Rather long low temperature series exist in the literature for the Ising model.

In four dimensions, the Z_2 and Z_3 (and, in the generalized sense explained in section 2.3.3, the Z_n) models are self-dual. This means that strong coupling expansions may also be regarded as low temperature expansions of the dual quantities. For example, the low temperature expansion of the Z_2 model starts from a ground state where all links are +1, up to a gauge transformation, and flips successively one, two, ... links. To any configuration of overturned spins corresponds by duality a permissible strong coupling diagram of the free energy, the plaquettes of which are dual to the frustrated plaquettes of the former. However, in the weak coupling expansion, some care has to be exercised, since gauge equivalent configurations have to appear only once. For example, a configuration of 4 flipped links emanating from the same site is equivalent to the complementary configuration of

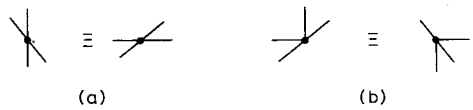


Fig. 13. Examples of low-temperature configurations of overturned links in four dimensions that may be drawn in two equivalent ways. a) is dual to 16.2 of table 6, b) to 14.3.

links incident on that site (fig. 13). This equivalence corresponds to the symmetry factor 1/2 attached to such a diagram. In short, duality is useful to visualize diagrams in an alternative way and to check the configuration numbers of four-dimensional theories.

– for large dimension d and $d\beta^4$ of order one, lattice gauge theories become exactly solvable (see section 3.3). It is moreover possible to compute a few correction terms which are suppressed by powers of $\beta \sim d^{-1/4}$. It is then possible to expand these exact expressions in powers of β, and to check in this way the terms of largest power of d for a given order in β.

– for large N and $\bar{\beta} = \beta/2N^2$ fixed, the U(N) or SU(N) gauge theory goes to the limit mentioned above. In this limit, the free energy per site is proportional to N^2, while mass scales as the string tension or the mass gap reach finite limits. However, in the character expansion used here, the existence of this limit results from cancellations [103, 104]; individual diagrams contributing to F may go as N^4 or higher powers of N. This provides an additional cross-check on the weights of diagrams.

3.2. Series

3.2.1. Free energy

The free energy has been derived to sixteenth order (diagrams up to 16 plaquettes) for general gauge groups and arbitrary dimension d, in terms of the character coefficients β_r. For a specific form of the action $\beta S(U_p)$, the latter may be computed as functions of β_f, and the expansion re-expressed in powers of β. For an action of the simple form $\beta \chi_r(U_p)$, it is more convenient to choose β_r as the expansion variable. The resulting expansions are somewhat more universal. Indeed the lowest order diagrams are only made of plaquettes in the representation r, and the first coefficients are thus group independent. Notice that β_r has a simple interpretation; it is the internal energy (plaquette energy) of the one-plaquette model with the same action.

Let us first present the general expression of the free energy per site F to 16th order. Fig. 14 displays a hopefully suggestive, though not faithful representation of the coefficients by means of a cross-section of a typical diagram. We define

$$\text{a) } S_n = \sum_{r \neq 0} \beta_r^n \, d_r^2 \qquad \text{(sphere)} \tag{3.53}$$

$$\text{b) } T_n = \sum_{r \neq 0} \beta_r^n \qquad \text{(torus)} \tag{3.54}$$

$$\text{c) } \Theta_{n_1, \ldots, n_p} = \sum_{r_1, \ldots, r_p \neq 0} \beta_{r_1}^{n_1} \cdots d_{r_1} \cdots N_{r_1, \ldots, r_p} \tag{3.55}$$

with

$$N_{r_1, \ldots, r_p} = \int \mathrm{D}U \chi_{r_1}(U) \cdots \chi_{r_p}(U)$$

Table 6
Table of diagrams with up to 16 plaquettes contributing to the free energy. In ambiguous cases, plaquettes are shaded to indicate that they have been selected, and broken lines are drawn for the eye's help

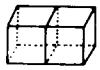
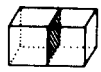
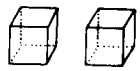
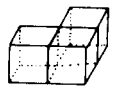
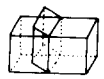
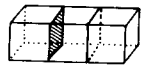
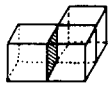
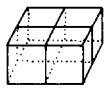
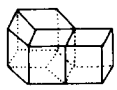
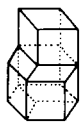
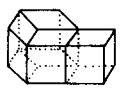
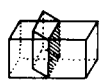
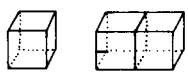
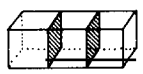
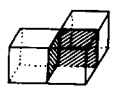
Number of plaquettes and diagram number	Diagram	r.n.c.	Group theoretic factor
6.1		$\frac{1}{6}d(d-1)(d-2)$	S_6
10.1		$\frac{1}{2}d(d-1)(d-2)(2d-5)$	S_{10}
11.1		$\frac{1}{2}d(d-1)(d-2)(2d-5)$	θ_{551}
12.1		$-\frac{1}{12}d(d-1)(d-2)(12d-29)$	S_6^2
12.2		$\frac{4}{3}d(d-1)(d-2)(d-3)$	S_{12}
12.3		$\frac{1}{3}d(d-1)(d-2)(d-3)$	S_{12}
14.1		$\frac{1}{2}d(d-1)(d-2)(2d-5)^2$	S_{14}
14.2		$2d(d-1)(d-2)(4d^2-22d+31)$	S_{14}
14.3		$d(d-1)(d-2)(d-3)$	S_{14}
14.4		$4d(d-1)(d-2)(d-3)$	θ_{842}
15.1		$\frac{1}{3}d(d-1)(d-2)(d-3)(2d-5)$	θ_{555}
15.2		$\frac{1}{3}d(d-1)(d-2)(d-3)$	θ_{933}
15.3		$\frac{4}{3}d(d-1)(d-2)(d-3)$	ϕ_{444}^{111}
15.4		$d(d-1)(d-2)(2d-5)^2$	θ_{951}

Table 6 (continued)

Number of plaquettes and diagram number	Diagram	r.n.c.	Group theoretic factor
15.5		$4d(d-1)(d-2)(4d^2-22d+51)$	Θ_{951}
16.1		$\frac{1}{2}d(d-1)(d-2)(4d^2-24d+37)$	S_{16}
16.2		$\frac{1}{8}d(d-1)(d-2)(d-3)$	T_{16}
16.3		$8d(d-1)(d-2)(d-3)^2$	S_{16}
16.4		$16d(d-1)(d-2)(d-3)^2$	S_{16}
16.5		$16d(d-1)(d-2)(d-3)^2$	S_{16}
16.6		$\frac{1}{3}d(d-1)(d-2)(d-3)(2d-5)$	Θ_{5551}
16.7		$-d(d-1)(d-2)(22d^2-113d+147)$	S_6S_{10}
16.8		$2d(d-1)(d-2)(d-3)$	Θ_{664}
16.9		$\frac{1}{2}d(d-1)(d-2)(2d-5)^2$	ζ_{51451}
16.10		$2d(d-1)(d-2)(4d^2-22d+31)$	ζ_{51451}
16.11		$4d(d-1)(d-2)(d-3)$	$\Theta_{10,4,2}$

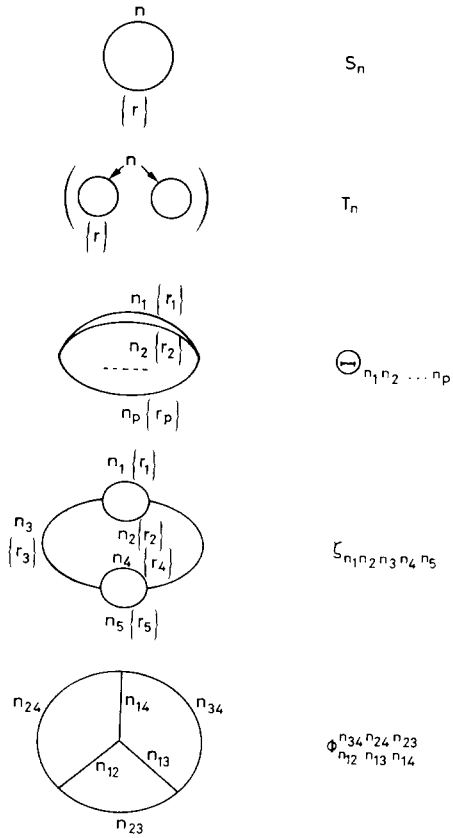


Fig. 14. Group theoretical factors representation for diagrams of the free energy.

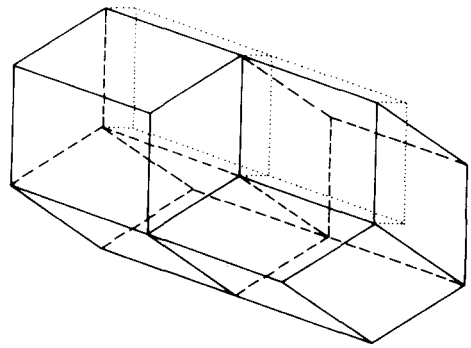


Fig. 15. The lowest order non-orientable diagram occurring to order 24: topology of a projective plane, represented by its three-dimensional projection. To avoid a four-fold singular line, two plaquettes have been shifted in a new direction, as indicated by the dotted lines.

$$\text{d)}\ \zeta_{n_1 n_2 n_3 n_4 n_5} = \sum \beta_{r_1}^{n_1} \cdots d_{r_1} d_{r_2} d_{r_4} d_{r_5} N_{r_1 r_2 r_3} N_{r_3 r_4 r_5} \tag{3.56}$$

(diagrams with two 3-fold singular lines each bounding two disks with n_1, n_2 (resp. n_4, n_5) plaquettes, and joined with an annulus of n_3 plaquettes),

$$\text{e)}\ \phi^{n_{34} n_{24} n_{23}}_{n_{12} n_{13} n_{14}} = \sum \prod_{i<j} (d_{r_{ij}} \beta_{r_{ij}}^{n_{ij}}) \int \prod_i \mathrm{D}U_i \prod_{i<j} \chi_{r_{ij}}(U_i U_j^{-1}) \tag{3.57}$$

(diagrams with four 3-fold singular lines joining two points A and B, and with six disks bounded by pairs of these lines).

In these expressions, the n's denote the number of plaquettes in each regular component, and the summation runs over non-trivial representations. For the sake of completeness, all diagrams with 16 plaquettes or less have been displayed in table 6. No non-orientable diagram contributes to this order (the first one occurs to order 24 and has the topology of a projective plane: fig. 15).

In terms of these coefficients, the free energy per site in d dimensions reads

$$F = \tfrac{1}{2}d(d-1)\{\ln\tilde{\beta}_0 + (d-2)[\tfrac{1}{3}S_6 + (2d-5)S_{10} + \tfrac{10}{3}(d-3)S_{12} + (20d^2-106d+143)S_{14}$$
$$+(84d^2-504d+757)S_{16} + \tfrac{1}{4}(d-3)T_{16} - (2d-29/6)S_6^2 - (44d^2-226d+294)S_6S_{10}$$
$$+(2d-5)\Theta_{551} + \tfrac{2}{3}(d-3)(2d-5)(\Theta_{555}+\Theta_{5551}) + \tfrac{8}{3}(d-3)(3\Theta_{842}+3\Theta_{1042}+2\Theta_{933}+\tfrac{3}{2}\Theta_{664}+\phi_{444}^{111})$$
$$+(20d^2-108d+149)(2\Theta_{951}+\zeta_{51415})+\cdots]\}. \tag{3.58}$$

For the specific form of the Wilson's action, one derives the expansions of F to order t^{16} or β^{16}. In table 7, we have only given results for some groups of interest. We think that expansions in t are well behaved and useful in the search of singularities, especially for "even" groups. Expansions in β, on the other hand, may be directly confronted to Monte-Carlo data. As all of them have already appeared in the literature, we only give references and correct a mistake to order β^{16} for the case of Z_3 and SU(3). (In the published series of Z_3 and SU(3), diagram 16.8 of table 6 had been omitted. We just give its additional contribution, using the same conventions as the original reference.)

Z_2	[7] (cf. erratum)
Z_3	[7] (erratum, add $\Delta F = 4d(d-1)(d-2)(d-3)(\beta/2)^{16}$)
U(1)	[7] (cf. erratum)
SU(2)	[7] (cf. erratum)
SU(3)	[7] (erratum, add $\Delta F = 108d(d-1)(d-2)(d-3)(\beta/3)^{16}$)
SU(4)	[105]
SU(5)	[106]
SU(6)	[105]
$SU(N)/Z_N$	[107]
SO(3)	[108]

For special values of d, longer expansions may be available. This is of course the case of the Z_2

Table 7
Strong coupling coefficients f_n for the free energy in terms of the first character coefficient t for Wilson's action $F = \frac{1}{2}d(d-1)\{\ln\beta_0 + (d-2)\sum f_n t^n\}$

	n						
G	6	8	10	12	14	16	Remarks
Z_2	$\frac{1}{3}$	0	$2d-5$	$\frac{2}{3}\left(2d-\frac{31}{4}\right)$	$20d^2-106d+143$	$\frac{124}{3}d^2-\frac{3421}{12}d+\frac{1889}{4}$	longer series exist at $d=3,4$
U(1)	$\frac{2}{3}$	0	$2(2d-5)$	$\frac{1}{3}\left(2d-\frac{543}{32}\right)$	$2\left(20d^2-\frac{317}{3}d+\frac{13649}{96}\right)$	$40d^2-\frac{2153}{6}d+\frac{784355}{1152}$	longer series exist at $d=4$
SU(2)	$\frac{4}{3}$	0	$4(2d-5)$	$-\frac{4}{3}\left(2d+\frac{146}{81}\right)$	$4\left(20d^2-\frac{314}{3}d+\frac{33971}{243}\right)$	$-\frac{16}{3}d^2-\frac{503147}{1620}d+\frac{12787481}{14580}$	longer series exist at $d=4$
U(∞) or SU(∞)	$\frac{2}{3}N^2$	0	$2(2d-5)N^2$	$\frac{2}{3}(10d-23)N^2$	$2(20d^2-106d+143)N^2$	$2(84d^2-432d+569)N^2$	only the leading term in N^2 has been written

Table 8
The coefficients f_n for the four-dimensional free energy, computed by Wilson

	n								
G	6	8	10	12	14	16	18	20	22
Z_2	$\frac{1}{3}$	0	3	$\frac{1}{6}$	39	$\frac{-27}{4}$	$\frac{5533}{9}$	$\frac{-345}{2}$	11017
U(1)	$\frac{2}{3}$	0	6	$\frac{-287}{96}$	$\frac{3793}{48}$	$\frac{-131869}{1152}$	$\frac{1788585347}{1399680}$	-2974.75173	24390.66152
SU(2)	$\frac{4}{3}$	0	12	$\frac{-3176}{243}$	$\frac{39980}{243}$	$\frac{-6569971}{14580}$	$\frac{95936872}{32805}$	$\frac{-79327880}{6561}$	65379.87679

theory in $d = 3$ dimensions, where duality with the Ising model enables one to write

$$F_{Z_2\,\text{gauge}}(\beta) = 3 \ln \cosh \beta + \sum b_r^{(0)} t^{2(r+3)}$$

in terms of the "zero field coefficients" given by Domb [4].

In four dimensions, Wilson [102] has been able to compute three additional terms in the expansion of the Z_2, U(1) and SU(2) internal energy. The corresponding results for the free energy appear in table 8. For other forms of the action, few explicit series are available. For the SU(2) mixed action, however, a double expansion in $\beta\gamma_f$ and $\beta\gamma_a$ to 16th order has been derived [108].

Finally, the expansion of a simplicial lattice in d dimensions has been recently derived [31]. We recall that this lattice, which generalizes the triangular lattice in two dimensions and the face-centred cubic lattice in three dimensions, may be considered as the cross-section of a $(d+1)$-hypercubic lattice by the plane $x_1 + \cdots + x_{d+1} = 0$. The merit of this lattice is to have large coordination numbers (a site has $d(d+1)$ neighbours).

3.2.2. Wilson loops and string tension

The quantity of interest is $W_r(C)$ defined in (1.16). The curve C here is rectangular. In section 3.1.5, it has been explained under which conditions an area law was expected for $W_r(C)$. Then, to leading order,

$$W_r(C) = d_r \beta_r^A + \cdots \tag{3.59}$$

hence the corresponding string tension K_r behaves as

$$K_r = -\ln \beta_r + \mathcal{O}(\beta)\,. \tag{3.60}$$

We now turn to the computation of higher order corrections. Earlier works on the subject can be found in [109–111]; we discuss here the most recent results [87, 112, 113]. The corrections may be classified according to the number of extra plaquettes added to the minimal surface. The topologies of these diagrams correspond to those of diagrams for the free energy drilled along the contour C. In the following, we restrict ourselves to Wilson loops in the fundamental representation $r = f$. Up to 12th order, all the topologies are obtained in that way from the closed diagrams symbolized on fig. 14. This results in the following list (see fig. 16)

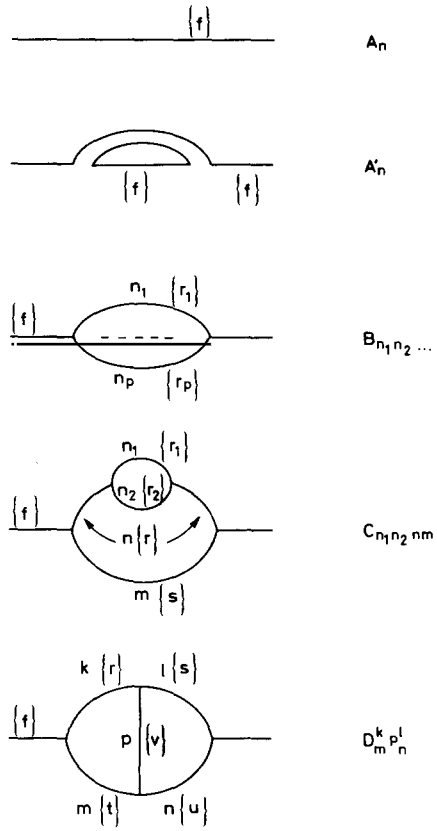


Fig. 16. Representation of group theoretic factors for the Wilson loop.

a) a disk with $A+n$ plaquettes $A_n = \beta_f^n$.
b) a disk with a handle, $A+n$ plaquettes: $A'_n = \beta_f^n d_f^{-2}$.
c) a hole in the topology Θ of fig. 14,

$$B_{n_1,\ldots} = d_f^{-1} \sum (d_{r_1}\beta_{r_1}^{n_1})(d_{r_2}\beta_{r_2}^{n_2})\ldots$$

d) a hole in the topology ζ of fig. 14,

$$C_{n_1 n_2 nm} = d_f^{-1} \sum N_{frs} N_{rr_1r_2} d_{r_1} d_{r_2} d_s \beta_{r_1}^{n_1} \beta_{r_2}^{n_2} \beta_r^n \beta_s^m \,.$$

e) a hole in the topology ϕ

$$D^m_k{}^n_l = d_f^{-1} \sum d_r d_s d_t d_u d_v \beta_r^k \beta_s^l \beta_t^m \beta_u^n \beta_v^p \int \mathrm{D}R\,\mathrm{D}S\,\mathrm{D}T\,\mathrm{D}U \chi_r(RT^{-1})$$
$$\times \chi_s(TS^{-1})\chi_t(SR^{-1})\chi_t(RU^{-1})\chi_u(SU^{-1})\chi_v(TU^{-1}) \,.$$

Using these notations and those of the preceding section, the string tension for arbitrary gauge groups and dimensions reads

$$K_f = -\ln t - 2(d-2)\sum \hat{K}_n , \tag{3.61}$$

with $\hat{K}_n$, the contributions of diagrams with $A+n$ plaquettes, given by

$$\begin{aligned}
\hat{K}_4 &= t^4 \\
\hat{K}_5 &= B_{51}t^{-1} \\
\hat{K}_6 &= 2t^6 - A_6 \\
\hat{K}_7 &= 0 \\
\hat{K}_8 &= (9d-22)t^8 + 2B_{82}t^{-2} + 4t^2 B_{42} \\
\hat{K}_9 &= (10d-29)B_{91}t^{-1} + (d-5/2)B_{55}t^{-1} + 2D^4_{11}t^{-2} + (8d-29)t^3 B_{51} \\
\hat{K}_{10} &= (48d-132)t^{10} - (11d-59/2)A_{10} + (d-5/2)B_{551}t^{-1} + (10d-29)(C_{1451}t^{-1} - t^4 A_6) \\
&\quad - dB^2_{51}t^{-2} + 12t^4 B_{42} \\
\hat{K}_{11} &= (24d-88)t^5 B_{51} - (11d-59/2)A_{551} - (10d-29)B_{51}A_6 t^{-1} + 20(d-3)B_{11,1}t^{-1} + 12tB_{73} \\
&\quad + 6B_{11,3}t^{-3} + 10t^5 B_{33} \\
\hat{K}_{12} &= (364/3d^2 - 1666/3d + 1831/3)t^{12} + (72-24d)t^6 A_6 + (11d-29)A_6^2 - 20(d-3)A_{12} + 12D^7_{11}t^{-3} \\
&\quad + 12D^4_{11}t^3_2 t + 6B^2_{42} + B_{12,4}t^{-4} + (16d-40)t^2 B_{64} + (48d-160)t^2 B_{82} + (60d-168)t^6 B_{42} \\
&\quad + 8(d-3)B_{84} + 4(2d-5)B_{86}t^{-2} + 4(9d-26)B_{12,2}t^{-2} .
\end{aligned} \tag{3.62}$$

The corresponding diagrams have been displayed in ref. [87] (actually, Münster has given diagrams relevant for his cluster expansion which differs from the expansion described above only by a reshuffling of the disconnected contributions). The interested reader may also find the explicit expression of K_f as a function of the first character coefficients for Z_n, U(1) and SU(N) gauge groups in [87, 112]. The series has been extended to 14th order for the groups Z_2, U(1) and SU(2) [113], but the formulae below actually apply to all SU(N), $N \neq 3$. For these groups, diagrams with 13 plaquettes have one and only one plaquette carrying a higher representation and contribute

$$\begin{aligned}
\hat{K}_{13} &= (144d^2 - 864d + 1261)B_{51}t^7 + (80d^2 - 538d + 893)B_{91}t^3 + (140d^2 - 822d + 1211)B_{13,1}t^{-1} \\
&\quad + (8d-20)C_{8151}t^{-2} + (16d-44)C_{4151}t^2 + 8(d-3)t^{-1}C_{11,8,4} + \mathcal{O}(t^{16}) .
\end{aligned}$$

Diagrams with fourteen plaquettes are more numerous and contribute

$$\begin{aligned}
\hat{K}_{14} &= [(1080d^2 - 5968d + 8288) + 2\tau(d-3) + (\tfrac{86}{3}d^2 - \tfrac{545}{3}d + 283)\alpha \\
&\quad + (-450d^2 + 2531d - 3568)\gamma]t^{14} + \mathcal{O}(t^{16}) .
\end{aligned}$$

In that expression, the four contributions come respectively from the topologies A, A′, B and from disconnected terms. The weights are $\tau = \alpha = \gamma = 1$ for Z_2, $\tau = \frac{1}{4}$, $\alpha = 8$, $\gamma = 4$ for SU(2), $\tau = 1$, $\alpha = 3$, $\gamma = 2$ for U(1), $\tau = 0$, $\alpha = 6N^2$, $\gamma = 2N^2$ for SU(N) ($N \to \infty$).

The final results for the coefficients K_n of

$$K = -\ln t - \sum K_n t^n$$

are listed in table 9. The reader may find more data in [112, 113]. We also quote the longer series

Table 9
Coefficients K_n for the strong coupling expansion of the string tension in four dimensions. $K = -\ln t - \Sigma K_n t^n$

	n										
G	4	5	6	7	8	9	10	11	12	13	14
Z_2	4	0	4	0	56	0	144	0	$\frac{3616}{3}$	0	4276
U(1)	4	0	2	0	$\frac{170}{3}$	0	$\frac{2125}{24}$	0	$\frac{862619}{720}$	0	$\frac{5754751}{2160}$
SU(2)	4	0	0	0	$\frac{176}{3}$	0	$\frac{10936}{405}$	0	$\frac{1532044}{1215}$	0	$\frac{3596102}{5103}$
SU(3)	4	12	−10	−36	$\frac{391}{2}$	$\frac{1131}{10}$	$\frac{2550837}{5120}$	$\frac{-5218287}{2048}$	$\frac{285551579}{61440}$		
SU(∞)	4	0	8	0	56	0	344	0	$\frac{4588}{3}$	0	11688

computed for the three-dimensional Z_2 model by Weeks [114] from the dual Ising model

$$K = -\ln t - 2t^4 - 2t^6 - 10t^8 - 16t^{10} - \tfrac{242}{3}t^{12} - 150t^{14} - 734t^{16} - \tfrac{4334}{3}t^{18} - \cdots. \tag{3.63}$$

It is also of interest to compute the string tension for generalized actions. The SU(2) string tension with a mixed action has been computed to the 11th order in β_f and β_a [115].

To study the U(N) gauge theory, Green and Samuel [104] have introduced the Wilson loop in the determinant representation $\langle \det U_C \rangle$. This order parameter actually probes the U(1) part of U(N) ~ U(1) $\otimes$ (SU(N)/Z_N). They show that, at $d = 4$, $\langle \det U_C \rangle$ has an area (resp. perimeter) fall-off for large loops at strong (resp. weak) coupling, very much as in the four-dimensional U(1) theory. On the contrary, in three dimensions, it is only in the $N \to \infty$ limit that $\langle \det U_C \rangle$ has two regimes. The computation of the string tension of this loop operator has been carried out to twelfth order.

3.2.3. *Glueball mass*

As explained in section 3.1.8, the computation of strong coupling expansions of mass gaps is quite delicate, and only a few terms have been computed so far.

Inverse correlation lengths along an axis have been computed to order β^8 for the three combinations of plaquettes that transform irreducibly under discrete cubic rotations and under C-parity (see section 3.1.8). They form a singlet $\mathcal{O}_1$, a doublet $\mathcal{O}_2$ and a triplet $\mathcal{O}_3$

$$\mathcal{O}_1 = \mathrm{Re\,tr}(U_{12} + U_{23} + U_{31}) \qquad (C = +1) \tag{3.64}$$

$$\mathcal{O}_2 = \{\mathrm{Re\,tr}(U_{13} - U_{23}); \mathrm{Re}(-2\,\mathrm{tr}\,U_{12} + \mathrm{tr}\,U_{13} + \mathrm{tr}\,U_{23})\} \qquad (C = +1) \tag{3.65}$$

$$\mathcal{O}_3 = \{\mathrm{Im\,tr}\,U_{12}; \mathrm{Im\,tr}\,U_{23}; \mathrm{Im\,tr}\,U_{31}\} \qquad (C = -1). \tag{3.66}$$

The latter does not exist for groups with a real fundamental character (Z_2, SU(2), ...). In the continuum limit, these states may be expanded onto states with definite angular momentum, parity and C-parity. Assuming that the smallest angular momentum gives the dominant contribution to each of the multiplets, we expect them to become respectively states of $J^{PC} = 0^{++}$, 2^{++} and 1^{+-}.

The coefficients of

$$m_{J^{PC}} = -4 \ln t + \sum_k m_k t^k \tag{3.67}$$

are presented in table 10 [97, 96] for some gauge groups in four dimensions. Also available from [87] are the series for Z_3, and for U(1) with the heat kernel action. The energy-momentum dispersion relation for the singlet state has also been computed in three and four dimensions for Z_2 and SU(2) [92]: this is a useful quantity if one wants to see how Lorentz invariance is restored in the approach of the continuum limit.

For the same purpose, it is useful to compute off-axis correlation lengths. Though different at small values of β, they should approach the on-axis quantities in the scaling region. Their computation is technically involved (see section 3.1.8) and fewer orders are known. In that case, states must be

Table 10
Table of on-axis glueball masses

	k							
G	1	2	3	4	5	6	7	8
Z_2	0	1	0	$\frac{-67}{2}$	0	$\frac{-287}{3}$	0	$\frac{-2279}{4}$
U(1)	0	$\frac{3}{2}$	0	$\frac{-793}{24}$	0	$\frac{-445}{6}$	0	$\frac{-1659829}{2880}$
SU(2)	0	2	0	$\frac{-98}{3}$	0	$\frac{-20984}{405}$	0	$\frac{-151496}{243}$
SU(3)	−3	9	$\frac{-27}{2}$	−7	$\frac{-297}{2}$	$\frac{858827}{10240}$	$\frac{47641149}{71680}$	$\frac{-183140613}{40960}$
SU(∞)	0	0	0	−34	0	−164	0	−546

0^{++} states

	k							
G	1	2	3	4	5	6	7	8
Z_2	0	1	0	$\frac{-19}{2}$	0	$\frac{37}{3}$	0	$\frac{-359}{4}$
SU(2)	0	2	0	$\frac{-26}{3}$	0	$\frac{13036}{405}$	0	$\frac{-28052}{243}$
SU(3)	−3	9	$\frac{-27}{2}$	17	$\frac{-153}{2}$	$\frac{1104587}{10240}$	$\frac{29577789}{71680}$	$\frac{-90611973}{40960}$

2^{++} states

	k							
G	1	2	3	4	5	6	7	8
SU(3)	3	0	$\frac{9}{2}$	$\frac{-99}{4}$	$\frac{33}{4}$	$\frac{-36771}{1280}$	$\frac{117897}{448}$	$\frac{-1559}{2}$

1^{+-} states

classified according to irreducible representations of other discrete subgroups of O(4) [116]; for example, correlation functions along a diagonal direction must be projected onto representations of the tetrahedral group T_d, and form three multiplets: a singlet, expected to become a 0^{++}, and two triplets (2^{++} and 1^{+-}). Their masses have been computed to order β^4 for Z_2, Z_3, SU(2) and SU(3) (the 1^{+-} only exists for Z_3 and SU(3)) [96, 99].

We finally recall that again in the three-dimensional Z_2 model, a longer series has been computed in the dual Ising model [117].

3.3. Resummation of diagrams

3.3.1. Large dimension

The large dimension limit is not uniform and different techniques must be used depending on the region under study. Here, we are interested by a rearrangement of the strong coupling series, which can be performed only when it is convergent, that is, in the strong coupling phase, including possible metastable regions. This method is different from others, studied later; comparison and final conclusions about the large dimension limit are therefore postponed to the end of section 4.2.3 devoted to the $1/d$ expansion.

A simple inspection of the strong coupling expansion for the free energy (table 7) reveals that the highest power of d increases by one every fourth term. It is therefore tempting to reorder the series as an expansion in $d^{-1/4}$ at fixed $\beta^4 d$. This rearrangement can be performed explicitly [118].

