THE BETHE APPROXIMATION FOR LATTICE GAUGE THEORIES

J B ZUBER

Institute for Theoretical Physics*, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

Received 31 October 1983

The Bethe approximation is defined for general lattice gauge theories. It amounts to solving the model on an infinite Cayley lattice of cubes. The approximation is tested on the 4-d Z_4 model, where it is shown to reproduce accurately most of the phase diagram. It also suggests which mass vanishes in the Coulomb phase

1. Introduction

Lattice gauge theories have achieved remarkable results, in particular through Monte Carlo simulations [1]. In the pure gauge sector, things are almost settled. However, the nature of the U(1) phase transition and of the nearby singularities in SU(N) theories and their possible influence on the onset of scaling are not yet fully understood; accurate determinations of all the mass scales are still the object of controversies. Analytical methods adequate to study these problems are not available. Strong coupling and mean field methods are not appropriate or accurate enough, at least in their current state [2], while the approximate renormalization group – à la Migdal-Kadanoff – has not yet reached its full predictive capability. This leaves some room for other approximation schemes, or resummation methods. The purpose of this paper is to study such a method, an extension of the Bethe approximation [3].

The Bethe approximation is well known and has been extensively studied in spin models [4], but, surprisingly, has not been applied to lattice gauge systems until recently [5]. There, it was defined in the Z_2 theory and shown to reproduce correctly its phase structure in 3 and 4 dimensions. It was not totally clear, however, whether it would generalize to more complicated models, and account for their richer structure.

In sect. 2, the Bethe approximation is defined and basic equations are derived for a general gauge group and action in arbitrary dimension. Sect. 3 discusses how singularities and phase transitions may arise, and sect. 4 shows how to compute the

^{*}Permanent address: SPhT, CEN Saclay, F 91191 Gif-sur-Yvette

correlation lengths. Sect. 5 discusses the application of this approximation to the general $4 - d Z_4$ model. The latter has been recently studied numerically [6], and an unexpected new (Coulomb?) phase, bounded by second-order transition lines, found in the lower-half plane of the two-constant phase diagram. This makes this model a good testing ground of the Bethe approximation, simple enough to make the equations manageable, but still with a rather complex phase structure. We shall see that the Bethe approximation reproduces quite accurately most of the phase diagram, including one of the phase boundaries of the Coulomb phase, although it fails to describe consistently the latter phase. It also suggests a finer characterization of that phase. These results are summarized and the outlook of the Bethe approximation discussed in sect. 6.

2. General formalism

In the context of spin models, the Bethe approximation may be introduced in various ways. The low-temperature expansion may be reorganized as an irreducible cluster expansion. Truncated to lowest order (cluster = single spin), it gives the mean-field approximation, whereas truncation to the next order (cluster = pairs of neighboring spins) yields the so-called Bethe-Peierls approximation [4]. Alternatively, one may, for a given spin, keep the direct interaction terms with its neighbors, but replace the interactions of the latter by a self-consistent field. It may be seen that these procedures amount to solving the model on an infinite Cayley tree, i.e. a lattice with no loop, where each spin has the same number q of neighbors. The subtle boundary effects [7] on such a lattice are thus discarded in this approach (for a discussion of those effects in lattice gauge theories see ref. [8]).

In the generalization of this method to lattice gauge theories, one may try to follow several routes: reorganization of the low-temperature expansion, or self-consistent introduction of two-link interactions, or consideration of a Cayley tree of plaquettes. However, all these methods either are restricted to discrete groups or have at some stage to introduce a non-gauge invariant coupling to an external field, and to compute the response of the system in the limit of a vanishing field, therefore in blatant contradiction with Elitzur theorem [9]. It is not clear whether such a Bethe approximation may be regarded as a saddle point method, as is the case of mean field where averaging over gauge degeneracies removes the conflict with Elitzur theorem (see ref. [2] and further references therein).

For these reasons, it seems better to introduce a Cayley lattice of *cubes* which allows a consistent gauge-invariant treatment. This is suggested by the case of the 3-dimensional Z_2 model, where such a tree lattice of cubes is the dual of the ordinary Cayley tree of sites. In *d* dimensions, this abstract lattice is made by pilling up "cubes" on top of each other, with the same coordination number (each face is common to 2(d-2) "cubes"), but in such a way that there is no cycle nor cluster of cubes. For both visualizing the resulting pattern and solving the model, it is useful to



Fig 1 Cross section of a Cayley tree of cubes in the case d = 3 each plaquette is shared by two cubes, l, l+1, are the generation labels

construct the lattice in a sequential way: a cube of generation l will support (2d - 5) cubes of generation l + 1 on each of its five free faces: see fig. 1 for a cross section of this lattice. We then introduce a gauge theory of group G in the ordinary way: to each link is attached a group element U_l , and each plaquette carries a Boltzmann weight that we character expand [2] as

$$\exp s(U_{\rm P}) = \tilde{\beta}_0 \sum_r d_r t_r \chi_r(U_{\rm P}).$$
(2.1)

The sum runs over all irreducible representations r of dimension d_r and character χ_r , including the trivial one for which $t_0 \equiv 1$, $d_0 = 1$, $\chi_0 \equiv 1$.

For a plaquette P belonging to a cube C of generation l (see fig. 1), let us consider the sum $x_r^{(l)}$ of contributions to

$$\int \chi_r^*(U_{\mathbf{P}}) \mathrm{e}^{\sum_{\mathbf{P}'} \mathcal{S}(U_{\mathbf{P}'})} \mathrm{D} U,$$

from strong coupling diagrams made of plaquettes of C excluding P or of plaquettes of subsequent generations; let $y_r^{(l)}$ be the sum of contributions to the same quantity made possibly with P and all its descendants. Notice that contributions to $y_r^{(l)}$ come from the 2d - 5 cubes above P (and their descendants), whereas those to $x_r^{(l)}$ come from only one cube. One may write

$$x_r^{(l)} = \left[y_r^{(l+1)} \right]^5.$$
 (2.2a)

