Commun. Math. Phys. 121, 351–399 (1989)

Communications in
Mathematical
Physics
© Springer-Verlag 1989

Quantum Field Theory and the Jones Polynomial *

Edward Witten **

School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, NJ 08540, USA

Abstract. It is shown that $2+1$ dimensional quantum Yang-Mills theory, with an action consisting purely of the Chern-Simons term, is exactly soluble and gives a natural framework for understanding the Jones polynomial of knot theory in three dimensional terms. In this version, the Jones polynomial can be generalized from S^3 to arbitrary three manifolds, giving invariants of three manifolds that are computable from a surgery presentation. These results shed a surprising new light on conformal field theory in $1+1$ dimensions.

In a lecture at the Hermann Weyl Symposium last year [1], Michael Atiyah proposed two problems for quantum field theorists. The first problem was to give a physical interpretation to Donaldson theory. The second problem was to find an intrinsically three dimensional definition of the Jones polynomial of knot theory. These two problems might roughly be described as follows.

Donaldson theory is a key to understanding geometry in four dimensions. Four is the physical dimension at least macroscopically, so one may take a slight liberty and say that Donaldson theory is a key to understanding the geometry of space-time. Geometers have long known that (via de Rham theory) the self-dual and anti-self-dual Maxwell equations are related to natural topological invariants of a four manifold, namely the second homology group and its intersection form. For a simply connected four manifold, these are essentially the only classical invariants, but they leave many basic questions out of reach. Donaldson's great insight [2] was to realize that moduli spaces of solutions of the self-dual Yang-Mills equations can be powerful tools for addressing these questions.

Donaldson theory has always been an intrinsically four dimensional theory, and it has always been clear that it was connected with mathematical physics at least at the level of classical nonlinear equations. The puzzle about Donaldson theory was whether this theory was tied to more central ideas in physics, whether it could be interpreted in terms of quantum field theory. The most important

* An expanded version of a lecture at the IAMP Congress, Swansea, July, 1988

** Research supported in part by NSF Grant No. 86-20266, and NSF Waterman Grant 88-17521

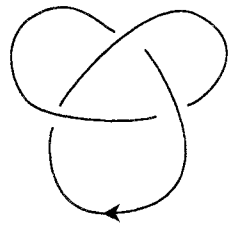

Fig. 1. A knot in three dimensional space

evidence for the existence of such a connection had to do with Floer's work on three manifolds [3] and the nature of the relation between Donaldson theory and Floer theory. Also, the "Donaldson polynomials" had an interesting formal analogy with quantum field theory correlation functions. It has turned out that Donaldson theory can indeed be given a physical interpretation [4].

As for the Jones polynomial and its generalizations [5–11], these deal with the mysteries of knots in three dimensional space (Fig. 1). The puzzle on the mathematical side was that these objects are invariants of a three dimensional situation, but one did not have an intrinsically three dimensional definition. There were many elegant definitions of the knot polynomials, but they all involved looking in some way at a two dimensional projection or slicing of the knot, giving a two dimensional algorithm for computation, and proving that the result is independent of the chosen projection. This is analogous to studying a physical theory that is in fact relativistic but in which one does not know of a manifestly relativistic formulation – like quantum electrodynamics in the 1930's.

On the physical side, the puzzle about the knot polynomials was the following. Unlike the Donaldson theory, where a connection with quantum field theory was not obvious, the knot polynomials have been intimately connected almost from the beginning with two dimensional many body physics. In fact, constructions of the knot polynomials have related them to two dimensional (or $1+1$ dimensional) many-body physics in a bewildering variety of ways, mainly involving soluble lattice models [7], solutions of the Yang-Baxter equation [8], and monodromies of conformal field theory [11]. In the latter interpretation, the knot polynomials are related to aspects of conformal field theory that have been particularly fruitful recently [12–16]. On the statistical mechanical side, studies of the knot polynomials have related them to Temperley-Lieb algebras and their generalizations, and to other aspects of soluble statistical mechanics models in $1+1$ dimensions. For physicists the challenge of the knot polynomials has been to bring order to this diversity, find the unifying themes, and learn what it is that is three dimensional about two dimensional conformal field theory.

