Method of loops applied to lattice gauge theory

Bernd Brügmann

Physics Department, Syracuse University, Syracuse, New York, 13244

The eigenvalue problem for the Hamiltonian of SU(2) lattice gauge theory is formulated in the loop representation, which is based on the fact that the physical Hilbert space can be spanned by states which are labeled by loops. Since the inner product between loop states can be calculated analytically, the eigenvalue problem is expressed in terms of vector components and matrix elements with respect to the loop basis. A small-scale numerical computation in 2+1 dimensions yields results which agree with results obtained from other methods.

I. INTRODUCTION

Loops have been recognized as playing an important role in gauge theories. First of all, the field tensor of the gauge field is directly related to parallel transport around a closed curve. This basic observation is encouragement enough for an attempt to formulate gauge theory in terms of quantities associated with loops, the ultimate goal being a nonperturbative approach to Yang-Mills theory. Mandelstam[1] made a first step in this direction even though he used open lines instead of closed loops. To varying extent and success a space of loops figured in the work of Kogut and Susskind,[2] Polyakov,[3] Makeenko and Migdal,[4] 't Hooft,[5] Gambini, Leal, and Trias,[6,7] Furmanski and Kolawa,[8] Rovelli and Smolin,[9] and others.

Because of their origin, loops correspond to the gauge-invariant objects of the theory. In fact, one can show that the physical Hilbert space of gauge-invariant states can be spanned by states which are labeled by loops [see Giles[10] for the SU(N) case]. The above-mentioned works differ mainly in how they propose to systematically utilize these observations.

Let me mention that the property of loop states to automatically solve the constraints of the theory has applications in quantum gravity. Here the diffeomorphism invariance of the loop states can be realized by labeling states by knot classes of loops, which are just the equivalence classes of loops under diffeomorphisms. The most important example arises in full (3+1)-dimensional quantum gravity. The formulation of general relativity in terms of the Ashtekar connection admits a transition to a loop representation, and a large space of solutions to all the constraints of quantum gravity can be obtained.[11]

In Refs. 7 and 9 a loop representation for Hamiltonian lattice gauge theory is constructed, complete with an inner product, operators, and dynamics. One can analyze this approach on two levels, asking whether loops are physically fundamental and whether one has found a useful mathematical formalism. I want to address these questions in SU(2) Hamiltonian lattice gauge theory in the loop representation of Rovelli and Smolin.[9] To this end I develop a straightforward formulation of the eigenvalue problem for the Hamiltonian, which is well suited for a numerical solution.

That this can be done supports the claim that the method of loops is useful. Furthermore, the results obtained suggest that the method of loops should be further investigated as a new method for computations in Hamiltonian lattice gauge theory. Even though Hamiltonian lattice gauge theory possesses inherent computational advantages over Lagrangian formulations—the lattice is only three dimensional and the gauge degrees of freedom can be eliminated—so far the Hamiltonian methods are inferior for technical reasons. However, the successes of lattice computations for QCD (e.g., fermion masses) as well as current limitations imposed by computing power make any improvement of the Hamiltonian approach attractive.

In response to the question whether loops are physically fundamental, loops are found to be "physical" in the sense that physical statements become statements about geometric properties of loops. For example, the action of the Hamiltonian operator on a loop state changes the loop according to simple geometric rules. Also, from strong coupling down to some finite value of the coupling constant, the physical Hilbert space can be spanned by states corresponding to short loops. This last fact is also important for a technical reason, since it suggests a simple way to truncate the infinite loop basis to obtain a finite problem.

Apart from the loop representation, the key observations underlying my approach to the eigenvalue problem for the Hamiltonian operator is that one can perform the integration in the inner product analytically (e.g., see Creutz[12]). A new element of the proposed method is the explicit use of the inner product.

The method can be described as follows. Consider SU(2) Hamiltonian lattice gauge theory on a finite lattice with periodic boundary conditions. (1) From the infinite set of loop states select the finite number of loop states corresponding to loops up to a certain length. These loops will in general not be linearly independent, which creates problems for many applications of the loop representation. (2) Use the inner product to find a loop basis of linearly independent loop states (which could be done, for example, by the Gram-Schmidt procedure). (3) Formulate the eigenvalue problem for the Hamiltonian $\hat{H}|\psi\rangle = \epsilon|\psi\rangle$ in terms of vector components and matrix

elements with respect to the loop basis. This is efficient since the matrices involved are sparse. (4) Feed the resulting matrix eigenvalue problem to some program-library routine for large sparse eigensystems.

As a small-scale numerical computation on a Vax 8810 I could handle a 4×4 lattice with loops up to length 12. Of immediate physical interest are the ground state, its energy, and the mass gap (or glueball mass) of the model for different values of the coupling constant. These quantities are obtained, and they agree with the results of standard variational methods. The components of the ground state indicate for which range of coupling the ansatz of short loops is valid, and which particular type of loop is important. What is unusual is that, in addition, some thirty of the lowest-energy states can be obtained; i.e., one can discuss the "glueball" (the first excited state) and the lowest part of the energy spectrum.

The organization of this paper is as follows. In Sec. II, Hamiltonian lattice gauge theory is reviewed. In Sec. III, I introduce the loop representation and discuss the linear dependencies among loop states. In Sec. IV, the inner product between loop states is presented. In Sec. V, I compute the action of the Hamiltonian operator on loop states. In Sec. VI, the ansatz of short loops is made and the eigenvalue problem is formulated. In Sec. VII, I briefly discuss the program to clarify its strong and weak points. In Sec. VIII, the numerical results are presented and discussed. I conclude with a summary and an outlook in Sec. IX. Readers not interested in the details of the method can skip Secs. IV, V, and VII.

II. HAMILTONIAN LATTICE GAUGE THEORY

In this section the theory is defined in the standard representation giving as much detail as is needed for the transition to the loop representation. See Refs. 2 and 13 for a complete treatment.

In Hamiltonian lattice gauge theory time is continuous while space is represented by a cubical lattice of "sites" (Fig. 1). The scale of the lattice is defined by the lattice spacing a. In what follows the lattice is a 2- or 3-dimensional box of $S \times S$ or $S \times S \times S$ sites with periodic boundary conditions. A "link" l is a pair of neighboring sites $s_1 s_2$ and is thought of as the line connecting them. By choosing an orientation for each link l which I call positive ($l > 0$), one obtains oriented links l and their inverses denoted as l^{-1}.

The classical configuration space is the Cartesian product $G \times \cdots \times G$ of a Lie group G, one copy for each positively oriented link of the lattice. Here I consider $G = \mathrm{SU}(2)$. A configuration U is determined by a choice of a 2×2 SU(2) matrix $U(l)_A{}^B$ for each link $l > 0$:

$$U \equiv (U(l_1)_A{}^B, \ldots, U(l_m)_C{}^D) . \tag{1}$$

With each negatively oriented link $l^{-1} < 0$ one associates

$$U(l^{-1})_A{}^B = U^{-1}(l)_A{}^B . \tag{2}$$

In quantum theory the configuration operators are 2×2 SU(2)-matrix-valued operators $\hat{U}(l)_A{}^B$, i.e., three for each positively oriented link l. The momentum operators on each link l are denoted as $\hat{p}_m(l)$, $m = 1, 2, 3$. The canoni-
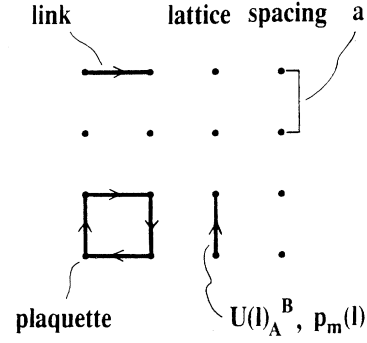

FIG. 1. A 4×4 lattice.

cal commutation relations for $l > 0$ and $l' > 0$ are

$$[\hat{U}(l)_A{}^B, \hat{U}(l')_C{}^D] = 0 , \tag{3}$$

$$[\hat{p}_m(l), \hat{U}(l')_A{}^B] = \tfrac{1}{2} \hat{U}(l)_A{}^C \sigma_{mC}{}^B \delta_{ll'} , \tag{4}$$

$$[\hat{p}_m(l), \hat{p}_n(l')] = i \epsilon_{mnk} \hat{p}_k(l) \delta_{ll'} , \tag{5}$$

where $\delta_{ll'}$ is 1 if $l = l'$ and 0 otherwise, and σ_m are the Pauli matrices satisfying $[\sigma_m, \sigma_n] = 2i\epsilon_{mnk}\sigma_k$.