We have first to understand this property on the maximum degree of d in the terms of the series and to identify all the contributing diagrams. Note that the degree in d for a given diagram is its "dimensionality", i.e. the maximum number of different dimensions it uses. Let us consider now the following reduction process on a connected diagram. Suppress a slab of the lattice limited by two consecutive hyperplanes $k < x_j < k+1$ and containing at least one plaquette of the diagram. As the diagram is closed, at least four plaquettes disappear (in the construction described in section 3.1.9, the reduction process consists in suppressing a closed oriented curve); other plaquettes (within the hyperplanes $x_j = k$ or $x_j = k+1$) may be duplicated and must be replaced by either a single plaquette or no plaquette at all. In any case, this operation decreases the number of plaquettes by at least 4, while the dimensionality decreases at most one, whence the quoted property (see details on the subdominant role of disconnected diagrams in [118]). It is also clear that the maximum dimensionality terms are obtained whenever exactly four plaquettes disappear at each reduction step, until the simplest cube diagram is obtained. So the contributing diagrams are all connected trees made with three-dimensional cubes, with all plaquettes in the same representation.

The next problem is to sum their contributions. Since each cube has six faces, it has up to six neighbours in the tree diagram. In a step-by-step construction, the addition of a new cube on a given plaquette of a cube leads to $2d-5$ ($\approx 2d$) possibilities. Some of these may be forbidden by excluded volume effects, but their number is negligible with respect to $2d$, when d is large. The expectation value of a single plaquette in the representation r, $p_r = \langle \chi_r(U_p) \rangle / d_r$ satisfies the self-consistency equation

$$p_r = \beta_r + 2d p_r^5 . \tag{3.68}$$

Introducing $p_r = \beta_r f_r$, $x_r = 2d\beta_r^4$, one parametrizes (3.68) as

$$f_r = (1-u_r)^{-1}, \qquad x_r = u_r(1-u_r)^4$$

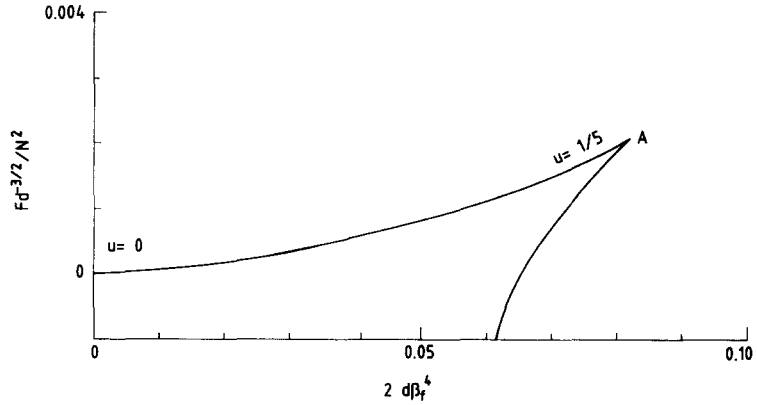


Fig. 17. The free energy in the strong coupling phase for large dimensions.

and the free energy is easy to integrate to

$$\tilde{F} \equiv F - \tfrac{1}{2}d(d-1)\ln\tilde{\beta}_0$$
$$= \frac{d^{3/2}}{12\sqrt{2}} \sum_{r\neq 0} d_r^2\, x_r^{1/2}\, g(x_r)\,(1+\mathcal{O}(d^{-1/4})) \tag{3.69}$$

with

$$g(x_r) = u_r(1-3u_r)/(1-u_r)^2\,.$$

With the Wilson action, only the fundamental representation and its conjugate give a dominant contribution to (3.69). Up to an overall factor N^2 (or $2N^2$ depending whether f and $\bar{f}$ are equivalent or not), the limiting free energy $\tilde{F}(x_f)$ is universal.

A graphical representation of the resulting free energy is displayed in fig. 17. It presents a cusp at the point A for $x_f = 2d\beta_f^4 = 4^4/5^5$ (for $u = 1/5$). As the rearrangement of the series is only possible within the disk of convergence, and as A is obviously a singular point, only the arc OA is physically acceptable. We shall see in section 4.2.3 that a part of this arc ending at A represents a metastable phase. At point A, the second derivative of the free energy has a singular behaviour $\partial^2 F/\partial\beta_f^2 \sim (\beta_{fc} - \beta_f)^{-1/2}$; this looks like a second order phase transition, since the corresponding correlation length becomes infinite, with a critical index 1/4 (zero mass "boxciton").

The Wilson loop may also be computed in the same approximation. It amounts to replacing each plaquette of the minimal surface by the expectation value p_r. Hence

$$\langle \chi_r(U_C)\rangle \sim p_r^A = \exp(-K_r A)$$
$$K_r = -\ln p_r = -\tfrac{1}{4}\ln(u_r/2d) \tag{3.70}$$

from which the leading terms $(dt^4)^n$ in the expansion of K may be recovered.

Corrections to the formula (3.69)–(3.70) can also be computed; the first ones are presented in ref. [118] for the partition function, in ref. [113] for the Z_2 string tension. For groups like Z_2, U(1), SU(2) or SU(∞) where only even powers of t are involved, the corrections in (3.69) is $\mathcal{O}(d^{-1/2})$. The first terms of the expansion are given in table 11.

Table 11
Coefficients of the free energy $\bar{F}/d^2$ in the large d limit

	dt^6	dt^8	d^2t^{10}	d^2t^{12}	d^3t^{14}	d^3t^{16}	d^4t^{18}	d^4t^{20}
Z_2	$\frac{1}{6}$	0	1	$\frac{2}{3}$	10	$\frac{62}{3}$	$\frac{380}{3}$	$\frac{7298}{15}$
U(1)	$\frac{1}{3}$	0	2	$\frac{1}{3}$	20	20	$\frac{760}{3}$	$\frac{2762}{5}$
SU(2)	$\frac{2}{3}$	0	4	$-\frac{4}{3}$	40	$-\frac{8}{3}$	$\frac{1520}{3}$	$\frac{3952}{15}$
SU(∞)	$\frac{1}{3}$	0	2	$\frac{10}{3}$	20	84	$\frac{760}{3}$	$\frac{9072}{5}$

3.3.2. Partial resummation. Bethe–Cayley lattice

The educated reader has recognized a Cayley tree in the relevant diagrams of the preceding section. This suggests a consideration of the same pattern in finite dimensions. We recall that, in the Ising model, one may perform a low temperature expansion in terms of irreducible clusters; truncating to first order yields the mean field approximation, and to second order amounts to solving the problem on a Cayley tree (Bethe approximation) [4]. In the gauge theory, we have unfortunately no gauge invariant equivalent of the magnetic field and magnetization, and the Bethe approximation cannot be justified in such a way. Let us just consider it as a partial resummation of strong coupling diagrams, by solving the model on a Bethe–Cayley tree of cubes [119].

This Bethe–Cayley lattice is defined recursively by building successive generations of cubes. Given a cube of generation l, we add $(2d-5)$ cubes of generation $(l+1)$ on each of its five free faces. On the resulting lattice, $2(d-2)$ cubes are incident on each face, and there is no cluster nor cycle of cubes. Accordingly, the solution on this lattice resums strong coupling diagrams of the original model which may be considered as trees of cubes, but misses, for example, diagrams 12.2, 16.1, 16.2 of table 6, or overcounts the diagram of fig. 19. On the other hand, every link is shared by an infinite number of plaquettes, and there is no weak coupling expansion in models with a discrete gauge group. Let us solve the model in its strong coupling phase, in the simplest case of a Z_2 gauge group. Consider a cube C of generation l, and the plaquette p separating it from its ancestor. Let x_l be the sum of closed diagrams

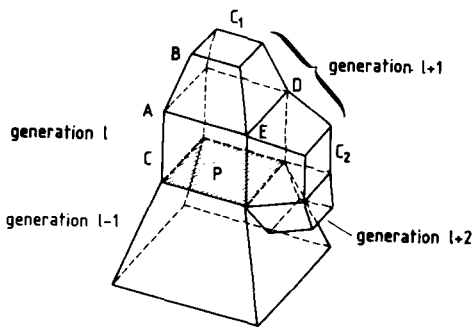


Fig. 18. Construction of a Cayley tree of cubes. Each face as ABDE has $2(d-2)$ adjacent cubes (only two are represented here). None of the cubes built on the faces of C_1 and C_2, nor their descendants will ever meet.

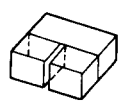


Fig. 19. A diagram overcounted in the Bethe approximation.

made of C and its descendants not containing p, ty_l the sum of such diagrams containing p. One may write in dimension d the following recursion relations

$$x_l = [(x_{l+1} + ty_{l+1})^{\text{"}2d-5\text{"}}]^5$$
$$y_l = [(tx_{l+1} + y_{l+1})^{\text{"}2d-5\text{"}}]^5 . \tag{3.71}$$

The quotation marks mean that, in the binomial expansion, one sets $t^{2k} \equiv 1$, $t^{2k+1} \equiv t$. Assuming the $\rho_l = y_l/x_l$ goes to a limit ρ as $l \to \infty$, we derive the self-consistent equation

$$\rho = [(t+\rho)^{\text{"}2d-5\text{"}}/(1+\rho t)^{\text{"}2d-5\text{"}}]^5 . \tag{3.72}$$

At $d = 3$, this is easily seen to be equivalent to the well-known Bethe approximation of the dual Ising model. At $d = 4$, the following parametrization of the free energy may be written

$$F = 6 \ln \cosh \beta + 20 \ln(1 + 3\rho t + 3\rho^2 + \rho^3 t) - 14 \ln(1 + 4\rho t + 6\rho^2 + 4\rho^3 + \rho^4) \tag{3.73}$$

with

$$\rho = p^5, \qquad p = t + (1 - pt)p^5(3 + p^{10})/(1 + 3p^{10}) .$$

To have a sensible limit as d goes to infinity, one has to rescale t so as to make $dt^4 \sim dp^4$ finite. This is nothing but the regime studied in the previous subsection.

At $d = 3$ or 4, the Bethe approximation has a second order phase transition at a physical value of t^2 ($t^2 = 0.44$ at $d = 3$, $t^2 = 0.19$ at $d = 4$). However, in dimension four, this singular point takes place in a metastable phase because of an earlier first order transition (at $t^2 = 0.178$) to a weak coupling phase of free energy $F_w = 6\beta - 3 \ln 2$. All these figures and phase patterns agree well with the properties of the original model. This makes the Bethe approximation attractive. It would be desirable to have a systematic and efficient way of computing corrections to this approximation.

3.4. Analysis and extrapolation

3.4.1. Generalities

Given a series expansion of some function $f(z)$

$$f(z) = \sum_0^n f_n z^n + \cdots \tag{3.74}$$

truncated to some order N, one may be interested in two related problems:

i) finding the singularities z_i of f close to the origin, and studying the behaviour of f in their vicinity; in particular, determining whether the singular behaviour is algebraic, and, in the latter case, what is the critical exponent α

$$f(x) \sim A_i(z - z_i)^{\alpha} . \tag{3.75}$$

ii) extrapolating the function to a region lying outside the radius of convergence, for instance computing its limit $f(z = \infty)$.

It may happen that additional information on the function (such as analyticity properties, behaviour near some singularities or at infinity, ...) is available. This is of course of utmost importance in the study of these two problems, and may suggest the most appropriate method. In the most common and unfortunate case where no such information is available, one proceeds more blindly and tests various possible techniques.

We are not going to make a detailed study of those methods and refer the reader to the abundant literature [120]. We only review rapidly some of them, to be used in the following applications, in order to fix the notations.

i) The simplest method to set up is the ratio method. Suppose that f_n are real, that the closest singularity z_0 is unique, hence real, and that f has an algebraic singularity (3.75). Then, for large n

$$f_n/f_{n+1} \sim z_0[1 + (1+\alpha)/n + \mathcal{O}(1/n^2)] . \tag{3.76}$$

Therefore, forming the ratios and looking at the way they approach their limit should yield both z_0 and α. The use of Neville tables may accelerate the convergence. However, it may happen that short series are not yet dominated by their closest singularity. Moreover, this method is restricted to real singularities and can handle only one singularity at a time.

ii) Padé approximants are rational fractions $[L/M](z) \equiv P_L(z)/Q_M(z)$ (P, Q of degrees L and M, with $L + M + 1 = N$) such that the Taylor expansion at $z = 0$ reproduces (3.74) up to $\mathcal{O}(z^N)$

$$f(z) - P_L(z)/Q_M(z) = \mathcal{O}(z^N) . \tag{3.77}$$

Padé approximants have the nice property of approximating a meromorphic function within a large domain restricted by non-pole singularities. In particular, they enable an extrapolation outside the circle of convergence of the series. They can also accommodate power behaviour at infinity; for example, if $f(z)$ behaves at infinity as z^{-p}, p a positive or negative integer, then the $[L/(L+p)]$ approximants are appropriate for studying the large $|z|$ region.

Some problems of Padé approximants lie in a sometimes erratic convergence and the possible occurrence of spurious singularities in some approximants. It is observed that they reconstruct a branch-point by an accumulation of interlaced poles and zeroes simulating the cut. Hence they are not of direct use for a singularity of the form (3.75). However, if there is no confluent subdominant singularity, i.e. if $A(z) = (z - z_c)^{-\alpha} f(z)$ is regular at z_c, the logarithmic derivative

$$(1/f)\, \mathrm{d}f/\mathrm{d}z = \alpha/(z - z_c) + A'/A \tag{3.78}$$

has an isolated pole and the corresponding Padé approximant ("D-log Padé") is well-suited; its pole and residue yield z_c and α. Practical cases (non-factorizable singularity $(z - z_c)^\alpha A(z) + B(z)$, confluent singularity $(z - z_c)^\alpha A(z) + (z - z_c)^\beta B(z)$, or even worse) can make the method less effective. In general, the location of the singularity is reliable, whereas the estimate of the exponent is more unstable.

iii) Integral approximants [121–122] generalize Padé approximants. One writes a linear differential equation with polynomial coefficients satisfied by f. We limit ourselves to a first order equation

$$Q_K(z) f'(z) + P_M(z) f(z) + R_L(z) = \mathcal{O}(z^N) . \tag{3.79}$$

The coefficients of the polynomials are determined uniquely (up to a common factor) provided

$K+M+L+2=N$. The $[L/M]$ Padé approximant corresponds to the case $Q\equiv 0$, while $R\equiv 0$ yields the $[M/K]$ D-log Padé. In the vicinity of a simple zero z_0 of Q, f behaves as

$$f(z)\sim A(z-z_0)^{\alpha}+B\,, \tag{3.80}$$

with

$$\alpha=-P(z_0)/Q'(z_0), \qquad B=-R(z_0)/P(z_0)\,.$$

Such approximations may sometimes give accurate estimates of α. On the other hand, as there is a large variety of approximants, the spread of the results may be important and puzzling. Then one tries to find a priori reasons to select subclasses of stable approximants.

iv) These integral approximants may also be generalized to handle double series [122, 123]†

$$F(z,w)=\sum f_{nm}z^n w^m\,. \tag{3.81}$$

One is particularly interested in singular points (z_0, w_0) where F exhibits the scaling behaviour

$$F(z,w)\sim(\widetilde{\Delta z})^{\alpha}\,\phi(\widetilde{\Delta w}\,(\widetilde{\Delta z})^{-\beta})+B \tag{3.82}$$

with $\Delta w=w-w_0$, $\Delta z=z-z_0$, and $\widetilde{\Delta w}$ and $\widetilde{\Delta z}$ two linear combinations of Δz and Δw:

$$\widetilde{\Delta w}=\Delta w-e_1\,\Delta z, \qquad \widetilde{\Delta z}=\Delta z-\Delta w/e_2\,. \tag{3.83}$$

From the available data, namely a finite number of coefficients f_{nm}, one determines four polynomials P, Q, R, S in both variables w and z so that

$$Q\,\partial F/\partial z+R\,\partial F/\partial w+PF+S=0\,. \tag{3.84}$$

The vanishing right hand side means that the matching is actually done to a certain order, i.e. for a given pattern $\{(n,m)\}$ of terms $z^n w^m$, and the pattern of terms for the polynomials are chosen so as to make the resulting linear system solvable.

The singular behaviour (3.82) is encountered at every common zero (z_0, w_0) of $Q(z,w)$ and $R(z,w)$, with

$$B=-S/P\big|_{z_0,w_0}\,,$$

(e_1, e_2) roots of

$$(\partial Q/\partial w)e^2+(\partial Q/\partial z-\partial R/\partial w)\,e-\partial R/\partial z\big|_{z_0,w_0}=0$$

and

$$\beta=(\partial R/\partial w-e_1\,\partial Q/\partial w)/(\partial Q/\partial z-e_2^{-1}\,\partial R/\partial z)\big|_{z_0,w_0}\,,$$

$$\alpha=-P/(\partial Q/\partial z-e_2^{-1}\,\partial R/\partial z)\big|_{z_0,w_0}\,.$$

†The generalization of Padé approximants to the multidimensional case is not unique and other techniques can be applied [124].

These approximants have been used with success in the study of bicritical spin systems [123]. We report below on a less successful trial; in lattice gauge theories, they have been found to be unstable and therefore unreliable. The origin of these instabilities may lie in a more complicated singular behaviour, or in a more complicated pattern of singularities.

3.4.2. Analysis of the free energy and related quantities

It is important to know where the singularities of the free energy F (or of its derivatives, the internal energy E, the specific heat $C, \ldots$) lie. Real physical singularities may signal the occurrence of a phase transition; complex ones are relevant insofar as they limit the convergence of series expansions or affect the physics at real β.

Strong coupling expansions seem well suited for this research of singularities. However, we recall that generically, above their lower critical dimension, lattice gauge theories undergo a first order phase transition. At some value β_c of the inverse coupling, the free energy changes from its strong coupling determination $F_{sc}(\beta)$ to a thermodynamically more favourable one $F_{wc}(\beta)$. $F_{sc}(\beta)$ may be analytic up to $\beta_2 > \beta_c$ while the true free energy $F = \max(F_{sc}, F_{wc})$ is singular at β_c. An example will be provided below by the four-dimensional Z_2 theory. F_{sc} and F_{wc} may correspond to the same analytic function in different Riemann sheets and continue to describe metastable phases beyond the transition point. Explicit examples of such behaviour are constructed in section 4. It may even happen that two models have the same strong coupling behaviour, but undergo first (or higher) order transitions at different points. The "naive" Eguchi–Kawai model offers an example of this puzzling behaviour. Both the full $U(\infty)$ model and its one-site partner have the same strong coupling expansion. However, they are known [78, 79] to have transitions at different points, towards different weak coupling phases.

It is clear from this discussion that strong coupling series are a priori unable to see some features of the phase structure. When a first order transition is present, the strong coupling expansion ignores it and describes the metastable phase beyond the transition. If this phase terminates, however, at some β_2, this coupling is seen as a singularity and gives therefore an upper bound on β_c. If sufficient analyticity is granted, it might be theoretically possible to continue analytically F_{sc} to F_{wc}, hence to determine β_c, but, in practice, the series are much too short for such an extrapolation.

Such a simple scenario seems to happen in the four-dimensional Z_2 theory. Self-duality tells us that the transition, if unique, must happen at $t_c = \tanh \beta_c = \sqrt{2} - 1 = 0.414$, and indeed Monte-Carlo simulations [125] show clearly that the internal energy is discontinuous at this point. This is corroborated by the fact [126] that the series for E (either truncated or Padé-extrapolated) gives a stable value at t_c (0.48–0.49) which is far from the value predicted by self-duality if the transition keeps E continuous ($\sqrt{2}/2 \simeq 0.71$). On the other hand, D-log Padé approximants for E (resp. C) indicate a stable singularity at 0.44 (resp. 0.42–0.43), in the vicinity, but significantly different from t_c. This analysis has been refined [127]; to deal with a single-valued function between t_c and t_2, it is proposed to invert the function $E(t)$ and to compute the series expansion of $t(E)$. The resulting series does not show any stable singularity for $0 \leq E \leq 1$, but dt/dE vanishes at $t \sim 0.424$. This seems to indicate that the change of variable has unfolded the singularity at t_2; such a property is known to be shared by the infinite d limit of the strong coupling phase where E has a square root singularity at t_2 (see section 4.2.3). To summarize, the picture that emerges from this analysis looks like fig. 20, and agrees remarkably well with Monte-Carlo data [125].

In the case of the four-dimensional U(1) theory, a continuous phase transition is expected between a confined and a Coulomb phase, and indeed observed in Monte-Carlo simulations [128]. The expansion of $U(t^2) \equiv E(t)/t$ in powers of t^2, on the other hand, displays alternating signs (as in the cases of Z_2 or SU(2)) which indicates a real negative nearest singularity. This unphysical singularity t_u^2, to which we

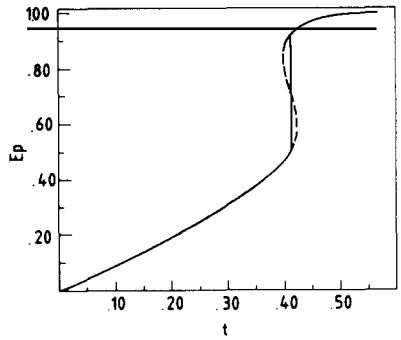


Fig. 20. The internal energy E in terms of $t = \tanh \beta$ for Z_2 gauge theory.

shall return below, may be mapped away by a conformal change of variables $z = t^2/(t^2 - t_u^2)$ [127] (diagonal and sub-diagonal Padé approximants are invariant in this mapping, but D-log approximants are not). Depending on the method, one finds, assuming a form $(t_c^2 - t^2)^{-\alpha}$ for the singularity of the specific heat, $t_c \sim 0.42$–0.45 (i.e. $\beta_c \sim 0.98$–1.01), with an exponent α ranging between 0.35 and 0.65. This must be compared with the Monte-Carlo data $\beta_c \sim 1.005$, $\alpha \sim \frac{2}{3}$ [129]. In spite of the spread of the estimates, coming from the shortness of the series, one sees that strong coupling expansions are quite consistent with the observed behaviour.

The case of the four-dimensional SU(2) theory seems more elusive. On the one hand, one does not expect any deconfining transition. However, a sharp structure in the specific heat has been observed, and modifications of the action may create first order transitions (e.g. mixed fundamental-adjoint actions). We may expect strong coupling series to have a hard time reproducing such a complicated pattern. Indeed the Padé analysis finds a singularity in the region $t \sim 0.45$–0.5 ($\beta \sim 2.1$–2.4); sometimes, this singularity splits into two complex conjugate singularities in the highest available approximants. It is, however, not clear whether this is a genuine effect or an artefact. The authors of [130] have applied the same method as in Z_2, inverting the expansion of $E(t)$ into an expansion of $t(E)$, and used the additional information on the first term of the weak coupling limit

$$E \sim 1 - \tfrac{3}{4}\beta^{-1} + \cdots$$

to constrain further the approximants. They find the nearby singularities of E in the complex plane at $t \sim 0.47 \pm i\, 0.05$ (i.e. $\beta \sim 2.2 \pm i\, 0.32$), in such a way that [131] they also reproduce well the bump in the specific heat. Direct evidence of the presence of these complex singularities has also been derived from a Monte-Carlo investigation of the complex zeroes of the partition function [132]. If one uses only the strong coupling expansion, the distinction between nearby conjugate singularities and a single real one might require as many as 30 terms (11 only are available).

One may try to use these techniques for the mixed SU(2) action of eq. (1.10), in order to reproduce and understand better the phase structure. Several methods have been attempted on the double series in (β_f, β_a) [133]. One may look at singularities of one-variable Padé approximants along lines of constant slope in the (β_f, β_a) plane. The results are consistent with the pattern of fig. 6, but not stable enough to locate accurately the (critical?) end-point of the first order line; incorporating information on the weak coupling behaviour does not help. We have also tried [134] to use the two variable Padé approximants previously defined; they have given very unstable results. It may be that the form of the singularity at the end-point is more complicated than the one assumed in (3.82).

These methods also apply to other groups. For SU(3), results are similar to SU(2), with a singularity of $F(\beta)$ at $\beta \sim 6.2$, a little beyond the cross-over region seen at $\beta \sim 5.6$. For most other groups, where a first order transition occurs, the results look like those of Z_2. One finds a singularity beyond the expected transition. This is the case of $SU(N)/Z_N$ [107], and of $SU(N)$, $N = 4, 5, 6, \infty$ [105, 106]. For SU(4) and SU(6), the occurrence of complex poles in the Padé approximants makes the analysis less reliable. It is worth mentioning that the series $E(t)$ are generally better behaved than $E(\beta)$. They are less often plagued with complex singularities, and more universal in the sense that they approach faster their limit at infinite d (section 4.2.3) or, for $SU(N)$, at infinite N. The $SU(\infty)$ series itself shows some evidence of a singularity at $t^2 = (\beta/2N)^2 \sim 0.13$–$0.14$, a little beyond the first order transition at $t^2 \sim 0.12$. The latter number is extrapolated from the SU(4)–SU(6) transition points. Notice that there is a substantial discrepancy with the corresponding figure extrapolated from the $U(N)$ groups $t^2_{U(\infty)} \sim 0.14$, the origin of which remains obscure.

We have mentioned above that the nearest singularity in t is often complex. In the three cases of Z_2, SU(2) and U(1), the signs of the expansion of E/t in powers of t^2 alternate, pointing to a singularity at negative t^2. The pattern of singularities is summarized in fig. 21. The three-dimensional counterpart of this phenomenon is well-known in the dual language of spin models [135]. The low temperature expansions of the Ising model display nearby complex singularities, which may depend on the nature of the lattice. Moreover, it has been shown in that case, by direct numerical investigation of the model, that these unphysical singularities are not an artefact of the expansion, but a genuine feature of the model, and that they show a good evidence of a non-trivial universal critical behaviour [136, 119]. In the other extreme situation, as $d \propto \infty$, we have seen in section 3.3 that the model is solvable and exhibits a singularity at $2dt^4 = 4^4/5^5$; hence, in that limit, there are four singularities, degenerate in magnitude, in the t-plane, and only one is real positive and physical (end of the metastable phase). At intermediate dimensions, as $d = 4$, the Bethe approximation described in section 3.3 lifts that degeneracy and reproduces well the singularity pattern of fig. 21. For example, for Z_2, it predicts a first order transition at $t^2 = 0.178$, and two singularities at $t^2 = 0.190$ and $t^2 = -0.151$. On the other hand, analysis of the free energy series at the unphysical singularities (Z_2, U(1) and SU(2)) shows some evidence of a universal critical behaviour, with the specific heat diverging with an exponent α' ranging from 0.45 to 0.6 (Bethe approximation predicts a classical value of $\frac{1}{2}$) [119]. The interpretation and role of the complex singularities and the nature of the corresponding continuous theory remain open problems.

So far, we have discussed the analysis of the singularity pattern from series expansions. The other application of these expansions, namely the accurate computation of $F(\beta)$, $E(\beta)$ or $C(\beta)$ in the strong coupling region should not be underestimated. It is quite valuable in the discussion of Monte-Carlo computations. In practice, because of possible nearby complex singularities, Padé approximants are used in this extrapolation; this yields accurate determinations of $E(\beta)$ up to the vicinity of the first real singularity (transition point for U(1), "cross-over" for SU(2), end of the metastable phase for models with a first order transition). The case of Z_2 has been illustrated here, and similar results are available in the literature for other groups.

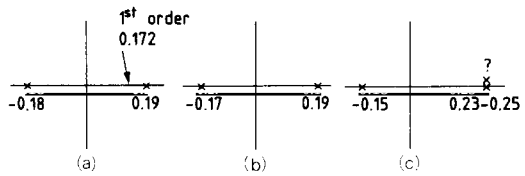


Fig. 21. Singularities in the complex t^2 plane of four-dimensional gauge theories: a) Z_2, b) U(1), c) SU(2).

To summarize, strong coupling expansions enable one to reproduce well the strong coupling phase. They also allow an investigation of the singularity pattern. Although the series are still too short to achieve a precision comparable to the case of three-dimensional spin models, they give a consistent picture of the possible singularities. Improving qualitatively this picture would presumably require much longer series.

3.4.3. The string tension and the roughening transition

In four-dimensional non-Abelian theories, where the only deconfining phase transition takes place at $\beta \propto g_0^{-2} = \infty$, one would like to make contact with the scaling limit of the string tension and of other physical mass scales, for example to compute the constant σ in (2.12). It seems that, if strong coupling expansions may be extrapolated in a suitable way from $\beta = 0$ to β large, so as to reproduce this scaling limit, one should be able to determine σ and to compute ratios of mass scales.

It turns out that this program encounters two kinds of difficulties. First, as discussed in the previous section, most non-Abelian models undergo a discontinuous non-deconfining phase transition at finite β. The string tension and mass gaps do not vanish there, but are reasonably expected to be discontinuous. This makes the extrapolation from small β to the asymptotic freedom regime quite hazardous.

Even in the cases of SU(2) or SU(3), where no such first order transition occurs, another difficulty arises. The string tension, which, loosely speaking, is an observable related to a two-dimensional surface, is sensitive to the "roughening transition" of that surface. We shall review briefly some aspects of that transition, referring the reader to the literature [137, 138] for more details.

i) Roughening in the three-dimensional Ising model.

Consider a three-dimensional Ising lattice at low temperature with appropriate boundary conditions creating an interface between two phases of opposite orientations (fig. 22). The interfacial free energy ΔF is the difference between this new free energy and the usual one; it is proportional to the cross-section A of the lattice

$$\Delta F = -KA \tag{3.85}$$

and K is the surface tension. At zero temperature, the interface is planar. As T grows, it becomes more and more chaotic, until a value T_R, the roughening temperature, lower than T_c (the ordinary "bulk" critical temperature), where its properties change qualitatively. Beyond this point, fluctuations are so large that one can no longer talk of a localized surface. If h_i denotes the altitude of the surface above its position at point r_i of the minimal surface, one may define the probability distribution of $h_i - h_j$ for distant points

$$p(z, r_i - r_j) = \langle \delta(h_i - h_j - z) \rangle \,. \tag{3.86}$$

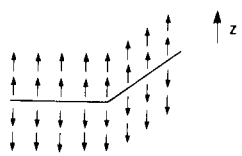


Fig. 22. An interface in the Ising model.

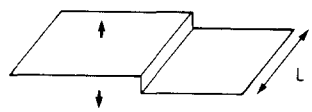


Fig. 23. A step of length L costs an extra free energy fL.