The power 5 in eq. (2.2a) comes from the five plaquettes of generation (l + 1) above the plaquette P, and this factorization reflects the fact that on a Cayley lattice, diagrams constructed on these plaquettes have no further contact. To lowest order in the strong coupling expansion, $y_r^{(l)} = t_r$, and $x_r^{(l)} = t_r^5$ represents the open box C above P. In general, using the rules of strong coupling expansions of ref. [2], one may write in d = 3 dimensions

$$y_r^{(l)} = \sum_{s,t} N_{rst} \frac{d_s d_t}{d_r} t_s x_t^{(l)}, \qquad (2.2b)$$

where N_{rst} is the number of times the trivial representation occurs in the direct product $r \otimes s \otimes t \otimes ...$ This relation expresses $y_r^{(l)}$ in terms of all possible representation assignments on the cube above P. In four dimensions there are 2d - 5 = 3 cubes above P and the relation reads:

$$y_r^{(l)} = \sum_{s, t, u, v} N_{rstuv} \frac{d_s d_t d_u d_v}{d_r} t_s x_t^{(l)} x_u^{(l)} x_v^{(l)}.$$
 (2.2c)

Eqs. (2.2b,c) together with (2.2a) form a system of exact recursive relations on the Cayley lattice.

One then assumes that the infinite lattice is homogeneous, i.e. that the ratio $x_r^{(l)}/x_0^{(l)}$ takes a value independent of l, denoted ρ_r . They satisfy:

$$\rho_{r} = \left(\frac{\sum N_{rsu_{1}} \quad u_{2d-5} \, d_{s} \, d_{u_{1}} \dots \, d_{u_{2d-5}} t_{s} \rho_{u_{1}} \dots \rho_{u_{2d-5}}}{d_{r} \sum N_{su_{1}} \quad u_{2d-5} \, d_{s} \, d_{u_{1}} \dots \, d_{u_{2d-5}} t_{s} \rho_{u_{1}} \dots \rho_{u_{2d-5}}}\right)^{5},$$

$$\rho_{0} \equiv 1,$$
(2.3)

in d dimensions.

As the action satisfies $s(U) = s(U^{\dagger})$, the character coefficients corresponding to conjugate representations are equal:

$$t_r = t_{\bar{r}},\tag{24}$$

and it seems natural to impose the same condition on the ρ 's:

$$\rho_r = \rho_{\bar{r}}.\tag{2.5}$$

We then introduce new notations:

$$\rho_r = p_r^5, \tag{2.6}$$

$$A_{n} = \sum_{r, s_{1}, \dots, s_{n}} N_{rs_{1} \dots s_{n}} d_{r} d_{s_{1}} \dots d_{s_{n}} t_{r} \rho_{s_{1}} \dots \rho_{s_{n}}.$$
 (2.7)

In words, ρ_r represents the contribution of an open box of five dressed plaquettes with a boundary in the representation r and p_r is the contribution of each plaquette. It is easy to show that A_n satisfies

$$\sum_{r\neq 0} \frac{1}{n} \rho_r \frac{\partial}{\partial \rho_r} A_n = A_n - A_{n-1}, \qquad (2.8)$$

a relation to be used soon. For example, for the gauge group Z_2 :

$$A_{n} = \sum_{s=0}^{n} t^{[s]} {n \choose s} \rho^{s}, \qquad (2.9)$$

where [s] = 0, 1 if s is even, odd. In that particular case, one has a symmetry

$$\hat{A}_n(\rho,t) = \frac{1}{n+1} \frac{\partial}{\partial \rho} A_{n+1}(\rho,t) = t A_n\left(\rho,\frac{1}{t}\right), \qquad (2.10)$$

as a consequence of $t(1/t)^{[s]} = t^{[s-1]}$, and eq. (2.8) implies

$$A_n - A_{n-1} = \rho \hat{A}_{n-1},$$

$$\hat{A}_n - \hat{A}_{n-1} = \rho A_{n-1}.$$
 (2.11)

In general eq. (2.3) may be recast as

$$\rho_r = p_r^5 = \left(\frac{1}{2d-4} \frac{1}{d_r^2} \frac{\partial A_{2d-4}/\partial \rho_r}{A_{2d-5}}\right)^5.$$
(2.12)

Simple diagrammatic arguments show that the average plaquette in the r representation is, in d = 4:

$$\langle \chi_r(U_{\rm P}) \rangle = \frac{\sum N_{rstuvw} d_s d_t d_u d_v d_w t_s \rho_t \rho_u \rho_v \rho_w}{d_r \sum N_{stuvw} d_s d_t d_u d_v d_w t_s \rho_t \rho_u \rho_v \rho_w}, \qquad (2.13)$$

or more generally, in d dimensions:

$$\langle \chi_r(U_{\rm P}) \rangle = \frac{1}{2d-3} \frac{\partial A_{2d-3}/\partial \rho_r}{d_r A_{2d-4}}.$$
 (2.14)

Now, using (2.8), (2.12), this may be rewritten as

$$\langle \chi_r(U_{\rm P}) \rangle = \frac{\sum_{s,t} N_{rst} \, d_s \, d_t \, p_s \rho_t}{1 + \sum_{s \neq 0} d_s^2 p_{\bar{s}} \rho_s} \,, \tag{2.15}$$

which fits the previous interpretation of p_r , namely the value of a dressed plaquette in the representation r: to compute $\langle \chi_r(U_{\rm P}) \rangle$ for some plaquette, it is sufficient to consider the diagrams on a single cube above the plaquette, made of these dressed plaquettes with all possible representation assignments, and to divide them by a similar contribution to the partition function, namely $1 + \sum_{s \neq 0} d_s^2 p_{\bar{s}} \rho_s$.