In the U representation of this algebra, $\hat{U}$ becomes a multiplication operator and $\hat{p}$ becomes a certain derivative operator. The Hilbert space is the direct product of spaces of square-integrable functions over the gauge group SU(2). A state $|\psi\rangle$ in this space is specified by a wave function $\psi(U)$ depending on a configuration U as given in (1). To be explicit I choose a (non-normalizable) basis of states $|U\rangle$ such that

$$\int [dU] |U\rangle \langle U| = \hat{I}, \quad \text{completeness} , \tag{6}$$

$$\langle U|U'\rangle = \prod_{l>0} \delta(U(l), U'(l)), \quad \text{normalization} . \tag{7}$$

The measure is $[dU] \equiv \prod_{l>0} dU(l)$, where $dU(l)$ is the Haar measure on the compact group SU(2), and the δ function is defined with respect to $dU(l)$. Then

$$|\psi\rangle = \int [dU] |U\rangle \psi(U), \quad \text{where } \psi(U) \equiv \langle U|\psi\rangle , \tag{8}$$

and the inner product defining square integrability is

$$\langle \phi|\psi\rangle = \int [dU] \phi^*(U)\psi(U) . \tag{9}$$

On this Hilbert space the operators $\hat{U}(l)$ and $\hat{p}(l)$, $l > 0$, act according to

$$\hat{U}(l)_A{}^B \psi(U) = U(l)_A{}^B \psi(U) , \tag{10}$$

$$\hat{p}_m(l)\psi(U) = \tfrac{1}{2} U(l)_A{}^B \sigma_{mB}{}^C \frac{\partial}{\partial U(l)_A{}^C} \psi(U)$$

$$\equiv \tfrac{1}{2} \mathrm{tr}\, U(l) \sigma_m \frac{\partial}{\partial U(l)} \psi(U) . \tag{11}$$

Note that the $|U\rangle$'s are the eigenstates of $\hat{U}$: $\hat{U}(l)_A{}^B |U\rangle = U(l)_A{}^B |U\rangle$. Finally, I extend the above to negative links $l^{-1} < 0$ by defining

$$\hat{U}(l^{-1})_A{}^B = U^{-1}(l)_A{}^B \ , \tag{12}$$

which implies for the momentum operator

$$\hat{p}_m(l^{-1}) = -\tfrac{1}{2}\mathrm{tr}\sigma_m U(l)\frac{\partial}{\partial U(l)} \ . \tag{13}$$

Note the difference between this expression and expression (11) for $\hat{p}_m(l)$.

Now I can give the constraints and the Hamiltonian to specify the physical system. The physical states have to be invariant under the local SU(2) rotation which is generated by the Gauss-law constraint of the theory at each site s:

$$\hat{C}_m(s) = \sum_{l(s)}\hat{p}_m(l), \quad \hat{C}_m(s)\psi(U)=0 \ , \tag{14}$$

where $l(s)$ runs over all oriented links emanating from s.

The dynamics of SU(2) Hamiltonian lattice gauge theory in $3+1$ dimensions is defined by the Hamiltonian operator (Kogut and Susskind[2])

$$\hat{H} = \frac{g^2}{2a}\sum_{l>0}\hat{p}_m(l)\hat{p}_m(l) + \frac{2}{ag^2}\sum_{\square}(2-\mathrm{tr}U_{\square}) \ , \tag{15}$$

where g is the coupling constant and l runs over all positive links in the lattice. $\sum_{\square}$ denotes the sum over all generic "plaquettes" $\square$ of the lattice, a plaquette being four positive links l_1,l_2,l_3,l_4 forming a square. $U_{\square}$ stands for $U(l_1)U(l_2)U(l_3^{-1})U(l_4^{-1})$ assuming that the orientation of the links $l_1,l_2,l_3^{-1},l_4^{-1}$ defines a consistent overall orientation for the plaquette. At this point let me mention that in the literature there are versions of $\hat{H}$ which differ from (15) in coefficients and type of summation—the latter is often not even specified. For the purpose of quantitative comparison one has to be careful about which $\hat{H}$ is used.

III. THE LOOP REPRESENTATION

The loop representation is motivated by the direct relationship between loops and the physical (gauge-invariant) states of gauge theory. Therefore let me start below by presenting this relationship. The attempt to systematically label states by loops leads to the definition of the loop representation as given previously by Gambini, Leal, and Trias[7] and Rovelli and Smolin,[9] the latter of which I want to follow. Finally, I will discuss the identities between loop states, which have to be incorporated when one wants to construct a loop basis.

In lattice gauge theory the fundamental object is the trace

$$\mathrm{tr}U(l_1)U(l_2)\cdots U(l_n) \tag{16}$$

of matrices $U(l_i)$ which represent elements of the gauge group (Wilson[14]). This term is gauge invariant if and only if the links $l_1,l_2,\ldots,l_n$ form an oriented closed line in the lattice, i.e., a loop.

Let γ be the loop formed by $l_1,l_2,\ldots,l_n$: $\gamma \equiv [l_1 l_2 \cdots l_n]$. Define $U_\gamma \equiv U(l_1)U(l_2)\cdots U(l_n)$ if $\gamma = [l_1 l_2 \cdots l_n]$. In general, a gauge-invariant state $\psi(U)$ could depend on $\mathrm{tr}U_\gamma$ for any loop γ, $\psi(U)=\psi(\mathrm{tr}U_{\gamma_1},$

$\mathrm{tr}U_{\gamma_2},\ldots)$, where γ_i, $i=1,\ldots,\infty$, denotes all the possible loops on the lattice. The importance of loops is due to the fact that wave functions of the form

$$\psi(U)=h(\gamma,U)\equiv \mathrm{tr}U_\gamma \tag{17}$$

span the space of physical states.[10]

Actually, it turns out to be more natural to work with "multiple loops." Let me call a multiloop any unordered collection of single loops (possibly containing a single loop several times) and write for example $\gamma=\alpha\cup\beta$, where "$\cup$" indicates a collection of loops in the sense of one discontinuous parametrization and does not mean a union of sets of links, i.e., $\alpha\cup\alpha\neq\alpha$. For such a multiloop the definition of the loop wave function h is

$$h\left[\alpha\cup\beta,U\right]=h(\alpha,U)h(\beta,U)=\mathrm{tr}U_\alpha \mathrm{tr}U_\beta \ . \tag{18}$$

From here on the term loop refers to multiloops.

Now I can define the loop representation as follows. Loop states, or just loops, are special vectors $|\gamma\rangle$ in the Hilbert space of states, which can be labeled by loops. Their components in the U representation are defined to be $h(\gamma,U)$:

$$|\gamma\rangle = \int [dU]|U\rangle\langle U|\gamma\rangle \ , \tag{19}$$

$$\langle U|\gamma\rangle = h(\gamma,U)=\mathrm{tr}U_{\alpha_1}\mathrm{tr}U_{\alpha_2}\cdots \mathrm{tr}U_{\alpha_n} \ , \tag{20}$$

where α_i are single loops and $\gamma=\alpha_1\cup\cdots\cup\alpha_n$. The span of $\{|\gamma\rangle; \gamma \text{ any loop}\}$ is the physical Hilbert space. This constitutes a major advantage of the loop representation—I do not have to deal with the gauge degrees of freedom.

Note that $h(\gamma,U)$ is real for SU(2). Therefore the inner product between loop states is symmetric. In terms of h it becomes

$$\langle\alpha|\beta\rangle=\langle\beta|\alpha\rangle=\int [dU]h\left[\alpha\cup\beta,U\right] \ . \tag{21}$$

For completeness I include the zero-length loop γ_0 defined by $h(\gamma_0,U)=\mathrm{tr}I=2$. [Clearly, $h(\gamma_0,U)=\text{const}$ is a "cyclic" loop of lattice gauge theory. That is, by applying the Wilson operator $W(\gamma)\equiv \mathrm{tr}U_\gamma$ on a constant, one can generate a basis for the Hilbert space of physical states. W plays the role of the creation operator on the space of loops. In this paper no use is made of this fact.]

The next logical step is to determine how the operators $\hat{U}$ and $\hat{p}$ act on loop states $|\gamma\rangle$. From (10) and (11) it follows that even though $\hat{U}$ and $\hat{p}$ act naturally on $h(\gamma,U)$, this does not translate into a natural action on loops via the transformation (19). Indeed, there is no good way to represent $\hat{U}$ and $\hat{p}$ in the loop representation. On the other hand, the Hamiltonian operator $\hat{H}$ of (15) maps loops to a finite linear combination of loops in a simple geometric fashion. Hence the loop representation is certainly a promising approach for issues involving $\hat{H}$ or other suitably defined operators.

Before I calculate the action of $\hat{H}$ in the next section, I have to point out that the loop states as defined above are not linearly independent, and for practical purposes one

usually needs a basis. Notice that since $|\gamma\rangle$ is defined in terms of traces, there exist several relations between loop states, which I describe below as rules (i)–(v). Rules (i)–(iv) will be treated as equivalence relations between *loops* defining what I want to consider as geometrically equivalent loops. Rule (v), however, introduces a non-trivial linear dependency between loops. Since the loop states are completely labeled by loops, one can represent each state by a picture of the loop. For example, $|\gamma\rangle=\square$ if γ is a plaquette, where the position of the loop in the lattice has to be clear from the context. Figure 2 gives some examples involving the following rules.

For any oriented lines u, v, and w, which are connected with consistent orientation when written as uv, and form loops when enclosed as $[uv]$, one finds (equivalences denoted by $=$)

(i) $\begin{aligned}[u]\cup[w]&=[w]\cup[u]\\ [u]&=[v]\Longrightarrow[u]\cup[w]=[v]\cup[w]\end{aligned}$ multiloops , (22)

(ii) $[uww^{-1}]=[u]$ spikes , (23)

(iii) $[uw]=[wu]$ origin , (24)

(iv) $[w]=[w^{-1}]$ orientation , (25)

(v) $[u]\cup[w]=[uw]+[uw^{-1}]$ spinor identity , (26)

where in (v) u and w are closed lines. (i) follows from the definition of $h(\gamma,U)$ for multiloops, expressing the fact that a loop is an unordered collection of single loops, and shows how equivalences between single loops generalize. (ii) follows from $U(l)U^{-1}(l)=I$, indicating that backtracking does not lead to new loops; i.e., "spikes" can be removed [see Fig. 2(a)]. (iii) follows from the cyclic property of the trace; loops do not have a preferred origin. (iv) follows from $\text{tr}\,U=\text{tr}\,U^{-1}$. In the case of SU(2), loops are not distinguished by orientation. For general SU(N) one has $\text{tr}\,U^{-1}=\text{tr}\,U^{\dagger}=(\text{tr}\,U)^*$ so the orientation of loops matters, although this alone is not too great a change.