The second moment of that distribution behaves for large separations as

$$\sum z^2 p(z, r_{ij}) = \langle (h_i - h_j)^2 \rangle \sim (\Delta h)^2 \qquad (\text{as } r_{ij} \to \infty) . \tag{3.87}$$

Δh measures the width of the interfacial region. At T_R, Δh diverges for an infinite system. The roughening transition is caused by long-wave fluctuations of the surface; the step free energy f (the free energy per unit length required to create a single step in the interface, see fig. 23) vanishes for $T \geq T_R$, and therefore the large scale deformations of the surface are costless. The surface tension K of eq. (3.85) does not vanish at T_R, but is singular.

There is a converging evidence for this singular behaviour coming from the study of exact models, series analysis, Monte-Carlo simulations and, recently, real experiments in crystal growth. Theoretically, a convenient approach is the solid-on-solid (SOS) model, where considering a simple class of surface deformations leads to an equivalence with the discrete Gaussian model $\beta\mathcal{H} = \beta \Sigma (h_i - h_j)^2$, or with the (dual) two-dimensional xy-model. To the transition of the latter corresponds the roughening temperature of the SOS model. The surface tension K, which is the free energy of the xy-model, is expected to have an essential singularity of the form

$$K = A(\beta) + B(\beta) \exp(-C/\sqrt{T_R - T}) \tag{3.88}$$

while the step free energy identifies with the inverse xy correlation length

$$f \sim \xi_{xy}^{-1} \sim \exp(-C'/\sqrt{T_R - T}) . \tag{3.89}$$

The length scale relevant in the discussion of surface fluctuations is actually ξ_{xy} for $T < T_R$, i.e. in the high-temperature phase of the xy-model. For $T > T_R$, ξ_{xy} vanishes, and the relevant scale is the size L of the lattice. Let $\lambda = \ln(\xi_{xy}/a)$ (resp. $\ln(L/a)$) for $T < T_R$ (resp. $T > T_R$). The function $p(z, r)$ reads for large separations $r \sim L$

$$p(z, L) \sim (1/\sqrt{4\pi\lambda}) \exp(-z^2/4\lambda) . \tag{3.90}$$

In particular, the width Δh of eq. (3.87) is

$$\Delta h = \sqrt{2\lambda}$$

while the probability that the surface is, at some point, at the same height as on its boundary is

$$p(0) = 1/\sqrt{4\pi\lambda} .$$

From the behaviour of ξ_{xy}, one derives that

$$\begin{aligned} p(0) \sim (\Delta h)^{-1} \quad & \sim (T_R - T)^{1/4} \qquad \text{for } T < T_R \\ & \sim 1/\sqrt{\ln L} \qquad \text{for } T > T_R . \end{aligned} \tag{3.91}$$

This singular behaviour derived in the SOS model is presumably universal, and should also take place in the original Ising model. According to the standard lore, the roughening singularity does not affect

"bulk" quantities like the free energy or its derivatives, or like the spin–spin correlation functions. The non-analyticities show up only in surface (non-local) observables.

To conclude, we recall from section 2.3.5 that by duality, the surface tension K of the three-dimensional Ising model is nothing but the string tension of a Wilson loop lying at the boundary of the interface. The previous discussion suggests that the same phenomenon may occur in higher-dimensional lattice gauge theories.

ii) Evidence of roughening in higher dimensions.

The two-dimensional surface spanned by the Wilson loop may no longer be viewed as an interface between two regions. However, long wave length fluctuations are still expected to make surface observables non-analytic at some coupling. At strong coupling $\beta < \beta_R$, the surface is smooth, it is rough for $\beta \geq \beta_R$. Remember that this roughening transition is not a deconfining transition. The string tension is non-analytic, but does not vanish.

To get a clear evidence of the roughening transition, we need a good indicator. The string tension itself [139] is only expected to present a weak singularity at β_R, as in 3 dimensions. However other quantities may diverge at β_R and give a clearer signal of roughening. For example, consider [140]

$$\eta^2 = \left[\sum_{x_\perp} E(x)\, x_\perp^2\right] \Big/ \left[\sum E(x)\right] \tag{3.92}$$

where

$$E(x) = (\langle W(C)\, \chi(U_p)\rangle - \langle W(C)\rangle \langle \chi(U_p)\rangle)/\langle W(C)\rangle\,. \tag{3.93}$$

$E(x)$ measures the chromo-electric energy density at the location x; the test plaquette p of centre x is parallel to the minimal plane of the Wilson loop. This η^2 is a generalization of Δh^2 in eq. (3.87). The dimensionless product $\sigma\eta^2$ should diverge at the roughening transition for an infinite loop. Another possibility [141] is to characterize the roughening transition as the point where the system has lost its memory of the boundary condition, viz. the Wilson loop. For any observable $\mathcal{O}$, e.g. $\chi(U_p)$ for a plaquette of the minimal surface, we expect the difference

$$\langle \mathcal{O} W(C)\rangle_{\text{conn}} = \langle \mathcal{O} W(C)\rangle - \langle \mathcal{O}\rangle\langle W(C)\rangle \tag{3.94}$$

to vanish at β_R. Both η^2 and $\langle \mathcal{O} W\rangle^{-1}_{\text{conn}}$ may be computed in strong coupling expansions and are expected to have a power law divergence at β_R; this is what happens in the three-dimensional SOS model for the similar quantities Δh^2 and $p(0)^{-1}$. It is thus quite satisfactory to see that both quantities give quite consistent determinations of β_R for a variety of gauge groups and dimensions [112, 142, 113]. One observes a remarkable group-independence of the roughening point measured in the variable t; this reflects in fact the nature of the roughening transition, caused by large scale fluctuations, whereas group dependent effects are short scale details,

$$\begin{aligned} d &= 3 \quad t_R \simeq 0.46 \\ d &= 4 \quad t_R \simeq 0.40 \\ d &= 5 \quad t_R \simeq 0.37\,. \end{aligned}$$

The estimated error is of order ± 0.01. We should mention that the series of SU(3), which contains odd

powers of t, leads to less stable results. One also notices that t_R may be below or above t_c, the deconfining transition, depending on the group and the dimension. For Z_2 for example, $t_R < t_c$ ($d = 3$), $t_R \simeq t_c \equiv \sqrt{2} - 1$ ($d = 4$) and $t_R > t_c$ ($d = 5$). In the latter case, the roughening transition takes place in the metastable region. It has been argued by Gliozzi [143] that it is natural for the phase transition to be first order as soon as $t_c \leq t_R$. Indeed, if the deconfining transition is second order, the interquark potential has the form [144]

$$V(R) = \text{Const.} + C/R \qquad (t = t_c \text{ second order}) \tag{3.95}$$

whereas it is natural to expect an exponential fall-off of the corrections to the leading behaviour up to the roughening point

$$V(R) = kR + \text{const.} + e^{-mR} \qquad (t < t_R)\,. \tag{3.96}$$

Consistency of these last two equations at $t = t_c$ requires that the transition t_c be first order. Notice, however, that this argument seems to connect roughening and the order of the bulk transition, in contrast with the standard lore.

It must also be emphasized that the location of the roughening transition is by no means universal. Changing the form of the action, or using the Hamiltonian formalism [145] may modify the location of t_R substantially. Even more dramatically, if we take a twisted Wilson loop as in fig. 24 instead of a planar one, it is easy to see that there is a large number of minimal surfaces. This results in a delocalization of the surface, and roughening takes place in the strong coupling limit $\beta = 0$. Therefore such an off-axis string tension should have better analyticity properties than the conventional planar one. This phenomenon has been exposed particularly clearly in the Hamiltonian formalism [146, 147]. It also manifests itself on special lattices as the simplicial lattice [31] where, for some directions, the minimal surface is highly degenerate (even for planar loops), or as the random lattice [32] where the mere concept of minimal surface does not make sense.

iii) Implications of roughening.

What is the critical theory at the roughening point? In three dimensions, the roughening transition has been shown to be in the universality class of the two-dimensional xy-model. Similar arguments [141] suggest that in d dimensions, the effective theory is given by $(d-2)$ decoupled two-dimensional xy-models. This should give a critical exponent $-\frac{1}{4}$ to quantities like η or $\langle \mathcal{O}W\rangle^{-1}$ and make the step free energy vanish with an essential singularity. Series analyses are in rough agreement with these expectations. Typically, the exponent of the zero of $\langle \mathcal{O}W\rangle$ is found to lie between 0.3 and 0.15 depending on the group and dimension [138]. The step free energy has also been computed for the Z_2 theory at strong coupling to eighth order in t [148]; it is found to vanish at t_R, more likely with an

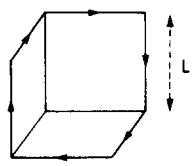


Fig. 24. A twisted loop, the boundary of a large number of minimal surfaces.

essential singularity $\exp[-C(t_R - t)^{-x}]$ than with a power law. Estimates of the exponent x are $\frac{1}{2} \leq x \leq \frac{2}{3}$ at $d = 3$, $\frac{1}{3} \leq x \leq \frac{2}{3}$ at $d = 4$. A similar calculation has been performed in Hamiltonian formalism [146]; there, one looks at the lowest excitation of the string binding two static charges (this is nothing but the cross-section of fig. 23). The mass of this "kink" excitation is also found to vanish at t_R with a similar behaviour. Clearly, longer series would be required to confirm fully the equivalence with the xy-model, but it is gratifying that series analyses support this equivalence reasonably well.

The vanishing of the step free energy or "kink" mass at and above β_R is physically important. It means that the surface is free to move in directions orthogonal to its minimal plane; in other words, it acts like a continuous surface, and part of the translation and rotation invariances of the continuum theory is restored for surface observables. For example, the potential between static charges, hence the string tension, should become rotational invariant in the rough phase. This has been nicely illustrated by strong coupling calculations of the off-axis potential within the Hamiltonian formalism [146] and by Monte-Carlo simulations (see fig. 5); in both cases, one sees that, in a two-dimensional space-like plane, equipotential lines are essentially circular at and above the roughening point. As the latter occurs for both four-dimensional SU(2) and SU(3) just before or inside the cross-over region where scaling is observed, one sees that roughening has the agreeable property of restoring rotational invariance of the string tension in the region of interest.

In the rough confined phase, a description of the surface in terms of a continuous $(d-2)$-vector field h_i is possible, with an effective action proportional to $\Sigma (h_i - h_j)^2$, plus irrelevant terms. As in the SOS model of paragraph i), h_i describes the position of the surface above the point i of the minimal surface, but now h is a continuous variable. From this effective theory, one deduces immediately that the squared width of the surface $\langle \Delta h^2 \rangle$ diverges like $\int d^2k/k^2 \sim \ln L$ as the size L of the loop grows. This is of course in agreement with (3.91), and expresses the impossibility of spontaneously breaking a continuous invariance (here translation) in a two-dimensional theory [149]. Another consequence of the existence of massless translation modes is the creation of a $1/r$ term with a universal coefficient in the static potential [150]

$$V(r) = \sigma r + \text{Const.} - (d-2)\,\pi/24r + \mathcal{O}(1/r^2)\,. \tag{3.97}$$

This appears as a Casimir effect: introducing the boundary condition that h vanishes along the loop modifies the energy of the system.

Returning to the strong coupling methods, we see that roughening jeopardizes the extrapolation of the string tension to large β. It is actually difficult to predict the quantitative effect of the non-analyticity at β_R. In the three-dimensional Z_2 model, neglecting roughening and extrapolating the string tension gives the fallacious impression that it vanishes for $t < t_c$. In the four-dimensional SU(2) and SU(3) theories, roughening takes place at the edge or within the cross-over region, and, in a first attempt, one may assume that it is still meaningful to use strong coupling expansions in that region. If one faces serious inconsistencies, the blame will be on roughening. We only recall or mention for completeness various proposals to avoid the problem of roughening:

- work with non-planar, off-axis Wilson loops, but at the expense of a much more difficult strong coupling computation (see, however, the Hamiltonian calculations [146]).
- as a variant of that strategy, work on a simplicial lattice [31].
- use small loops [151], large enough to obey an area law decay, small enough to be insensitive to roughening.
- use variant actions, for which roughening takes place deeper in the weak coupling regime.

iv) The string tension in four-dimensional theories.

The first theories on which we may probe our truncated series of table 9 are the Z_2 and U(1) theories. In both cases, there is a critical point where the bulk correlation length diverges and where the string tension should also vanish continuously. In the U(1) theory, it is the deconfining transition at $\beta \sim 1.$ ($t_c \sim 0.45$); in Z_2, it is the end-point of the metastable region at $t_c \sim 0.43$, a fair evidence of which has been found in the analysis of section 3.4.2. Also, in both cases, roughening occurs at a stronger coupling $t \sim 0.41$. It is not a priori very clear whether the series of K have to be used as they stand, i.e. truncated to a certain order, or extrapolated. Padé approximants are not obviously the best way to extrapolate them beyond the essential roughening singularity. When truncated, the series give a zero at $t \sim 0.48$–0.5, depending on the order. This is well beyond the expected zero at t_c, and shows that, if the previous picture is correct, neglecting the lack of convergence of the series tends to overestimate the string tension. We recall that the opposite effect is observed on the three-dimensional Z_2 theory, where the naive series seem to indicate a vanishing of the tension before t_c. These puzzling observations show that the forthcoming results in the non-Abelian case have to be taken with a grain of salt...

In the case of SU(2) with the conventional Wilson action, the Monte-Carlo results and the approximations resulting from truncating the series to various orders have been plotted in fig. 3. The curves represent estimates from various truncations of the series; from top to bottom: $-\ln t$, orders t^4, t^8, t^{10}, t^{12}, t^{14}. The expected position of the roughening transition is also indicated. The fact that orders 12 and 14 (the lowest curves) give very close estimates should not deceive the reader. We recall that orders t^{4n} give the major contribution to K. The convergence of the method should therefore be determined from the comparison of 8th or 10th order to the 12th or 14th one. The apparently remarkable agreement of the latter with Monte-Carlo data might be spoiled at a higher order. Even the sign of the error in the blind use of the strong coupling series in the cross-over region cannot be asserted, as we have just discussed. On the other hand, it might be that Monte-Carlo simulations have not yet had their last word, i.e. that much larger lattices would be required to take full account of finite size effects and of roughening, with a resulting tension reduced by a small amount.

The next step is to try to extract the physical, renormalization group invariant ratio σ/Λ_L^2 (with Λ_L defined in (2.10)) from the strong coupling computations. One observes that strong coupling results have the right slope predicted by asymptotic freedom in a narrow range of values of β

$$K = \sigma a^2|_{\text{order } n} \simeq K_n \exp\{-6\pi^2(\beta - \beta^{(n)})/11\} \qquad \text{for } \beta \sim \beta^{(n)}\,. \tag{3.98}$$

Dividing by the renormalization group expression

$$\Lambda a^2 = (6\pi^2\beta/11)^{121/102} \exp\{-6\pi^2\beta/11\}\,,$$

one gets an estimate of σ/Λ^2

$$\sigma/\Lambda^2|_{\text{order } n} \simeq K_n\,(6\pi^2\beta^{(n)}/11)^{-121/102} \exp[6\pi^2\beta^{(n)}/11]\,. \tag{3.99}$$

Clearly, such an estimate is highly sensitive to the value of $\beta^{(n)}$. One finds $\Lambda/\sqrt{\sigma} = (0.9,\ 0.93,\ 1.36,\ 1.39) \times 10^{-2}$ at resp. orders 8, 10, 12, 14. These values have to be compared with Creutz's fit $(1.3 \pm 0.2) \times 10^{-2}$. Although the twelfth and fourteenth order estimates look quite satisfactory, it seems difficult to estimate the error made in this procedure. We think that here longer series would not help much and that the roughening singularity is the stumbling block of strong coupling methods.

The case of SU(3) presents some differences. First the truncated strong coupling estimates seem to converge in a more erratic way. This is due to the irregular signs in the expansion (table 9) and to a rather big coefficient at order 11 (see fig. 25). By the same procedure as for SU(2), one finds $\Lambda/\sqrt{\sigma} \approx (2.6,\ 3.1,\ 3.7,\ 3.5) \times 10^{-3}$ from the 8th, 9th, 10th and 12th approximants respectively, to compare with the Monte-Carlo estimate [152] $(6 \pm 1) \times 10^{-3}$. On the other hand, roughening takes place in the present case at $\beta \simeq 5.8$–5.9, which is already in the weak coupling region, according to Monte-Carlo data (the same phenomenon was observed, even more pronounced, in the Hamiltonian computations [153]). Therefore, it might seem that, in SU(3), a strong coupling determination of $\Lambda/\sqrt{\sigma}$ is more reliable than for SU(2), and that longer series improve the situation. In the SU(3) Hamiltonian calculation [153], however, it was found that the ratio of the spatial cut-off Λ_S to $\sqrt{\sigma}$ takes the values $(4.4,\ 5.4,\ 6.2) \times 10^{-3}$ to orders 8, 10 and 12 in g_0^{-2} respectively. This was extrapolated to a number ranging between 12 and 18.5×10^{-3}, which is in clear disagreement with the previous estimates of $\Lambda/\sqrt{\sigma}$ and the computed ratio $\Lambda_S/\Lambda = 0.91$. This apparent lack of convergence remains a puzzle. To make this discussion even more confused, we mention that recent Monte-Carlo computations [154] on a larger lattice have given a value of $\Lambda/\sqrt{\sigma}$ twice larger.

Checking the universality of the ratio $\Lambda/\sqrt{\sigma}$ for variant actions should be an important consistency test of our current understanding and numerical accuracy. Monte-Carlo simulations have been run for SU(2) Manton and heat-kernel actions [156] and for the mixed fundamental–adjoint action [155], with

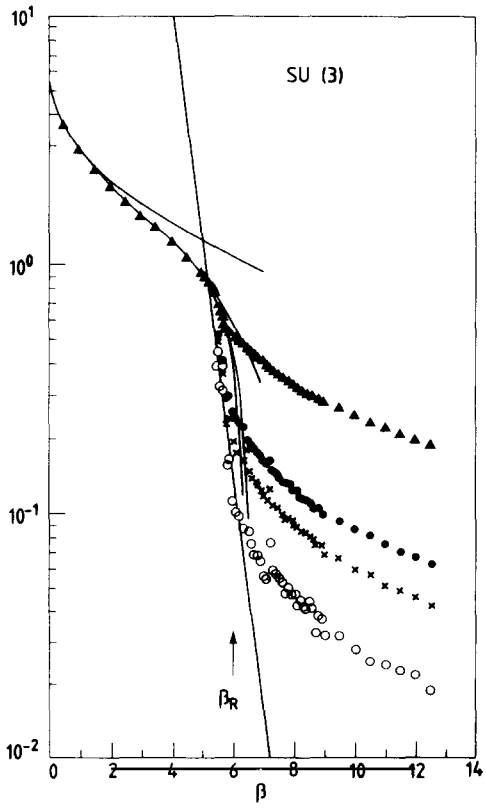


Fig. 25. SU(3) string tension. Strong coupling estimates are superimposed on Creutz–Moriarty data [152]. The solid lines are from top to bottom zeroth, 11th, 12th and 10th order. Also indicated is the roughening coupling.

rather puzzling results. The mixed action has also been analyzed theoretically in [115]; the strong coupling expansion in β_f and β_a is derived to eleventh order, for a fixed ratio β_a/β_f. The string tension estimated in that way is found to have, for increasing β_a/β_f, a more and more pronounced shoulder between its strong and weak coupling regimes. Moreover, by comparing the curves of constant string tension computed by strong coupling methods with the weak coupling, asymptotic freedom prediction, these authors argue that the Monte-Carlo determination of $\Lambda/\sqrt{\sigma}$ has been done in a region of couplings which are still strong, and where scaling has not yet set in. The strong and weak coupling estimates seem to match only in a small region around $\beta_a/\beta_f \approx 0$, $\beta_f \approx 2.2$–2.3. However, this pessimistic conclusion might be affected by an improved treatment of the weak coupling behaviour in the lower half-plane, where the expression (2.3) may be too naive. Such an improvement, abstracted from the large N limit, has been proposed in [159–162]. Other mechanisms have also been proposed to explain the apparent violation of universality [156–158]. Clearly the situation still needs clarification. On the numerical side, a more accurate determination of $\Lambda/\sqrt{\sigma}$, possibly through action improvement, is highly desirable; on the theoretical side, strong coupling computations will remain useful only if the roughening effects are somehow mastered.

We finally discuss shortly the Green–Samuel order parameter [104] in U(N) theories, namely $\langle \det W(C) \rangle$. We recall from section 3.2.3 that this probes the U(1) part of U(N) and that it is expected to have an area law decay at strong coupling, and a perimeter law at weak coupling ($d = 4$). Therefore the point where this U(1)-like string tension vanishes should be a critical point of the system. From the 12th order series, Green and Samuel find a zero at $t_{GS} = \beta \approx 0.4$ in the $N \to \infty$ limit. This is again beyond the first order phase transition at $t \sim 0.33$–0.35, as well as the end of the metastable region estimated at 0.37. As in the case of the string tension of the U(1) theory discussed at the beginning of this section, this may signal that the roughening singularity spoils the series extrapolation.

3.4.4. The glueball masses

The glueball mass spectrum seems to be the ideal field of application of strong coupling series. It should not suffer from spurious effects like roughening, and, provided the series are long enough and converge reasonably, one might be able to match strong coupling and asymptotic freedom predictions.

One should first recall the pioneering work in Hamiltonian formalism [90]. The masses of the lowest-lying scalar, axial-vector and tensor states were computed to fourth order in g_0^{-4} and the diagonal Padé approximants enabled one to extrapolate the mass ratios to infinity, with the result

$$m_T/m_S = 1.003, \qquad m_A/m_S = 1.575\,. \tag{3.100}$$

In Euclidean formalism, the same three masses have been computed to eighth order in $\beta \sim g_0^{-2}$ for the SU(2) and SU(3) gauge theories, as reviewed in sections 3.1.8 and 3.2.3.

In SU(2), the results for the scalar 0^{++} state are encouraging. One uses a truncated series to 6th or 8th order, or a diagonal Padé approximant to the non-trivial part $F(t)$ of the series

$$ma = -4 \ln t + F(t)\,. \tag{3.101}$$

The results display, as the string tension, a "shoulder" between the strong and weak coupling regimes. Using again the empirical rule that the part of this curve with the asymptotic freedom behaviour is reliable, one gets the following estimates [91]

$$m_S/\Lambda = 193, 127, 92 \tag{3.102}$$

for the 6th, 8th orders and Padé approximant respectively. This gives $m_S/\sqrt{\sigma} = 1.8 \pm 0.8$ which compares well with the Monte-Carlo determination 2.0 ± 0.4 [163] and similar results of other groups [164–167].

The same analysis performed in SU(3) gives $m_S/\Lambda = 490$, 310 to 6th and 8th order [91]. The results obtained by Padé approximants are not very stable. This suggests the use of more elaborate extrapolation techniques. Following the suggestion of [130] discussed in section 3.4.2, Smit [168] has re-expressed the expansion of m in the variable E, the internal plaquette energy, and constructed an approximate $E(\beta)$ valid for both small and large β. The resulting estimates of $m(\beta)$ display a behaviour consistent with scaling in a much larger region than the original series, and give

$$m_S/\Lambda \approx 340 \pm 40$$

or

$$m_S/\sqrt{\sigma} \approx 2.1 \pm 0.6\,. \tag{3.103}$$

This is in good agreement with the Monte-Carlo analysis $m_S/\Lambda = 280 \pm 40$ [169] (see also [170]). Let us mention that, if one uses the value $\sqrt{\sigma} = 400$ MeV (derived from the Regge slope in the real world with quarks), this glueball has a mass of about 800 MeV.

The case of higher states is more embarrassing. In numerical studies, a rich spectrum has been found by some groups [170, 164], whereas others [169] could not find any scaling for the possible candidates. This may either mean that these states do not exist in the continuum limit, or that the scaling "window" is still too narrow to observe them. It is quite conceivable that, in the approach to the critical point at $g_0 = 0$, some of the states decouple and keep a fixed product ma. Clearly, the strong coupling methods are biased and implicitly assume the existence of a state (in the continuum limit) by trying to fit its mass with the renormalization group formula. The masses of the $A = 1^{+-}$ and $T = 2^{++}$ states in SU(3) have been extrapolated by the same method [168] as the 0^{++} with a resulting mass ratio $m_T/m_S = 1.1 \pm 0.1$, whereas only weak indications were observed for the axial-vector glueball, with possibly $m_A/m_S \sim 2$. Münster [171] uses a different technique of extrapolation, implementing automatically the scaling limit. He introduces the expression

$$s = [\exp(\tfrac{1}{4} m_S a) - 1]^{-1} \tag{3.104}$$

as a new expansion variable. Indeed

$$s = \beta + \mathcal{O}(\beta^2) \qquad \text{for } \beta \sim 0$$

$$s \sim (m_S a)^{-1} \to \infty \qquad \text{for } \beta \to \infty$$

and scaling means that s_A and s_T should also behave as s for large s. In the expansion of

$$s_{A,T} \equiv [\exp(\tfrac{1}{4} a m_{A,T}) - 1]^{-1} = s\Big[1 + \sum \mu_k s^k\Big]\,,$$

the bracket is thus extrapolated by a diagonal Padé approximant. This gives for SU(3)

$$m_T/m_S \simeq 1, \qquad m_A/m_S \simeq 1.8 \pm 0.3 .$$

In spite of the words of caution at the beginning of this paragraph, the agreement between these numbers and the Hamiltonian estimates looks very encouraging.

In the case of the string tension, roughening is expected to restore part of the rotation and translation symmetries far from the critical point. There is no such phenomenon for plaquette–plaquette correlations, and it is important to see to which extent these symmetries are approximated in the cross-over region. Strong coupling methods have been used in two different ways.

i) The energy-momentum dispersion of glueballs has been computed [92] at strong coupling. The function $E(p)$ may be expanded for small $\boldsymbol{p}$ as

$$a\,E(\boldsymbol{p}) = F(\beta) + a^2 F_1(\beta) \sum p_i^2 + a^4 F_2(\beta) \left[\sum p_i^2 \right]^2 + a^4 F_3(\beta) \sum p_i^2 p_j^2 + \cdots \tag{3.105}$$

In the approach of a critical point, we expect to recover a Lorentz invariant spectrum $E(p) = \sqrt{m^2 + p^2}$; hence

$$F(\beta) \to am; \qquad 2FF_1 \to 1; \qquad -8F^3F_2 \to 1; \qquad F^3F_3 \to 0; \ldots . \tag{3.106}$$

These combinations may be computed in strong coupling expansions, extrapolated, and the relations (3.106) tested. This program works reasonably well in three-dimensional gauge models, but suffers from the lack of convergence of the (short) series in four dimensions.

ii) The off-axis (diagonal) glueball masses were computed to fourth order in β, and their extrapolation compared to the on-axis masses. In view of the low convergence of the latter to 8th order, it is not surprising that the shorter series of the former do not give any very conclusive signal of rotation restoration. Nevertheless the results are encouraging; for example, the ratio of the diagonal scalar mass (to 4th order) to the on-axis one (8th order) is about 1.2 [96].

To conclude, we mention that, in the four-dimensional Z_2 gauge theory, the extrapolation of the mass gap series gives a rather small mass, consistent with zero, at the self-dual point. A vanishing mass would be an unconventional, though not impossible, situation at a first order transition. This has been the object of some speculations [172]. It is also conceivable that the mass is non-zero there and vanishes at the nearby end-point of the metastable region, as in the three-dimensional 3-state Potts model [173].

3.4.5. Conclusions

In the three preceding subsections, we have seen that strong coupling methods have achieved remarkable results in spite of various pitfalls. To improve their accuracy, it is of course tempting to suggest the computation of a few additional terms in the various series. The situation, however, is more subtle.

We have shown, in the discussion of the free energy, that strong coupling methods are blind to first order transitions, a very ordinary circumstance in lattice gauge theories. In cases free of such discontinuities, it seems that much longer series are necessary to see and analyse the intricate details – nearby complex singularities, end-points of first order lines – of the phase structure. Clever matchings with other expansions or approximations might be more rewarding.

Similarly, in the case of the string tension, the exploitation of the series expansions is hampered by the occurrence of the roughening singularity. It seems that neglecting the lack of convergence of the series in that region leads to systematic errors. Therefore, a better way of extrapolating the string tension in that region is needed more than longer series.

On the contrary, it seems that the glueball states are awaiting courageous people willing to extend their series. Suitably extrapolated, these longer series might offer a real alternative to Monte-Carlo simulations.

4. Mean field approximation

4.1. Generalities

This subsection presents generalities about the mean field approximation and related questions [1, 5, 7, 174–193]; the application to lattice gauge theories will be treated in the next subsection. We use the general notations defined in subsection 1.2 for arbitrary fields (excluding, however, fermions) parametrized by a set of (real) variables $\{\phi_x^\alpha\}$. We are interested in the evaluation of the generating functional of connected Green functions $W(J)$. The generating functional for the irreducible correlation functions $\Gamma(\Phi)$, obtained by the Legendre transform of W

$$\Gamma(\Phi) + W(J) = J \cdot \Phi ,$$

with

$$\Phi = \delta W / \delta J , \tag{4.1}$$

is useful in the critical region to identify the relevant continuum field theory.

4.1.1. Variational analysis

There are many standard equivalent ways of deriving the mean field approximation. One of them is a particular case of a variational estimate, and we describe here the general method used in Lagrangian formulation of lattice models. Although this is not well adapted to a systematic computation of corrections, it is interesting both for the simplicity of the derivation and for the rigour of the resulting inequalities.

The idea is to compare the system under study to a set of reference models, depending on external parameters h ("collective coordinates"); these models are assumed to be exactly solvable, or, at least, much information about their properties is available. Denoting by the subscript h all quantities relative to these reference models, the partition function of the original system can be viewed as a mean value

$$Z = \int \exp\{S(\phi) - S_h(\phi) + S_h(\phi)\}\, \mathrm{D}\phi = Z_h \langle \exp\{S(\phi) - S_h(\phi)\}\rangle_h . \tag{4.2}$$

The convexity property of the exponential function ($\langle \exp X\rangle \geq \exp\langle X\rangle$) yields the inequality

$$F \equiv W(J=0) \geq F_h + \langle S(\phi) - S_h(\phi)\rangle_h , \tag{4.3}$$

and the following variational estimate for F

$$F_{\text{var}} = \operatorname*{Max}_{h} [F_h + \langle S(\phi) - S_h(\phi) \rangle_h] , \tag{4.4}$$

bounding from below the exact free energy. Eq. (4.4) shows that the "effective coupling" h is self-consistently determined.