If the action reads $s(U_p) = \beta_1 \operatorname{Re} \chi_1(U_p) + \cdots$ we must integrate

$$\frac{1}{d_1} \langle \chi_1(U_{\rm P}) \rangle = \frac{\partial f}{\partial \beta_1}$$
(2.16)

to get the free energy per plaquette f. It is not difficult to show that

$$f = \ln \tilde{\beta}_0 + \frac{5}{3}(d-2)\ln A_{2d-5} - \frac{1}{3}(5d-13)\ln A_{2d-4} + \text{constant}$$
(2.17)

is the solution to (2.16). First, differentiating (2.17) and using (2.12), one shows that

$$\frac{\partial f}{\partial \beta_1} = t_1 + \sum_{r \neq 0} \frac{\mathrm{d}t_r}{\mathrm{d}\beta_1} \frac{\partial}{\partial t_r} \ln A_{2d-4}$$
$$= \frac{\sum N_{us_1 \quad s_{2d-4}} d_u \left(t_u t_1 + \frac{\mathrm{d}t_u}{\mathrm{d}\beta_1} \right) d_{s_1} \dots d_{s_{2d-4}} \rho_{s_1} \dots \rho_{s_{2d-4}}}{A_{2d-4}} . \tag{2.18}$$

But the character coefficients satisfy

$$d_1 d_u \left(t_u t_1 + \frac{\mathrm{d} t_u}{\mathrm{d} \beta_1} \right) = \frac{1}{2} \sum t_s \left(N_{1r\bar{s}} + N_{\bar{1}r\bar{s}} \right) d_s \tag{2.19}$$

and (2.16) then follows from (2.18), (2.19). We notice that the Bethe approximation yields an unambiguous answer for the internal energy but determines the free energy up to a constant. We shall return to the choice of this constant in sect. 3.

At this point, several remarks are in order.

(i) So far the derivations have been very formal. For finite groups, which possess only a finite number of representations, the expressions certainly make sense. For example, in the $3 - d Z_2$ model, the whole construction is equivalent, by duality, to the Bethe approximation for the Ising model, and all the expressions (2.3)-(2.17) reduce to well-known results [3,4]. The case of continuous groups with infinitely many representations, should require more care, since the convergence of expressions like A_n in (2.7) is not guaranteed. At strong coupling, i.e. small β or t, convergence is ensured by the power law behavior of t_1 and ρ_1 . For example, for the U(1) Wilson theory, with $s(U_P) = \beta \cos U_P$,

$$t_l = \frac{I_l(\beta)}{I_0(\beta)} \sim \frac{t_1^l}{l!}, \qquad \rho_l \sim t_l^5.$$

At larger values of β , however, divergences of these series might correspond to phase transitions of the system. We shall not address any longer this question in the present paper.

(ii) For abelian groups, irreducible representations form a group. As a consequence, $\rho_l = 1$ for all *l* is always a solution. For example, for d = 4, in eq. (1.3),

$$\frac{\sum N_{rstuv} t_s}{\sum N_{stuv} t_s} = 1 \,,$$

(assuming again the convergence of the numerator and denominator). This is because for any r, s, t, u, there is one and only one v such that $N_{rstuv} = N_{stuv} = 1$, and terms in the numerator and denominator are in one-to-one correspondence. For that trivial solution, the plaquette expectation value is one: $\langle \chi_r(U_P) \rangle = 1$.

(iii) Physically, we expect the solution to eqs. (2.3)–(2.13), i.e. the solution on the Cayley tree, to give a good description of the strong coupling regime of the original model. Indeed, it amounts to summing up a large fraction of the strong coupling diagrams, namely those diagrams which are trees of cubes (with, however, some overcounting, see ref. [5] for more details). On the other hand, the weak coupling phase (if not the analytic continuation of the strong coupling one) is expected to be represented for discrete abelian groups by the seemingly crude approximation $\rho = 1$. However it is well known from numerical simulations (see ref. [1] and further references therein) that the low-temperature phase of discrete group models is essentially frozen, with plaquette energies very close to one. This suggests that $\rho = 1$ may be a reasonable approximation after all, and explains why the phase structure of the simple model with Z_2 symmetry is well reproduced by the Bethe approximation [5]. The more complicated case of Z_4 will be examined below.

(iv) The large dimension, strong coupling limit of lattice gauge theories has been studied by Drouffe, Parisi and Sourlas (DPS) [10]. When $d \to \infty$, $\beta \to 0$ with $d\beta^4 \sim O(1)$, they showed that only diagrams made of trees of cubes survive. This is of course apparent in the previous equations. In the limit $\beta \to 0$, $d \to \infty$, $\rho_u \sim t_u^5$, only two terms survive in the numerator of eq. (2.3) and the trivial one ($s = u = \cdots = 0$) in the denominator. Hence the dressed plaquette satisfies the self-consistent equation:

$$p_r = \rho_r^{1/5} = t_r + (2d - 5) p_r^5 + \cdots$$

$$\approx t_r + 2dp_r^5,$$

which was the key equation of ref. [10]. In this limit, DPS found that the gauge system has a singularity at some β , but they had to resort to a different larger d approximation (mean field) to describe the weak coupling phase. Here, in contrast, we work at finite dimension, and want to see to which extent the Bethe approximation is capable of describing the low-temperature phase(s).

3. Singularities and transitions for finite gauge groups

We now turn to the search of singularities of the solutions to eqs. (2.12). For a finite group, we look for singularities of the mapping $(t_1, \ldots, t_n) \rightarrow (\rho_1, \ldots, \rho_n)$ where n is the number of inequivalent irreducible nontrivial representations. Singularities are associated with vanishing of the jacobian:

$$\frac{D(t_1,\ldots,t_n)}{D(\rho_1,\ldots,\rho_n)}$$

For example, in the particular case of Z_2 , the equation for singularities $dt/d\rho = 0$, leads, by logarithmic differentiation of

$$\rho = \left(\frac{\hat{A}_{2d-5}}{A_{2d-5}}\right)^5$$

with respect to ρ , to the condition:

$$1 = 5(2d-5) \left(\frac{A_{2d-6}}{A_{2d-5}} - \frac{\hat{A}_{2d-6}}{\hat{A}_{2d-5}} \right), \tag{3.1}$$

where use has been made of (2.11). The two ratios in the bracket may be eliminated between (2.11) and (2.12), i.e. $p = (\hat{A}_{2d-5})/(A_{2d-5})$, to give the equation:

$$1 + p2 + p4 + p6 + p8 = 5(2d - 5) p4.$$
 (3.2)