What really captures the signature of SU(2) is the spinor identity (v). For 2×2 SU(2) matrices A and B, the formula for the inverse matrix (det $A=1$) is

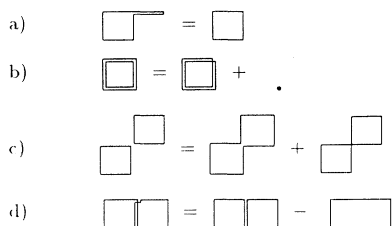$$A^{-1}{}_A^{\,B}=\epsilon_{AC}\epsilon^{BD}A_D{}^C , \qquad (27)$$

FIG. 2. Examples for SU(2) loop identities. (a) Rule (ii); (b)–(d) rule (v) or spinor identity. When loops cross, they are meant to cross and not to just touch. The dot denotes the zero-length loop.

and in two dimensions

$$\epsilon_{AC}\epsilon^{BD}=\delta_A^B\delta_C^D-\delta_A^D\delta_C^B , \qquad (28)$$

which imply

$$\text{tr}\,A\,\text{tr}\,B=\text{tr}\,AB+\text{tr}\,AB^{-1} . \qquad (29)$$

For $A=U_{[u]}$ and $B=U_{[w]}$ one has (v). Basically, whenever a loop intersects itself, there is a triplet of linearly dependent loops which corresponds to the three possible ways to route a loop through its intersection [see Figs. 2(b)–2(d)]. Equation (29) can also produce relations between loops and "nonloops" if AB corresponds to a closed line but A does not. Obviously, it is not trivial to find a complete set of identities between loops, but it can be done [see Ref. 10 for SU(N)]. These identities are called Mandelstam identities, and in the case of SU(2) they reduce to rules (iii) and (v).

Already in the case of SU(3) there exist relations which have a completely different structure. The analogous equations for the inverse and epsilon imply

$$\text{tr}\,AB^{-1}=\tfrac{1}{2}(\text{tr}\,A\,\text{tr}\,B\,\text{tr}\,B-\text{tr}\,A\,\text{tr}\,BB$$
$$-2\,\text{tr}\,AB\,\text{tr}\,B+2\,\text{tr}\,ABB) , \qquad (30)$$

so, in a sense, one loop (B^{-1}) can be worth two loops (B). Although this goes against one's intuition, it is not an obstacle for computations.

The issue is to find a basis of loop states. A possibility considered in the literature is to impose rules (i)–(v) including the Mandelstam identities as equivalence relations on the space of loops. For example, rule (ii) leads to an obvious reduction inside equivalence classes of loops. (v) is an invitation to eliminate all single loops with two identical links l, e.g., $[ulwl]=[ul]\cup[wl]-[uw^{-1}]$ as in Fig. 2(d), although this does not exhaust the whole content of (v). It is remarkable that a reduction scheme of this sort could be carried out systematically.[7] As another example let me give a formula for $\text{tr}\,U^n$. $U_\gamma{}^n$ corresponds to going n times around the same loop γ. Equation (29) gives

$$\text{tr}\,U^n=\text{tr}\,U\,\text{tr}\,U^{n-1}-\text{tr}\,U^{n-2} . \qquad (31)$$

After repeated application of (29), one obtains $\text{tr}\,U^n$ as a polynomial of $\text{tr}\,U$ (with not so simple coefficients). Equation (31) defines a doubly recursive sequence $x_n=cx_{n-1}-x_{n-2}$ which can be summed using techniques similar to those that apply to the Fibonacci series:

$$\text{tr}\,U^n=\frac{1}{2^n}\{[\text{tr}\,U+\sqrt{(\text{tr}\,U)^2-4}\,]^n$$
$$+[\text{tr}\,U-\sqrt{(\text{tr}\,U)^2-4}\,]^n\} . \qquad (32)$$

In this paper I propose what I think is a more straightforward method to construct a loop basis using the inner product. Applying a standard algorithm (Crout's algorithm) the inner product can handle the identities in a uniform way. In the next section I will discuss how one can compute the inner product (21) between loop states.

IV. THE INNER PRODUCT

In Sec. II the choice for the inner product was the Haar measure on the group SU(2). The Haar measure is natural in the sense that given a compact Lie group G, there is a unique measure dU such that

$$\int dU \, 1 = 1 \, , \tag{33}$$

$$\int dU \, f(U) = \int dU \, f(V_1 U V_2) \quad \forall V_1, V_2 \in G, f : G \to \mathbb{C} \, . \tag{34}$$

Equation (33) expresses normalization and (34) the left and right invariance of the measure. From (33) and (34) it follows that

$$\int dU \, f(U) = \int dU \, f(U^{-1}) \, , \tag{35}$$

so the overall orientation of a loop γ will not change the integral $\int dU \, h(\gamma, U)$.

For $G = \mathrm{SU}(N)$ and the fundamental representation of $U \in G$ by $N \times N$ matrices $U_A{}^B$, Creutz[12] was able to use the invariance of the measure to perform the integration of products explicitly. Here I need the result for SU(2), which can be extracted from the graphical notation used

FIG. 3. Examples for the inner product between loop states. The loops under the integrals symbolize whether or not links are doubled.

in Ref. 12. When n is odd, symmetry arguments imply that the integral has to vanish:

$$\int dU \, _{A_1}{}^{B_1} U_{A_2}{}^{B_2} .. U_{A_n}{}^{B_n} = 0 \, . \tag{36}$$

When n is even,

$$\int dU \, U_{A_1}{}^{B_1} U_{A_2}{}^{B_2} \cdots U_{A_n}{}^{B_n} = \frac{1}{(n/2+1)!} \sum_\pi \epsilon_{A_{\pi(1)} A_{\pi(2)}} \cdots \epsilon_{A_{\pi(n-1)} A_{\pi(n)}} \epsilon^{B_{\pi(1)} B_{\pi(2)}} \cdots \epsilon^{B_{\pi(n-1)} B_{\pi(n)}} \, , \tag{37}$$

where π runs over all permutations of the indices $(1,2,\ldots,n)$ which do not lead to identical terms under the sum. Note that $\epsilon_{A_1 A_2} \epsilon_{A_3 A_4} = \epsilon_{A_3 A_4} \epsilon_{A_1 A_2}$ and $\epsilon_{A_1 A_2}{}^{B_1 B_2} = \epsilon_{A_2 A_1}{}^{B_2 B_1}$; therefore, there are

$$\frac{n!}{2^{n/2}(n/2)!} = (n-1)(n-3) \cdots 5 \times 3 \times 1 \tag{38}$$

terms to sum over.

The inner product between two loop states $|\alpha\rangle = |\alpha_1 \cup \cdots \cup \alpha_m\rangle$ and $|\beta\rangle = |\beta_1 \cup \cdots \cup \beta_m\rangle$ can be written schematically as

$$\langle \alpha | \beta \rangle = \int \prod_l dU(l) \mathrm{tr} U_{\alpha_1} \cdots \mathrm{tr} U_{\alpha_m} \mathrm{tr} U_{\beta_1} \cdots \mathrm{tr} U_{\beta_n}$$

$$= (\delta\text{'s}) \left[\int dU(l_1) U \cdots U^{-1} \right] \cdots \left[\int dU(l_p) U \cdots U^{-1} \right] \, . \tag{39}$$

In the second equality I have used (33) for links not in $\alpha \cup \beta$ and collected U's on the same link l or l^{-1}. (δ's) denotes the product $\delta_{A_1}{}^{B_1} \delta_{A_2}{}^{B_2} \cdots$ inducing the products and traces between the matrices $U(l_i)_{A_i}{}^{B_i}$ whose indices I have suppressed. Substituting for the inverse of U

$$U^{-1}{}_A{}^B = \epsilon_{AC} \epsilon^{BD} U_D{}^C \tag{40}$$

we see that $\langle \alpha | \beta \rangle$ is equal to a product of δ's, ϵ's, and integrals over products of U's as dealt with in (36) and (37). There are as many U's in a particular integral as there are occurrences of l and l^{-1} in $\alpha \cup \beta$. Hence, (36) implies the main feature of the inner product:

$$\langle \alpha | \beta \rangle \neq 0 \quad \text{only if } \alpha \cup \beta \text{ contains any (unoriented) link } l$$

an even number of times ;

i.e., for nonzero result the combination of two loops must result in doubled links (see Fig. 3 for examples). This, of course, does not require equal support for α and β, e.g., $\alpha = \beta \cup \gamma$ results in $\langle \alpha | \beta \rangle \neq 0$ when γ has doubled links.

If the inner product is not trivially zero, one can apply (37) for the integration of products of components $U_A{}^B$. The final statement is that one can evaluate $\langle \alpha | \beta \rangle$ by performing a sum over products of contracted δ's and ϵ's, which—as discussed later—can be left to an algebraic manipulation program.

Having found a possibility to compute the inner product, let me now formally construct a loop basis. After restriction to a finite basis this construction can be carried out explicitly. Enumerate all loops as $\tilde\alpha_i$, $i = 1, 2, \ldots, \infty$. The condition for linear independence is

$$\sum_i c_i |\tilde\alpha_i\rangle = 0 \Longrightarrow c_i = 0 \forall i \, . \tag{41}$$

Defining the "metric" on loops by $\tilde{G}_{ij} := \langle \tilde{\alpha}_i | \tilde{\alpha}_j \rangle$, (41) is equivalent to the requirement that the rows of $\tilde{G}_{ij}$ be linearly independent, or that $\tilde{G}_{ij}$ be nondegenerate. Given a finite set $\{|\tilde{\alpha}_i \rangle\}$, I obtain a loop basis $\{|\alpha_i \rangle\} \subset \{|\tilde{\alpha}_i \rangle\}$ by eliminating loops $\tilde{\alpha}_i$ which give rise to degeneracies in $\tilde{G}_{ij}$ (Crout's algorithm; see Sec. VII). $G_{ij} := \langle \alpha_i | \alpha_j \rangle$ is a positive-definite symmetric matrix, but it is not diagonal since linearly independent loops need not be orthogonal [see example (d) in Fig. 3]. One could orthogonalize the basis using linear combinations of loop states as basis states, but one would then lose the direct relation "loop $\leftrightarrow$ basis state" which I want to retain.