The variational solution F_{var} presents, of course, the singularities of the reference models F_h. This is what happens when the reference system is known to have phase transitions of its own. However, another source of singularities exists in (4.4) and is in fact the great interest of the method for the determination of the phase structures. Indeed the value h^* maximizing the right-hand side of (4.3) is a function of the coupling constants of the original model and may present singularities. These singularities have been mathematically classified (this is the object of "catastrophe theory" [194]) according to the number of coupling constants and the number of degrees of freedom left for h. We only give a sketch for one coupling constant g (typically the temperature) and one parameter h. Fig. 26 illustrates the two possible patterns.

In the first one, h^* jumps from one value to another; F_{var} is continuous, but not its first derivative, nor the physical observables. However, the contribution of the subdominant extremum may be followed in (4.3) by continuity, corresponding physically to a metastable phase (dotted solution in fig. 26a). This behaviour is typical of a first order transition and, as will be seen, is often encountered in lattice gauge theories.

The second scheme (fig. 26b) leaves h^*, F_{var} and its first derivatives continuous and characterizes a second order transition. The second derivative and the correlation lengths ξ are singular. As the continuum limit of lattice theories is obtained whenever ξ/a becomes infinite, this kind of transition is of particular interest.

It should be noticed at this point that such a behaviour for F_{var} is only suggestive of the real behaviour of F. It is indeed easy to construct examples where F_{var} has a transition (first or second order) whereas F is regular, and conversely. If both present a transition, the orders may differ; and finally, in the case of a second order transition in both cases, the critical indices characterizing the critical behaviour may be different (due to the generic behaviour $(g - g_c)^{1/2}$ of h^*, F_{var} always has "trivial" critical exponents in the variational approach). In any case, the phase structure and singular behaviour of F_{var} must be confirmed by a careful study of corrections.

The choice of the reference models has not yet been made explicit. The first criterion is that the

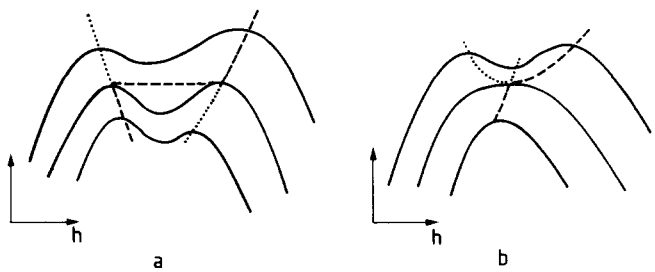


Fig. 26. Sketch of the function to be maximized (right hand side of eq. (4.3)) as a function of h for different couplings; a) first order transition, b) second order transition.

variational solution (4.4) may be easily estimated. There must also exist some physical arguments explaining why the reference models are reasonable representatives of the system under study, so that the variational approximation is expected to simulate the exact solution fairly well. Sufficient flexibility for the parameters h must be introduced to allow the computation of corrections. If the set of parameters is large enough to provide a complete basis for all possible field configurations, it is indeed clear that corrections involving fluctuations around the value h^* can be computed. From now on, we restrict ourselves to a particular choice leading to the so-called mean field approximation.

Remark: Re-expressed in the Hamiltonian formalism by taking the limit of the transfer matrix as the lattice spacing in one direction vanishes, this variational method amounts to finding the ground state energy by minimizing $\langle \Phi_h | \mathscr{H} | \Phi_h \rangle / \langle \Phi_h | \Phi_h \rangle$ for trial states $|\Phi_h\rangle$.

4.1.2. Mean field approximation (variational)

A very simple, exactly solvable model is provided by non-coupled independent fields in an external field H

$$S_H(\phi) = H \cdot \phi = \sum_x \sum_\alpha H_x^\alpha \phi_x^\alpha \,. \tag{4.5}$$

This model, with a suitable "mean field" H, may give a reasonable representation of a given system. H represents the global effect of other fields interacting with a given field ϕ_x. Of course, this approximation is accurate only if the fluctuations around the self-consistently determined mean field are small and can be neglected. This is expected when a given field interacts with degrees of freedom which are either very numerous (large d, i.e. large number of interacting neighbours) or almost frozen in an ordered state (large β). We shall see that the corrections can be reordered for lattice gauge theories as a $1/d$ expansion and that they are small at large β.

According to this discussion, the replacement of fields interacting with ϕ_x by H_x leads to the action (4.5) only if the original action S is linear in the set of variables ϕ_x relative to a fixed location x. We assume now that this is the case. Note that this is true for the Wilson action in the fundamental representation when the corresponding matrix elements are chosen for the field parametrization; see remark 3 below about this constraint.

Using the particular form (4.5) with $H = h + J$ in the variational principle, it is now easy to complete the above calculations for the mean field approximation. Introducing the test free energy per field location

$$u(h) = \ln \int \mathrm{e}^{h \cdot \phi} \, \mathrm{d}\mu(\phi) \,, \tag{4.6}$$

which gives

$$F_h = \sum_x u(h_x + J_x) \,, \tag{4.7}$$

we find

$$W(J) \geq W_{\mathrm{MF}}(J) = \operatorname*{Max}_h \left\{ \sum_x u(h_x + J_x) + S(\Phi) - h \cdot \Phi \right\} , \tag{4.8}$$

with

$$\Phi \equiv \langle \phi \rangle_h = \partial u(h+J)/\partial h\,, \tag{4.9}$$

while the stationarity conditions on h read

$$h = \partial S(\Phi)/\partial \Phi\,. \tag{4.10}$$

The Legendre transform (4.1) can also be performed and yields for the irreducible functional

$$\Gamma_{\mathrm{MF}}(\Phi) = H \cdot \Phi - u(H) - S(\Phi)\,,$$

with H solution of

$$\Phi = \partial u(H)/\partial H\,. \tag{4.11}$$

The mean field is often searched for as a constant field H independent of x, to be consistent with translation invariance; however, this restriction is not necessary and may even be questionable for lattice gauge models where gauge transformations do not preserve translation invariance of these solutions.

Remark 1: The linear form (4.5) chosen for the test action implies that the result depends on the parametrization $\{\phi_x^\alpha\}$ of the fields in terms of real parameters. Again, physical intuition is the guide in this choice. It must be kept in mind that average values $\langle \phi \rangle$ have to be estimated using this parametrization. For instance, the U(1) parametrization $e^{i\theta}$ in terms of only one real parameter θ misses the high temperature phase where all θ are equally probable and the average field $\langle e^{i\theta} \rangle$ vanishes. Furthermore, the action is not linear (as assumed) in the parameter θ. This parametrization is nevertheless valid at low temperature; but representation as a complex number in terms of two real parameters is more adapted for a global study at any temperature, although it requires a δ-function for the modulus in the measure.

Remark 2: The support of the measure $d\mu(\phi)$ may be finite and a prescription is necessary to extend the definition of the action $S(\phi)$ from the original domain of variations of the fields to all values of the real parameters ϕ_x^α. The variational principle allows any extension, but this affects the results. However, we shall see that, as soon as one wants to compute systematic corrections, or even to treat correctly the effects of degenerate solutions, the extension $S(\Phi)$ must be as analytical as possible. The linear form assumed for $S(\Phi)$ obeys this criterion at best.

Remark 3: Some actions (as composite gauge actions involving both the fundamental and adjoint representations of the group) do not seem to satisfy the linear assumption for the action. A first strategy consists in modifying slightly (4.8)–(4.10) to get correct equations. In the simple case where S also uses squared fields $S = S(\phi, \phi^2)$, both $\langle \phi \rangle_h$ and $\langle \phi^2 \rangle_h$ are required and the modified equations follow

$$W(J) \geq W_{\mathrm{MF}}(J) = \operatorname*{Max}_h \left\{ \sum_x u(h_x + J_x) + S(\Phi, \Phi^{(2)}) - h \cdot \Phi \right\}, \tag{4.8'}$$

with

$$\Phi \equiv \langle \phi \rangle_h = \partial u(h+J)/\partial h\,, \tag{4.9'a}$$

and

$$\Phi^{(2)} \equiv \langle \phi_x^\alpha \phi_x^\beta \rangle_h = \frac{\partial^2 u(h+J)}{\partial h_x^\alpha \, \partial h_x^\beta} + \frac{\partial u(h+J)}{\partial h_x^\alpha} \frac{\partial u(h+J)}{\partial h_x^\beta} . \tag{4.9'b}$$

which differ from the preceding result by the second derivative of u in (4.9'b). The difference between $\langle \phi \rangle^2$ and $\langle \phi^2 \rangle$ due to the fluctuations of ϕ is now taken into account. Note that, in the saddle point method presented below, one obtains at lowest order eqs. (4.8–9) and not (4.8'–9'); the second derivative appearing in (4.9'b) will appear in corrections. A second strategy is to linearize the action by considering the quadratic terms $\phi_x^\alpha \phi_x^\beta$ as additional and independent real parameters, imposing the constraint by a suitable δ-function in the measure $d\mu(\phi)$. This leads to independent mean fields associated both with fields and their squares, and therefore the function u in (4.6) is more complicated; on the other hand, the system (4.8–9) remains valid.

Remark 4: The method is not applicable to the fermionic degrees of freedom since inequalities do not make sense in Grassmann algebras. Similarly, the saddle point approach cannot be directly applied.

4.1.3. Saddle point method and corrections

The variational approach presented in the preceding subsections is not well adapted to a systematic computation of corrections around the mean field result. A more powerful and systematic approach is based on an integral formula estimated through the saddle point method [195, 174, 181]. The system is equivalent to a problem of independent degrees of freedom in a random external field; the mean field approximation has neglected the fluctuations of this random field and thus appears as a classical solution. We now construct the "quantum" corrections to this approximation.

The usual exponential representation of the δ-function

$$\delta(\phi_x^\alpha - \Phi_x^\alpha) = (1/2i\pi) \int_{-i\infty}^{i\infty} \exp\{h_x^\alpha(\phi_x^\alpha - \Phi_x^\alpha)\} \, dh_x^\alpha \tag{4.12}$$

is inserted in the definition of the generating functional,

$$\begin{aligned} Z(J) &= \int\int \exp\{S(\Phi) + J \cdot \phi\} \prod \delta(\Phi - \phi) \, d\Phi \, d\mu(\phi) \\ &= \int\int\int \exp\{S(\Phi) + (J+h) \cdot \phi - h \cdot \Phi\} \prod (1/2i\pi) \, d\Phi \, dh \, d\mu(\phi) . \end{aligned} \tag{4.13}$$

Integration over the decoupled fields ϕ is performed according to (4.6) and we obtain

$$Z(J) = \int\int \exp\left\{ S(\Phi) + \sum_x u(J_x + h_x) - h \cdot \Phi \right\} \prod (1/2i\pi) \, d\Phi \, dh . \tag{4.14}$$

The measure is now flat in Φ and h, and the integrand is regular. This allows a computation using the saddle point method in both variables Φ and h

$$W_{sp}(J) = \sum_x u(h_x^* + J_x) + S(\Phi^*) - h^* \Phi^* , \tag{4.15}$$

with the saddle point determined by the stationarity conditions

$$\Phi_x^* = \partial u(h_x^* + J_x)/\partial h_x^* , \tag{4.16}$$

$$h_x^* = \partial S(\Phi_x^*)/\partial \Phi_x^* . \tag{4.17}$$

These relations coincide with the mean field variational solution (4.8)–(4.10) derived previously. The Legendre transform can be recast as

$$\Gamma_{\rm sp}(\Phi) = \sum_x \gamma(\Phi_x) - S(\Phi) , \tag{4.18}$$

where $\gamma(\Phi)$ is the Legendre transform of the one-site integral (4.6)

$$\gamma(\Phi) + u(h) = h \cdot \Phi, \qquad \text{with } \Phi = \partial u(h)/\partial h . \tag{4.19}$$

The integral representation (4.14) now allows the computation beyond the saddle point contribution. One writes

$$h_x = h_x^* + \mathrm{i}\eta_x$$
$$\Phi_x = \Phi_x^* + \phi_x$$

and expands the effective action of (4.14) in the fluctuations η, ϕ. Keeping only the quadratic terms in η and ϕ builds up the first correction

$$Z = Z_{\rm sp} \int \prod (\mathrm{d}\phi_x \, \mathrm{d}\eta_x/2\pi) \exp\Big\{ -\tfrac{1}{2} \sum u''(h^* + J)\eta_x^2 - \mathrm{i}\eta \cdot \phi - \tfrac{1}{2} \sum (\partial^2 S/\partial\phi_x \, \partial\phi_y)\, \phi_x \phi_y \Big\}$$
$$= Z_{\rm sp} \det{}^{-1/2}\{\delta_{xy} - u''(J + h^*)\, \partial^2 S/\partial\phi_x \, \partial\phi_y\} \tag{4.20}$$

$$W^{(1)}(J) = W_{\rm sp}(J) - \tfrac{1}{2} \operatorname{Tr} \ln[\delta_{xy} - u''(J + h^*)\, \partial^2 S/\partial\phi_x \, \partial\phi_y] \tag{4.21}$$

$$\Gamma^{(1)}(\Phi) = \Gamma_{\rm sp}(\Phi) + \tfrac{1}{2} \operatorname{Tr} \ln[\delta_{xy} - (\partial^2 S/\partial\Phi_x \, \partial\Phi_y)\{\partial^2 \gamma(\Phi_x)/\partial\Phi_x^2\}^{-1}] . \tag{4.22}$$

Notice that the factors $(2\mathrm{i}\pi)^{-1}$ have disappeared in the Gaussian integration. Eq. (4.22) is the one-loop approximation to the field theory of action (4.18), where $\gamma(\Phi)$ represents the free part and S the interaction. A diagrammatic interpretation of higher order corrections is of course possible, with Feynman rules derived from the expansion of (4.18).

One may wonder about the justification of this saddle point approximation. What is the large parameter which suppresses successive corrections? It turns out that a large dimension d may play such a role; corrections build up a $1/d$ expansion (see below). On the other hand, we shall see that the saddle point expansion appears as a rearrangement of a large β, i.e. small coupling expansion: only a finite number of terms of the loop-wise expansion contribute to a given order β^{-n}. We finally recall for completeness that mean field is also justified in some cases where the number of internal degrees of freedom grows large: large N-vector models [7], large q-Potts models [196].

4.1.4. Diagrammatic interpretation of corrections

In this section, we show that the mean field approximation and its corrections may be regarded as a resummation of the strong coupling diagrams discussed in section 3. This is, of course, expected in the strong coupling phase of the mean field approximation where corrections are nothing but a reshuffling of the strong coupling diagrams. It may also be understood to be the case in general in the large d interpretation of mean field. As is well known ([7] and subsection 4.2.3 below), the natural scale for β is $1/d$, i.e., one seems to be in the strong coupling regime. However, powers of d come into the configuration numbers of diagrams and make the argument more involved. One has to use functional techniques that we describe now to work out the correct large d limit. The notations and formalism of subsection 3.1.6 are used (notice that the character expansion method cannot be used here).

A connected diagram is reducible with respect to a given k-vertex, if deleting this vertex cuts the diagram into exactly k parts. Any connected diagram takes the tree structure displayed in fig. 27a,

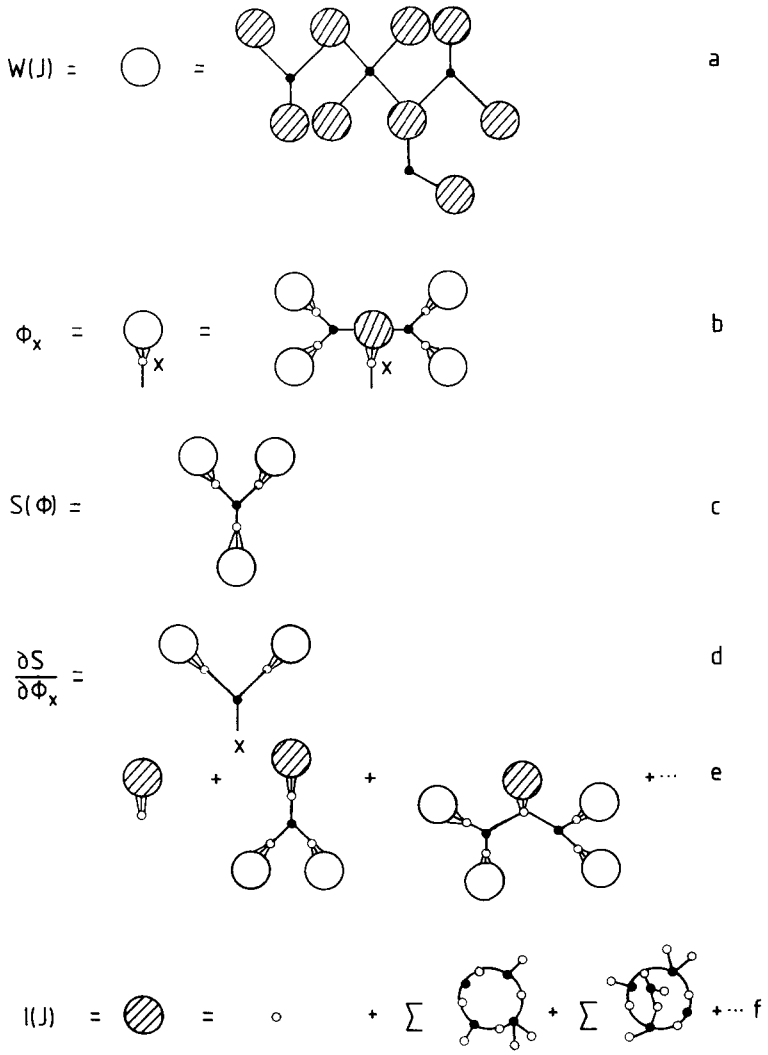


Fig. 27. Diagrammatic construction of the free energy from irreducible diagrams.

where only the n_k k-vertices ($k = 2, 3, \ldots$) with respect to which the diagram is reducible (reducibility vertices) have been drawn, all other parts being gathered in hatched irreducible blobs.

Let $I(J)$ be the sum of contributions of all irreducible diagrams, computed according to the rules of subsection 3.1.6. In particular, $I(J)$ contains the trivial contribution given by the single-site diagram with no vertex, namely $u(J)$. Our problem is to reconstruct $W(J)$ (the sum of contributions of all connected diagrams) from $I(J)$. We denote by Φ the sum of all diagrams with a marked site ($\Phi = \partial W(J)/\partial J$, see fig. 27b). We may construct reducible diagrams in three different ways:

i) since "vertices" are defined by the expansion of the functional $S(\,.\,)$, the quantity $S(\Phi)$ generates any reducible diagram with a multiplicity $\Sigma_k n_k$ (see fig. 27c);

ii) from the diagrammatic interpretation of $\partial S/\partial \Phi$ (fig. 27d), it follows that $\Phi\, \partial S(\Phi)/\partial \Phi$ counts all reducible diagrams with the multiplicity $\Sigma\, k n_k$;

iii) reducible diagrams can also be generated from irreducible blob by "dressing" the sites. The particular dressing of one of these sites is diagrammatically shown in fig. 27e. For this particular site, the initial contribution $u^{(p)}(J_x)$ must be replaced by

$$u^{(p)}(J_x) + u^{(p+1)}(J_x)\, \partial S/\partial \Phi_x + \frac{1}{2!}\, u^{(p+2)}(J_x)\, (\partial S/\partial \Phi_x)^2 + \cdots = u^{(p)}(J_x + \partial S/\partial \Phi_x)\,.$$

Therefore the replacement $J \to J + \partial S/\partial \Phi$ in $I(J)$ performs the simultaneous dressing of all the sites of the irreducible blobs. All connected diagrams are constructed in this way, with a multiplicity n_I, the number of irreducible blobs in each diagram.

Taking into account these different estimates together with the topological relation

$$n_I - \sum (k-1) n_k = 1\,,$$

which expresses that $W(J)$ is a tree made of blobs and reducibility vertices (see fig. 27a), each diagram is counted just once in the combination

$$W(J) = I(J + \partial S/\partial \Phi) + S(\Phi) - \Phi\, \partial S/\partial \Phi\,. \tag{4.23}$$

The similar construction of Φ (fig. 27b) yields the relation

$$\Phi = \partial I/\partial J\big|_{J + \partial S/\partial \Phi}\,, \tag{4.24}$$

which means that the right hand side of (4.23) is an extremum in Φ.

$I(J)$ appears as a sum of diagrams with $0, 1, \ldots$ loops (see fig. 27f). It is easy to check that the restriction to no loop (i.e. $I(J) = u(J)$) reproduces the mean field result, and that the first correction derived for $\Gamma(\Phi)$ from the steepest descent method is obtained by the loop expansion of $I(J)$ up to 1-loop terms. Note that the saddle point expansion brings a supplementary information to this diagrammatic derivation; among the extrema of the right hand side of (4.23), the maximum must be selected. In the case of several maxima, the considerations of the following subsections will apply to (4.23).

This diagrammatic interpretation of corrections to the mean field approximation is useful for the computation of these corrections and for reorganizations of the series (as the $1/d$ expansion discussed later).

4.1.5. Multiple solutions of the stationarity equations

Stationarity conditions may have, and generally do have, more than one solution. This occurs in models obeying a symmetry (as gauge models) as soon as the solutions are not invariant under the corresponding transformations. Let us first discuss the case of well separated solutions.

As the saddle point method is based on the expansion of the integrand around one point, followed by a term-by-term integration, it may be thought that the full result is obtained as the sum of the series of corrections around just one arbitrarily chosen saddle point. This is not correct, however, because these series are asymptotic and Stokes discontinuities are encountered when this term-by-term integration is performed instead of the right procedure of integrating the summed series for the integrand.

We use the following prescription in the computation of corrections to the mean field approximation. The contribution from the different saddle points to $Z(J)$ (and not to $W(J)$!) must be summed up. It must be recognized from the beginning that there is no guarantee that contributions from expansions around these saddle points are exactly additive, for the expansion performed around one point might extend to the vicinity of a second one, resulting in spurious duplication if the expansion of the same function around the second saddle point is added. A hint on the correctness of this prescription comes from the mathematical study of the one-variable steepest descent method [197]. Indeed, the additivity is justified by appealing to Cauchy's theorem, according to which the value of an integral is unaffected by deformation of the path in the complex plane, provided no singularity is crossed in the process and no divergences are introduced. Let us explain this mechanism on a simple example. The integral $\int \exp\{f(x)\}\,dx$, where $f(x)$ is an analytical function as shown in fig. 28a, is computed using the new variable $y = f(x)$. This leads to the integral $\oint_C e^y f'(x)^{-1}\,dy$ along the path displayed on fig. 28b. This contour is deformed up to $-\infty$ as shown by the broken arrow and breaks into two independent paths turning around the singularities y_1 and y_2, each of which yields the corresponding saddle point expansion, as may be easily verified. Hence the summation prescription follows, as long as "sufficient" analyticity properties allow the necessary path deformations.

In the cases where complete reduction to single saddle points proves impossible (for instance, if between two such points the function cannot be diverted out to infinity because of some branch point), there is no general way to determine to which extent contributions from one saddle point are included in higher terms of the expansion around the other.

For the present case of several variables, very little is known about the correctness of the

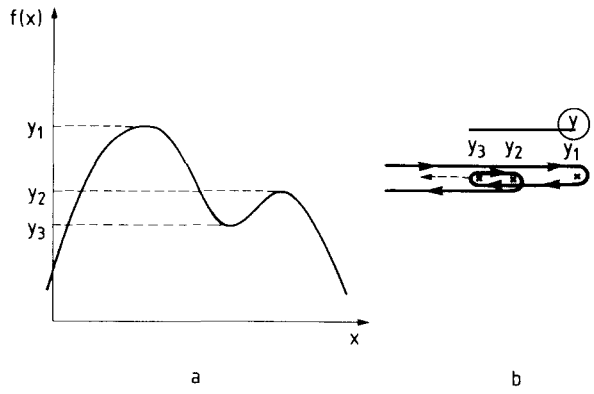


Fig. 28. a) A function with several maxima, the exponential of which must be integrated. b) Corresponding contour of integration in the complex plane $y = f(x)$ and its deformation.

prescription. We have assumed that $S(\Phi)$ is analytic in Φ and even linear in the set of variables Φ_x for fixed x. Under these conditions, we have done our best and we can reasonably hope that the prescription is correct. Note that some results depending on this prescription have been checked to be correct in lattice gauge systems; for instance, the weak coupling expansion of the Z_2 mean field gauge theory is correct only if the prescription is applied.

In the thermodynamic limit, the extensivity of the free energy leads to a new simplification. Only the highest contributions (including, however, all corrections) of the saddle points have to be summed, the other ones vanishing exponentially in this limit.

Because of the plausible occurrence of a Stokes discontinuity when two saddle points coalesce, separate investigation is needed for each range of parameters over which the number of saddle points in the complex plane is constant. Continuity of asymptotic forms should not be presumed when this number changes. This remark is essential in the case of a second order transition; the approach of the critical point must be studied independently in the different phases.

Remark: Similar saddle points expansions around more general variational estimates can be performed, using δ-function integral representations involving more general test actions. However, for the reasons explained in this subsection, the mathematical meaning of the resulting series may be questionable.

4.1.6. Continuously degenerate saddle points

The above considerations remain valid when the saddle points are continuously degenerate. The technique to be used is, however, slightly different. One has to perform a change of variables in the integral so that the integration variables split into two subsets. The first one must parametrize the manifold of the degenerate maxima; it corresponds to zero mass excitations, and the integration must be carried out separately. The integration over the second subset is treated using the usual steepest descent method since no degeneracy is left.

In general, all such saddle points are generated from one of them by applying a Lie group of (gauge) transformations. In this case, gauge fixing terms are introduced and compensated by Faddeev–Popov ghosts. We do not recall this well-known method here.

4.2. Applications to pure lattice gauge theory

We now apply the general formalism described in the previous section to the particular case of pure gauge theories.

4.2.1. General features of the mean field solution

Restricting ourselves to the Wilson action, the mean field approximation introduces a one-parameter constant external field h linearly coupled to the gauge fields which represents the mean interaction with other fields

$$S_h(U_l) = (h/N) \sum_l \operatorname{Re} \operatorname{Tr} U_l . \tag{4.25}$$

This trial action is not adequate for the calculation of corrections and is inconsistent with gauge invariance. We shall see, however, that a more elaborate treatment reduces to this simple case in the saddle point approximation. A detailed study of the corresponding solution is thus useful.

The functions $u(h)$ introduced in eq. (4.6) coincide with $\ln \tilde{\beta}_0$ of eq. (3.28) for $\beta \equiv h$. They have been listed for various groups of interest in table 4, and their general properties are recalled in section 3.1.3. In the absence of external sources, the variational solution (4.8) leads to a free energy per link

$$(\ln Z)/\mathcal{N}d = F/d \geq F_{\mathrm{mf}}/d \equiv \operatorname*{Max}_{h} \{u(h) - h\,u'(h) + \beta^*\,u'^4(h)/4\} \tag{4.26}$$

with a rescaled coupling constant

$$\beta^* = 2\beta(d-1)\,. \tag{4.27}$$

The extremality conditions determine the mean field h as the solution of

$$h = \beta^*\,u'^3(h)\,. \tag{4.28}$$

In this approximation, every link field is an independent random variable, with an average

$$\langle U_l \rangle = u'(h)\,. \tag{4.29}$$

As it stands, this formula violates Elitzur's theorem. This disease will be examined and cured later on.

The free energy expressed in eq. (4.26) has the characteristic shape displayed in fig. 29, irrespective of the gauge group. Its qualitative features are the following.

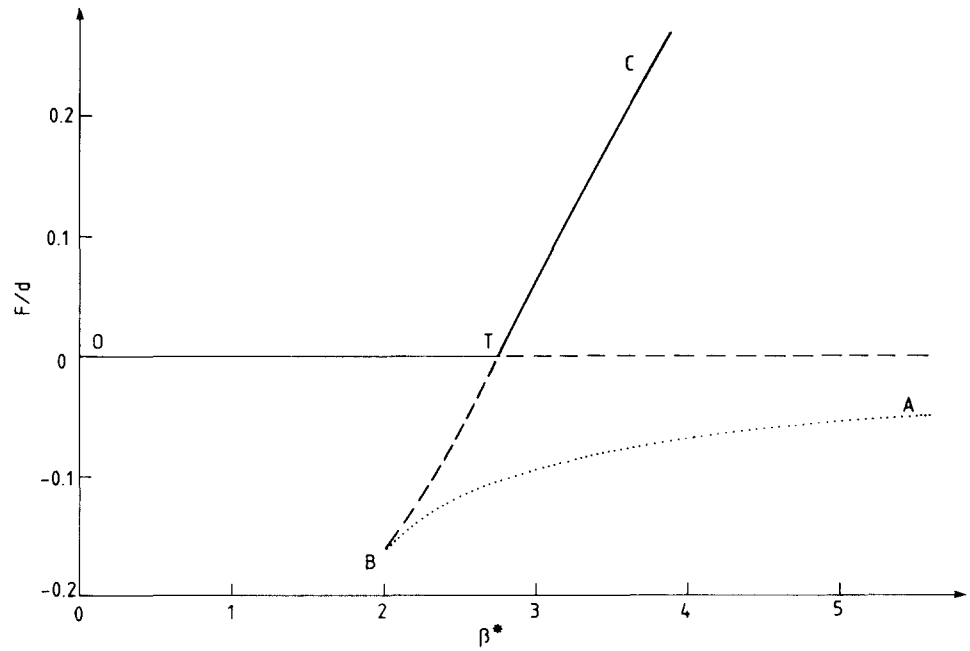


Fig. 29. Mean field solution for the free energy. The dashed part is metastable, the dotted arc is unphysical. The curve corresponds to the Z_2 case, but all gauge groups display a similar shape.

$h = 0$ is always a local maximum and dominates in the strong coupling region. Indeed the energy term $\beta^* u'^4$ is quartic in h (we recall that $u(h)$ is regular near $h = 0$ and behaves there as h^2) and is subdominant with respect to the entropy term $u - hu'$. As this entropy part is concave with only one maximum at $h = 0$ (the exponential in the definition (4.6) of $u(h)$ is indeed a convex function of its argument Tr U_l, maximal at $U_l = \mathbb{1}$), the aforementioned property is true.

As β^* increases, a shoulder appears in the right hand side of eq. (4.26) as a function of h and splits into an unphysical minimum and another local maximum. This is the behaviour illustrated in fig. 26a, whereas the case of fig. 26b does not occur. Therefore, the mean field approximation always provides a first order transition, separating the strong coupling region, where the maximum at $h = 0$ dominates, from the weak coupling region ($h \neq 0$). This transition occurs at point T of fig. 29, when the two maxima are equal. The dotted part BA corresponds to the minimum rather than to a maximum in the right hand side of eq. (4.26), and is unphysical. Note that the end-point A of the strong coupling metastable region lies at infinity.