In general, if the ρ 's are given by implicit equations: $F_r(\{\rho\}, \{t\}) = 0, r = 1, ..., n$, then the equation for singularities is

$$J = \frac{D(F_1, \dots, F_n)}{D(\rho_1, \dots, \rho_n)} = 0.$$
 (3.3)

In our case, $F_r = \rho_r - R_r(\{\rho\}, \{t\})$ and

$$J = \det\left(\delta_{rs} - \frac{\partial R_r}{\partial \rho_s}\right). \tag{3.4}$$

How to solve the equation for ρ , in practice? The form of the equation:

$$\rho_r = R_r(\{\rho\}, \{t\})$$

suggests an iterative procedure:

$$\rho_r^{(n+1)} = R_r(\{\rho^{(n)}\}, \{t\}).$$

Of course, $\rho_r^{(n)}$ converges to the solution ρ_r as long as the eigenvalues of the linearized mapping at the fixed point are of modulus less than one. These eigenvalues satisfy the equation:

$$\det\left(\frac{\partial R_r}{\partial \rho_s} - \lambda \delta_{rs}\right) = 0.$$
(3.5)

Comparing (3.4) and (3.5), we see that singularities are to be found among the points

where the iterative mapping turns from convergent to divergent. This is a very useful method in practice to locate points or lines (or, in general, manifolds) of singularities in the *t*-space. For small *t*, the iterative mapping converges to the right solution $\rho \sim t^5$ because the eigenvalues are small. At some larger *t*, the fixed point may become unstable, and an additional study is required to determine whether it is an actual singularity.

These singularities are, of course, algebraic and generically square-root branch points. In most cases, two real solutions, ρ and ρ' , merge at the critical t_c . The solution ρ picked by the system beyond the singularity may be continuous at t_c , or discontinuous. Accordingly, the system undergoes a second- or a first-order transition. An example of the former case may be found in the $3 - d Z_2$ theory, where the critical value of ρ turns out to be 1 (cp. eq. (3.2) for d=3, p=1): the strong coupling and weak coupling phases connect continuously. The second instance, (first-order transition) is more common in gauge theories of dimension $d \ge 4$: see e.g. Z_2 in d = 4 in ref. [5]. Notice that if ρ is discontinuous at t_c , the actual transition point of the system is given by the thermodynamic criterion that free energies of the two phases be equal. This usually occurs at a smaller value t_1 of t, and the region $t_1 < t < t_c$ is the metastable region of the strong coupling phase. In this comparison of free energies, the yet undetermined constant of eq. (2.17)becomes of some relevance. When a first-order transition takes place, we choose to fix this constant independently in each phase by looking at some appropriate small or large coupling-limit and adjusting the constant to match the behavior of the original model. For example, in the $d = 4 Z_2$ theory, the free energy per plaquette reads:

$$f = \ln \tilde{\beta}_0 + \frac{10}{3} \ln(1 + 3\rho t + 3\rho^2 + \rho^3 t) - \frac{7}{3} \ln(1 + 4\rho t + 6\rho^2 + 4\rho^3 t + \rho^4) + C.$$

At strong coupling, $\rho \sim t^5$ and f must match the true behavior:

$$f \sim \ln \tilde{\beta}_0$$
, hence $C_{st} = 0$.

For large β , $t \simeq \rho \simeq 1$, $\ln \tilde{\beta}_0 = \ln \cosh \beta \sim \beta - \ln 2$,

$$f \simeq \beta - \ln 2 + \frac{2}{3}\ln 2 + C_w,$$
$$C_w = -\frac{1}{2}\ln 2.$$

A similar procedure will be used below in the Z_4 model.

Besides this class of singularities, a different mechanism is also observed in some cases: a solution that exists in a subspace of the ρ -space may become unstable under fluctuations in orthogonal directions. For example, a real solution may tend to develop an imaginary part beyond a critical point. This creates a second-order transition in the system. This mechanism will be illustrated below in the Z₄ model.

4. Correlation length

We now turn to the computation of the correlation length in the Bethe approximation. The general idea, adapted from similar computations in spin models [12], is to write that in the long-distance limit, the plaquette-plaquette connected correlation function satisfies an equation of the form:

$$\left(\Delta_{\mathbf{P}} - \mu^2\right) \langle \chi_r(U_{\mathbf{P}}) \chi_r(U_{\mathbf{P}_0}) \rangle_{\mathbf{c}} = 0, \qquad (4.1)$$

where $\Delta_{\rm P}$ is some equivalent on the Cayley lattice of the laplacian. It turns out to be:

$$\Delta_{\mathbf{P}}\chi(U_{\mathbf{P}}) = \sum_{\mathbf{P}'} \left[\chi(U_{\mathbf{P}'}) - \chi(U_{\mathbf{P}}) \right] + \chi(U_{\mathbf{P}''}) - \chi(U_{\mathbf{P}}), \qquad (4.2)$$

where the sum runs over the 5(2d-5) "daughter" plaquettes P', i.e. belonging to the (2d-5) cubes of the next generation, and P'' is the "mother" of P (see fig. 2). The advantage of the formulation (4.1) is that it avoids a direct computation of $\langle \chi \chi \rangle_c$, and only deals with its local variations, which are calculable in the Bethe approximation by simple diagrammatic considerations.

For the sake of simplicity, we only present here this computation in the case of the Z_2 theory; the extra complications arising in a more complex case like Z_4 will be mentioned below.

If P_0 is taken as a plaquette of the first generation and P of the *l*th generation, there is a unique tree of cubes of minimal length connecting them: let us denote \mathfrak{T} the corresponding set of plaquettes, excluding P_0 and P, and all their descendants. Let X, Y_0, Y and W be the sum of diagrams, connected or not, built from \mathfrak{T} , hence containing neither P_0 nor P and whose boundary is respectively:

$$\partial X = \emptyset$$
, $\partial Y_0 = P_0$, $\partial Y = P$, $\partial W = P_0 \cup P$, (fig. 3).