V. THE ACTION OF THE HAMILTONIAN ON LOOP STATES

As remarked before, the loop representation is well adapted to the Hamiltonian formulation since the Hamiltonian operator $\hat{H} = (g^2/2a)\hat{H}_K + (2/ag^2)\hat{H}_P$ maps loop states to a simple linear combination of loop states. In this section I will calculate $\hat{H}|\gamma\rangle$ by finding the action of $\hat{H}$ in the U representation and then using the transform

(19), $\hat{H}|\gamma\rangle = \int [dU]|U\rangle \hat{H} h(\gamma, U)$.

First, let me consider the potential term in the Hamiltonian which acts as a multiplication operator. When acting on a loop wave function $h(\gamma, U)$, the potential term augments γ by the plaquette $\square$. More precisely,

$$\hat{H}_P h(\gamma, U) \equiv \sum_{\square} (2 - \mathrm{tr}\,\hat{U}_{\square})\mathrm{tr}\,U_\gamma$$

$$= \sum_{\square} \left[2h(\gamma, U) - h\left[\square \cup \gamma, U\right] \right] . \quad (42)$$

Therefore

$$\hat{H}_P|\gamma\rangle = \sum_{\square} \left[2|\gamma\rangle - |\square \cup \gamma\rangle \right] . \quad (43)$$

The kinetic term acts by applying the derivative operator $\hat{p}$ twice. To write down the result of the product rule, I abbreviate the loop wave function $h(\gamma, U) = \mathrm{tr}\,U_{\alpha_1} \cdots \mathrm{tr}\,U_{\alpha_m}$ as $(\delta\text{'s})U \cdots U$, where γ is a multiloop and $(\delta\text{'s})$ is the product $\delta_A^B \delta_C^D \cdots$ which induces the multiplications and traces between the U's. Then

$$\hat{H}_K h(\gamma, U) = \sum_{l>0} \sum_{j=1}^{n} \sum_{k=1}^{n} (\delta\text{'s})U \cdots [\hat{p}_m(l)U(l_j)] \cdots [\hat{p}_m(l)U(l_k)] \cdots U . \quad (44)$$

From the definition (11) of $\hat{p}$ it follows that

$$\hat{p}_m(l)U(l')_A{}^B = \begin{cases} \frac{1}{2}U(l)_A{}^C \sigma_{mC}{}^B & \text{if } l'=l, \\ -\frac{1}{2}\sigma_{mA}{}^C U(l)_C{}^B & \text{if } l'=l^{-1}, \\ 0 & \text{otherwise.} \end{cases} \quad (45)$$

Hence the only nonzero terms in (44) occur when l_j is equal to l_k up to orientation. In particular, this will be the case when $j = k$. There are six different pairings of links, and the summation reduces to

$$\sum_{l>0} \sum_{j=1}^{n} \sum_{k=1}^{n} = \sum_{j=k, l_j>0} + \sum_{j=k, l_j^{-1}>0} + 2 \left[\sum_{j<k, l_j=l_k>0} + \sum_{j<k, l_j^{-1}=l_k^{-1}>0} + \sum_{j<k, l_j^{-1}=l_k>0} + \sum_{j<k, l_j=l_k^{-1}>0} \right] . \quad (46)$$

Now, $\hat{p}_m$ inserts into the product of U's a Pauli matrix σ_m to the left or right of the matrix U to which it is applied. To perform the implicit summation in the index m remember that

$$\sigma_{mA}{}^B \sigma_{mB}{}^C = 3\delta_A^C , \quad (47)$$

$$\sigma_{mA}{}^B \sigma_{mC}{}^D = \delta_A^D \delta_C^B + \epsilon_{AC}\epsilon^{DB} . \quad (48)$$

It is this last equation which, when inserted between products of U's, leads to a reordering of the U's and therefore to a change of the associated loops. $\delta_A^D \delta_C^B$ has its index structure "crossed," while the second term inverts any SU(2) matrix according to (40). It is not obvious that the resulting contractions correspond to loops and not just to arbitrary arrangements of links. One can distinguish between ten cases according to whether the $\hat{p}_m(l)$'s act in the same or different single loops and how the links l_j and l_k are oriented. Notice, however, that since $\hat{p}_m(l)\hat{p}_m(l) = \hat{p}_m(l^{-1})\hat{p}_m(l^{-1})$, only the relative orientation of l_j with respect to l_k can matter.

Let $U = U(l)$, $\hat{p}_m = \hat{p}_m(l)$, and A and B be any SU(2)

matrices. Then the generic term for the two $\hat{p}$'s hitting the same U is, with (47),

$$\mathrm{tr}\,A\hat{p}_m\hat{p}_m U = \mathrm{tr}\,A\hat{p}_m\hat{p}_m U^{-1} = \frac{3}{4}\mathrm{tr}\,AU . \quad (49)$$

Since the result does not depend on the relative orientation of the link in $\hat{p}$ and U, the first two sums in (46) combine to form a sum which contains each link in the loop exactly once:

$$\left[\sum_{j, l_j>0} + \sum_{j, l_j^{-1}>0} \right] U \cdots \hat{p}_m\hat{p}_m U(l_j) \cdots U$$

$$= \frac{3}{4}\|\gamma\| h(\gamma, U) , \quad (50)$$

where $\|\gamma\| \equiv (\text{length of } \gamma)$ denotes the number of links of γ. The kinetic part of the Hamiltonian therefore contains a diagonal piece.

It will suffice to give an example for the other terms that can appear. Suppose that the $\hat{p}$'s act in different single loops on two links $l_j = l_k^{-1}$. One finds

$$\operatorname{tr} A\,(\hat{p}_m U)\operatorname{tr} B\,(\hat{p}_m U^{-1})=-\tfrac{1}{4}\operatorname{tr} A U\sigma_m\operatorname{tr} B\sigma_m U$$

$$=\tfrac{1}{4}\operatorname{tr} A U B^{-1}U-\operatorname{tr} AB\ . \qquad (51)$$

To begin with, $\operatorname{tr} AU\,\operatorname{tr} BU$ corresponds to a loop, and one should convince oneself that the links and lines connect in such a way that $\operatorname{tr} AUB^{-1}U$ and $\operatorname{tr} AB$ also represent loops.

To describe the action of the Hamiltonian in general I can define a map f which, given a loop γ and two of its links l_j and l_k with $j\neq k$, produces a linear combination of two loops $f(j,k,\gamma)$. Let $\gamma=\alpha\cup\bar{\gamma}$ where α is the single or double loop containing l_j and l_k [rule (i)]. Then $f(j,k,\gamma)=f(j,k,\alpha)\cup\bar{\gamma}$. If α is single, then

$$[ulwl]\overset{f}{\mapsto}[ul]\bigcup[wl]+[uw^{-1}]\quad\text{if } l=l_j=l_k\ ,$$

$$[ulwl^{-1}]\overset{f}{\mapsto}-[u]\bigcup[w]-[ulw^{-1}l^{-1}]\quad\text{if } l=l_j=l_k^{-1}\ . \qquad (52)$$

If α is double and l_j and l_k fall in different single loops, then

$$[ul]\bigcup[wl]\overset{f}{\mapsto}[ulwl]-[uw^{-1}]\quad\text{if } l=l_j=l_k\ ,$$

$$[ul]\bigcup[wl^{-1}]\overset{f}{\mapsto}[ulw^{-1}l]-[uw]\quad\text{if } l=l_j=l_k^{-1}\ . \qquad (53)$$

The last two cases are not really independent since $\alpha\cup\beta=\alpha\cup\beta^{-1}$. As before, the result only depends on the relative orientation of l_j and l_k, and one can reduce the last four sums in (46) to a single summation over all pairs of links which are equal up to orientation, $|l_j|=|l_k|$.

Collecting everything one obtains

$$\hat{H}_K|\gamma\rangle=\tfrac{3}{4}\|\gamma\|\,|\gamma\rangle+\tfrac{1}{2}\sum_{j<k,\,|l_j|=|l_k|}|f(j,k,\gamma)\rangle\ . \qquad (54)$$

Figure 4 shows, as an explicit example involving the ma-

FIG. 4. An example for the action of the kinetic part of the Hamiltonian. Only if a loop has a doubled link are there more than the diagonal terms. The spinor identity from Fig. 2(d) is used.

nipulation of loops, how the action of $\hat{H}_K$ commutes with the spinor identity (v). This is a nontrivial check since the length $\|\gamma\|$ of γ changes under (v).

The conclusion of this section is that as claimed $\hat{H}$ maps loops to loops, and explicit rules for this action are available.