The Wilson loop parameter shows that this transition is deconfining. The mean field picture approximates

$$W(\mathrm{C}) = \left\langle \mathrm{Tr} \prod U_l \right\rangle$$

as

$$W(\mathrm{C}) = \mathrm{Tr} \left\langle \prod U_l \right\rangle = \mathrm{Tr} \prod \langle U_l \rangle + \text{connected corrections},$$

and there are arguments – though no proof to our knowledge – to the effect that the corrections fade away as the exponential of the area. The first term $N(u'(h))^P$ vanishes for $h = 0$, leaving only an area law; it gives a perimeter law for the weak coupling solution $h \neq 0$.

In spite of its simplicity, this approximation is qualitatively good; it is expected to be accurate in high dimensions, where a first order transition is always observed. Even in rather low dimensions, the location of the point T is reproduced up to 10–20%; we cannot expect more precision without the $1/d$ correction. Some numerical results are

Z_2	$\beta_c^* = 2.755$,	$d = 3$:	$\beta_c = 0.689$	(0.7613 from dual Ising)
		$d = 4$:	$\beta_c = 0.459$	(0.4407 exact)
Z_3	$\beta_c^* = 4.383$,	$d = 3$:	$\beta_c = 1.096$	(1.085 from dual Potts)
		$d = 4$:	$\beta_c = 0.730$	(0.6700 exact).

Here, $d = 3$ is the lower critical dimension, where a transition first appears (this critical dimension limits the validity of the approximation).

Continuous groups are expected to have their critical dimension at $d = 4$; even if the phenomena at this dimension are peculiar, something happens (cross-over phenomenon) which might be considered as the shadow of first order transition occurring at higher dimensions. The comparison gives

U(1)	$\beta_c^* = 7.293$,	$d = 4$:	$\beta_c = 1.216$	(1.01 Monte-Carlo)
SU(2)	$\beta_c^* = 1.060$,	$d = 4$:	$\beta_c = 2.826$	(2.2 Monte-Carlo)
		$d = 5$:	$\beta_c = 2.120$	(1.64 Monte-Carlo).

Calculations can also be performed in the axial gauge. Temporal links are set to $\mathbb{1}$ and are no longer dynamical variables. Only spatial links are involved in the trial action (4.25), and the free energy per site reads

$$F \geq F_{\mathrm{mf,ag}} = (d-1) \operatorname*{Max}_h \left\{ u(h) - h\,u'(h) + \frac{\beta^*(d-2)}{(d-1)} \frac{u'^4(h)}{4} + \frac{\beta^*}{(d-1)} \frac{u'^2(h)}{2} \right\} \tag{4.30}$$

and this formula coincides with (4.26) in the large d limit. Results obtained with this estimate are quite similar to the preceding ones. The point A in fig. 29 is now at finite distance, because the energy term has a quadratic part and can compete with the entropy term near $h = 0$; this maximum disappears at point A

$$\beta^*_{\mathrm{A}} = (d-1)/u''(0)\,. \tag{4.31}$$

As d decreases, the points A and T come close to each other and merge at a critical dimension

$$d_{\mathrm{c}} = 2 - \tfrac{1}{6} u^{(\mathrm{iv})}(0)/u''^3(0) \tag{4.32}$$

where the transition becomes a second order one.

In this gauge fixing formulation, the problem of violation of Elitzur's theorem is avoided, but replaced by the problem of long-range correlation of spatial links in the temporal direction. For instance, a temporal Wilson loop $R \times T$ has an exponential fall-off only in R ($\sim N(u'(h))^{2R}$), rather than the expected perimeter behaviour in $(R + T)$. For a loop of large T and finite R, this seems to violate a rigorous bound [198]. The approximation has also a wrong weak coupling behaviour. However, both diseases can be cured as will now be illustrated in the Z_2 case.

The weak coupling expansion of the actual model starts from a configuration with all gauge fields $U_l \equiv \mathbb{1}$. The first correction comes from flipping one link; $2(d-1)$ plaquettes are frustrated in this process and the relative weight factor is $\exp\{-4(d-1)\beta\}$, yielding a total correction $(1 + \mathcal{N}d \exp\{-4(d-1)\beta\} + \cdots)$ to the partition function. The same calculation, carried out in the axial gauge, only yields $1 + \mathcal{N}(d-1)\exp\{-4(d-1)\beta\}$. It is clear that this comes from the now forbidden overturn of a temporal link, which amounts to flipping an infinite number of spatial links (see fig. 30). Adding these extra configurations restores the right low temperature limit. The same problem occurs in the mean field approximation, where F_{mf} of eq. (4.26) reproduces the correct large β behaviour, whereas $F_{\mathrm{mf,ag}}$ of eq. (4.30) does not. This suggests that we should consider the configurations of fig. 30, where now all spatial links read $\pm u'(h)$. Such a "defect" configuration is not generated by any finite order correction to the original mean field, and does not make the action stationary. In principle, it is possible to "broaden" the

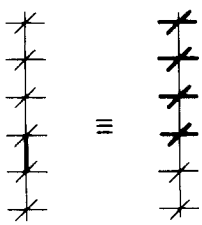


Fig. 30. Flipping one temporal link is equivalent to flipping an infinite number of spatial links.

defect so as make it a saddle point (see [199] for details). For our present purpose, however, the simple picture is good enough. The contribution of the defects is exponentially smaller than the original mean field configuration, but, in a dilute gas approximation [182], they contribute to the free energy and cure the large β limit, at least to leading order. It may also be seen that they reinstate the correct perimeter law of the Wilson loop [199]. The fact that contributions of two different saddle points have to be added in the axial gauge formulation for discrete groups makes the computation of corrections more hazardous (see the discussion of section 4.1.5).

4.2.2. Generalized mean field: restoration of gauge invariance, corrections

As noticed in section 4.1.3, the one parameter action (4.25) is not adapted to a systematic computation of corrections. It is necessary to couple a "mean field" to each gauge degree of freedom. For instance, we use the trial action

$$S_{\{h\}}(U_l) = \sum_l (1/N) \operatorname{Re} \operatorname{Tr}(h_{ij} U_{ji}) \tag{4.33}$$

where the generalized mean field h_{ij} is now an arbitrary complex $N \times N$ matrix associated with each link, and we recall that i and j label the sites. For convenience, we conventionally write $h_{ji} = h_{ij}^\dagger$. It is hopeless to try to solve the general saddle point equations for the mean field configuration $\{h_l\}$; however, a rather natural ansatz for the solution is

$$h_{ij} = h g_i g_j^{-1} \qquad h \in \mathrm{R},\ g_i \in \mathrm{G} \tag{4.34}$$

depending on a real constant h and on a local gauge transformation $\{g_i\}$. In the saddle point eqs. (4.16)–(4.17), the g_i's disappear and the mean field solution of the preceding subsection is recovered. Nevertheless this formulation offers some advantages.

i) The possibility of computing corrections according to the general formalism of sections 4.1.3–4. Illustrations of this technique are presented in the following sections.

ii) Restoration of gauge invariance and recovery of Elitzur's theorem. Solution (4.34) is degenerate with respect to $\{g_i\}$; the procedure of section 4.1.5 must therefore be applied, and the mean field solution for the partition function Z is summed over the g_i's. In particular,

$$\langle U_{ij} \rangle = \int \int u'(h)\, g_i\, g_j^{-1}\, \mathrm{D}g_i\, \mathrm{D}g_j \equiv 0 \tag{4.35}$$

and Elitzur's theorem is satisfied. The reader may object that such an argument, when applied to a model with a global symmetry, a spin model, say, seems to imply the vanishing of any order parameter – e.g. the magnetization – and rules out the possibility of a spontaneous breakdown of the symmetry. To understand fully the differences between models with global and local symmetries, one has to break that symmetry by an external source J, and look at the behaviour of the (infinite) system when the breaking term is removed. The discussion follows very closely the lines of section 2.4. In a gauge theory, if a given mean field is the true saddle point in the presence of an infinitesimal source J, there is also a large set of gauge transforms of that mean field which are almost saddle points with a very close contribution. This set conspires to give a vanishing expectation value to any non-gauge invariant observable, in the limit $J \to 0$. A similar argument is not possible for a global symmetry, because of the lack of commutativity of the limits $\mathcal{N} \to \infty$ and $J \to 0$ (cf. section 2.4).

iii) The entropy of the gauge degenerate mean field must be taken into account. When the gauge group is discrete, the saddle points are separated and their equal contributions must be summed; there are $\|G\|^{\mathcal{N}}$ such configurations when h is non-zero. Hence a term $d^{-1}\ln\|G\|$ must be added to the right hand side of eq. (4.26) for the free energy per link. In a model with a global symmetry, such a term $\mathcal{N}^{-1}\ln\|G\|$ in the free energy per site would play no role in the thermodynamic limit $\mathcal{N}\to\infty$. The case of continuous groups, where gauge fixing is necessary, is discussed below.

iv) We finally mention here that this saddle point approach is also useful in identifying the continuous theory at the critical point. This remark is of no use in this section, but will be relevant in section 5 for the critical point of the Higgs model.

The parametrization (4.33) and ansatz (4.34) may sometimes be too restrictive. This is the case of the mixed SU(2) fundamental–adjoint system that we review briefly here [177, 178, 180, 191, 193]. In the large β_a region, the action has two local maxima for $U_p = \pm 1$; corresponding link-configurations fluctuate around $U_l = \pm 1$ (up to a gauge). The role of these "fluxon" configurations in the SU(2) theory has been the object of much interest [200, 69]. Within the mean field approach, they cannot be handled by (4.33) which allows fluctuations around one value only. Therefore, the Z_2 degrees of freedom of the fields must be explicitly extracted for a separate treatment if one wants to describe correctly the large β_a region. A practical way is to write

$$U_l = \varepsilon_l V_l, \qquad \varepsilon_l = \pm 1, \qquad V_l \in \mathrm{SU}(2). \tag{4.36}$$

The Z_2-degrees of freedom ε_l may be treated as defects in a dilute gas approximation [193] and the SU(2) variables in a mean field picture. Alternatively, one may introduce mean fields for both $\{\varepsilon_l\}$ and $\{V_l\}$. We shall discuss here this second approach [177]. The parametrization (4.36) is not one-to-one and introduces a new "gauge"-symmetry ($\varepsilon_l \to -\varepsilon_l$, $V_l \to -V_l$). A gauge fixing procedure might be introduced to suppress these spurious degrees of freedom, such as, e.g., a constraint $\mathrm{tr}\, V_l \geq 0$. It is nevertheless more convenient to keep this discrete symmetry unbroken in the mean field procedure, bearing in mind that non-zero mean field solutions are $2^{\mathcal{N}d}$ times degenerate. The resulting phase diagram displays three phases:

a) strong coupling: β_a, β_f small ($\langle V_l\rangle = 0, \langle\varepsilon_l\rangle = 0$).

b) SO(3) weak coupling phase, β_a large, β_f small ($\langle V_l\rangle \neq 0$, $\langle\varepsilon_l\rangle = 0$). The gauge field U_l fluctuates around two separated values ± 1 which play the same role, because β_f is too small to break the Z_2 symmetry between these states.

c) SU(2) weak coupling phase, β_a, β_f large ($\langle V_l\rangle \neq 0$, $\langle\varepsilon_l\rangle \neq 0$). The phase diagram of fig. 6 is reproduced, but the first order line does not end and separates completely the regions a and c.

The configurations (4.34) set all plaquette terms $h_p = h_{ij}h_{jk}h_{kl}h_{li}$ proportional to the unit matrix (up to a gauge). In general, this maximizes the action and such solutions are expected to dominate in the weak coupling regions. However, the energy is no longer maximal [201] at $U_p = \mathbb{1}$ when β_a is negative; the ansatz (4.34) is therefore certainly incorrect in this region and other solutions of the stationary equations must be looked for. We do not report here on these solutions which generate rather complicated phase diagrams, with second order lines which may be related with the continuum limit in dimensions strictly greater than 4 [180, 191].

4.2.3. $1/d$ expansion for discrete models

We have repeatedly stated that the mean field technique and its extensions are expected to provide an accurate description of high dimensional statistical systems. The saddle point approximation yields a

critical inverse coupling β of order d^{-1}. The diagrammatic method of section 4.1.4 is well suited to show that, for large d and βd fixed, the loopwise expansion yields a $1/d$ series.

The variational formula (4.22) is used together with the knowledge of the irreducible functional $I(J)$ as an expansion in $1/d$ at fixed βd. Let us first show that the loop expansion of fig. 27f can be used for this reordering. For a pure lattice theory, the "vertex" (black dot) is geometrically the centre of a plaquette and is joined to four "sites" (open dots), here the middles of the bordering links. The line joining a vertex to a site is therefore geometrically a segment of length $a/2$ parallel to an axis. In order to construct a closed loop, it must be possible to associate these segments in parallel pairs of opposite directions; therefore a single loop with n vertices ($2n$ segments) has at most n different orientations, and the choice of these orientations leads to a behaviour d^n for its configuration number. With each vertex bringing a factor $\beta \sim d^{-1}$, the final contribution of a single loop at fixed βd is of order $1 + \mathcal{O}(1/d)$. Every additional loop yields at least an extra factor $1/d$ because at least one link must be shared with the old loop and its orientation in the new loop cannot be freely chosen. On the other hand, the lowest order term in the loop expansion, i.e. a single "site" (link) is of order d (there are $\mathcal{N}d$ links on the lattice).

Therefore, we have characterized the diagrams which contribute to the $1/d$ expansion of $I(J)$. Note that, in contradistinction with spin systems [202], an infinite number of diagrams contribute to each term of this expansion. The first correction involves only one-loop diagrams. Plaquettes must be arranged as shown in fig. 31 in order to allow the maximal number of orientations. With n plaquettes, such diagrams yield a contribution

$$d!\,(2\beta u'^2 u'')^n/[(d-n)!\,2n]$$

and their asymptotic contributions in the limit $d \to \infty$ at fixed βd can be easily summed up over $n \geq 3$. In addition, the irreducible diagrams with two plaquettes at the same location and linked in various ways must also be added. Details and second order corrections can be found in [174].

This method works correctly with discrete gauge groups, for which the variational solution is isolated. Similar calculations for continuous groups have not yet been attempted. We recall that, in addition to the variational estimate (4.22), a supplementary term $\mathcal{N} \ln\|G\|$ must be added to the weak coupling

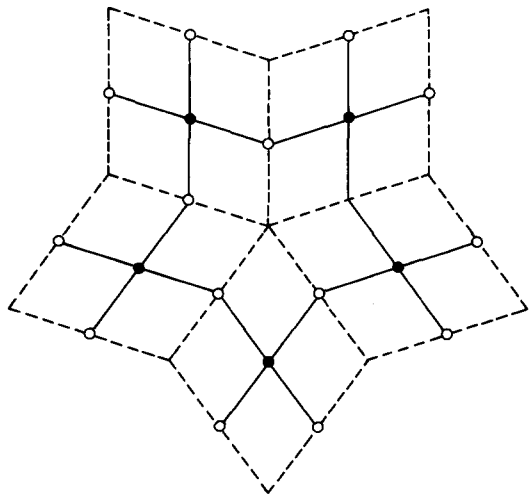


Fig. 31. Diagrams contributing to the first $1/d$ correction to the mean field result. Here $n = 5$ plaquettes are drawn.

branch, to take into account the degeneracy $\|G\|^N$ of the saddle point. This term is of order d^0 and thus contributes to the first order correction of the $1/d$ expansion.

Fig. 32 displays the results for the Z_2 gauge group. The first two corrections have been included in $I(J)$, before using the variational principle (4.22). As the dimension decreases, the weak coupling metastable phase shrinks; it disappears, in this approximation, at $d \simeq 2.9$, very near the true critical dimension 3 where the first order transition is replaced by a second order one.

The end-point A of the strong coupling metastable region remains, in this expansion, at infinity. At such cusp points, one sees a unique maximum exploding into a large number of degenerate maxima. What happens there is badly reproduced in the crudest steepest descent method; moreover, the sudden occurrence of the additional entropy terms taking into account the multiplicity of the saddle points yields an incorrect discontinuity. The complementary description made in section 3.3.1 for the strong coupling phase must be used instead. In particular, the position of the end point was found to be $d\beta = \frac{4}{5}10^{-1/4}d^{3/4}$, which is at infinity for $d = \infty$, but which cannot be obtained as a $1/d$ expansion. This is reflected by a decreasing radius of convergence of this $1/d$ expansion in the strong coupling phase as βd increases.

Similarly, the $d^{-1/4}$ expansion of section 3.3.1 cannot handle the weak coupling phase correctly. In the variables of fig. 17, the weak coupling region ($\beta d \sim \mathcal{O}(1)$) corresponds at $d = \infty$ to the vertical axis. A detailed study of the $d^{-1/4}$ expansion shows that corrections develop more and more singularities at $u = 1$ to construct the end-point of the weak coupling metastable phase. The two methods appear as complementary in the description of the large d limit.

We finally quote the $1/d$ expansion of the first order transition. Numerically, in the Z_2 case,

$$2(d-1)\beta_c = 2.755205 - 0.912561d^{-1} + 0.601169d^{-2} + \mathcal{O}(d^{-3}) \tag{4.37}$$

(the saddle point entropy $\ln 2$ is responsible for the sign of the $1/d$ correction).

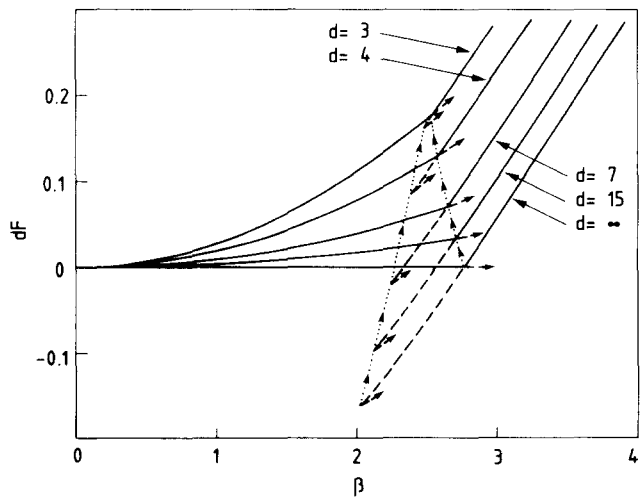


Fig. 32. Behaviour of the Z_2 free energy as the dimension decreases (from $1/d$ expansion cut to 3 terms). Solid lines are the physical parts. The metastable and unphysical branches have been represented as dashed, or indicated by an arrow. The dotted lines show the trajectories of the transition points as d varies.

4.2.4. Corrections to mean field for continuous groups

In the case of a continuous gauge group, gauge fixing is mandatory for the computation of corrections. Otherwise, zero modes corresponding to infinitesimal gauge transformations (eq. (4.34)) of the solution to the saddle point equations would give rise to infinite fluctuations. The most convenient choice is again the axial gauge (see, however, the end of this subsection for a brief discussion of other choices). The axial gauge also has the merit of already yielding the right low temperature limit of the free energy at the mean field "tree" approximation, as we shall see now. We consider for definiteness the SU(N) or U(N) Wilson action in d dimensions [182, 193, 184, 203]. Following the same lines as in sections 4.1.3 and 4.2.2, one considers the partition function in the axial gauge

$$Z = \int \prod_{\substack{\text{spatial}\\ \text{links}\\ l}} \mathrm{D}U_l \exp\Big\{(\beta/N)\,\mathrm{Re}\Big[\sum_{\substack{\text{space-}\\ \text{like}\\ \text{plaquettes}}} \mathrm{tr}\, U_p + \sum_{\substack{\text{time-}\\ \text{like}\\ \text{plaquettes}}} \mathrm{tr}\, U_l U_{l'}^{\dagger}\Big]\Big\} \tag{4.38}$$

and one inserts in Z the following representation of 1

$$\begin{aligned} 1 &= \prod_{\substack{\text{spatial}\\ \text{links } l}} \int \mathrm{d}^{2N^2} V_l\, \delta^{2N^2}(V_l - U_l) \\ &= \prod_l \int \mathrm{d}V_l\, \mathrm{d}H_l\, (2\mathrm{i}\pi N)^{-2N^2} \exp\{(1/N)\,\mathrm{Re}\, H(U-V)\} \end{aligned} \tag{4.39}$$

where V is a $N \times N$ complex matrix, and $\mathrm{d}V$ is the flat measure $\Pi_{\alpha,\beta}\, \mathrm{d}\,\mathrm{Re}\, V_{\alpha\beta}\, \mathrm{d}\,\mathrm{Im}\, V_{\alpha\beta}$, and likewise for $\mathrm{d}H$. As in section 4.1.3, the $2\mathrm{i}\pi N$ factor disappears ultimately in the Gaussian integration, and we omit it in the following. One gets

$$Z = \int \prod_{\substack{\text{spatial}\\ \text{links } l}} \mathrm{d}V_l\, \mathrm{d}H_l \exp\{S_{\mathrm{eff}}[V_l, H_l]\} \tag{4.40}$$

$$S_{\mathrm{eff}} = (\beta/N)\,\mathrm{Re} \sum_p \mathrm{tr}\, V_p + \sum_l (u(H_l) - (1/N)\,\mathrm{Re}\, H_l V_l) \tag{4.41}$$

where we recall that

$$u(H) \equiv \ln \int \mathrm{D}U \exp(1/N)\,\mathrm{Re}\,\mathrm{tr}(UH)\,. \tag{4.42}$$

The saddle point of the functional is assumed to be uniform: H_l, V_l independent of l, real and proportional to the unit matrix:

$$H_l = H\mathbb{1}, \qquad V_l = V\mathbb{1}\,. \tag{4.43}$$

Of course, we have no a priori guarantee that such c-number solutions are the dominant saddle points, but we shall verify that they exist for small enough coupling and that they describe well the physics in that region. Then H, V must be a saddle point of the real function

$$[1/\mathcal{N}(d-1)]\, S_{\text{eff}}[V\mathbb{1}, H\mathbb{1}] = \beta[\tfrac{1}{2}(d-2)V^4 + V^2] + u(H) - HV \tag{4.44}$$

hence

$$\begin{cases} V^* = u'(H^*) \\ H^* = \beta V^*(2 + 2(d-2)V^{*2}) . \end{cases} \tag{4.45}$$

These equations have only the trivial solution $H^* = V^* = 0$ for $\beta < \beta_B$, and a non-trivial solution appears at $\beta \geq \beta_B$. For investigating the large β limit of this solution and of the related free energy, we have to expand the function $u(H)$ at large argument H. It turns out that such an expansion is also sufficient in the whole weak coupling phase, and that we do not need the exact expression of $u(H)$ [204]: this is because the critical coupling β_c is large enough so that V is of order one and H large for $\beta > \beta_c$. One may write

$$u(H) \sim H + C - A_0 \ln H + A_1/H + A_2/2H^2 + \cdots \tag{4.46}$$

A_0 counts the number of infinitesimal generators of the Lie algebra, and the constants A_1 and A_2 have been computed in [176]. The constant C is worked out using the explicit form of the integration measure given in appendix A (see table 12).

The mean field free energy is the saddle point value of S_{eff}. The internal (plaquette) energy is therefore

$$E = \{2/d(d-1)\}\partial F/\partial\beta = [(d-2)V^{*4} + 2V^{*2}]/d . \tag{4.47}$$

For $\beta \to \infty$, $H^* \sim 2\beta(d-1)$, $V^* \sim 1 - A_0/\{4\beta(d-1)\}$, hence

$$E = 1 - A_0/d\beta - \cdots \tag{4.48}$$

which is the correct large β behaviour, counting the number of degrees of freedom. Had we not fixed the gauge, we would have the wrong result

$$E = V^4 = 1 - A_0/\{\beta(d-1)\} - \cdots$$

which would have to be amended by the corrections.

Table 12
Asymptotic behaviour of $u(H) = \ln \int DU \exp\{(H/N)\,\mathrm{Re\,Tr}\, U\}$

SU(N)	$u(H) \sim H + \left[\ln \frac{\prod_1^{N-1} k!}{(2\pi)^{(N-1)/2}} + \frac{N^2-2}{2}\ln N\right] - \frac{N^2-1}{2}\ln H - \frac{N^2-1}{8H} - \frac{(2N^2-5)(N^2-1)}{48H^2} - \cdots$
U(N)	$u(H) \sim H + \left[\ln \frac{\prod_1^{N-1} k!}{(2\pi)^{N/2}} + \frac{N^2}{2}\ln N\right] - \frac{N^2}{2}\ln H + \frac{N^2}{8H} + \frac{N^4}{16H^2} + \cdots$
SU(N)/Z_N	$u(H) \sim H + \left[\ln \frac{\prod_1^{N-1} k!}{(2\pi)^{(N-1)/2}} + \frac{N^2-1}{2}\ln\frac{N^2-1}{2N} + \frac{1}{2}\ln N\right] - \frac{N^2-1}{2}\ln H + \frac{(N^2-1)^2}{16H} + \frac{(N^2+2)(N^2-1)^3}{64N^2H^2} + \cdots$

Computation of the one-loop corrections to the non-trivial saddle point proceeds in a standard way. It requires the functions

$$\partial^2 u(H)/\partial H_{\beta\alpha}\,\partial H_{\xi\gamma}|_{H=H_1} = \partial^2 u(H)/\partial H^\dagger_{\beta\alpha}\partial H^\dagger_{\xi\gamma}|_{H=H_1} = A\,P^{(0)}_{\alpha\beta,\gamma\xi} + B\,P^{(A)}_{\alpha\beta,\gamma\xi} \tag{4.49}$$

and

$$\partial^2 u(H)/\partial H_{\beta\alpha}\partial H^\dagger_{\xi\gamma}|_{H=H_1} = A'\,P^{(0)}_{\alpha\beta,\gamma\xi} + B'\,P^{(A)}_{\alpha\beta,\gamma\xi} \tag{4.50}$$

where $P^{(0)}$ and $P^{(A)}$ are the projectors on the trivial and adjoint representations

$$\begin{aligned} P^{(0)}_{\alpha\beta,\gamma\xi} &= (1/N)\delta_{\alpha\beta}\delta_{\gamma\xi} \\ P^{(A)}_{\alpha\beta,\gamma\xi} &= \delta_{\alpha\xi}\delta_{\gamma\beta} - (1/N)\delta_{\alpha\beta}\delta_{\gamma\xi}\,. \end{aligned} \tag{4.51}$$

As a consequence of the invariance of $u(H)$ under a unitary transformation on H and $H^\dagger$, it is easy to see that, for all H,

$$A' - A = B' - B = u'(H)/2NH \tag{4.52}$$

(the first equality does not hold for SU(N)). Moreover, for large H,

$$B \sim -B' \sim -1/4NH$$

and

$$\begin{aligned} &A \sim -A' \sim -1/4NH \qquad \text{for U}(N) \\ &A \sim A' \sim \mathcal{O}(1/H^2) \qquad \text{for SU}(N)\,. \end{aligned} \tag{4.53}$$

It is thus natural to project the fluctuations of H and V about their mean field values H^* and V^*, on the trivial and adjoint representations. Moreover, their Hermitian and anti-Hermitian parts are not coupled to each other in the Gaussian approximation: the anti-Hermitian parts, coupled to $A - A'$ or $B - B'$, give the dominant contribution to the large β (large H) limit.

The case of SU(2) deserves a special mention. If one uses the representation of SU(2) matrices in terms of Pauli matrices

$$U = u_0 + \mathrm{i}\boldsymbol{u}\cdot\boldsymbol{\sigma} \qquad u_0^2 + \boldsymbol{u}^2 = 1,\ u \text{ real}$$

and write

$$H = h_0 - \mathrm{i}\boldsymbol{h}\cdot\boldsymbol{\sigma}, \qquad V = v_0 + \mathrm{i}\boldsymbol{v}\cdot\boldsymbol{\sigma}$$

with complex h, v, one sees that

$$u(H) = \ln \int \mathrm{D}U \exp \mathrm{Re}(h_0 u_0 + \boldsymbol{h}\cdot\boldsymbol{\sigma})$$

does not depend on the imaginary part of h. In the loop expansion, integration over Im h gives a delta-function for Im V and these degrees of freedom decouple to all orders.

Finally, the second derivatives of Wilson's action with respect to the link variables introduces two bond-shifting operators, the Fourier transforms of which read

$$\Delta_{\mu\nu}(k) = \left[2\cos k_d + 2V^2 \sum_{\rho\neq\mu,d} \cos k_\rho\right] \delta_{\mu\nu} + 4V^2(1-\delta_{\mu\nu}) \exp[i(k_\mu - k_\nu)/2] \sin(k_\mu/2) \sin(k_\nu/2)$$

$$(\mu, \nu = 1, \ldots, d-1) \quad (4.54)$$

and

$$\bar{\Delta}_{\mu\nu}(k) = \left[2\cos k_d + 2V^2 \sum_{\rho\neq\mu,d} \cos k_\rho\right] \delta_{\mu\nu} + 4V^2(1-\delta_{\mu\nu}) \exp[i(k_\mu - k_\nu)/2] \cos(k_\mu/2) \cos(k_\nu/2)$$

$$(\mu, \nu = 1, \ldots, d-1). \quad (4.55)$$

Δ and $\bar{\Delta}$ are relevant for the anti-Hermitian and Hermitian parts of the fluctuations, respectively. Straightforward algebra then leads to the one-loop result

$$Z = \exp\{S_{\text{eff}}(H^*, V^*)\} \times \{\det{}^{-1/2}(\mathbb{1} - 2\beta N(A'-A)\Delta) \times \det{}^{-(N^2-1)/2}(\mathbb{1} - 2\beta N(B'-B)\Delta)$$

$$\times \det{}^{-1/2}(\mathbb{1} - 2\beta N(A'+A)\bar{\Delta}) \times \det{}^{-(N^2-1)/2}(\mathbb{1} - 2\beta N(B'+B)\bar{\Delta})\}^{\mathcal{N}}. \quad (4.56)$$

Because of eq. (4.53), only the first two determinants contribute to order β^0 in the free energy. Using eqs. (4.52), (4.45), one may write

$$2\beta N(A'-A) = 2\beta N(B'-B) = \beta V/H = 1/[2 + 2(d-2)V^2] \quad (4.57)$$

(for SU(N), the first equation is absent), so that the one-loop free energy reads

$$F = F_{\text{mf}} - \begin{Bmatrix} N^2/2 \\ (N^2-1)/2 \end{Bmatrix} \operatorname{tr} \ln(\mathbb{1} - (\beta V/H)\Delta) + \mathcal{O}(1/\beta) \quad (4.58)$$

where $N^2/2$ refers to U(N) and $(N^2-1)/2$ to SU(N).