By the homogeneity assumption explained in sect. 2, we have $Y_0 = Y$. We also introduce the ratios $\tilde{Y} = Y/X$, $\tilde{W} = W/X$. It is then a pure matter of combinatorics to express the partition function and various plaquette averages in terms of contributions of plaquettes "between" P_0 and P, viz. X, Y, W and of contributions "below"



Fig. 2 A plaquette P, its "mother" P'', and its five daughters P' (in d = 3)



Fig. 3 Artist's view of typical contributions to. (a) X, (b) Y, (c) W

$$P_{0} \text{ or "above" } P, \text{ viz. } x_{0}, x_{1} = p^{5}x_{0}, A_{2d-5} \text{ and } \hat{A}_{2d-5} = pA_{2d-5} \text{ (cp. sect. 2).}$$

$$Z = x_{0}^{2(2d-5)} \left[XA_{2d-5}^{2} + 2YA_{2d-5}\hat{A}_{2d-5} + W\hat{A}_{2d-5}^{2} \right],$$

$$\langle U_{P} \rangle = Z^{-1}x_{0}^{2(2d-5)} \left[(X+W)A_{2d-5}\hat{A}_{2d-5} + Y \left(A_{2d-5}^{2} + \hat{A}_{2d-5}^{2} \right) \right]$$

$$= \frac{(1+\tilde{W})p + \tilde{Y}(1+p^{2})}{1+2\tilde{Y}p + \tilde{W}p^{2}},$$

$$\langle U_{P_{0}}U_{P} \rangle = Z^{-1}x_{0}^{2(2d-5)} \left[X\hat{A}_{2d-5}^{2} + 2YA_{2d-5}\hat{A}_{2d-5} + WA_{2d-5}^{2} \right]$$

$$= \frac{p^{2} + 2\tilde{Y}p + \tilde{W}p^{2}}{1+2\tilde{Y}p + \tilde{W}p^{2}}.$$
(4.3)

By identifying the above expression for $\langle U_P \rangle$ with $(2.15) = \langle U_P \rangle = (p + p^5)/(1 + p^2)$, one derives a relation between \tilde{Y} and \tilde{W} :

$$p^5 = \tilde{Y}(1-p^6) + \tilde{W}p.$$
 (4.4)

The connected correlation function may be expressed in terms of \tilde{Y} only as.

$$\langle U_{\mathrm{P}}U_{\mathrm{P}_{0}}\rangle_{\mathrm{c}} = \langle U_{\mathrm{P}}U_{\mathrm{P}_{0}}\rangle - \langle U_{\mathrm{P}}\rangle^{2}$$
$$= \frac{(1-p^{2})^{2}(\rho-\tilde{Y})}{p(1+p^{6})^{2}(1+p\tilde{Y})}.$$
(4.5)

Now it is easy to write the relation between X, Y, W, and X', Y', W' relative to a

daughter plaquette P':

$$\begin{aligned} X' &= x_0^{10d-26} A_{2d-5}^4 [XA_{2d-6} + Y\hat{A}_{2d-6}], \\ Y' &= x_0^{10d-26} A_{2d-5}^4 [YA_{2d-6} + W\hat{A}_{2d-6}], \\ &= x_0^{10d-26} \hat{A}_{2d-5}^4 [YA_{2d-6} + XA_{2d-6}], \\ W' &= x_0^{10d-26} \hat{A}_{2d-5}^4 [WA_{2d-6} + YA_{2d-6}], \end{aligned}$$
(4.6)

where the factors A_{2d-5}^4 or \hat{A}_{2d-5}^4 come from the four "sisters" of P' belonging to the same cube, and A_{2d-6} and \hat{A}_{2d-6} from the "step sisters" belonging to other cubes built on P.

The consistency of (4.4) and (4.6) is readily checked and it is easy to see that the fixed point of the transformation $(\tilde{Y}, \tilde{W}) \rightarrow (\tilde{Y}', \tilde{W}')$ is $\tilde{Y} = \rho$, $\tilde{W} = \rho^2$. This is the long-distance limit of \tilde{Y} and \tilde{W} , with the obvious consequence that $\langle U_P U_{P_0} \rangle_c$ vanishes at large separation. To compute the inverse correlation length squared μ^2 , it is thus sufficient to form the same combination as in (4.1)–(4.2) with $\Delta \tilde{Y} = \tilde{Y} - \rho$. One finds

$$\Delta \tilde{Y}' = p^4 \frac{A_{2d-6} - p\hat{A}_{2d-6}}{A_{2d-5}} \Delta \tilde{Y} + O(\Delta \tilde{Y}^2)$$
$$\equiv \alpha \Delta \tilde{Y} + O(\Delta \tilde{Y}^2), \qquad (4.7)$$

and for the mother plaquette

$$\Delta \tilde{Y}^{\prime\prime} = \alpha^{-1} \Delta \tilde{Y} + \cdots$$

Hence,

$$\mu^{2} = 5(2d - 5)(\alpha - 1) + (\alpha^{-1} - 1)$$

= (1 - \alpha)(\alpha^{-1} - 5(2d - 5)). (4.8)

From (2.11), it follows that

$$\alpha = \frac{\rho}{p} \frac{1 - p^2}{1 - \rho^2} = \frac{p^4}{1 + p^2 + p^4 + p^6 + p^8},$$
(4.9)

$$\mu^{2} = p \frac{(1+p^{2})(1+p^{6})}{1+p^{2}+p^{4}+p^{6}+p^{8}} (1+p^{2}+p^{4}+p^{6}+p^{8}-5(2d-5)p^{4}). \quad (4.10)$$