VI. THE EIGENPROBLEM FOR THE HAMILTONIAN

Of immediate physical interest are the lowest eigenvalues and corresponding eigenvectors of the Hamiltonian operator $\hat{H}$; that is, I am looking for solutions $|\psi\rangle$ and ϵ of

$$\hat{H}|\psi\rangle=\epsilon|\psi\rangle\ , \qquad (55)$$

where ϵ is close to the ground-state energy. There are several ways to approach this problem. The simple method of my choice is to approximate the Hilbert space of states by a finite-dimensional subspace spanned by loop states associated with *short* loops. After showing that this subspace contains a reasonable ansatz for the low-energy states of the model, I can proceed and use the inner product to formulate the eigenproblem in terms of vector components and matrix elements with respect to the finite loop basis. Equation (55) then becomes an equation of finite size vectors and matrices, and there are standard numerical techniques to solve a matrix eigenproblem.

Why are the short loops a reasonable ansatz for the ground state? Consider the strong-coupling limit $(g\to\infty)$ of Hamiltonian lattice gauge theory. In this limit the asymptotic ground-state wave function for a finite lattice is known to be

$$\psi_0(U)=\exp\left[\lambda\sum_{\square}\operatorname{tr} U_{\square}\right]\ , \qquad (56)$$

where $\lambda=\lambda(g)$ can be determined, for example, by minimizing $\langle\psi_0|\hat{H}|\psi_0\rangle/\langle\psi_0|\psi_0\rangle$ by varying λ. This variational method can be applied analytically to yield results in $2+1$ dimensions,[15] and numerical implementations exist which are applicable to both $2+1$ and $3+1$ dimensions.[16] Most variational methods that are used to solve (55) take ψ_0 (or a refined version) as a starting point. It is crucial that $\psi_0(U)$ also turns out to be a reasonable ansatz for the ground state in the physically interesting scaling region of the theory where roughly $g\approx1$.

To see the significance of (56) in the loop representation expand the exponential:

$$\psi_0(U)=1+\lambda\sum_{\square_1}\operatorname{tr} U_{\square_1}+\frac{\lambda^2}{2}\sum_{\square_1}\sum_{\square_2}\operatorname{tr} U_{\square_1}\operatorname{tr} U_{\square_2}+\cdots\ . \qquad (57)$$

Note that the nth term in this Taylor series consists of a sum over loop wave functions corresponding to loops of length $4n$. As defined, the series always converges since the lattice is finite—there is a finite number M of plaquettes $\square$ and $0\leq\operatorname{tr} U\leq2$ for U an SU(2) matrix. In designing a length related cutoff, however, I have to include loops up to a sufficiently large length $4n$ so that

$$\frac{\lambda^n}{n!}\left[\sum_\Box \mathrm{tr} U_\Box\right]^n \le \frac{\lambda^n}{n!}(2M)^n \ll 1 \ . \tag{58}$$

In actual variational calculations for SU(2) in $2+1$ dimensions one has $0 < \lambda < 1$ and $g \to \infty$ (see Ref. 16). Therefore, the smaller g, the larger the maximal length of loops should be.

The above should be motivation enough for the following ansatz. The loop basis is reduced to the finite basis consisting of all basis loops up to a certain length $L_{\max}$. The fact that the loop representation admits this simple approximation of the Hilbert space of states by short loops can be considered one additional important feature of the loop representation.

Remember that the length of a loop, defined as the number of its links, is no invariant under the loop identities (ii) and (v). When choosing a basis I will pick loops which are of minimal length, i.e., those which cannot be replaced by shorter loops. In what follows, I will consider a finite lattice with periodic boundary conditions; i.e., the lattice forms a torus. In general there will be "topological" loops which wrap around the lattice, but since I do not want to deal with topological effects, let me exclude topological loops even though they do not classify as long loops in the size of lattice I am going to study.

In effect, this approach will be equivalent to a variational ansatz with as many free parameters as there are loops in the basis, which is certainly an improvement over ψ_0 as in (56). In addition, I expect to find good approximations to excited states close to the ground state. In any case, after computation of any eigenvalues, I will have to check whether the components of the eigenvectors are largest on the short loops in order to justify the initial assumption. How this simple ansatz of including all short loops can be improved will be discussed later when there are specific eigenvectors at hand.

Now the stage is set for the formulation of the eigenvalue problem (55) in terms of vector components and matrix elements with respect to a finite loop basis $\{|\alpha\rangle, i=1,\ldots,N\}$. The use of a nonorthogonal basis as discussed in Sec. V does not lead to a standard eigenvalue problem. Instead, by substituting $|\psi\rangle = \sum_{j=1}^N \psi^j|\alpha_j\rangle \equiv \psi^j|\alpha_j\rangle$ into (55) and taking the inner product with $|\alpha_i\rangle$ one obtains

$$\langle \alpha_i|\hat{H}|\alpha_j\rangle \psi^j = \epsilon\langle \alpha_i|\alpha_j\rangle \psi^j \ . \tag{59}$$

Denoting the matrix elements $\langle \alpha_i|\hat{H}|\alpha_j\rangle$ of $\hat{H}$ by H_{ij}, one has

$$H_{ij}\psi^j = \epsilon G_{ij}\psi^j \ . \tag{60}$$

Hence, what I want to solve is a generalized eigenproblem with two symmetric real matrices, one of which is positive definite (G_{ij}). Under these conditions the general eigenproblem is not substantially harder to solve numerically than the standard one since there is a direct relation between the two. Notice, however, that a naive attempt like inverting G_{ij} and solving $H^i{}_j\psi^j = \epsilon\psi^i$ fails since $H^i{}_j \equiv (G^{-1})^{ik}H_{kj}$ is in general not a symmetric matrix, and the asymmetric eigenvalue problem is numerically quite unstable.

This concludes the theoretical formulation of the eigenproblem for the Hamiltonian. Of course, this particular application of the loop representation and the explicit use of the inner product were designed to produce a setup which was well suited for numerical computations. To what extent this was achieved is the topic of the remaining sections.

VII. COMPUTATIONS

So far I have said very little about the computational feasibility of the approach to the eigenvalue problem as presented in the preceding sections. As usual, if one finally wants to compute numbers with a computer, the setup of a problem is influenced by the hardware and software available. In this case the overall volume of computations is determined by the number of basis loops, which should be as large as possible for a good approximation of the Hilbert space. The specific model considered below is SU(2) Hamiltonian lattice gauge theory on a periodic 4×4 spatial lattice in a loop basis of loops up to length 12. In a sense this is the smallest nontrivial scenario if one wants to include loops consisting of more than two plaquettes which should not be crowded into too small a lattice. As explained later, the method does not depend significantly on the number of dimensions. Furthermore, for $a=1$, Eq. (15) gives the Hamiltonian for both $2+1$ and $3+1$ dimensions. It turns out that on a 4×4 lattice there are 8660 linearly independent loops up to length 12, and this translates into a general eigenvalue problem of sparse 8660×8660 matrices (sparse meaning that most entires are zero). A problem of this size can be comfortably managed on a Vax 8810 (allocated fast memory 8 MB, 10 MB on hard disk).

I will briefly discuss the important parts of the program to highlight its merits and weaknesses. But first I present an overview of what has to be done and in what order: (1) Find all loops up to a given length, (2) select a basis via the inner product, (3) compute G_{ij}, (4) compute $(H_P)_{ij}$ and $(H_K)_{ij}$, and finally (5), for $a=1$ and different values of g feed G_{ij} and $H_{ij} = (g^2/2)(H_K)_{ij} + (1/2g^2)(H_P)_{ij}$ to a program-library subroutine for sparse eigenproblems.

A. Find all loops

Clearly, this is a combinatorical problem. The difficulty is that one has to expect the number of loops to grow exponentially with length. In what ways can I select links in a lattice such that they form closed lines? To begin with, let me construct all single loops emanating from a given point (in three dimensions for generality). Denote by 1,2,3 a step in the positive x,y,z direction and by $-1,-2,-3$ a step in the opposite direction. A loop with a given origin can be represented as an (ordered) list of directions such as $(1,2,-1,-2)$ for a plaquette in the x-y plane. Let n_i be the number of steps in direction $i=\pm1,\pm2,\pm3$. Then the condition for a closed loop of length L is

$$\sum_i n_i = L \quad \text{and} \quad n_i = n_{-i} \quad \text{for } i=1,2,3 \ . \tag{61}$$

(Remember that in Sec. VI I have excluded "topological" loops for which $n_i = n_{-i} + \text{const.}$) Therefore finding all loops of length L is equivalent to finding all different permutations of generic lists:

$$(1, \ldots, 1, -1, \ldots, -1, 2, \ldots, -2, 3, \ldots, -3) \ . \quad (62)$$

At this point one can easily incorporate rules (ii), (iii), and (iv); i.e., pick one representative for each equivalence class. Rule (ii) can be taken to forbid a step in direction i followed by one in direction $-i$, and so on. Only rule (v) will be left for the inner product to handle. Actually, it is not hard to find formulas for the number of loops $n(L)$, but the summation over generic lists cannot be carried out explicitly. In Fig. 5(a) some values for $n(L)$ are tabulated.

Having found all single loops through a given point, I now have to specify general multiloops at all possible positions in the lattice. One can enumerate the links of the lattice arbitrarily by $\pm 1, \pm 2, \ldots, \pm M$ and represent a single loop by a list of links: Starting from each point of the lattice I transcribe a list of directions representing a single loop into a list of numbers encoding particular links which form that single loop. A multiloop is formed as a selection of several single loops. In the end a general loop is represented by a list of lists of numbers. I have given all these details to make the following point. Once all the loops of a particular model are encoded in lists of numbers, the program is independent of the geometric aspect of the loops. For example, there is no difference in the treatment of a set of lists of numbers representing loops in three dimensions or two dimensions. The number of dimensions is hidden in the statistics of the loops.