Several comments are now in order.

i) It is of course possible to pursue the loop-wise expansion to higher orders. The important point is that only a finite number of terms in the expansion contribute a given order in $1/\beta$. This results from the behaviour of successive derivatives of $u(H)$ (see e.g. (4.53)) and from a simple power-counting argument. For example, only one-loop and two-loop diagrams contribute to order $1/\beta$ in F, and so on ... This has to be contrasted with the case of discrete gauge groups, say Z_2 for definiteness, where one may show that an infinite number of terms are expected to contribute to the second correction $\exp\{-8(d-1)\beta\}$. Therefore, one may say that, for continuous groups, the mean field approximation (including the corrections) resums the weak coupling expansion, and moreover does this resummation in a very efficient way. Fig. 33 displays the plaquette energy of the four-dimensional SO(3) theory (the present discussion may be easily extended to SO(3) or any other group).

ii) The spectrum of excitations in the mean field approximation may be read off the expressions (4.54)–(4.56), since the argument of the determinants is their propagator. For instance, $1 - (\beta V/H)\Delta$ has

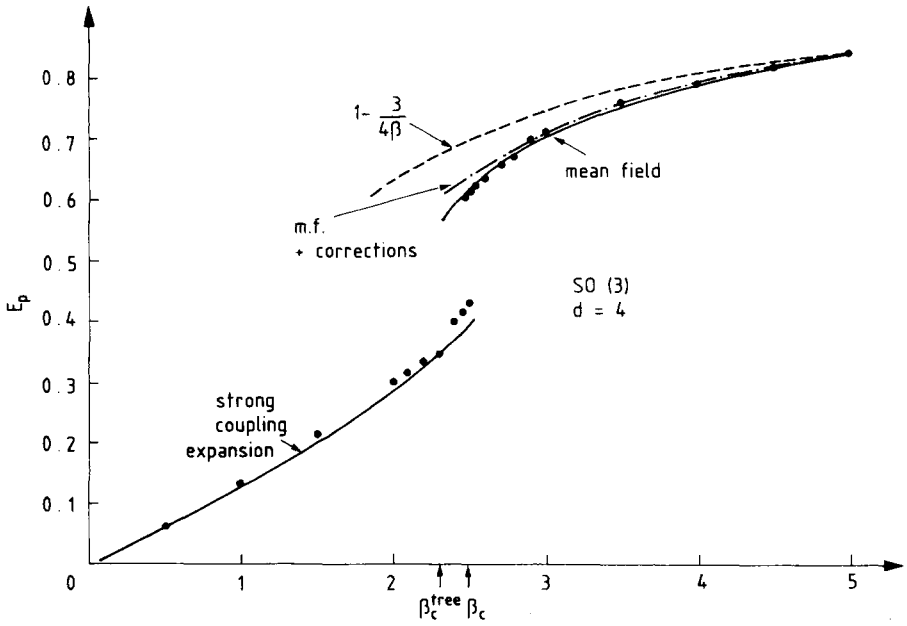


Fig. 33. The SO(3) plaquette energy. Monte-Carlo data [42] (full circles) compared to the strong and weak coupling expansions and to the mean field solution.

two eigenvalues, $\lambda_1 = (4\beta V/H)\sin^2(k_d/2)$, and $\lambda_2 = (4\beta V/H)\,(\sin^2(k_d/2) + V^2 \Sigma_1^{d-1} \sin^2(k_\mu/2))$, the latter $(d-2)$ times degenerate; λ_1 vanishes at $k_d = 0$, which corresponds to "time"-independent gauge transformations allowed in the axial gauge, whereas λ_2 vanishes only at $\boldsymbol{k} = 0$, and represents the other global transformations. Both are massless excitations. On the contrary, the other (Hermitian) channels have no massless excitations. That the most relevant channels in the large β limit correspond to the anti-Hermitian parts of the fluctuations, and give a zero mass to these excitations should be of no wonder to us, and fits with the continuum theory.

iii) Had we not fixed the gauge, we would have found $(1 - \{1/2(d-1)V^2\}\Delta')$ as the argument of the determinants in eqs. (4.54–56), with $\Delta'_{\mu\nu}$ a $d \times d$ matrix obtained from Δ by replacing $\cos k_d$ by $V^2 \cos k_d$. This matrix has a zero mode $\sin(k_\nu/2)\exp(ik_\nu/2)$, corresponding of course to local gauge transformations.

We now come to the determination of the transition point between strong and weak coupling phases. This is achieved by comparing the various free energy solutions provided by mean field plus corrections. In the strong coupling phase, where the mean field is trivial, we have seen (section 4.1.4) that loop corrections reconstruct the strong coupling expansion discussed at length in section 3. At weak coupling, one may take the one-loop expression of (4.56); the "Hermitian" channels play a minor role, and the determinant is a slowly varying function of β. For most gauge groups SU(N), $N \geq 4$, U(N), all N, SO(3), ..., this method gives a value to the (first order) phase transition in very good agreement with the Monte-Carlo simulations. The results are presented in table 13. The character expansions of section 3 have been used and suitably extrapolated. The small, but systematic discrepancy between the mean-field estimate and Monte-Carlo data should be corrected if a small positive two-loop contribution is added to F in the weak coupling phase. For illustration, the two free energy determinations are plotted in fig. 34 for U(∞). Of course, for U(1), the phase transition is incorrectly predicted to be first

Table 13
Comparison of mean-field and Monte-Carlo results for the position of the transition

		Mean field	Monte-Carlo	
$d = 4$	U(1)	1.03	1.01	[205]
	U(2)	3.45	3.30 ± 0.05	[205]
	U(3)	7.30	6.88 ± 0.5	[205]
	U(4)	12.5	12.14 ± 0.7	[205]
	U(5)	19.5	18.8 ± 1.1	[205]
	U(6)	28.0	27.0 ± 2.6	[205]
	U(∞)	$\beta/2N^2 = 0.38$	0.375–0.38	(extrap.)
	SU(4)	11.6	10.2	[206, 207]
	SU(5)	18	16.4 ± 0.2	[206, 208]
	SU(6)	26.5	24.0 ± 1.0	[209]
	SO(3)	2.6	2.48	[42]
$d = 5$	SU(2)	1.77	1.64	[210]

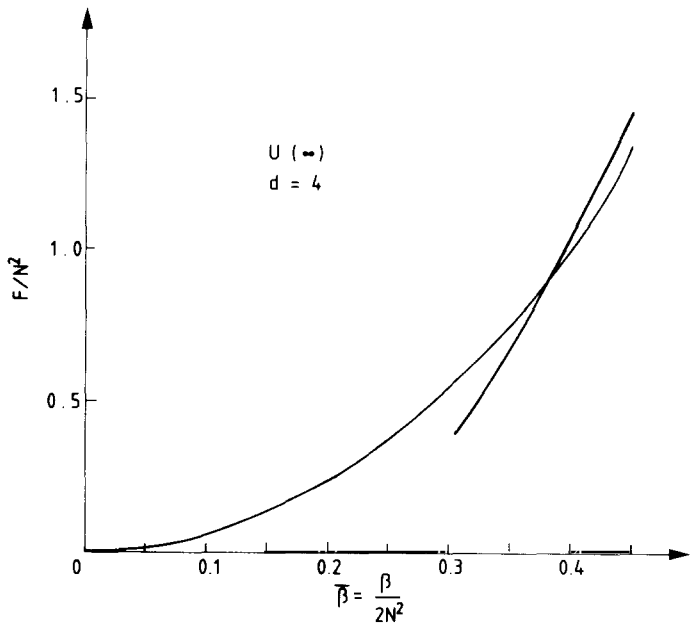


Fig. 34. Four-dimensional free energy for U(∞) gauge group.

order, but we cannot expect mean field to reproduce the physics of monopole loop condensation. The most striking effect takes place in SU(2) or SU(3) where the strong and weak coupling determinations do not cross and are almost degenerate in a range of β (fig. 35); the corrections to mean field have removed the first order transition predicted to lowest order. Of course, it is an exaggeration to say that mean field predicts the absence of a phase transition, because the bad convergence of the strong coupling expansion and possible higher order contributions in the weak coupling phase limit the accuracy. However, it is quite gratifying to see the consistency of the method with the "experimental" situation.

The computation of corrections has also been applied to the SU(2) model with a mixed action. The mean field incorporates the fluxon configurations discussed at the end of section 4.2.2. The resulting

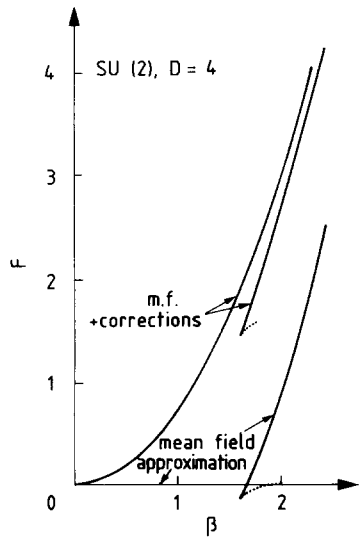


Fig. 35. Mean field plus corrections for four-dimensional SU(2) case.

phase diagram reproduces very well all the features of fig. 6, with all the transition lines shifted a little to the right, as in the previous case [193].

The computation described so far has been carried out in the axial gauge. Gauge invariant quantities should of course not be affected by the choice of gauge, but gauge invariance does not hold order by order in the loop expansion. For aesthetical reasons (covariance) or for their suitability in the computation of higher order corrections, other choices of gauge have been proposed [188, 211]. The stationarity equations are solved in the absence of gauge fixing and the gauge condition is imposed on the fluctuations; for example, the Landau gauge is a constraint on the anti-Hermitian part of the V fluctuations

$$\sum_{\mu} [(v_{n\hat{\mu}} - v_{n-\hat{\mu},\hat{\mu}}) - \text{h.c.}] = 0 .$$

Such a constraint must generally be accompanied by the appropriate Faddeev–Popov determinant, which also contributes to the corrections. So far, actual computations have only been carried out for U(1). It has been verified that the correct large β behaviour of $E(\beta)$ is recovered, although the original mean field approximation seems worse than in the axial gauge.

Finally, the mean field approach has been recently applied to the puzzling Higgs–Coulomb transition of Z_n gauge models. We recall from section 2.3.3 that these models possess three phases. For n large, Z_n looks like U(1), and the confinement-Coulomb phase is well reproduced in a standard mean field picture. However, the second transition requires a more elaborate discussion. It has been shown [186] that those fluctuations around mean field which are U(1) gauge transformations play an important role. When $n \to \infty$, Z_n is supposed to approach U(1), and these excitations should become a zero mode. Actually, at β finite, their eigenvalue (mass) is exponentially small in n, whereas at $\beta \sim \mathcal{O}(n^2)$, it is large. This suggests to separate these U(1) degrees of freedom from the Gaussian fluctuations, and treat them as collective coordinates. Using this method, Alessandrini has been able to reproduce qualitatively the behaviour for n large of the transition point $\beta \sim 0.02n^2$ observed in Monte-Carlo simulations [128].

4.2.5. Conclusions

The mean field approximation in lattice gauge theories has been considered with much suspicion for many years, because of its seemingly irremediable conflict with Elitzur's theorem. We hope to have convinced the reader that this was not founded, and that this approximation gives an accurate way of analyzing the local properties of gauge systems. The phase diagram and local observables [189] are well reproduced and it seems that grinding out a few higher order corrections would still improve the agreement. On the other hand, it is clear that long distance physics, the approach to the continuum limit and its scaling properties are not expected to be within the reach of the approximation. However, it might be that mean field with its corrections reproduces this limit to some extent, as strong coupling approximations do. Also, the fact that scaling already seems to take place at rather small distances leaves open this possibility.

5. Higgs fields

5.1. Generalities

The preceding sections have been devoted to the pure gauge models. However, a realistic theory necessarily includes matter fields. As bosons appear to be simpler and perhaps better understood than fermions, we restrict ourselves to this case.

It is known in field theory that the inclusion of matter fields can exert a dramatic effect on the behaviour of gauge theories. The Higgs mechanism, where scalar fields interact with gauge bosons, makes them massive and cuts the forces to a short range. In the lattice regularization, the occurrence of such a phenomenon is expected to be seen in the phase diagram of the combined matter-gauge model. This section is devoted to some discussion of this system.

The lattice Higgs model is defined as follows. We introduce on each lattice site a scalar field ϕ_i with d_r components. It transforms under a gauge transformation according to some d_r-dimensional irreducible representation of G

$$\phi_i \to D^r(g_i)\,\phi_i\,. \tag{5.1}$$

In view of this gauge transformation, it is natural to split these fields into a subset of gauge parameters for the orbit, and a subset of complementary parameters ("radial excitations"). A little thought shows that only the first subset interacts with the gauge fields, and that only the second subset can be involved in the local (= depending on only one site) gauge invariant potential terms of the action. Even in the absence of gauge interactions, i.e. with global transformations (5.1) g_i independent of i, the interaction between the two subsets leads to a very rich structure (related in particular with σ- and CP(n)-models in the continuum limit). When gauge fields are added, we expect an even more complex structure which has remained unexplored so far. Therefore, we restrict here the variations of the Higgs fields to a single orbit, without radial excitations. For instance, with the Z_2 gauge group, ϕ_i can take only the discrete values ± 1 instead of any real value; with the U(1) group, ϕ_i is constrained to be a unimodular complex number (or, equivalently, a two-dimensional unit vector); more generally, with the metric preserving groups SO(N) (resp. SU(N)), ϕ_i is a unit real (resp. complex) N-component vector. Also, we do not treat here the case of several Higgs fields in the system.

A new term γS_H involving Higgs fields must be added in the action to the pure gauge action βS_g, extensively studied in the previous sections (we add now a subscript g for gauge). The restriction to a

single gauge orbit suppresses the terms in S_H depending on only one site; keeping only nearest neighbour terms, we are led to the following gauge invariant form

$$\gamma S_H = \gamma \sum_{(i,j)} \phi_i D^r(U_{ij}) \phi_j , \tag{5.2}$$

and we now consider the generating functional

$$Z(J,j) = \int \prod \mathrm{D}\phi_i \int \prod \mathrm{D}U_l \exp\left\{\beta S_g + \gamma S_H + \sum_l \mathrm{tr}(J_l U_l) + \sum_i j_i \cdot \phi_i\right\} . \tag{5.3}$$

The invariant measure $\mathrm{D}\phi_i$ on Higgs fields is naturally induced by the Haar measure on G, for ϕ_i can be parametrized by an element g_i of the gauge group

$$\phi_i = D^r(g_i)\,\phi_0 \tag{5.4}$$

using a fixed, arbitrarily chosen vector ϕ_0; in this parametrization, the measure is clearly $\mathrm{D}g_i$.

The g_i's of the parametrization (5.4) define a transformation, which fixes the gauge in a given configuration of Higgs fields. This is called the unitary gauge. Performing this transformation on the U_l's, the partition function reduces to

$$Z = \int \prod \mathrm{D}U_l \exp\left\{\beta S_g + \gamma \sum \phi_0 D^r(U_{ij})\,\phi_0\right\} . \tag{5.5}$$

The integrand depends no longer on ϕ_i's and the corresponding integration has thus been carried out. Hence, the model is equivalent to a (non-gauge-invariant!) pure gauge system in an external constant field.

Before discussing the (β, γ) phase diagram, it is instructive to study various limiting cases.

a) $\gamma = 0$ is, of course, the well known pure gauge model.

b) $\beta = 0$ is a trivial model of non-interacting gauge fields (as seen in the unitary gauge formula (5.5)) and has no transition when γ varies.

c) $\beta = \infty$ is the pure statistical spin system with an action $\gamma \Sigma\, \phi_i \cdot \phi_j$. Indeed, this limit selects only the gauge field configurations maximizing S_g, i.e. with all U_p set to unity; in an appropriate gauge, all U_l are set to unity, hence the result. We recall quickly some properties of these models, extensively studied in statistical mechanics. Among them are the Ising model ($\phi = \pm 1$), the xy-model (2-dimensional real ϕ), the Heisenberg and generalized models (3- and more dimensional real ϕ). They generally present a second order transition for sufficiently high dimension: 2 and above for Ising, 3 and above for the others. The 2-dimensional xy-model is rather peculiar and exhibits a continuous order transition line for $\gamma > \gamma_c$ [212]. The phases are characterized by the mean value of the Higgs field, vanishing identically only in the low coupling (high temperature γ^{-1}) phase. The large distance behaviour of the correlation function is

$$\langle \phi_0 \phi_r \rangle \sim \mathrm{Cst} \exp(-r/\xi) \qquad \gamma < \gamma_c \qquad (\langle \phi \rangle \equiv 0) \tag{5.6}$$

$$\langle \phi \rangle^2 + \mathrm{Cst} \exp(-r/\xi) \qquad \gamma > \gamma_c, \text{ discrete groups} \tag{5.7}$$

$$\langle \phi \rangle^2 + \mathrm{Cst} \exp(\mathrm{Cst}'/r^{d-2}) \qquad \gamma > \gamma_c, \text{ continuous groups} . \tag{5.8}$$

The last behaviour (5.8) is a consequence of the existence of Goldstone's modes (spin waves) in the spontaneously broken model. Finally, note that the phase structure may be richer, for instance, Z_n models ($n > 4$) exhibit three phases, separated by two second order transitions and characterized by the successively encountered behaviours (5.6), (5.8) and (5.7) when γ increases.

d) $\gamma = \infty$ selects, in the unitary gauge formula (5.5), the values of U_l maximizing the matter part of the action. They form a subgroup $H \subset G$ and thus the limit is the pure gauge model for H. When the representation r is the fundamental one, H contains only the unity; the system is trivial and, in particular, does not undergo phase transitions. Other representations, hereafter referred to as multiply charged Higgs systems, lead to non-trivial cases. Examples are provided by the U(1) system with charge 2 Higgs field

$$\beta \sum \cos(\theta_{ij} + \theta_{jk} + \theta_{kl} + \theta_{li}) + \gamma \sum \cos 2(\phi_i + \theta_{ij} - \phi_j) \tag{5.9}$$

or by the SU(2) system with isospin 1 matter

$$\beta \sum \operatorname{tr} U_p + \gamma \sum \phi_i D^1(U_{ij}) \phi_j \tag{5.10}$$

(ϕ_i are 3-dimensional unit vectors onto which acts the SO(3) rotation D^1). The unitary gauge selects, at large γ, the gauge configurations $\theta_{ij} = 0$ or π, or $U_l = \pm 1$; the system thus reduces to the pure Z_2 gauge model for both cases, with its first order transition.

5.2. The phase diagram

Fig. 36 displays the expected phase diagrams in Higgs systems for various cases. These qualitative sketches are suggested by the limiting cases studied in the previous subsection and will be confirmed by the results obtained later. The three expected phases are

a) β and γ both large: the Higgs field takes a non-zero mean value and the gauge field fluctuates around 1 (in the unitary gauge). This is a Higgs mechanism, with massive gauge bosons. The force is short ranged and the Wilson loop exhibits a perimeter fall-off.

b) β large, γ small: finite energy states representing free charges are observed. For continuous groups, the gauge bosons have massless excitations and cause a Coulomb force between static charges.

c) β, γ both small: this is the confinement phase. One might expect an area law for the Wilson loop; however, the introduction of matter fields has modified drastically the situation. Indeed, the separation of static charges with a linearly growing potential is balanced by the energetically favoured creation of a particle–antiparticle pair from the vacuum. This is always possible if the Higgs fields are in the fundamental representation; a gauge-invariant singlet can be constructed from the static source and Higgs particles, and this results in a perimeter law for the Wilson loop. Conversely, with multiply charged Higgs fields, fundamental sources cannot be screened and the fundamental Wilson loop keeps an area fall-off. In this case, the confinement phase is characterized by Wilson's criterion, and is certainly non-connected with other phases (fig. 36, c–d).

The problem of finding order parameters that discriminate between these phases has been recently reconsidered. Mack and Meyer [213] propose to look at the vortex free energy, while Bricmont and Fröhlich [214] have pointed out that only in phase (b), the gauge invariant correlation function has a

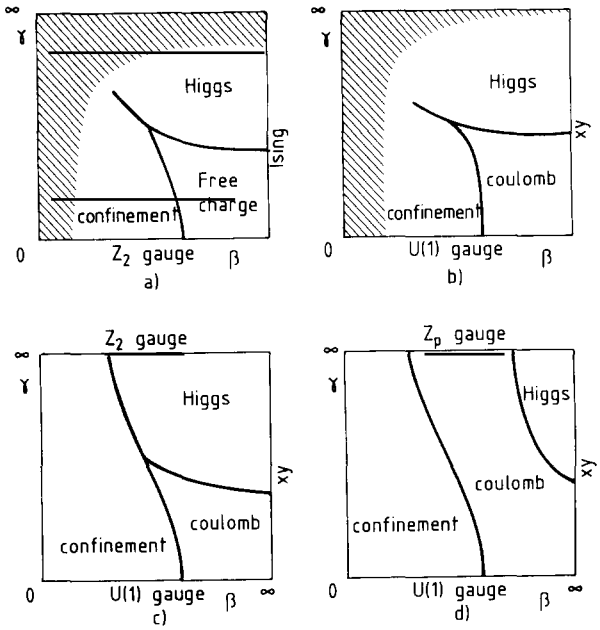


Fig. 36. The expected phase diagrams of the Higgs system. a) Z_2, b) U(1), charge 1 Higgs, c) U(1), charge 2 Higgs, d) U(1), charge $p > 4$ Higgs. The analyticity domain resulting from eqs. (5.14–5.15) is shaded.

power law correction on top of its exponential fall-off

$$\left\langle \phi_0 \prod U_l \phi_x \right\rangle \underset{|x|\to\infty}{\sim} |x|^{-(d-1)/2} \, e^{-m|x|} .$$

With singly charged matter fields, the failure of Wilson's criterion leads one to wonder about the differences between phases a and c. In fact, the two phases are continuously connected by an analyticity domain, which proves their identity. The proof [215, 216] uses an extension of the convergence theorem presented in section 3.1.2. The introduction of Higgs fields is performed in the unitary gauge by just modifying the measure

$$DU_l \to D_H U_l = e^{\gamma S(U)} DU_l \Big/ \int e^{\gamma S(U)} DU_l , \tag{5.11}$$

and the convergence proof is done for the strong coupling expansion around $\beta = 0$ for fixed γ. The inequality (3.7) has to be improved to get a sensible result, and the bound (3.2) is too crude. For the free energy, Hölder inequality allows one to write

$$\left| \int \prod D_H U_l \prod \Omega(U_p) \right| \le C_2 \left\{ \left[\int D_H U \, |\Omega(U)|^p \right]^{1/p} \right\}^{|D_1|} \tag{5.12}$$

where p is a given positive integer. This improved inequality shows immediately that the analyticity domain includes the region

$$16(d-1)\left[\int \mathrm{D}_{\mathrm{H}}U\,|\Omega(U)|^p\right]^{1/p} < 1 \tag{5.13}$$

in which the series are absolutely bounded from above by a convergent geometric series. In the Z_2 case, this yields

$$8(d-1)\{|\mathrm{e}^{-2\beta}-1|^p \cosh 2\gamma \cosh^{-4}\gamma\}^{1/p} < 1 \tag{5.14}$$

hence the shaded analyticity domain displayed in fig. 36a. Note in particular the finite width on both extremities of this domain, observed for any discrete gauge group. This completes the proof of connectivity† between the Higgs and confinement phases for the Z_2 gauge group and, similarly, for all discrete gauge groups.

The estimation of (5.13) is more complex for continuous groups as γ increases, and it is no longer sure that the analyticity domain extends up to ($\gamma = \infty$, $\beta = \infty$). However, this result still holds for U(1) (see also for this group [217]), where the simple upper bound $\Omega(U) < \text{Const.}\,\beta S_{\mathrm{g}}(U)$ leads to the result displayed in fig. 36b

$$\beta^{2p}\int_0^{2\pi} \mathrm{d}\theta\, \mathrm{e}^{\gamma\cos\theta}|\cos\theta - 1|^{2p} \Big/ \int_0^{2\pi} \mathrm{d}\theta\, \mathrm{e}^{\gamma\cos\theta} < \text{Const}. \tag{5.15}$$

Note that this region of analyticity has no longer a finite width near ($\beta = \infty$, $\gamma = \infty$).

Another important exact result concerns the transitions observed in the limiting models: they are not isolated, but connected to transition lines inside the phase diagram. This stability theorem is again based on convergence properties of expansions. Let us first consider the small γ region. In a fixed configuration of gauge fields, the Higgs system appears as a spin model with varying couplings. The first diagram appearing in the small γ expansion is the closed curve bounding a plaquette; at order γ^4, the resulting effective action for the gauge fields is therefore

$$S_{\mathrm{eff}} = (\beta + \tanh^4\gamma)\sum_p \mathrm{Tr}\, U_p + \mathcal{O}(\gamma^6)\,. \tag{5.16}$$

We specialize to the Z_2 case, although the result is quite general. Hence the system at small fixed γ reduces to the pure gauge model with a shifted effective coupling $\beta_{\mathrm{eff}} = \beta + \gamma^4$. Therefore, the transition enters the phase diagram, with the curvature shown in fig. 36.

Next corrections do not modify this result, since the high temperature expansion in γ is convergent. However, even with these convergence properties, induced interactions might destabilize the system near a second order transition. But, first, the effective action has the same symmetries as the pure gauge model; hence the situation differs from the case where an external field removes the transition. Secondly, the long range couplings contained in the effective action are exponentially damped and are not expected to affect the transition.

†In fact, it seems that the proof is not complete. A first order transition line might end at the point $\beta = 0$, $\gamma = \infty$, and the analyticity domain just derived might apply to the metastable phase. A complete proof would show, for instance, the identity (after re-ordering) of the expansions around $\beta = 0$ at fixed γ and around $\gamma = \infty$ at fixed β.

We do not reproduce the similar reasoning for the transition near $\beta = \infty$. In the three-dimensional Z_2 model, this is particularly obvious, because the phase diagram is symmetric under the duality transformation $(\beta, \gamma) \to (\beta^*, \gamma^*)$:

$$\sinh 2\beta \sinh 2\gamma^* = \sinh 2\beta^* \sinh 2\gamma = 1 , \tag{5.17}$$

as discussed in subsection 2.3.

It seems that strong coupling expansions should be an effective tool for investigating the quantitative features of the phase diagrams, viz. predicting the location of the transition lines, and studying the critical behaviour at the end-point of the confinement-Higgs line. For this purpose, two variable expansions have been derived to 16th order in β and γ for the Z_2 Higgs model [218]. Unfortunately, the analysis of these series by either of the methods described in section 3.4.1 turned out to be quite delicate and did not yield any stable result. Even in high dimensions, where, according to the mean field picture, the critical behaviour should become algebraic, we could not pin point accurately the location of the end-point, nor extract reliable exponents. It seems that our expansions were not long enough to allow a detailed analysis of the complicated pattern depicted on fig. 36. It may be worth mentioning at this point that strong and weak coupling methods have been successfully applied in a class of Higgs models, with a Z_q symmetry, and Potts-like interaction [196]. Expanding about the limit $q = \infty$ made it possible to resum in a non-trivial way strong and weak coupling expansions. The resulting phase diagrams were found to be in good agreement with Monte-Carlo data, even for values of q as low as 2.

5.3. Mean field analysis

The mean field techniques can easily be applied to the Higgs system. In the early days of lattice gauge theory, they were used to predict the qualitative features of the phase diagram [5]. In the parametrization of the degrees of freedom, it might seem interesting to use the unitary gauge, where only one dynamical field remains. However, this parametrization, which gives, as one will see, good results in the large γ region, fails for small γ. Indeed, in this region, the Higgs field fluctuates wildly and so does the unitary gauge transformation. It is therefore natural to treat separately these degrees of freedom and to introduce mean fields for both the U_l and ϕ_i fields. On the other hand, this other technique uses redundant degrees of freedom, which may create spurious transitions; this is the case, as we shall see, for the Z_2 Higgs model: the transition line between the confined and Higgs phases does not end and separates completely these two regions.

For simplicity, we only describe the explicit calculations for the Z_2 Higgs model. Using first two mean fields h_i and H_l for the two fields ϕ_i and U_l, we look for a constant, translation invariant solution $h_i = h$, $H_l = H$. The variational free energy (4.8) per site reads

$$F_{\mathrm{MF1}} = \operatorname*{Max}_{h,H} \left\{ \mathrm{d}(u(H) - H\,u'(H)) + v(h) - h\,v'(h) + \beta \frac{d(d-1)}{2} u'^4(H) + \gamma d\,u'(H)\,v'^2(h) \right\} . \tag{5.18}$$

The test free energies u and v are the same in this Z_2 case

$$u(H) = \ln \cosh H, \qquad v(h) = \ln \cosh h , \tag{5.19}$$

and the extremality conditions read

$$\begin{cases} H = 2\beta(d-1)\,u'^3(H) + \gamma\, v'^2(h) \\ h = 2\gamma d\, u'(H)\, v'(h)\,. \end{cases} \tag{5.20}$$

The three previously described phases are easily identified in this system. The general solution $H \neq 0$, $h \neq 0$ yields non-zero average values for both gauge and Higgs fields, and therefore corresponds to the Higgs phase. There are also two other particular solutions $h \equiv 0$, $H \neq 0$ and $h \equiv 0$, $H \equiv 0$ corresponding to the free charge and confined phases. It is also clear that the solution $h \equiv 0$, $H \equiv 0$ is always an isolated maximum and therefore the confinement region is, in this approximation, separated from the other phases by a first order transition line; this is not the case of the two other phases, separated by a second order transition line obtained by requiring that the second derivative with respect to h of the right hand side of (5.18) vanishes. The numerical result of this study is displayed in fig. 37, curve MF1. As stated previously, the first order transition line does not terminate. The results are particularly good in the low γ region; we also display the numerical results [219] obtained using Monte-Carlo techniques for the triple point and the end-point.