At d = 3, this expression agrees with the formula given in [11]. Comparing eqs. (4.10) and (3.2), one sees that μ^2 vanishes at the singular point, as expected. As $d\rho/dt$ has a square root singularity at t_c , μ which is proportional to $[(dt/d\rho)(1-p^2)^{2d-5}]^{1/2}$ vanishes with the critical exponent $\nu = \frac{1}{2}$ at d = 3, $\nu = \frac{1}{4}$ at $d \ge 4$:

$$\xi^{-1} = \mu \sim |t - t_{\rm c}|^{\nu}.$$

5. The Z_4 model in four dimensions

The general plaquette action of a Z_4 theory depends on two couplings β_1 and β_2 [6]:

$$e^{s(U)} = \exp \beta_1 (U + U^3) + \beta_2 U^2$$

= $\tilde{\beta}_0 [1 + t_1 (U + U^3) + t_2 U^2],$ (5.1)

where

$$4\tilde{\beta}_{0} = e^{2\beta_{1}+\beta_{2}} + 2e^{-\beta_{2}} + e^{-2\beta_{1}+\beta_{2}},$$

$$4\tilde{\beta}_{0}t_{1} = e^{2\beta_{1}+\beta_{2}} - e^{-2\beta_{1}+\beta_{2}},$$

$$4\tilde{\beta}_{0}t_{2} = e^{2\beta_{1}+\beta_{2}} - 2e^{-\beta_{2}} + e^{-2\beta_{1}+\beta_{2}}.$$
(5.2)

In terms of the t variables, the physical region, β_1 , β_2 real corresponds to the triangle:

$$t_2 < 1, \qquad 1 + t_2 - 2t_1 \ge 0, \qquad 1 + t_2 + 2t_1 \ge 0$$
 (5.3)

(see fig. 4).

It is easy to define the Bethe approximation for the Z_4 model in four dimensions, following the lines of sect. 2. The relevant combinations $A_3(\rho)$ and $A_4(\rho)$ read:

$$A_{4}(\rho_{1},\rho_{2}) = 1 + 3\rho_{2}t_{2} + 3\rho_{2}^{2} + \rho_{2}^{3}t_{2} + 6\rho_{1}t_{1}(1+\rho_{2})^{2} + 6\rho_{1}^{2}(1+\rho_{2})(1+t_{2}) + 8\rho_{1}^{3}t_{1},$$

$$A_{4}(\rho_{1},\rho_{2}) = 1 + 4\rho_{2}t_{2} + 6\rho_{2}^{2} + 4\rho_{2}^{3}t_{2} + \rho_{2}^{4}$$

$$+ 8\rho_{1}t_{1}(1+\rho_{2})^{3} + 12\rho_{1}^{2}(1+\rho_{2})^{2}(1+t_{2})$$

$$+ 32\rho_{1}^{3}t_{1}(1+\rho_{2}) + 8\rho_{1}^{4}(1+t_{2}), \qquad (5.4)$$

in terms of which may be written the equations for ρ_1 and ρ_2 , and the expression of



Fig 4 The phase diagram in the $t_1 - t_2$ plane Full lines are the borders of the domains of existence of solutions (IV) (ACW and HX) and (III) (line BC) Along the broken line WTX the mass μ_3 vanishes Dot-dashed lines indicate where the actual first-order transition takes place The shaded area is where the Bethe approximation is inconsistent As nothing dramatic happens across the line $1 + t_2 = 2t_1$ (border of the physical region), some lines have been continued to the point X

the free energy

$$\rho_1 = \rho_{\bar{1}} = \left(\frac{\frac{1}{8}\partial A_4/\partial \rho_1}{A_3}\right)^3;$$

$$\rho_2 = \left(\frac{\frac{1}{4}\partial A_4/\partial \rho_2}{A_3}\right)^5.$$
(5.5)

There are some limiting cases where the model is known to simplify:

(i) for $t_1 = 0$, i.e. $\beta_1 = 0$, the model represents a double covering of Z_2 . This is clear on (5.4) where $A_3(\rho_1 = 0, \rho_2)$, $A_4(\rho_2)$ reduce to the corresponding expression for Z_2 .

(ii) for $t_2 = 1$, i.e. $\beta_2 = \infty$, only $U^2 = 1$ contribute in (5.1), and the model is again a Z_2 model. Again, this is consistent with (5.4)–(5.5).

(iii) for the Wilson action $\beta_2 = 0$, one sees that $t_2 = t_1^2$, and that (5.4) and the resulting equations for ρ_1 and ρ_2 are consistent with $\rho_2 = \rho_1^2$, where ρ_1 is given again by the Z_2 theory. It has been noted by Creutz and Roberts [12] that the Z_4 theory differs from a $Z_2 \times Z_2$ model only by the contribution of non-orientable closed diagrams. As the Cayley lattice does not contain any such surface, the reduction to a Z_2 system is natural.

(iv) for $t_1 = t_2$, i.e. for the Potts gauge model $\beta_1 = \beta_2$, the system admits a solution $\rho_1 = \rho_2$ which, however, is not the solution of the \mathbb{Z}_2 model:

$$\beta = \left(\frac{\left(t+3\rho+3\rho^2 t+\rho^3\right)+6\rho t+18\rho^2 t+6\rho^2+6\rho^3+20t\rho^3}{\left(1+3\rho t+3\rho^2+\rho^3 t\right)+6\rho t+18\rho^2 t+6\rho^2+6\rho^3+20t\rho^3}\right)^5.$$
 (5.6)

One can restrict the domain (5.3) to $t_1 > 0$, because as $t_1 \rightarrow -t_1$, $\rho_1 \rightarrow -\rho_1$ and $\rho_2 \rightarrow \rho_2$. The system (5.5) has a few simple solutions:

(I) $\rho_1 = \rho_2 = 1$ is always a solution, as discussed in sect. 2 and will represent the weak coupling phase.

(I') $\rho_1 = -1$, $\rho_2 = 1$ is also a solution, but does not seem to play any physical role for $t_1 > 0$.

(II) $\rho_1 = 0$, $\rho_2 = -1$ is a solution, and will represent the "antiferromagnetic" phase which develops at large negative β_2 .

(III) $\rho_2 = 1$, ρ_1 non-trivial is also possible; ρ_1 must satisfy:

$$\rho_1 = \left(\frac{t' + 3\rho_1 + 3\rho_1^2 t' + \rho_1^3}{1 + 3\rho_1 t' + 3\rho_1^2 + \rho_1^3 t'}\right)^5,$$
(5.7)

i.e. is a solution of a Z_2 model with an effective inverse temperature

$$t' = \frac{2t_1}{1+t_2} \,. \tag{5.8}$$

(IV) Finally, the system possesses, as usual, a strong coupling solution, such that $\rho_1 \sim t_1^5$, $\rho_2 \sim t_2^5$.