B. Inner product

As elaborated on in Sec. IV, the task of integration can be solved algebraically. It might be that there is an even simpler, faster numerical alternative—a possibility I have not explored. Instead all inner products are handled by a specialized routine for algebraic manipulations. I should mention that the number of terms which have to be evaluated can become quite large. The integration over six identical plaquettes (as an example for the inner product of a length 12 loop with itself) requires the contraction of 50 000 terms. Luckily, most nonzero inner products will involve lower multiplicity of links and therefore far fewer terms. And by far the most inner products between random loops will be trivially zero.

C. Metric G_{ij}

First, the degenerate metric $\tilde{G}_{ij} = \langle \tilde{\alpha}_i | \tilde{\alpha}_j \rangle$ with $\tilde{\alpha}_i$ as found in Sec. VII A is computed. As expected, $\tilde{G}_{ij}$ is "almost" diagonal. The next step is to find a loop basis $\{ | \alpha_i \rangle \}$ and the nondegenerate metric $G_{ij} = \langle \alpha_i | \alpha_j \rangle$.
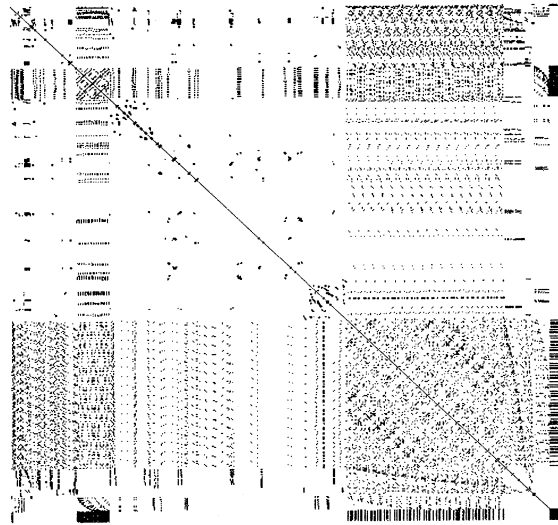
a)

length L	number	of loops n
	$2+1$	$3+1$
4	1	3
6	2	22
8	14	288
10	76	4,428
12	505	72,963
14	3,386	1,276,218
16	23,823	

b)

loop	occurences
0	1
4	16
6	32
8	148
4+4	120
10	710
4+6	512
12	3,745
4+8	2,477
6+6	479
4+4+4	420
total	8,660

FIG. 5. (a) Number of single loops n of length L through a given point in $2+1$ and $3+1$ dimensions. Loops are not distinguished by their origin or orientation, and are not allowed to have spikes [rules (iii), (iv), and (ii), respectively]. (b) Basis loops on a periodic 4×4 lattice. Type "4+4" denotes a loop consisting of two length 4 single loops, etc.

FIG. 6. The 8660×8660 metric $G_{ij} = \langle \alpha_i | \alpha_j \rangle$. Each nonzero entry is represented by a dot, but each dot covers roughly 20×20 entires. The rows and columns are ordered by length and type as in Fig. 5(b). Since the inner product is nondegenerate there are only nonvanishing diagonal entires. The banding corresponds directly to the loop types and shows how two loops can combine to a loop with doubled links resulting in a nonzero inner product.

Later the Cholesky decomposition of G_{ij} will be needed:

$$G_{ij} = L_{ik} L_{kj}^T \ ,$$

where L is a lower triangular matrix .   (63)

Crout's algorithm can be used to find the Cholesky decomposition of a nondegenerate symmetric matrix. I modified the version of Crout's algorithm given in Ref. 17 so that when it is applied to $\tilde{G}_{ij}$ it eliminates degeneracies in $\tilde{G}_{ij}$ by deleting loops from $\{\tilde{\alpha}_i\}$, thereby constructing $\{\alpha_i\}$, G_{ij}, and its decomposition at the same time. By far the most time consuming part of the program (several hours) is the computation of $\tilde{G}_{ij}$, since even a trivial check for doubled links consumes a lot of time when performed $\approx 10^8$ times. For the 8660 loops up to length 12 on a 4×4 lattice only $\approx 40\,000$ of the $8660^2 \approx 75\,000\,000$ entries of G_{ij} are nonzero; i.e., G_{ij} is sparse. See Fig. 5(b) for a table of the loops which form the basis. Figure 6 shows a graphical representation of G_{ij}.

D. $(H_P)_{ij}$ and $(H_K)_{ij}$

These computations are straightforward in the sense that Sec. V gives the explicit rules to find the loops $\hat{H}_P |\alpha_i\rangle$ and $\hat{H}_K |\alpha_i\rangle$. Of course, these loops will in general not be in the basis $\{|\alpha_i\rangle\}$, and I have to compute the matrix elements with the inner product. $(H_K)_{ij}$ is in structure very similar to G_{ij}, while the extra plaquette in $(H_P)_{ij}$ leads to $\approx 420\,000$ nonzero entries. Obviously, some sort of efficient storage scheme for large sparse matrices has to be used.

E. Solve eigenvalue problem

The general eigenvalue problem $H\psi = \epsilon G\psi$ is solved by decomposing G into its Cholesky factors as in (63) and solving

$$(L^{-1} H (L^T)^{-1})(L^T \psi) = \epsilon (L^T \psi) \ ,   (64)$$

which represents a standard symmetric eigensystem $Ax = \epsilon x$. Since solving an eigensystem is fairly complicated, usually the best thing to do is to use "canned" eigenroutines as found in several FORTRAN libraries. But even the best algorithms available for symmetric $n \times n$ matrices (e.g., the QR factorization method) carry an operation count of order n^3, which forbids $n > 1000$. Exploiting sparsity is hardly possible since these methods are based on transformations of the sparse matrix A which result in "fill-ins" that quickly destroy sparsity. A solution to this problem are the Lanczos algorithms if only a few extreme eigenvalues and eigenvectors are needed (see Ref. 18). These algorithms are sophisticated iteration methods involving the sparse matrix A only for multiplication of a vector as in Ax. The inherent numerical instability of Lanczos algorithms is that they require the eigenvectors which are computed to be mutually orthogonal. In practice, one often loses the orthogonality of the eigenvectors which in turn limits the number of eigenpairs that can be computed. I used the eigenroutines of Scott (1982, Ref. 19). The time required to find the five lowest eigenvalues and eigenvectors of the

8660×8660 eigensystem $H\psi = \epsilon G\psi$ amounts to 7 min.

As explained above, once one has arrived at an eigenvalue problem involving a large sparse matrix, the Lanczos algorithms are the method of choice. The fact that these large eigenvalue problems can actually be solved was utilized in the context of a Hamiltonian formulation of lattice models first in Ref. 20.

Let me summarize. The computer program to set up the eigenvalue problem is mostly algebraic and self-made. The canned Lanczos eigenroutines implement numerical algorithms which run sufficiently fast, but because of numerical instabilities are still subject to ongoing development. The limitations of the program arise in the size allowed for the matrices and the time required to compute them. Figure 5(a) indicates how rapidly the number of loops (and therefore the size of the problem) increases for longer loops. But as the next section will show, the numerical results suggest ways to drastically reduce the number of basis loops needed for a good approximation of the Hilbert space.

VIII. RESULTS

In what follows the model is SU(2) Hamiltonian lattice gauge theory on a periodic 4×4 lattice. By using a finite loop basis consisting of all loops up to length 12, I was able to compute some of the low-lying eigenvalues and eigenvectors of the Hamiltonian. That is, I partially solved the generalized eigenproblem (60) for 8660×8660 matrices H_{ij} and G_{ij}.

When discussing the results I want to make a clear distinction between the lattice-gauge-theory model, which is a statistical model in its own right, and statements about the continuum theory, which are obtained using a highly nontrivial limiting procedure. For the finite statistical model I can present definite results which agree with the literature. To make statements about the physics of the continuum, say, about the mass gap, one would like to detect in the computed data what is called scaling toward the continuum limit. This scaling behavior I do not find, but I will explain below why one cannot expect to find scaling given the present status of the program. In this context of a Hamiltonian formulation scaling was found so far only in certain nonlinear σ models, although the work of Gambini, Leal, and Trias[7] looks promising.

Let me quickly review the features of Hamiltonian lattice gauge theory which are relevant for the following discussion. Remember that in general one has a whole family of Hamiltonians $H_{ij} \equiv H_{ij}(a,g)$ (considering the number of lattice sites fixed). Typically one computes the mass gap $M(a,g)$ for several values of the coupling constant g at fixed lattice spacing $a = 1$. There are two interesting limiting situations. First, there is the strong-coupling regime where $g \rightarrow \infty$ while $a = $const. This is where my assumptions leading to the truncation of the loop basis originate. Second, there is the weak-coupling region where $g \rightarrow 0$. Now I have to distinguish between two cases. If a is kept constant and $a \neq 0$, I obtain valid results for the statistical model. On the other hand, $a \rightarrow 0$ is associated with the continuum limit of the theory, about which I want to make a few remarks.

In 3+1 dimensions, g is dimensionless, and a is the only quantity having the dimension of length ($c = \hbar = 1$). Hence

$$M(a,g) = \frac{1}{a} f(g) \tag{65}$$

for some dimensionless function f. One consistent way to give physical content to the lattice theory is to require that M be physical and therefore be *independent* of a and g. Equation (65) implies for $M = \text{const}$ and $a \rightarrow 0$ that $f(g) \rightarrow 0$. That is, as $a \rightarrow 0$ one has to take the (bare) coupling constant g to a critical value of f. For SU(N) lattice gauge theories, $g_{\text{critical}} = 0$. This is where perturbation theory becomes valid, and renormalization-group arguments determine the exact form of f. Scaling towards the continuum limit is observed when one can fit $f(g)$ to the data computed for $M = (1/a) f(g)$.