Let us turn now to the mean field approximation in the unitary gauge. The variational free energy is now

$$F_{\mathrm{MF2}} = d \operatorname*{Max}_{H} \left\{ u(H) - H\,u'(H) + \beta \frac{(d-1)}{2} u'^4(H) + \gamma\, u'(H) \right\} \tag{5.21}$$

with the extremality condition

$$H = 2\beta\,(d-1)\,u'^3(H) + \gamma\,. \tag{5.22}$$

This equation has one or three solutions (note that $H = 0$ is no longer a solution as $\gamma \neq 0$); one finds a first order transition line (curve MF2 of fig. 37) which ends on a second order point. As expected, this method is not accurate in the small γ region, but gives a correct description of the system when γ is large. The end-point corresponds to a triple solution of eq. (5.22) and the critical values are

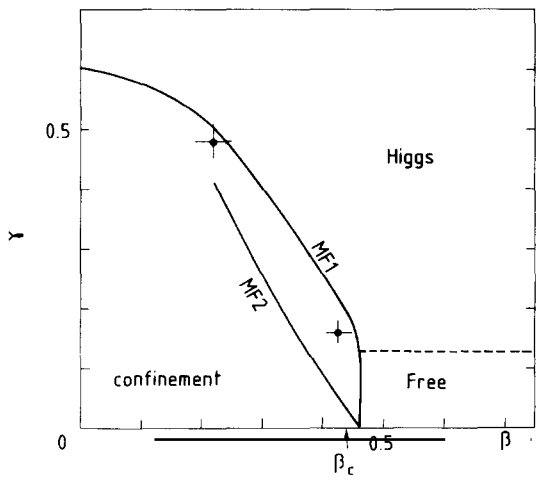


Fig. 37. The two mean field techniques applied to the Z_2 Higgs model. The numerically estimated locations of the triple point and of the end-point are also indicated with their error bars [219].

$$H_c = \ln(1+\sqrt{2})$$
$$(d-1)\beta_c = \tfrac{2}{3}$$
$$\gamma_c = -\sqrt{2}/3 + \ln(1+\sqrt{2})\,. \tag{5.23}$$

The continuous field theory describing the critical behaviour at long distance near this point has been identified [181]. One uses for this purpose the Legendre transform (4.18) which allows the calculation of the various irreducible Green functions

$$\Gamma(U) = \sum_l (H_l - \gamma)\, U_l - u(H_l) - \beta \sum_p U_p\,, \tag{5.24}$$

with H_l solution of

$$U_l = u'(H_l)\,. \tag{5.25}$$

The second derivative of this functional at the critical point gives the propagator

$$\Gamma^{(2)}_{l_1 l_2} = \delta^2\Gamma/\delta U_{l_1}\delta U_{l_2}|_c = 2\delta_{l_1 l_2} - \{1/3(d-1)\}P_{l_1 l_2}\,. \tag{5.26}$$

$P_{l_1 l_2}$ is one if l_1 and l_2 belong to the same plaquette, zero otherwise. In momentum space, this inverse propagator reads

$$\Gamma^{(2)}_{\mu\nu}(q) = \sum_r e^{iqr}\Gamma^{(2)}_{r\hat{\mu},r\hat{\nu}} = \frac{4d}{3(d-1)}\left(\delta_{\mu\nu} - \frac{1}{d}\right) + \frac{a^2}{3(d-1)}\left[\delta_{\mu\nu}(q^2 - 2q_\mu^2) + \tfrac{1}{2}(q_\mu^2 + q_\nu^2)\right] + \mathcal{O}(q^4a^4)\,. \tag{5.27}$$

$(d-1)$ of the eigenvalues of this matrix are degenerate and do not vanish at $q^2a^2 = 0$; they correspond to massive degrees of freedom which decouple in the continuum limit. The last eigenvalue does vanish at $a^2q^2 = 0$; using the projection operator P on the corresponding eigenvector $(1/\sqrt{d})(1, 1, \ldots, 1)$, we obtain the usual, rotation invariant massless inverse propagator

$$P\,\Gamma^{(2)}(q)\,P = (2a^2/3d^2)q^2 + \mathcal{O}(q^4a^4)\,. \tag{5.28}$$

The critical domain can therefore be described at long distance as a single component scalar field theory. Similar estimates of the 3- and 4-point functions show that only the latter does not vanish at the critical point; the continuum limit is therefore described by a ϕ^4 scalar field theory.

This simple example shows explicitly how the approach of a second order transition is related to a continuous field theory. In this case, the resulting limit is not particularly interesting; however, the method might be promising.

Appendix A. Group theory complements on U(N) and SU(N)

A.1. Representations of U(N) *and* SU(N)

Consider first the case of U(N). We use the following notations: U is a $N \times N$ unitary matrix with

eigenvalues $\varepsilon_j = \exp(i\alpha_j)$, $j = 1, \ldots, N$. The irreducible representations of U(N) are labelled by a set of N positive or negative integers

$$\{\lambda\} = \{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N\}, \qquad (\lambda_N \geq 0 \text{ or } < 0) \tag{A.1}$$

or alternatively if $l_i = \lambda_i + N - i$

$$l_1 > l_2 > \cdots > l_N . \tag{A.2}$$

The corresponding character reads [220]

$$\chi_{\{\lambda\}}(U) = \det(\varepsilon_i^{l_j})/\det(\varepsilon_i^{N-j}) \equiv \Delta_\lambda(\varepsilon)/\Delta_0(\varepsilon) \tag{A.3}$$

and its dimension is

$$d_{\{\lambda\}} = \prod_{i<j} (l_i - l_j)/(j - i) = \chi_{\{\lambda\}}(\mathbb{1}) . \tag{A.4}$$

If all the λ's are translated by the same integer s, we have of course

$$\chi_{\{\lambda+s\}}(U) = (\det U)^s \chi_{\{\lambda\}}(U) . \tag{A.5}$$

If all the λ's are non-negative (polynomial representations), it is natural to consider the associated Young tableau and the representation of the symmetry group Σ_n of $n = |\lambda| \equiv \Sigma \lambda_i$ objects. This leads to the Frobenius representation of $\chi_{\{\lambda\}}$ in terms of traces of powers of U. Let $[\rho]$ be a class of elements of Σ_n, denoted as usual [221]

$$[\rho] = [1^{\alpha_1} 2^{\alpha_2} \cdots n^{\alpha_n}] , \tag{A.6}$$

with $\Sigma p\alpha_p = n$, if it consists of α_1 1-cycles, α_2 2-cycles, ... There are

$$n_{[\rho]} = n!/(1^{\alpha_1} \alpha_1!\, 2^{\alpha_2} \alpha_2! \ldots) \tag{A.7}$$

elements in this class. We now introduce the notation

$$t_{[\rho]}(U) = (\operatorname{tr} U)^{\alpha_1} (\operatorname{tr} U^2)^{\alpha_2} \ldots (\operatorname{tr} U^n)^{\alpha_n} \tag{A.8}$$

and denote by $\hat{\chi}_{\{\lambda\}}([\rho])$ the character of Σ_n pertaining to the Young tableau $\{\lambda\}$. Frobenius relation reads

$$\chi_{\{\lambda\}}(U) = (1/n!) \sum_{[\rho]} n_{[\rho]} \hat{\chi}_{\{n\}}([\rho])\, t_{[\rho]}(U) , \tag{A.9}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0, \qquad n = \lambda_1 + \cdots + \lambda_N$$

or conversely, using the completeness relation of the $\hat{\chi}$,

$$t_{[\rho]}(U) = \sum_{\{\lambda\},|\lambda|=n} \hat{\chi}_{\{\lambda\}}([\rho])\, \chi_{\{\lambda\}}(U)\,. \tag{A.10}$$

In particular, for $[\rho] = 1^n$, $\hat{\chi}_{\{\lambda\}}(1) = \sigma_{\{\lambda\}}$,

$$\mathrm{tr}^n U = \sum_{\{\lambda\},|\lambda|=n} \sigma_{\{\lambda\}}\, \chi_{\{\lambda\}}(U)\,. \tag{A.11}$$

Therefore, the dimension $\sigma_{\{\lambda\}}$ of the representation $\{\lambda\}$ in Σ_n identifies with the number of times the representation $\{\lambda\}$ of U(N) appears in the nth tensor product of the fundamental representation $U^{\otimes|n|}$. One shows [220, 222] that

$$\sigma_{\{\lambda\}}/d_{\{\lambda\}} = |n|! \prod_{i=0}^{N-1} i!/(\lambda_{N-i} + i)!\,. \tag{A.12}$$

When some of the λ are negative, one may make use of (A.5) to write

$$\chi_{\{\lambda\}}(U) = (\det U)^{\lambda_N} \chi_{\{\lambda-\lambda_N\}}(U)$$

and then use (A.9) for $\chi_{\{\lambda-\lambda_N\}}$. However, this expression is not convenient in strong coupling expansions. In that context, one is rather interested in "low-lying" representations, so that χ may be expressed in terms of a small number of traces of powers of U and $U^\dagger$. If we denote

$$\{\lambda\} = m_1 \geq \cdots \geq m_p \geq 0 \geq \cdots \geq 0 \geq -n_1 \geq \cdots -n_q$$

which may be visualized by a double Young tableau (fig. 38), an expression of $\chi_{\{\lambda\}}$ in terms of $t_{[\rho]}(U)$ and $t_{[\rho]}(U^\dagger)$, involving no more than $\Sigma\, m_i$ U's and $\Sigma\, n_i$ $U^\dagger$'s can be given [223]. A list of low-lying representations with $|\lambda| \leq 3$ is given in table 14.

In the case of SU(N), the eigenvalues $\varepsilon_j = \exp(\mathrm{i}\alpha_j)$ have a product one. The representations are as in eqs. (A.1)–(A.3), but constrained by $\lambda_N = 0$. One can therefore use the expression (A.9) of χ, but, as noticed above, it is not always suited for strong coupling expansions.

A case of physical interest – may be the only one, ultimately!! – is the group SU(3). Its representations are commonly labelled by two integers λ, μ such that the notation of (A.1) reads

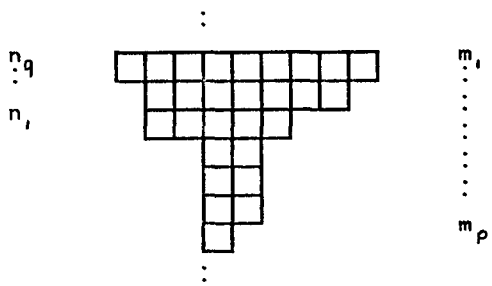


Fig. 38. Double Young tableau for U(N) representations.

Table 14
The presentations of U(N) with $|\lambda| \leq 3$. For the polynomial representations, we have also displayed $\sigma_{\{\lambda\}}$ defined in (A.11)

Representation		Dimension	σ	Character
{1}	(f)	N	1	tr U
{−1}	conjugate	N		tr $U^\dagger$
{2}		$N(N+1)/2$	1	$(\mathrm{tr}^2\, U + \mathrm{tr}\, U^2)/2$
{−2}	conjugate	$N(N+1)/2$		$(\mathrm{tr}^2\, U^\dagger + \mathrm{tr}\, U^{\dagger 2})/2$
{1, 1}		$N(N-1)/2$	1	$(\mathrm{tr}^2\, U - \mathrm{tr}\, U^2)/2$
{−1, −1}	conjugate	$N(N-1)/2$		$(\mathrm{tr}^2\, U^\dagger - \mathrm{tr}\, U^{\dagger 2})/2$
{1, −1}	adjoint	N^2-1		tr U tr $U^\dagger - 1$
{3}		$N(N+1)(N+2)/6$	1	$(\mathrm{tr}^3\, U + 2\,\mathrm{tr}\, U^3 + 3\,\mathrm{tr}\, U\,\mathrm{tr}\, U^2)/6$
{−3}	conjugate	$N(N+1)(N+2)/6$		$(\mathrm{tr}^3\, U^\dagger + 2\,\mathrm{tr}\, U^{\dagger 3} + 3\,\mathrm{tr}\, U^\dagger\,\mathrm{tr}\, U^{\dagger 2})/6$
{2, 1}		$N(N^2-1)/3$	2	$(\mathrm{tr}^3\, U - \mathrm{tr}\, U^3)/3$
{−2, −1}	conjugate	$N(N^2-1)/3$		$(\mathrm{tr}^3\, U^\dagger - \mathrm{tr}\, U^{\dagger 3})/3$
{1, 1, 1}		$N(N-1)(N-2)/6$	1	$(\mathrm{tr}^3\, U + 2\,\mathrm{tr}\, U^3 - 3\,\mathrm{tr}\, U\,\mathrm{tr}\, U^2)/6$
{−1, −1, −1}	conjugate	$N(N-1)(N-2)/6$		$(\mathrm{tr}^3\, U^\dagger + 2\,\mathrm{tr}\, U^{\dagger 3} - 3\,\mathrm{tr}\, U^\dagger\,\mathrm{tr}\, U^{\dagger 2})/6$
{2, −1}		$N(N-1)(N+2)/2$		$\mathrm{tr}\, U^\dagger(\mathrm{tr}^2\, U + \mathrm{tr}\, U^2)/2 - \mathrm{tr}\, U$
{1, −2}	conjugate	$N(N-1)(N+2)/2$		$\mathrm{tr}\, U\,(\mathrm{tr}^2\, U^\dagger + \mathrm{tr}\, U^{\dagger 2})/2 - \mathrm{tr}\, U^\dagger$
{1, 1, −1}		$N(N-2)(N+1)/2$		$\mathrm{tr}\, U^\dagger(\mathrm{tr}^2\, U - \mathrm{tr}\, U^2)/2 - \mathrm{tr}\, U$
{1, −1, −1}	conjugate	$N(N-2)(N+1)/2$		$\mathrm{tr}\, U\,(\mathrm{tr}^2\, U^\dagger - \mathrm{tr}\, U^{\dagger 2})/2 - \mathrm{tr}\, U^\dagger$

$$\{\lambda\} = \{\lambda + \mu, \mu, 0\} \equiv (\lambda, \mu)\,. \tag{A.13}$$

The dimension of that representation is

$$d_{(\lambda,\mu)} = \tfrac{1}{2}(\lambda + 1)(\mu + 1)(\lambda + \mu + 2)\,, \tag{A.14}$$

its triality is $2\mu + \lambda$ (mod. 3). One notices that

$$\begin{aligned}(\lambda, \mu) \otimes (1, 0) &= (\lambda + 1, \mu) \oplus (\lambda - 1, \mu + 1) \oplus (\lambda, \mu - 1) \\ (\lambda, \mu) \otimes (0, 1) &= (\lambda, \mu + 1) \oplus (\lambda - 1, \mu) \oplus (\lambda + 1, \mu - 1)\end{aligned} \tag{A.15}$$

where the second term appears only when $\lambda \neq 0$, and the third when $\mu \neq 0$. It is then easy to compute recursion formulae for the coefficients

$$T^{(n)}_{(\lambda,\mu)} = \int \mathrm{D}U\, (\chi_f + \chi_{\bar{f}})^n\, \chi^*_{(\lambda,\mu)} \tag{A.16}$$

which count the number of times the representation (λ, μ) occurs in the tensor product $\otimes^n\, (f + \bar{f})$:

$$T^{(n)}_{(\lambda,\mu)} = \sum_{\substack{\text{neighbours} \\ (\lambda',\mu')}} T_{(\lambda',\mu')} \tag{A.17}$$

where the sum runs over the neighbours of (λ, μ) that appear in (A.15). The contribution of those (λ', μ') for which λ' or μ' is negative is dropped.

A.2. Group integrals

If $U = \Omega \operatorname{diag}(e^{i\alpha}) \Omega^\dagger$, the measure of integration takes the form [224]

$$DU = \text{const.} \prod_i \frac{d\alpha_i}{2\pi} \prod_{j<k} |\exp(i\alpha_j) - \exp(i\alpha_k)|^2 \, D\Omega \tag{A.18}$$

for U(N), and

$$DU = \text{const.} \prod_i d\alpha_i \, \delta_P\Big(\sum \alpha_i\Big) \prod_{j<k} |\exp(i\alpha_j) - \exp(i\alpha_k)|^2 \, D\Omega \tag{A.19}$$

for SU(N), where $D\Omega$ is of no concern to us, since we only consider class functions. In (A.19), δ_P is the periodic delta function. The constants are adjusted by normalizing $\int DU = 1$. It is useful to notice that, with the notations of (A.3),

$$\prod_{j<k} (\exp(i\alpha_j) - \exp(i\alpha_k)) = \Delta_0(\varepsilon) \,. \tag{A.20}$$

Many integrals encountered in strong coupling expansions are easily computed using character orthogonality (3.32), (3.33), (3.35). However, more complicated integrals, e.g. $\int DU \chi_r(UV) \chi_s(UW) \chi_t(U)$ call for a different treatment. They may be expressed in terms of generalized Clebsh–Gordan coefficients, or using explicit expressions of the χ_r in terms of the fundamental representation (see section A.1), computed in terms of the generic integrals

$$\int DU \, U_{j_1 i_1} \cdots U_{j_n i_n} \, U^\dagger_{l_1 k_1} \cdots U^\dagger_{l_p k_p} \tag{A.21}$$

on which we now focus our attention.

In the case of U(N), invariance under the U(1) subgroup ensures that only $n = p$ integrals are non-vanishing. The generating function

$$F_n(A, B) = \int DU \, \mathrm{tr}^n AU \, \mathrm{tr}^n BU^\dagger \,, \tag{A.22}$$

with A and B two unitary matrices, is computed, using the expression (A.11) and the orthogonality of characters:

$$\begin{aligned} F_n(A, B) &= \int DU \sum_{\substack{\{\lambda\},\{\lambda'\} \\ |\lambda|=|\lambda'|=n}} \sigma_{\{\lambda\}} \sigma_{\{\lambda'\}} \chi_{\{\lambda\}}(UA) \chi_{\{\lambda'\}}(U^\dagger B) = \sum_{\substack{\{\lambda\} \\ |\lambda|=n}} (\sigma^2_{\{\lambda\}}/d_{\{\lambda\}}) \chi_{\{\lambda\}}(AB) \\ &= \sum_{\substack{[\rho] \\ \text{classes} \\ \text{of } \Sigma_n}} n_{[\rho]} \, t_{[\rho]}(AB) \sum_{\substack{\{\lambda\} \\ |\lambda|=n}} n!^{-1} (\sigma^2_{\{\lambda\}}/d_{\{\lambda\}}) \hat{\chi}_{\{\lambda\}}([\rho]) = n! \sum_{[\rho]} n_{[\rho]} \, C_{[\rho]} \, t_{[\rho]}(AB) \,. \end{aligned} \tag{A.23}$$

The polynomial nature of $t_{[\rho]}(AB)$ now permits a continuation to arbitrary $N \times N$ complex matrices A and B, and differentiation of the expression gives

$$\int \mathrm{D}U\, U_{j_1 i_1} \cdots U^{\dagger}_{l_1 k_1} \cdots = \sum_{\rho,\sigma \in \Sigma_n} C_{[\rho]}\, \delta_{i_1 l_{\rho\sigma_1}} \cdots \delta_{j_1 k_{\sigma_1}} \cdots \tag{A.24}$$

where the coefficient $C_{[\rho]}$ depends only on the class of the permutation ρ

$$C_{[\rho]} = n!^{-2} \sum (\sigma^2_{\{\lambda\}}/d_{\{\lambda\}})\, \hat{\chi}_{\{\lambda\}}([\rho])\,. \tag{A.25}$$

In this expression, as in (A.23), the summation $\Sigma_{\{\lambda\}}$ runs over all polynomial representations of U(N) with n boxes. If $n \leq N$, this may also be considered as a sum over all the representations of the symmetric group. The coefficients $C_{[\rho]}$ are easy to compute from the $\hat{\chi}$ given by Hamermesh [221]. We only list the first ones (see also [225])

$$\begin{aligned}
&n = 1 \quad && C_{[1]} = 1 \\
&n = 2 && C_{[1^2]} = 1/(N^2-1) \\
& && C_{[2]} = -1/N(N^2-1) \\
&n = 3 && C_{[1^3]} = (N^2-2)/N(N^2-1)(N^2-4) \\
& && C_{[1,2]} = -1/(N^2-1)(N^2-4) \\
& && C_{[3]} = 2/N(N^2-1)(N^2-4)\,.
\end{aligned} \tag{A.26}$$

They exhibit poles at $N^2 = 0, 1, \ldots, (n-1)^2$. These poles, known as de Wit–'t Hooft anomalies, come from the zeroes of $d_{\{\lambda\}}$, the dimensions of representations $\{\lambda\}$, and are, of course, not present in the actual result at $n > N$, obtained by keeping only representations with no more than N rows in (A.23), (A.25). Furthermore, for $n > N$, the tensors appearing in the right hand side of (A.24) are no longer independent, or, stated differently, tr U^p, $p > N$, may be re-expressed in terms of tr U^q, $q \leq N$. For example, in the case of U(2), with $n = 3$, $t_{[3]}(AB)$ may be re-expressed in terms of $t_{[1^3]}$ and $t_{[1\,2]}$ using the identity $\chi_{\{1,1,1\}}(AB) \equiv 0$. This gives

$$F_3(A,B) = 3!(n_{[1^3]}\, C^{\mathrm{U}(2)}_{[1^3]}\, t_{[1^3]}(AB) + n_{[1\,2]}\, C^{\mathrm{U}(2)}_{[1\,2]}\, t_{[1\,2]}(AB))\,, \tag{A.27}$$

with

$$C^{\mathrm{U}(2)}_{[1^3]} = \tfrac{1}{6}\,, \qquad C^{\mathrm{U}(2)}_{[1\,2]} = -\tfrac{1}{24}\,.$$

In the case of SU(N), the generic integral (A.21) vanishes unless $n - p$ is a multiple of N, as a consequence of Z_N-invariance. If $n = p < N$, as any polynomial representation of U(N) with $n < N$ boxes is also a representation of SU(N), the formulae (A.23)–(A.25) still apply. In the general case, however, one writes, with A, B two U(N) matrices,

$$\begin{aligned}
F_{n,p}(A,B) &= \int \mathrm{D}U\, \mathrm{tr}^n AU\, \mathrm{tr}^p BU^{\dagger} = \int \mathrm{D}U \sum_{\substack{\{\lambda\},|\lambda|=n \\ \{\lambda'\},|\lambda'|=n}} \sigma_{\{\lambda\}}\, \sigma_{\{\lambda'\}}\, \chi_{\{\lambda\}}(AU)\, \chi_{\{\lambda'\}}(BU^{\dagger}) \\
&= \int \mathrm{D}U \sum_{\{\lambda\},\{\lambda'\}} \sigma_{\{\lambda\}}\, \sigma_{\{\lambda'\}}\, \det{}^{\lambda_N} A\, \det{}^{\lambda'_N} B\, \chi_{\{\lambda-\lambda_N\}}(AU)\, \chi_{\{\lambda'-\lambda'_N\}}(BU^{\dagger}) \\
&= \sum \sigma_{\{\mu+\lambda_N\}}\, \sigma_{\{\mu+\lambda'_N\}}\, \det{}^{\lambda_N} A\, \det{}^{\lambda'_N} B\, \chi_{\{\mu\}}(AB)/d_{\{\mu\}}\,.
\end{aligned} \tag{A.28}$$

Now the sum runs over all representations of SU(N) (since $\mu_N = 0$) and over the non-negative integers λ_N and λ'_N, such that $|\mu| + \lambda_N \cdot N = n$, $|\mu| + \lambda'_N \cdot N = p$. In the case $p = 0$ (to which one might always restrict oneself thanks to the condition $\det U = 1$), one finds [226]

$$F_{n=qN}(A) = \int \mathrm{D}U \, \mathrm{tr}^{qN} AU = (qN!) \prod_{i=0}^{N-1} \frac{i!}{(q+i)!} \det{}^q A \,. \tag{A.29}$$

As before, these results extend to arbitrary $N \times N$ complex matrices A, and differentiating them yields the desired integral (A.21).

A.3. Character expansion coefficients

Using the previous machinery, calculations of character expansion coefficients are easy. For U(N), and Wilson's action, using (A.3), (A.11),

$$\begin{aligned}
\tilde{\beta}^{\mathrm{U}(N)}_{\{\lambda\}} &= \int \mathrm{D}U \exp\{(\beta/2N)\,\mathrm{tr}(U + U^\dagger)\}\, \chi^*_{\{\lambda\}}(U) \\
&= \text{const.} \int \prod_i (\mathrm{d}\alpha_i/2\pi)\, \Delta_0(\mathrm{e}^{\mathrm{i}\alpha})\, \Delta^*_l(\mathrm{e}^{\mathrm{i}\alpha}) \exp\Big\{(\beta/N) \sum \cos\alpha\Big\} \\
&= \text{const.}\, N! \sum_\sigma \varepsilon_\sigma \int \prod_j (\mathrm{d}\alpha_j/2\pi) \exp\{(\beta/N)\cos\alpha_j\} \exp\{\mathrm{i}\alpha(\lambda_{\sigma j} - \sigma j + j)\} \\
&= \det\{I_{\lambda_j - j + i}(\beta/N)\}
\end{aligned} \tag{A.30}$$

where the normalization constant has been determined by setting $\beta = 0$, $\{\lambda\} = 0$, $\tilde{\beta}_0 = 1$.

In the case of SU(N), we have an extra δ-function in the measure

$$2\pi \sum_k \delta\Big(\sum \alpha_i - 2k\pi\Big) = \sum_{n=-\infty}^{\infty} \exp\Big\{\mathrm{i}\, n \sum \alpha_i\Big\} \,. \tag{A.31}$$

The summation over n goes through the previous U(N) calculation and we are led to

$$\tilde{\beta}^{\mathrm{SU}(N)}_{\{\lambda\}} = \sum_{n=-\infty}^{\infty} \det\{I_{\lambda_j - j + i + n}(\beta/N)\} \,. \tag{A.32}$$

In the $N \to \infty$ limit, these expressions have a remarkably simple form, when $\bar{\beta} = \beta/2N^2$ is kept fixed. As shown by Gross and Witten,

$$\tilde{\beta}_0 = \mathrm{e}^w = \exp(\beta/2N)^2, \qquad \text{for } \bar{\beta} < 1 \,, \tag{A.33}$$

hence

$$N\beta_{\mathrm{f}} = N\, \mathrm{d}w/\mathrm{d}\beta = \beta/2N = N\bar{\beta} \,. \tag{A.34}$$

More generally, using the notations of section A.1, one may show that for the general representation $\{\lambda\} = \{m\,; n\}$

$$d_{\{\lambda\}}\beta_{\{\lambda\}} = (\sigma_{\{m\}}/|m|!)(\sigma_{\{n\}}/|n|!)(N\beta_{\mathrm{f}})^{|\lambda|}\,. \tag{A.35}$$

Moreover, as $N \to \infty$, the corrections to these expressions are exponentially small [227]. For example, for SU(N),

$$\ln \tilde{\beta}_0 = (\beta/2N)^2 + (2/N!)(\beta/2N)^N + \cdots .$$

However, after rescaling $\beta/2N = N\bar{\beta}$, this becomes

$$\ln \tilde{\beta}_0 = N^2\bar{\beta}^2 + 2(N^N/N!)\bar{\beta}^N\,, \tag{A.36}$$

and the coefficient in front of the "exponentially small" correction grows like $\mathrm{e}^N(2\pi N)^{-1/2}$.

Appendix B. Diagrammatic analysis of the large N limit

In order to prove that the free energy per site of the U(N) theory behaves as N^2 at strong coupling (see section 2.5), one may use the Dyson–Schwinger equations [77]. This infinite set of equations relating averages of loops of all sizes and shapes is sufficient to generate order by order the strong coupling expansion of all closed loops. For example, given the lowest order of the single plaquette

$$\langle N^{-1}\,\mathrm{tr}\; U_p\rangle = t + \cdots$$

it shows that all expectation values $\langle N^{-1}\,\mathrm{tr}\; U_C\rangle$ are also (at most) of order N^0 to any finite order in t. In the Wilson action for example, where the coupling is rescaled according to $\beta = 2N^2\bar{\beta}$, this entails that $F \propto N^2$

$$\langle N^{-1}\,\mathrm{tr}\; U_p\rangle = (2/d(d-1))\partial F/\partial\beta = (1/d(d-1))\partial(N^{-2}F)/\partial\bar{\beta}\,. \tag{B.1}$$

However, it may be interesting to have a purely diagrammatic derivation and to characterize the relevant diagrams. For this purpose, one may use the cumulant expansion of section 3.1.6. The partition function reads

$$Z = [1 + S(\partial/\partial J) + \tfrac{1}{2}S^2(\partial/\partial J) + \cdots]\exp\sum_l W(J_l, \bar{J}_l) \tag{B.2}$$

with

$$S(J) = \bar{\beta}N \sum_{\substack{\text{oriented}\\ \text{plaq.}}} \frac{\delta}{\delta J_{l_1}^{\alpha\beta}}\frac{\delta}{\delta J_{l_2}^{\beta\gamma}}\frac{\delta}{\delta \bar{J}_{l_3}^{\gamma\delta}}\frac{\delta}{\delta \bar{J}_{l_4}^{\delta\alpha}} \tag{B.3}$$

and

$$\mathrm{e}^{W(J,\bar{J})} = \int \mathrm{D}U \exp \mathrm{tr}(JU + \bar{J}U^{\dagger})\,. \tag{B.4}$$

The latter integral has been thoroughly studied by Brézin and Gross [228] in the large N limit. It is sufficient to say here that at strong coupling ($J\bar{J}$ small)

$$W(J,\bar{J}) = W(J\bar{J}) = \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\substack{\alpha,\ldots,\alpha_n \\ \sum p\alpha_p = n}} C_\alpha(N)\,[\mathrm{tr}\, J\bar{J}/N]^{\alpha_1}\,[\mathrm{tr}(J\bar{J})^2/N]^{\alpha_2}\cdots \tag{B.5}$$

where the cumulants $C_\alpha(N)$ are such that $N^{2n-2}\,C_\alpha(N)$ has a limit as $N\to\infty$. The first C_α have been computed explicitly by Bars [229], Brower and Nauenberg [230] and Samuel [225]. According to the previous rules, the free energy is computed in terms of connected diagrams made of P "4-vertices" $\mathrm{tr}((\delta/\delta J)(\delta/\delta J)(\delta/\delta\bar{J})(\delta/\delta\bar{J}))$ and of v_α "α-sites" $(\mathrm{tr}\, J\bar{J})^{\alpha_1}\cdots[\mathrm{tr}(J\bar{J})^n]^{\alpha_n}$. Contractions of matrix indices are conveniently visualized by using double lines following ref. [73] (fig. 39). For a gauge invariant quantity like F, all indices must be contracted, and a factor N is associated with each of the f closed loops of indices. For F, the total power of N is therefore

$$\# = P - \sum_\alpha v_\alpha \sum_k ((2k+1)\alpha_k - 2) + f\,.$$

One then uses the relation between v_α and the number l of lines joining sites to vertices

$$l = 2\sum_\alpha v_\alpha \sum k\alpha_k$$

and introduces the genus g of the c-connected surface built by the index loops

$$2c - g = f - l + \left(P + \sum_\alpha v_\alpha \sum_k \alpha_k\right) = f + P - \sum_\alpha v_\alpha \sum_k (2k-1)\,.$$

This enables one to write

$$\# = 2c - g + 2\sum_\alpha v_\alpha \left(1 - \sum_k \alpha_k\right) \le 2 - g \le 2\,. \tag{B.6}$$

The last inquality expresses the fact that the splitting of a "site" into $\Sigma\,\alpha_k$ traces may increase the

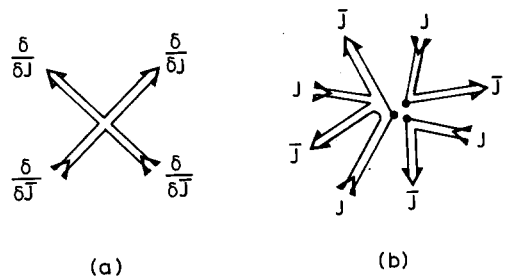


Fig. 39. Double-line representation of 4-vertices (a) and α-sites (b: $\alpha_1 = 2$, $\alpha_2 = 1$).