Obviously, besides these solutions (I)–(IV), the algebraic system of equations for ρ_1 , ρ_2 has a huge number of roots, physical or unphysical. The physical solutions have to pass a few tests of stability (the masses μ^2 computed as in sect. 4 have to be positive), but it is not very clear how to find all of them.

On fig. 4, the relevant solutions in various parts of the (t_1, t_2) plane have been displayed. The solid lines indicate where the solutions (IV) and (III) become singular and cease to exist. Near these lines, the system experiences a first-order transition.

Special cases have been indicated: P corresponds to the Potts model $(t_1 = t_2, \rho_1 = \rho_2)$, W to the Wilson action $(t_2 = t_1^2, \rho_2 = \rho_1^2)$. The broken line WX deserves a particular discussion. It is a singular line of the system, although the strong coupling solution (IV), ρ_1, ρ_2 real, exists on both sides and has no singularity. To understand what happens there, we have to compute the correlation lengths of the model.

The correlation functions $\langle \chi_r(U(0))\chi_s(U(R))\rangle$ and their asymptotic fall off may be studied by the same methods as in sect. 4. They are parametrized in terms of functions \tilde{Y}_r , \tilde{W}_{rs} , r, s = 1, 2, 3, three of which only are linearly independent. As a consequence, there are three independent correlation lengths, $\mu_1^2, \mu_2^2, \mu_3^2$. Typically, if $\mu_1 = \inf(\mu_1, \mu_2)$

$$\langle \operatorname{Re} U_{\mathbf{P}}(0) \operatorname{Re} U_{\mathbf{P}}(R) \rangle \underset{R \to \infty}{\sim} e^{-\mu_{1}R} + O(e^{-\mu_{2}R}),$$
$$\langle U_{\mathbf{P}}^{2}(0) U_{\mathbf{P}}^{2}(R) \rangle \sim e^{-\mu_{1}R} + O(e^{-\mu_{2}R}),$$
$$\langle \operatorname{Im} U_{\mathbf{P}}(0) \operatorname{Im} U_{\mathbf{P}}(R) \rangle \sim e^{-\mu_{3}R},$$

with

$$\mu_3^2 = (1 - \alpha_3) \left(\alpha_3^{-1} - 15 \right)$$
$$\alpha_3 = \frac{1 - p_2}{1 - \rho_2} \frac{\rho_1}{p_1}.$$

At special points of the phase diagram, these masses may be related. For instance, for both the Potts model and the Wilson action, $\mu_1 = \mu_3 \neq \mu_2$. But in general, μ_1^2 and μ_3^2 may vanish at different points. Along the broken line WX of fig. 4, the mass μ_3 vanishes but $\mu_1^2 > 0$. This means that if one follows the solution (IV) across the line, the system will be unstable for the modes of mass μ_3 . In turn, this suggests that $\langle \text{Im } U_P \rangle$ might be nonvanishing and that the solution $\rho_1 = \rho_{\bar{1}}$ real has to be abandoned in favor of a solution $\rho_1 \neq \rho_{\bar{1}}$. It seems natural to replace the condition (2.5) by the weaker one $\rho_{\bar{1}} = \overline{\rho_1}$. Indeed, if one returns to (2.3) for $\rho_1 \neq \rho_{\bar{1}}$, one sees that ρ_1 and $\rho_{\bar{1}}$ tend to develop an imaginary part beyond the line WX. Unfortunately, this solution is not physically reasonable: its free energy computed from (2.17) is *lower* than the free energy of the original real solution. This is clearly a paradoxical situation: on the one hand the real solution has a negative correlation length squared and is unstable towards a complex solution; on the other hand, the complex solution exists but is not thermodynamically favored.

Actually, the derivation of eq. (2.3) with $\rho_1 \neq \rho_{\bar{1}}$ assumes implicitly that a coherent choice of orientation of plaquettes has been done throughout the lattice. Such choices conflict with the assumption of homogeneity of the Cayley lattice necessary in eqs. (2.3) and (2.13). It seems that we have reached the limit of our model: going



Fig 5 The phase diagram in the $\beta_1 - \beta_2$ plane Full lines are first order, the broken line is second order and the phase boundaries found in Monte Carlo simulations [6] are indicated by dotted lines

beyond the broken line WX is not fully consistent. As a consequence, the description of that phase and its transition line to the weak coupling phase (I) are missing. This has been indicated by a shaded area on figs. 4 and 5.

We now return to the other transition lines (full lines of fig. 4). Across all these lines, the parameters ρ_1 , ρ_2 are discontinuous. The actual location of the first-order transition lines must be deduced from the comparison of free energies in the various phases, as explained in sect. 3. The arbitrary constant in the free energy is adjusted to reproduce the leading strong or weak coupling behaviour in each phase. One finds

$$C_{\rm IV} = 0$$
, $C_{\rm II} = C_{\rm III} = -\frac{1}{6}\ln 2$, $C_{\rm I} = -\frac{1}{3}\ln 2$,

which is consistent with the special limiting cases mentioned above and the similar procedure for Z_2 . In that way, the phase boundaries represented on fig. 4 (dot-dashed lines) and 5 (full lines) are obtained. On fig. 5, which displays the $(\beta_1 - \beta_2)$ plane, the results of Monte Carlo simulations [6] have also been plotted. There is clearly an excellent agreement, as far as the first-order lines are concerned. One can make two remarks:

(i) there is no built-in duality in the Bethe approximation since the Cayley lattice is not self-dual. The transition from the weak (I) to the strong (IV) coupling phase does not occur along the line $2t_1 + t_2 = 1$ but at slightly weaker coupling. Actually, the weak coupling free energy $f - \ln \tilde{\beta}_0$ (cp. 2.17) vanishes along the self-dual line: f_I is lower than it should be because the system is totally frozen in our approximation. For example, the Potts transition point P lies at $t_1 = t_2 = 0.3357$, i.e. $\beta_1 = \beta_2 = 0.276$ instead of the self-dual value $t_1 = t_2 = \frac{1}{3}$; the triple point C is at $\beta_1 = 0.224$, $\beta_2 = 0.397$, and the transition of the Wilson model at $\beta_1 = 0.450$, $\beta_2 = 0$.