The problem is that in order to make the theory well defined, one uses a finite lattice, and therefore the lattice spacing must have some finite value which renders the lattice large enough to contain the physics one is looking for. When looking for scaling one can only hope that it is possible to detect scaling for small but still finite values of a.

Let me now discuss the computational data I obtained for the statistical model. Figure 7 shows for $a = 1$ the energy $\epsilon_0(g^2)$ of the ground state, the energy $\epsilon_1(g^2)$ of the first excited state, and their difference which is the mass gap $M(g^2)$. Figure 8 shows a rescaled curve of M. The

numerical errors in the eigenvalues are less than 10^{-6}. What does one expect for strong and weak coupling, and what in the transition region? The Hamiltonian H_{ij} is a linear combination of matrices:

$$H_{ij} = \frac{g^2}{2} H_{Kij} + \frac{2}{g^2} H_{Pij} . \tag{66}$$

Assuming that H_K and H_P are nondegenerate and nonsingular, the coefficients g^2 and $1/g^2$ determine that for $g \rightarrow \infty$ the Hamiltonian is purely kinetic and has eigenvalues proportional to g^2, while for $g \rightarrow 0$ the Hamiltonian consists of the potential term and has eigenvalues proportional to $1/g^2$. Therefore, in Fig. 8, $M(g^2)$ is a hyperbola for small g and a straight line for large g. The transition from strong to weak coupling occurs at around $g^2 = 2$, and I expect to find a change in the structure of the ground state at this point.

The computed form of $M(g^2)$ is of the generic type one would find for a large class of symmetric matrices H_K and H_P. The characteristic signature of the model, i.e., scaling, should show in the region around $g^2 = 2$, but I could not detect any deviation from the generic dependence $M \sim g^2 + 1/g^2$. There is, however, an intuitive reason why one does not expect to find physics under these circumstances. Mass is measured in inverse lattice spacings, and in this case $M \geq 3.7$. The physical length scale under consideration is $1/M < 0.3$, which is much smaller than $a = 1$. The conclusion is that in order to see
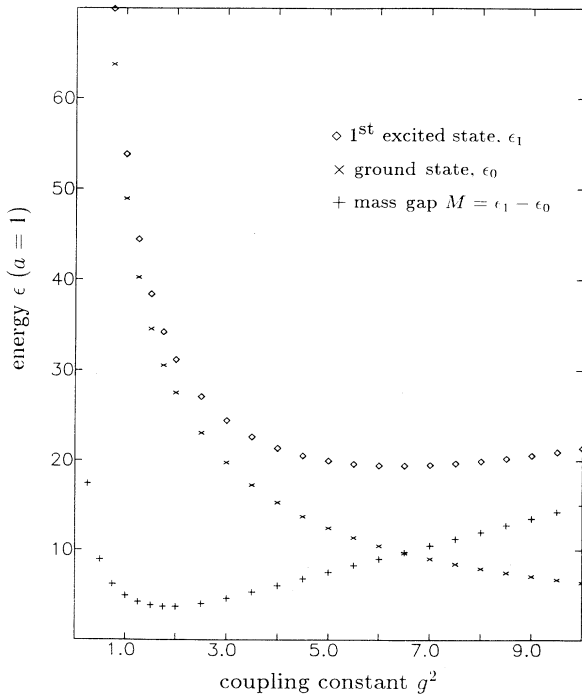

FIG. 7. The mass gap as the difference of the energies of the first excited state and the ground state.
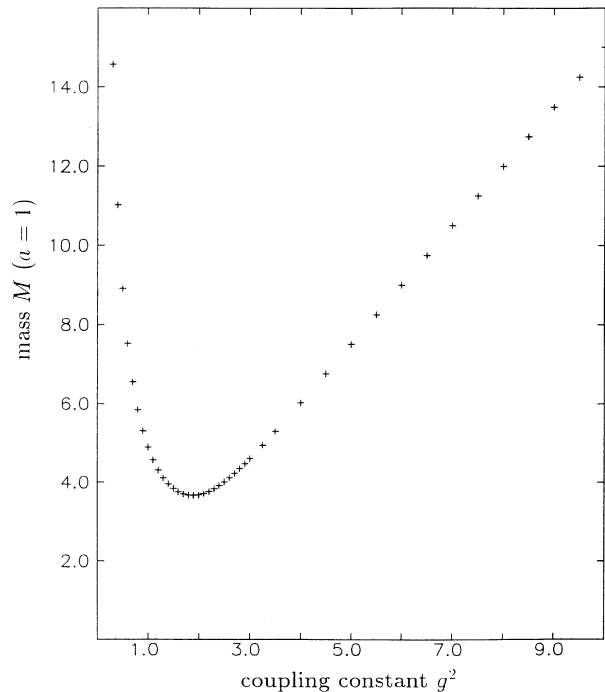

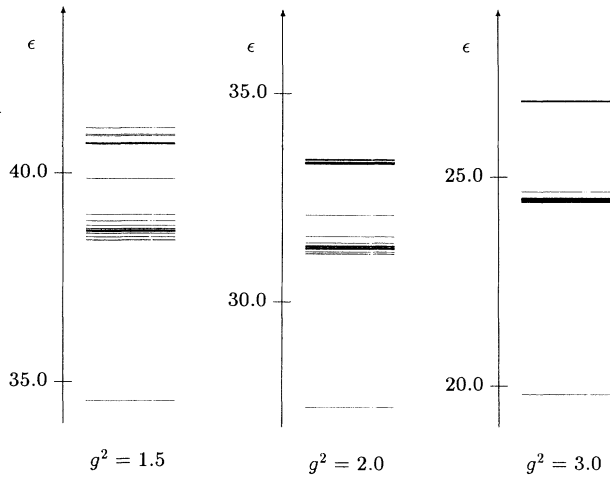FIG. 8. The mass gap or glueball mass vs the square of the coupling constant.

FIG. 9. The energy spectrum for three values of the coupling constant.

smaller values of M.

First I want to compare my results with a standard variational method for SU(2) Hamiltonian lattice gauge theory. In $2+1$ dimensions with the ansatz (56) for the ground state, Arisu, Kato, and Fujiwaru[15] have performed the minimization of the ground-state energy analytically. Based on Ref. 15 is the numerical work of Coldwell, Huang, and Katoot.[16] In both cases I find very good quantitative agreement in M for strong coupling. For example at $g^2 = 2.5$, Refs. 15 and 16 show $M = 3.95$ while I find $M = 4.01$, and this agreement holds also for larger g^2. For weak coupling the results still coincide qualitatively in that M diverges as $g^2 \rightarrow 0$, but the actual degree of divergence is different. This is expected since the variational method and the loop method no longer coincide in their approximation of the ground state.

The only other loop-based method for computations by Gambini, Leal, and Trias[7] gives similar results for the ground-state energy and the mass gap for strong and weak coupling, but their data seem to contain at least the hint that scaling occurs.

One advantage of formulating the eigenvalue problem in terms of finite matrices is that several of the lowest eigenvalues are obtained easily—as opposed to other methods which are hard pressed to produce more than the two lowest eigenpairs. To my knowledge this was not achieved before for SU(2) Hamiltonian lattice gauge

scaling, one requires $0.25 < M < 1$ so that the physical quantity neither falls through the cracks of the lattice nor is too large for the finite 4×4 lattice. It is likely that the inclusion of longer loops on larger lattices will lead to
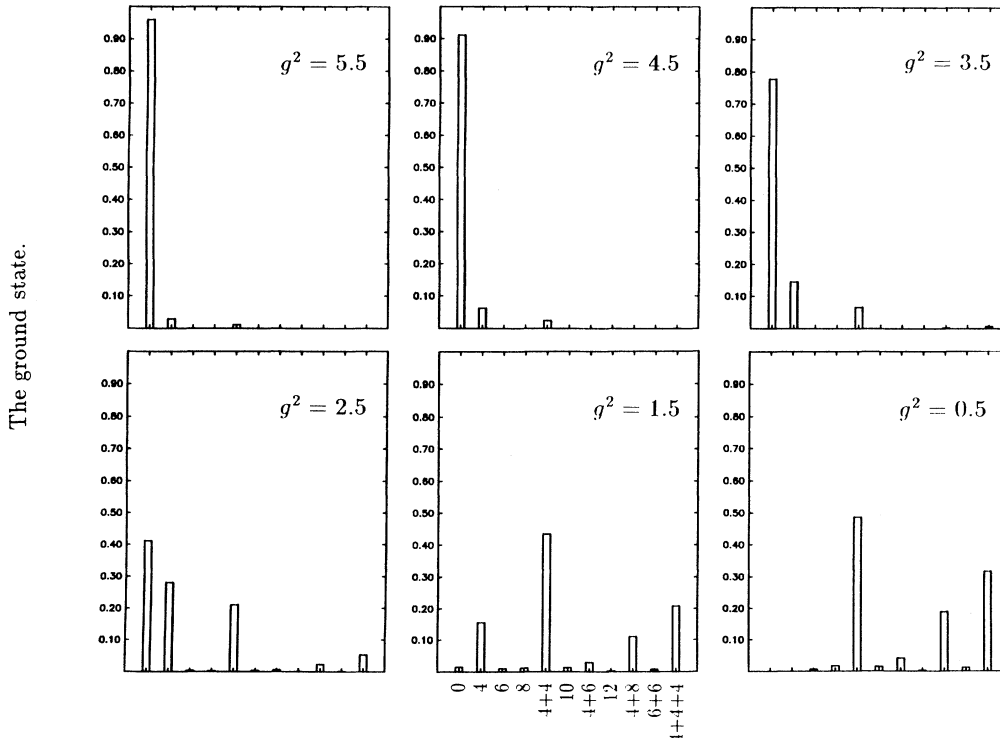

FIG. 10. The ground state. The bars measure the combined contribution of all loops of a given type to the norm of the ground state, which is 1.

theory. Figure 9 shows the energy spectrum consisting of eigenvalues from ϵ_0 up to ϵ_{34}, beyond which numerical instabilities were encountered. The diagram shows a nondegenerate ground state which is clearly separated by the mass gap from higher energies. These occur in groups of 16, which probably corresponds to the 16 sites in the lattice. The numerical accuracy of the Lanczos algorithm suggests that at low energies there are not more than twofold degeneracies; i.e., these states are invariant under all or half of the 16 translations. The stronger the coupling is, the closer together are the energy levels in a group of 16. For $g^2 = 1.5$, only the first and the beginning of the second group of 16 were obtainable. For $g^2 = 2.0$ and $g^2 = 3.0$, the first, second, and part of the third group are shown. Even though the difference between ϵ_{33} and ϵ_{34} is barely shown in the diagram, the gap is numerically distinct, and one can assume that the next higher levels all crowd together on the highest level shown.