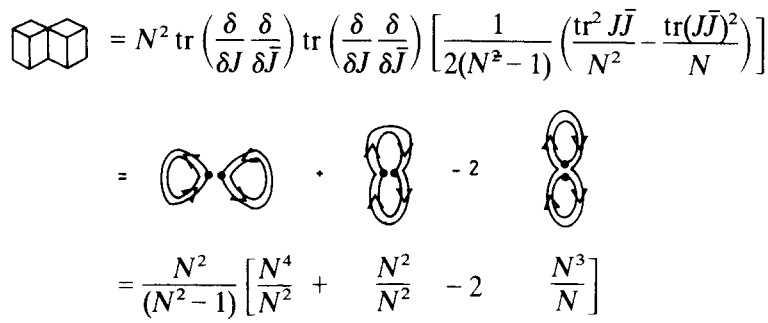


Fig. 40. Dominant diagrams in the large N limit. Computing the diagram depicted on the left amounts, after integration over the links common only to two plaquettes, to computing the expression displayed. In turn, this may be represented by graphs. The second one has $2[c + \Sigma v_a(1 - \Sigma \alpha_k)] = 0$, the two others have it equal to 2.

number of connected parts (fig. 40). From (B.6), we conclude that, as in the continuous theory, only planar graphs ($g = 0$) contribute to the leading ($\sim N^2$) behaviour of F. We emphasize that these graphs are abstract graphs expressing index contractions, and not diagrams made of plaquettes of the lattice. We also learn from (B.6) that corrections are suppressed by inverse powers of N^2.

This also implies the factorization property of the large N limit. Given two observables $\mathcal{O}_1$ and $\mathcal{O}_2$, one has

$$\langle \mathcal{O}_1 \mathcal{O}_2 \rangle = \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_2 \rangle (1 + \mathcal{O}(1/N^2)) . \tag{B.7}$$

Indeed, one may rescale $\mathcal{O}_1$ and $\mathcal{O}_2$ to make them of order N^2 and add them with sources in the action, without modifying the previous property

$$Z = \int \mathrm{D}U \exp\{\beta S + j_1 \mathcal{O}_1 + j_2 \mathcal{O}_2\} = \exp\{N^2 F(j_1, j_2)\} .$$

Then

$$\langle \mathcal{O}_1 \mathcal{O}_2 \rangle = Z^{-1} (\partial/\partial j_1)(\partial/\partial j_2) Z_{j_1 = j_2 = 0} = N^4 (\partial F/\partial j_1)(\partial F/\partial j_2) + N^2 (\partial^2 F/\partial j_1 \, \partial j_2)$$

hence the factorization property.

Acknowledgments

It is a pleasure to thank our colleagues and friends B. Berg, A. Billoire, C. Korthals-Altes, E. Marinari, G. Münster, A. Ukawa for interesting and stimulating discussions or correspondence during the preparation of this report. We also benefited from the critical reading and useful comments by P. Hasenfratz. We want to thank Professor K. Wilson for communicating us his unpublished strong coupling expansions. Special mention is reserved to our friend C. Itzykson who has been associated with most of our own research work reported here and contributed a lot to our understanding.

References

[1] K. Wilson, Phys. Rev. D10 (1974) 2445.
[2] A.M. Polyakov, unpublished.
[3] See e.g. H.E. Stanley, Introduction to phase transitions and critical phenomena (Clarendon Press, Oxford, 1971).
[4] See the series Phase transitions and critical phenomena, eds. C. Domb and M.S. Green (Academic Press, London).
[5] R. Balian, J.-M. Drouffe and C. Itzykson,, Phys. Rev. D10 (1974) 3376.
[6] R. Balian, J.-M. Drouffe and C. Itzykson, Phys. Rev. D11 (1975) 2098.
[7] R. Balian, J.-M. Drouffe and C. Itzykson, Phys. Rev. D11 (1975) 2104; D19 (1979) 2514(E).
[8] J.B. Kogut and L. Susskind, Phys. Rev. D11 (1975) 395.
[9] M. Creutz, L. Jacobs and C. Rebbi, Phys. Reports 95 (1983) 201.
[10] L.P. Kadanoff, Rev. Mod. Phys. 49 (1977) 267.
[11] J.-M. Drouffe and C. Itzykson, Phys. Reports 38C (1978) 133.
[12] J.B. Kogut, Rev. Mod. Phys. 51 (1979) 659.
[13] Cargese lectures 1976, New developments in quantum field theory and statistical mechanics, eds. M. Levy et al. (Plenum Press, 1977); Cargese lectures 1979, Recent developments in gauge theories, eds. G. 't Hooft et al. (Plenum Press, 1980).
[14] J.B. Kogut, in: Recent advances in field theory and statistical mechanics, Les Houches 1983 lectures, eds. J.B. Zuber and R. Stora (North-Holland) to appear.
[15] G.G. Batrouni, Nucl. Phys. B208 (1982) 467.
[16] D.J. Gross and E. Witten, Phys. Rev. D21 (1981) 446.
[17] N.S. Manton, Phys. Lett. 96B (1980) 328.
[18] J. Villain, J. Phys. (Paris) 36 (1975) 581.
[19] A.M. Polyakov, Phys. Lett. 72B (1978) 447.
[20] J.-M. Drouffe, Phys. Rev. D18 (1978) 1174.
[21] L. Susskind, Phys. Rev. D20 (1979) 2610.
[22] M. Stone, Nucl. Phys. B152 (1979) 97.
[23] M. Nauenberg and D. Toussaint, Nucl. Phys. B190[FS3] (1981) 217.
[24] P. Menotti and E. Onofri, Nucl. Phys. B190[FS3] (1981) 288.
[25] C.B. Lang, C. Rebbi, P. Salomonson and B.S. Skagerstam, Phys. Lett. 101B (1981) 173.
[26] K. Wilson, in: Recent developments in gauge theories, Cargese lectures 1979 (Plenum Press).
[27] R.C. Edgar, Nucl. Phys. B205 (1982) 345.
[28] K. Symanzik, 1981 Berlin Conf., Mathematical problems in theoretical physics, eds. R. Schrader et al., Springer 1982 Lecture notes in physics, Vol. 153.
[29] P. Weisz, Nucl. Phys. B212 (1983) 1.
[30] W. Celmaster, Phys. Rev. D26 (1982) 2955.
[31] J.-M. Drouffe and K.J.M. Moriarty, Nucl. Phys. B220[FS8] (1983) 253.
[32] N. Christ, R. Friedberg and T.D. Lee, Nucl. Phys. B202 (1982) 89; Nucl. Phys. B210[FS6] (1982) 310, 337.
[33] F. Wegner, J. Math. Phys. 12 (1971) 2259.
[34] S. Elitzur, Phys. Rev. D12 (1975) 3978.
[35] K. Osterwalder and E. Seiler, Ann. Phys. (NY) 110 (1978) 440.
[36] A. Guth, Phys. Rev. D21 (1980) 2291.
[37] G. 't Hooft, Nucl. Phys. B138 (1978) 1.
[38] M. Creutz, Phys. Rev. D21 (1980) 2308; Phys. Rev. Lett. 45 (1980) 463.
[39] B. Lautrup and M. Nauenberg, Phys. Rev. Lett. 45 (1980) 1755.
[40] C.B. Lang and C. Rebbi, Phys. Lett. 115B (1982) 137.
[41] G. Bhanot and M. Creutz, Phys. Rev. D24 (1981) 3212.
[42] J. Greensite and B. Lautrup, Phys. Rev. Lett. 47 (1981) 9;
I. Halliday and A. Schwimmer, Phys. Lett. 101B (1981) 327.
[43] H.D. Politzer, Phys. Rev. Lett. 30 (1973) 1346;
D.J. Gross and F. Wilczek, Phys. Rev. Lett. 30 (1973) 1343;
W. Caswell, Phys. Rev. Lett. 33 (1974) 344;
D.R. Jones, Nucl. Phys. B75 (1974) 531.
[44] A. Hasenfratz and P. Hasenfratz, Phys. Lett. 93B (1980) 165.
[45] R. Dashen and D.J. Gross, Phys. Rev. D23 (1981) 2340.
[46] H. Kawai, R. Nakayama and K. Seo, Nucl. Phys. B189 (1981) 40.
[47] P. Weisz, Phys. Lett. 100B (1981) 331.
[48] A. Hasenfratz and P. Hasenfratz, Nucl. Phys. B193 (1981) 210.
[49] A. Gonzalez-Arroyo and C.P. Korthals-Altes, Nucl. Phys. B205[FS5] (1982) 46.

[50] W. Celmaster and R. Gonsalves, Phys. Rev. D20 (1979) 1420.
[51] C. Itzykson, unpublished.
[52] R. Savit, Rev. Mod. Phys. 52 (1980) 453.
[53] L.P. Kadanoff and H. Ceva, Phys. Rev. B3 (1971) 3918.
[54] E. Fradkin and L. Susskind, Phys. Rev. D17 (1978) 2637.
[55] See, for instance, P.S. Aleksandrov, Combinatorial topology (Graylock, New York, 1956).
[56] S. Elitzur, R.B. Pearson and J. Shigemitsu, Phys. Rev. D19 (1979) 3698.
[57] D. Horn, M. Weinstein and S. Yankielowicz, Phys. Rev. D19 (1979) 3715.
[58] A. Ukawa, P. Windey and A.H. Guth, Phys. Rev D21 (1980) 1013.
[59] A.M. Polyakov, Nucl. Phys. B120 (1977) 429.
[60] T. Banks, J. Kogut and R. Myerson, Nucl. Phys. B129 (1977) 493.
[61] A. Guth, Phys. Rev. D21 (1980) 2291.
[62] M. Peskin, Ann. Phys. (NY) 113 (1978) 122.
[63] S. Mandelstam, Phys. Reports 23C (1975) 245.
[64] G. 't Hooft, in: Proc. EPS Intern. Conf., Palermo 1975.
[65] T.A. Degrand and D. Toussaint, Phys. Rev. D22 (1980) 2478.
[66] S.D. Drell, H.R. Quinn, B. Svetitsky and M. Weinstein, Phys. Rev. D19 (1979) 619.
[67] M. Göpfert and G. Mack, Comm. Math. Phys. 82 (1982) 545.
[68] L. Yaffe, Phys. Rev. D21 (1980) 1574.
[69] E. Tomboulis, Phys. Rev. D23 (1981) 2371.
[70] S. Samuel, Phys. Lett. 122B (1983) 287.
[71] G. 't Hooft, Nucl. Phys. B153 (1979) 141.
[72] J. Groeneveld, J. Jurkiewicz and C.P. Korthals-Altes, Acta Phys. Pol. B12 (1981) 615;
J. Groeneveld, J. Jurkiewicz and C.P. Korthals-Altes, Physica Scripta 23 (1981) 1022.
[73] G. 't Hooft, Nucl. Phys. B72 (1974) 461; B75 (1974) 461;
G. Veneziano, Nucl. Phys. B117 (1976) 519;
E. Witten, Nucl. Phys. B160 (1979) 57.
[74] S. Wadia, Chicago preprint EFI 80/15, March 1980, unpublished.
[75] C.B. Lang, P. Salomonson and B.S. Skagerstam, Phys. Lett. 107B (1981) 211; Nucl. Phys. B190 (1981) 337.
[76] T. Eguchi and H. Kawai, Phys. Rev. Lett. 48 (1982) 1063.
[77] D. Foerster, Phys. Lett. 87B (1979) 87;
T. Eguchi, Phys. Lett. 87B (1979) 91;
D.H. Weingarten, Phys. Lett. 87B (1979) 97.
[78] G. Bhanot, U.M. Heller and H. Neuberger, Phys. Lett. 113B (1982) 47;
M. Okawa, Phys. Rev. Lett. 49 (1982) 353.
[79] G. Bhanot, U.M. Heller and H. Neuberger, Phys. Lett. 115B (1982) 237;
M. Okawa, Phys. Rev. Lett. 49 (1982) 705.
[80] G. Parisi, Phys. Lett. 112B (1982) 463;
G. Parisi and Y.C. Zhang, Phys. Lett. 114B (1982) 319;
D.J. Gross and Y. Kitazawa, Nucl. Phys. B206 (1982) 440;
S.R. Das and S.R. Wadia, Phys. Lett. 117B (1982) 228.
[81] Y. Kitazawa and S.R. Wadia, Phys. Lett. 120B (1983) 377;
H. Neuberger, IAS preprint (1982).
[82] G. Parisi and Wu Yong-Shi, Sc. Sinica 24 (1981) 483.
[83] J. Alfaro and B. Sakita, CCNY-HEP-82/8 preprint (1982);
J. Greensite and M. Halpern, DCB-PTH-82/14 preprint (1982).
[84] A. Gonzalez-Arroyo and M. Okawa, Phys. Lett. 120B (1983) 174.
[85] U. Heller and H. Neuberger, Phys. Rev. Lett. 49 (1982) 621;
G. Bhanot, Phys. Lett. 120B (1983) 371;
I. Bars, M. Günaydin and S. Yankielowicz, CERN TH.3379 preprint (1982).
[86] J. Shigemitsu and J.B. Kogut, Nucl. Phys. B190[FS3] (1982) 365.
[87] G. Münster, Nucl. Phys. B180[FS2] (1981) 23.
[88] K.G. Wilson and J.B. Kogut, Phys. Reports 12C (1974) 75.
[89] M. Lüscher, Comm. Math. Phys. 54 (1977) 283.
[90] J.B. Kogut, D.K. Sinclair and L. Susskind, Nucl. Phys. B114 (1976) 199;
B. Berg and A. Billoire, Nucl. Phys. B221 (1983) 109.
[91] G. Münster, Nucl. Phys. B190[FS3] (1981) 439; B200[FS4] (1982) 536 (EA); B205[FS5] (1982) 648 (E).
[92] N. Kimura and A. Ukawa, Nucl. Phys. B205[FS5] (1982) 637.
[93] L. Onsager, Phys. Rev. 65 (1944) 117.

[94] See also M.E. Fisher and R.J. Burford, Phys. Rev. 156 (1967) 583.
[95] T.P. Cheng and T.T. Wu, Phys. Rev. 164 (1967) 719.
[96] K. Seo, Nucl. Phys. B209 (1982) 200.
[97] G. Münster, Phys. Lett. 121B (1983) 53.
[98] H. Kawai and R. Nakayama, Phys. Lett. 113B (1982) 329.
[99] R. Nakayama, UT-382 preprint.
[100] K. Seo and A. Ukawa, Phys. Lett. 114B (1982) 329.
[101] J. Hoek, J. Comp. Phys. 49 (1983) 265.
[102] K. Wilson, private communication.
[103] F. Green and S. Samuel, Nucl. Phys. B190[FS3] (1981) 113.
[104] F. Green and S. Samuel, Nucl. Phys. B194 (1982) 107.
[105] J.-M. Drouffe and K.J.M. Moriarty, Phys. Lett. 108B (1982) 333.
[106] J.-M. Drouffe and K.J.M. Moriarty, Phys. Lett. 105B (1981) 449.
[107] J.-M. Drouffe, K.J.M. Moriarty and G. Münster, Phys. Lett. 115B (1982) 301.
[108] R. Dashen, U. Heller and H. Neuberger, Nucl. Phys. B215[FS7] 360.
[109] J.-M. Drouffe, unpublished, preprint ITP.SB.78.35.
[110] A. Duncan and H. Vaidya, Phys. Rev. D20 (1979) 903.
[111] N. Kimura, Prog. Theor. Phys. 64 (1980) 310.
[112] G. Münster and P. Weisz, Nucl. Phys. B180[FS2] (1981) 13.
[113] J.-M. Drouffe and J.-B. Zuber, Nucl. Phys. B180[FS2] 253, 264 (the second one of these papers contains several misprints, corrected expressions are given in the present review).
[114] J.D. Weeks, unpublished.
[115] A. Gonzales-Arroyo, C.P. Korthals-Altes, J. Peiro and M. Perrottet, Phys. Lett. 116B (1982) 414.
[116] R.C. Johnson, Phys. Lett. 114B (1982) 147;
J.E. Mandula, G. Zweig and J. Govaerts, Los Alamos preprints 82-3585 and 82-3586.
[117] H.B. Tarko and M.E. Fisher, Phys. Rev. B11 (1975) 1217.
[118] J.-M. Drouffe, G. Parisi and N. Sourlas, Nucl. Phys. B161 (1979) 397.
[119] C. Itzykson, R.B. Pearson and J.-B. Zuber, Nucl. Phys. B220 [FS8] (1983) 415;
J.-B. Zuber, Stony Brook preprint ITP-SB-83-49.
[120] D.S. Gaunt and A.J. Guttmann, in: Phase transitions and critical phenomena, eds. C. Domb and M.S. Green (Academic, London, 1974) vol. 3, p. 181;
D.L. Hunter and G.A. Baker Jr., Phys. Rev. B7 (1973) 3346, and further refs. therein.
[121] D.L. Hunter and G.A. Baker Jr., Phys. Rev. B19 (1979) 3808.
[122] M.E. Fisher and H. Au-Yang, J. Phys. A12 (1979) 1677.
[123] M.E. Fisher and J.H. Chen, in: Phase transitions, Proc. 1980 Cargese Summer School, eds. M. Levy et al. (Plenum, N.Y., 1982).
[124] J.S.R. Chisholm, Math. Comput. 27 (1973) 841;
C. Alabiso and P. Butera, J. Math. Phys. 16 (1975) 840;
M.F. Barnsley and P.D. Robinson, SIAM J. Math. Anal. 9 (1978) 272.
[125] M. Creutz, L. Jacobs and C. Rebbi, Phys. Rev. Lett. 42 (1979) 1390.
[126] J.-M. Drouffe, Nucl. Phys. B170[FS1] (1980) 91.
[127] M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi and B. Taglienti, Phys. Lett. 105B (1981) 51.
[128] M. Creutz, L. Jacobs and C. Rebbi, Phys. Rev. D20 (1979) 1915.
[129] B. Lautrup and M. Nauenberg, Phys. Lett. 95B (1980) 63.
[130] M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi and B. Taglienti, Nucl. Phys. B190[FS3] (1981) 782.
[131] M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi and B. Taglienti, Phys. Lett. 102B (1981) 270.
[132] M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi and B. Taglienti, Phys. Lett. 108B (1982) 331.
[133] R. Dashen, U.M. Heller and H. Neuberger, Nucl. Phys. B215[FS7] (1983) 360.
[134] J.-M. Drouffe and J.-B. Zuber, unpublished.
[135] C.J. Thomson, A.J. Guttmann and B.W. Ninham, J. Phys. C2 (1969) 1889;
A.J. Guttmann, J. Phys. C2 (1969) 1900;
C. Domb and A.J. Guttmann, J. Phys. C3 (1970) 1652.
[136] R.B. Pearson, Phys. Rev. B26 (1982) 6285.
[137] J.D. Weeks and G.H. Gilmer, Dynamics of crystal growth, Adv. in chemical phys. vol. 40, eds. Prigogine and Rice (Wiley).
[138] J.-M. Luck, thèse de 3ieme cycle, Paris 1981, and further references therein.
[139] A. Hasenfratz, E. Hasenfratz and P. Hasenfratz, Nucl. Phys. B180[FS2] (1981) 353.
[140] M. Lüscher, G. Münster and P. Weisz, Nucl. Phys. B180[FS2] (1981) 1.
[141] C. Itzykson, M.E. Peskin and J.-B. Zuber, Phys. Lett. 95B (1980) 259.
[142] G. Münster and P. Weisz, Nucl. Phys. B180[FS2] (1981) 330.
[143] F. Gliozzi, Phys. Lett. 101B (1981) 180.

[144] M. Peskin, Phys. Lett. 94B (1980) 161.
[145] J.B. Kogut, R.B. Pearson and J. Shigemitsu, Phys. Lett. 98B (1980) 63.
[146] J.B. Kogut, D.K. Sinclair, R.B. Pearson, J.L. Richardson and J. Shigemitsu, Phys. Rev. D23 (1981) 2945.
[147] J.B. Kogut and D.K. Sinclair, Phys. Rev. D24 (1981) 1610.
[148] J.-M. Drouffe, J.-M. Luck and J.-B. Zuber, J. Phys. (Paris) 42 (1981) 653.
[149] N.D. Mermin and H. Wagner, Phys. Rev. Lett. 17 (1966) 1133;
S. Coleman, Comm. Math. Phys. 31 (1973) 259.
[150] M. Lüscher, K. Symanzik and P. Weisz, Nucl. Phys. B173 (1980) 365;
M. Lüscher, Nucl. Phys. B180[FS2] (1981) 317.
[151] J.P. Kowall and H. Neuberger, Phys. Lett. 106B (1981) 197; Nucl. Phys. B189 (1981) 535.
[152] M. Creutz and K.J.M. Moriarty, Phys. Rev. D26 (1982) 2166;
R.W.B. Ardill, M. Creutz and K.J.M. Moriarty, Phys. Rev. D27 (1983) 1956.
[153] J.B. Kogut, R.B. Pearson and J. Shigemitsu, Phys. Lett. 98B (1980) 63.
[154] H. Lipps, G. Martinelli, R. Petronzio and F. Rapuano, CERN preprint TH-3548.
[155] G. Bhanot and R. Dashen, Phys. Lett. 113B (1982) 299.
[156] C.B. Lang, C. Rebbi, P. Salomonson and B.S. Skagerstam, Phys. Rev. D26 (1982) 2028.
[157] R. Gavai, F. Karsch and H. Satz, Bielefeld preprint BI-TP 82/26 (1982).
[158] H.S. Sharatchandra and P.H. Weisz, DESY 81-083 preprint (1981).
[159] Y.M. Makeenko and M.J. Polikarpov, Nucl. Phys. B205[FS5] (1982) 386.
[160] Chen Tien-Lun, Chung-I Tan and Zheng Xi-te, Phys. Lett. 109B (1982) 383.
[161] S. Samuel, Phys. Lett. 112B (1982) 237.
[162] Y.M. Makeenko, M.I. Polikarpov and A.V. Zhelarkin, Moscow preprint ITEP-21 (1983).
[163] B. Berg, A. Billoire and C. Rebbi, Ann. Phys. (NY) 142 (1982) 185.
[164] M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi, B. Taglienti and Zhang Y. Chen, Phys. Lett. 110B (1982) 295.
[165] K. Ishikawa, M. Teper and G. Schierholz, Phys. Lett. 110B (1982) 399.
[166] K.H. Mütter and K. Schilling, Phys. Lett. 117B (1982) 75; Nucl. Phys. B200[FS4] (1982) 362.
[167] R.C. Brower, M. Creutz and N. Nauenberg, Nucl. Phys. B210[FS6] (1982) 133.
[168] J. Smit, Nucl. Phys. B206 (1982) 309.
[169] B. Berg and A. Billoire, Nucl. Phys. B221 (1983) 109 and Nucl. Phys. to appear.
[170] K. Ishikawa, M. Teper and G. Schierholz, Phys. Lett. 116B (1982) 429;
C. Michael and I. Teasdale, Nucl. Phys. B215[FS7] (1983) 433.
[171] G. Münster, Phys. Lett. 121B (1983) 53.
[172] G. Bhanot and D. Foerster, Phys. Lett. 118B (1982) 395.
[173] H.W.J. Blöte and R.M. Swendsen, Phys. Rev. Lett. 43 (1979) 299.
[174] J.-M. Drouffe, Nucl. Phys. B170[FS1] (1980) 211.
[175] J. Greensite and B. Lautrup, Phys. Lett. 104B (1981) 41.
[176] P. Cvitanovic, J. Greensite and B. Lautrup, Phys. Lett. 105B (1981) 197.
[177] J.-M. Drouffe, Phys. Lett. 105B (1981) 45.
[178] D.J. Pritchard, Phys. Lett. 106B (1981) 193.
[179] L. Masperi and C. Omero, Phys. Lett. 107B (1981) 207.
[180] J.-M. Drouffe, Nucl. Phys. B205[FS5] (1982) 27.
[181] E. Brézin and J.-M. Drouffe, Nucl. Phys. B200[FS4] (1982) 93.
[182] H. Flyvbjerg, B. Lautrup and J.-B. Zuber, Phys. Lett. 110B (1982) 279.
[183] V.F. Müller, Nucl. Phys. B205[FS5] (1982) 578.
[184] V.F. Müller and W. Rühl, Nucl. Phys. B210[FS6] (1982) 289.
[185] R. Horsley, Phys. Lett. 107B (1981) 237.
[186] V. Alessandrini, Phys. Lett. 117B (1982) 423; Nucl. Phys. B215[FS7] (1983) 337.
[187] V. Alessandrini, V. Hakim and A. Krzywicki, Nucl. Phys. B200[FS4] (1982) 355.
[188] V. Alessandrini, V. Hakim and A. Krzywicki, Nucl. Phys. B205[FS5] (1982) 253.
[189] N. Kawamoto and K. Shigemoto, Phys. Lett. 114B (1982) 42.
[190] M. Bononi and G. Marchesini, Phys. Lett. 110B (1982) 275.
[191] Chen T.-L., C.-I Tan and Zheng X.-T., Phys. Lett. 109B (1982) 383.
[192] Zheng X.-T., C.-I Tan and Chen T.-L., Phys. Rev. D26 (1982) 2843.
[193] J.M. Alberty, H. Flyvbjerg and B. Lautrup, Nucl. Phys. B220[FS8] (1983) 61;
H. Flyvbjerg and E. Marinari, Phys. Lett. 132B (1983) 385.
[194] R. Thom, Structural stability and morphogenesis (Benjamin, 1975).
[195] E. Brézin, J.-C. Le Guillou and J. Zinn-Justin, in: Phase transitions and critical phenomena Vol. 6, eds. C. Domb and M.S. Green (Academic, London).
[196] P. Ginsparg, Y. Goldschmidt and J.-B. Zuber, Nucl. Phys. B170[FS1] (1980) 388;
R. Kotecky, Comm. Math. Phys. 82 (1981) 391.

[197] See e.g. R.B. Dingle, Asymptotic expansions: their derivation and interpretation (Academic Press, London, 1973).
[198] B. Simon and L.G. Yaffe, Pasadena preprint CALT-68-912 (1982).
[199] B. Bunk, B. Lautrup, I.O. Stamatescu, W. Weinzierl and U. Wolff, Z. Phys. C18 (1983) 239.
[200] I. Halliday and A. Schwimmer, Phys. Lett. 101B (1981) 327;
G. Mack, Phys. Rev. Lett. 45 (1980) 1378;
G. Mack and V.B. Petkova, Ann. Phys. 125 (1980) 117; Z. Phys. C12 (1982) 177;
L.G. Yaffe, Phys. Rev. D21 (1980) 1574;
R.C. Brower, D.A. Kessler and H. Levine, Phys. Rev. Lett. 47 (1981) 621; Nucl. Phys. B205[FS5] (1982) 77;
L. Caneschi, I.G. Halliday and A. Schwimmer, Nucl. Phys. B200[FS4] (1982) 409;
G. Mack and E. Pietarinen, Nucl. Phys. B205[FS5] (1982) 141.
[201] C.P. Bachas and R.F. Dashen, Nucl. Phys. B210[FS6] (1982) 583.
[202] M.E. Fisher and D.S. Gaunt, Phys. Rev. 133 (1964) A224;
R. Abe, Prog. Theor. Phys. 47 (1972) 62.
[203] V.F. Müller, T. Raddatz and W. Rühl, Phys. Lett. 122B (1983) 148.
[204] R.C. Brower, P. Rossi and Chung-I Tan, Phys. Rev. D23 (1981) 942.
[205] M. Creutz and K.J.M. Moriarty, Phys. Rev. D25 (1982) 610.
[206] M. Creutz, Phys. Rev. Lett. 46 (1982) 1441.
[207] K.J.M. Moriarty, Phys. Lett. 106B (1981) 130.
[208] H. Bohr and K.J.M. Moriarty, Phys. Lett. 104B (1981) 217.
[209] M. Creutz and K.J.M. Moriarty, Phys. Rev. D25 (1982) 1724.
[210] M. Creutz, Phys. Rev. Lett. 43 (1979) 553.
[211] W. Rühl, Z. Phys. C18 (1983) 207;
B. Lautrup and W. Rühl, Kaiserlautern preprint (February 1983);
H. Flyvbjerg, Nordita preprint 83/45;
H. Flyvbjerg, P. Mansfield and B. Söderberg, Nordita preprint.
[212] J. José, L.P. Kadanoff, S. Kirkpatrick and D. Nelson, Phys. Rev. B16 (1977) 1217 and references therein.
[213] G. Mack and H. Meyer, Nucl. Phys. B200[FS4] (1982) 249.
[214] J. Bricmont and J. Fröhlich, Phys. Lett. 122B (1983) 73.
[215] E. Fradkin and S.H. Shenker, Phys. Rev. D19 (1979) 3682.
[216] G.F. deAngelis, D. deFalco, F. Guerra and R. Marra, Acta Phys. Austr. Supp. 19 (1978) 205.
[217] T. Banks and E. Rabinovici, Nucl. Phys. B160 (1979) 349.
[218] J.-M. Drouffe and J.-B. Zuber, unpublished.
[219] M. Creutz, Phys. Rev. D21 (1981) 1006.
[220] H. Weyl, The classical groups (Princeton, 1946).
[221] M. Hamermesh, Group theory and its application to physical problems (Addison-Wesley, 1962).
[222] C. Itzykson and J.-B. Zuber, J. Math. Phys. 21 (1980) 411.
[223] F. Green and S. Samuel, Nucl. Phys. B190[FS3] (1981) 113.
[224] M.L. Mehta, Random matrices and the statistical theory of energy levels (Academic, NY, 1967).
[225] S. Samuel, J. Math. Phys. 21 (1980) 2695.
[226] M. Creutz, J. Math. Phys. 19 (1978) 2043.
[227] Y.Y. Goldschmidt, J. Math. Phys. 21 (1980) 1842.
[228] E. Brézin and D.J. Gross, Phys. Lett. 97B (1980) 120.
[229] I. Bars, Phys. Lett. 116B (1982) 57.
[230] R.C. Brower and M. Nauenberg, Nucl. Phys. B180[FS2] (1981) 221.