(ii) in the lower half-plane, the second-order transition crosses the first-order line and becomes visible at the tricritical point T: $\beta_1 = 0.522$, $\beta_2 = -0.09$. This is a little higher than the tricritical point determined by Monte Carlo simulations [6].

Finally, as explained before, we cannot determine within the present scheme where the new phase terminates.

6. Summary and discussion

We have shown that the Bethe approximation may be defined in lattice gauge theories of arbitrary group and dimension. Although the approximation is essentially a resummation of (a large part of) strong coupling series, it is flexible enough to reproduce quite accurately phase diagrams and to describe correctly the order of phase transitions. This is true at least for discrete gauge groups as illustrated here by the $4 - d Z_4$ model. It is not the point here to compare this approach to others based on mean field [13] or on Migdal-Kadanoff recursion relations [12] but only to show that the Bethe approximation may successfully reproduce complicated phase patterns. Another instance, not discussed here, is the 3-d 3-state Potts model, where the approximation correctly predicts a first-order transition with a small latent heat. The approximation has failed to provide a consistent description of the intermediate ("Coulomb") phase. This may not be too surprising if one remembers that in the original model, closed monopole loops on the dual lattice play an important role: the dual of the Cayley lattice, however, has no closed loop. The approximation has been able to determine one boundary of this Coulomb phase. Moreover, it predicts that along this line, only the correlation length of $\langle Im U_P Im U_P \rangle$ vanishes, while the others remain finite. Such a decoupling is not uncommon and has been, for example, observed in the Higgs model [14]. In the present case, it should be easy to test in the original model. One may go further and speculate that in the intermediate phase, $\langle Im U_P \rangle$ takes a nonvanishing value, signaling the spontaneous breakdown of the discrete symmetry $U_l \rightarrow U_l^{\dagger}$. A similar phenomenon has been recently discussed in the closely related gauge Ashkin-Teller model [15]. On the other hand, if for other Z_N models only some of the correlation lengths vanish at the transition point, it will be interesting to see how the U(1) limit is approached: there, we expect both correlation functions $\langle \cos \theta_0 \cos \theta_R \rangle$ and $\langle \sin \theta_0 \sin \theta_R \rangle$ to have divergent correlation lengths at the critical point.

Returning to the Bethe approximation, it is clear that more work is needed to see how valuable it is for continuous gauge groups. It might, for example, describe correctly the phase structure and nearby singularities of SU(N) mixed actions. It should also offer an effective way to resum strong coupling series for glueball masses. Notice, however, that the mass defined as in (4.1) or in [12] (i.e. the first coefficient in the low-momentum expansion of the inverse correlation function) differs from the definition currently used [2], which is the pole in k^2 of the correlation function: the two definitions differ at strong coupling but, of course, merge at a critical point. Also, we observe that a computation of the string tension is a priori feasible in the Bethe approximation, although an area-law may hardly be distinguished from a perimeter law on a typical closed curve drawn on a Cayley lattice! Finally, I mention that the Bethe approximation, which ultimately derives self-consistent equations for an effective plaquette average, might have some relationship with the mean-plaquette approximation [16].

I want to thank M. Creutz, L. Jacobs, M. Ogilvie, J. Perk, C. Rebbi and R. Shrock for fruitful discussions. Part of this work has been done when the author was visiting the Institute for Theoretical Physics at Santa Barbara and the Physics Department at Brookhaven. The nice hospitality of these institutes and of the Institute for Theoretical Physics at Stony Brook is gratefully acknowledged. This work was supported in part by NSF Grant PHY 81-09110 A-01.

Note added in proof. A recent Monte Carlo simulation of M. Okawa (private communication) seems to rule out the possibility of a nonvanishing value of $\langle \text{Im} U_{\text{P}} \rangle$ in the $4 - d Z_4$ model.

References

- [1] M. Creutz, L Jacobs and C Rebbi, Phys Reports 95 (1983) 201
- [2] J M Drouffe and J-B Zuber, Phys Reports, 102 (1984) 1
- [3] H A Bethe, Proc Roy Soc A150 (1935) 552,
 R Peierls, Proc Roy Soc A154 (1936) 207
- [4] D M Burley, in vol 2, and C Domb, in vol 3 of Phase transitions and critical phenomena, eds C Domb and M S Green (Academic Press),
- J M Ziman, Models of disorder (Cambridge University Press, 1979)
- [5] C Itzykson, R B Pearson and J B Zuber, Nucl Phys B220[FS8] (1983) 415
- [6] M Creutz and M Okawa, Nucl Phys B220[FS8] (1983) 19,
 F C Alcaraz and L Jacobs, Phys Rev D27 (1983) 938;
 M Fukugita, M Kobayashi and T Kaneko, Nucl Phys B215[FS7] (1983) 289
- [7] E Muller-Hartmann, Z Phys B27 (1977) 161, and references therein
- [8] A Maritan and A Stella, Padova preprint
- [9] S Ehtzur, Phys Rev D12 (1980) 2291
- [10] J M Drouffe, G Parisi and N Sourlas, Nucl Phys B161 (1979) 397
- [11] R.J Elhott and W Marshall, Rev Mod Phys 30 (1958) 75, H B Tarko and M E Fisher, Phys Rev B11 (1975) 1217
- [12] M Creutz and L E Roberts, Nucl Phys B215[FS7] (1983) 447, L E Roberts, Brookhaven preprint
- [13] K Odaka and M Fukugita, Nucl Phys B220[FS8] (1983) 471,
 V Alessandrini and P Boucaud, B225[FS9] (1983) 303
- [14] E Brézin and J M Drouffe, Nucl Phys. B200[FS4] (1982) 93
- [15] F C Alcaraz and L Jacobs, Phys Rev Lett 51 (1983) 530
- [16] G G Batroum, Nucl Phys. B208 (1982) 12, 467