Together with the eigenvalues the corresponding eigenvectors are computed. Apart from the restriction to short loops, I have an unbiased ansatz and can now analyze which loops are important for the ground state and the first excited state. Notice that since the loop states are not orthogonal. I cannot compare just the components. Instead, as a measure for the contribution to the squared norm $\langle \psi | \psi \rangle = \sum_{j,k=1}^{n} G_{jk} \psi^j \psi^k$ from the ith

component, I consider $\psi^i \sum_{j=1}^{n} G_{ij} \psi^j$ (no sum over i). Figures 10 and 11 show for the ground state and the first excited state the sum of the contributions from all loops of a certain type for different values of the coupling constant. The length of the bars indicates on which loops a state is peaked.

The first general observation is that for $g^2 > 2$ the ground state is concentrated on the short loops, while for $g^2 < 2$ the components of the ground state are pushed to longer loops. As an example, consider how the contribution of the zero-length loop changes with g^2. From dominating the ground state for $g^2 = 5.5$, its contribution decreases until it is zero for $g^2 = 0.5$. On the other hand, the contribution of all 16 plaquettes is small for strong coupling, increases to reach a maximum at around $g^2 = 2.5$, and then falls to zero at weak coupling. If this feature persisted for models including longer loops—i.e., that a state is peaked on loops of a certain length and is not spread out over loops of up to infinite length—then the idea to truncate the loop basis would work even for weak coupling.

Second, multiloops are more important than single loops. Even though single loops of length 12 constitute more than $\frac{2}{5}$ of all loops [Fig. 5(b)] their collective contribution never grows over 0.01. Obviously one could fine-tune the ansatz of short loops by including only certain short loops, thereby making the inclusion of loops up to a
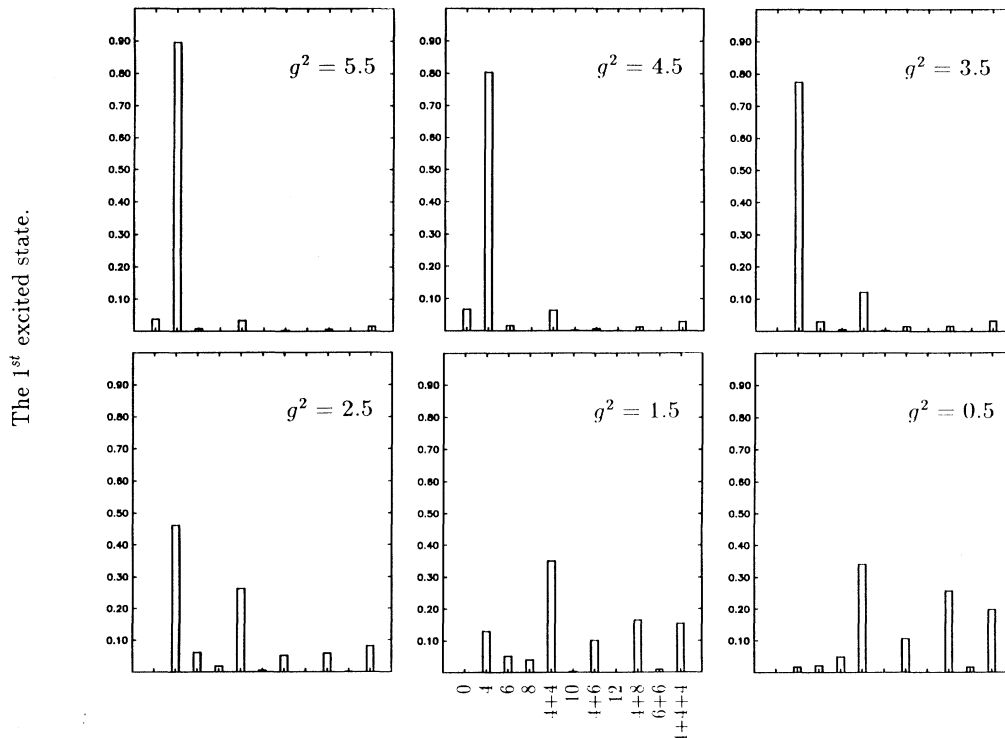

FIG. 11. The first excited state.

larger maximal length feasible.

The above results fit nicely with the original idea behind the Lanczos algorithms.[18] Given a Hilbert space and a Hamiltonian operator thereon whose eigenvalues are to be computed, a first step to simplify the problem is to construct a special basis for the Hilbert space. Basically, each state is approximated by a linear combination of states $|\phi_n\rangle$, $n = 1, \ldots N$, which are obtained from some starting vector $|\phi_0\rangle$ by repeated action of the Hamiltonian, $|\phi_n\rangle = \hat{H}^n|\phi_0\rangle$. Let $|\phi_0\rangle$ be the zeroth-length loop. The potential term of $\hat{H}$ adds *one* plaquette, while the kinetic term can join *two* single loops if they have a common link. Hence single loops of a certain length always belong to a higher order of approximation than multiloops of the same length.

All the statements about the components of the ground state are true in a similar way for the first excited state or "glueball." The main difference is that, for strong coupling, its largest components are distributed among the plaquettes, which are orthogonal to the zero-length loop.

IX. CONCLUSION

The purpose of this work was to test the viability of the method of loops in a simple example of gauge theory. The results are encouraging, and perhaps one can also say that they help to build confidence in loop-based methods for quantum gravity.

The particular approach, which I chose for its simplicity, was to formulate the eigenvalue problem for the Hamiltonian of lattice gauge theory in terms of vector components and matrix elements with respect to the loop basis. This was made possible by the following observations. The physical Hilbert space can be approximated by a finite basis of short loops, and the Hamiltonian maps

a loop to a simple, finite linear combination of loops. Explicit use was made of the fact that the inner product between loop states can be computed analytically. This greatly simplified the task of finding a basis of linearly independent loop states, and enabled me to write down the matrix eigenvalue problem. The nice feature of this formulation is that it constitutes not only a method that "works in principle," but one which seems to be an efficient new approach to Hamiltonian lattice gauge theory. The values obtained for the ground-state energy and the mass gap agree with the literature. In addition, some of the low-lying eigenvalues and eigenvectors could be obtained.

The obvious extensions are to use larger lattices and longer loops to find scaling. The only problem involved in a transition to $3 + 1$ dimensions lies in the larger number of loops, but new technical issues do not arise—in contrast with the case of standard variational methods. Speaking of variational methods, the availability of the inner product for general computations should enable one to make quite a sophisticated ansatz for the ground state.

The next big question is how one can incorporate fermions into any computational method based on loops. As of now, I cannot offer any analysis of the complexity of the problems involved. Notice, however, that fermions on the lattice are already built into the loop representation of Rovelli and Smolin.[9]

[1]S. Mandelstam, Ann. Phys. (N.Y.) **19**, 1 (1962).
[2]J. Kogut and L. Susskind, Phys. Rev. D **11**, 395 (1975).
[3]A. M. Polyakov, Nucl. Phys. **B164**, 171 (1979).
[4]Yu. M. Makeenko and A. A. Migdal, Phys. Lett. **88B**, 135 (1979).
[5]G. 't Hooft, Nucl. Phys. **B153**, 141 (1979).
[6]R. Gambini and A. Trias, Phys. Rev. D **23**, 553 (1981); Nucl. Phys. **B278**, 436 (1986).
[7]R. Gambini, L. Leal, and A. Trias, Phys. Rev. D **39**, 3127 (1989).
[8]W. Furmanski and A. Kolawa, Nucl. Phys. **B291**, 594 (1987).
[9]C. Rovelli and L. Smolin, Syracuse University report (unpublished).
[10]R. Giles, Phys. Rev. D **24**, 2160 (1981).
[11]C. Rovelli and L. Smolin, Nucl. Phys. **B331**, 80 (1990).
[12]M. Creutz, J. Math. Phys. **19**, 2043 (1978).

[13]M. Creutz, *Quarks, Gluons and Lattices* (Cambridge University Press, Cambridge, England, 1983).
[14]K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
[15]H. Arisu, M. Kato, and T. Fujiwaru, Prog. Theor. Phys. **70**, 229 (1983).
[16]M. C. Huang, R. L. Coldwell, and M. W. Katoot, Nucl. Phys. **B309**, 733 (1988).
[17]W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1989).
[18]B. Parlett, *The Symmetric Eigenvalue Problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
[19]D. Scott, LASO2 (University of Texas, Austin, 1982).
[20]H. H. Roomany, H. W. Wyld, and L. E. Holloway, Phys. Rev. D **21**, 1557 (1980).