Now, the Donaldson and Jones (and Floer and Gromov [17]) theories deal with topological invariants, and understanding these theories as quantum field theories involves constructing theories in which all of the observables are topological invariants. Some physicists might consider this to be a little bit strange, so let us pause to explain the physical meaning of "topological invariance". The physical meaning is really "general covariance". Something that can be computed from a manifold M as a topological space (perhaps with a

smooth structure) without a choice of metric is called a "topological invariant" (or a "smooth invariant") by mathematicians. To a physicist, a quantum field theory defined on a manifold M without any a priori choice of a metric on M is said to be generally covariant. Obviously, any quantity computed in a generally covariant quantum field theory will be a topological invariant. Conversely, a quantum field theory in which all observables are topological invariants can naturally be seen as a generally covariant quantum field theory. Indeed, the Donaldson, Floer, Jones, and Gromov theories can be seen as generally covariant quantum field theories in four, three, and two space-time dimensions. The surprise, for physicists, perhaps comes in how general covariance is achieved. General relativity gives us a prototype for how to construct a quantum field theory with no a priori choice of metric - we introduce a metric, and then integrate over all metrics. This example is so influential in our thinking that we tend to think of a generally covariant theory as being, by definition, a theory in which the metric is a dynamical variable. The lesson from the Donaldson, Floer, Jones, and Gromov theories is precisely that there are highly non-trivial quantum field theories in which general covariance is realized in other ways. In particular, in this paper we will describe an exactly soluble generally covariant quantum field theory in which general covariance is achieved not by integrating over metrics but because we begin with a gauge invariant Lagrangian that does not contain a metric.

1. The Chern-Simons Action

We have been urged [1] to try to interpret the Jones polynomial in terms of three dimensional Yang-Mills theory. So we begin on an oriented three manifold M with a compact simple gauge group G. We pick a G bundle E, which may as well be trivial, and on E we place a connection A_i^a, which can be viewed as a Lie algebra valued one form (a runs over a basis of the Lie algebra, and i is tangent to M). An infinitesimal gauge transformation is

$$A_i \to A_i - D_i \varepsilon\,, \tag{1.1}$$

where ε, a generator of the gauge group, is a Lie algebra valued zero form and the covariant derivative is $D_i\varepsilon = \partial_i \varepsilon + [A_i, \varepsilon]$. The curvature is the Lie algebra valued two form $F_{ij} = [D_i, D_j] = \partial_i A_j - \partial_j A_i + [A_i, A_j]$. Now we need to choose a Lagrangian. We will *not* pick the standard Yang-Mills action [1]

$$\mathscr{L}_0 = \int_M \sqrt{g}\, g^{ik} g^{jl} \operatorname{Tr}(F_{ij} F_{kl})\,, \tag{1.2}$$

as this depends on the choice of a metric g_{ij}. We want to formulate a generally covariant theory (in which all observables will be topological invariants), and to this aim we want to pick a Lagrangian which does not require any choice of metric.

[1] In what follows, the symbol "Tr" denotes an invariant bilinear form on the Lie algebra of G, a multiple of the Cartan-Killing form; we will specify the normalization presently

Precisely in three dimensions there is a reasonable choice, namely the integral of the Chern-Simons three form:

$$\begin{aligned}\mathscr{L} &= \frac{k}{4\pi} \int_M \mathrm{Tr}(A \wedge dA + \tfrac{2}{3} A \wedge A \wedge A) \\ &= \frac{k}{8\pi} \int_M \varepsilon^{ijk} \,\mathrm{Tr}(A_i(\partial_j A_k - \partial_k A_j) + \tfrac{2}{3} A_i [A_j, A_k]) . \end{aligned} \tag{1.3}$$

The Chern-Simons term in three dimensional gauge theory has a relatively long history. The abelian gauge theory with only a Chern-Simons term was studied by Schwarz [18] and in unpublished work by I. Singer. Three dimensional gauge theories with the Chern-Simons term added to the usual action (1.2) were introduced in [19–21]. The nonabelian theory with only Chern-Simons action was studied classically by Zuckerman [22]. The abelian Chern-Simons theory has recently been studied in relation to fractional statistics by Hagen [24] and by Arovas et al. [25] and in relation to linking numbers by Polyakov [23] and Fröhlich [15]. The novelty in our present discussion is that we will consider the quantum field theory defined by the nonabelian Chern-Simons action and argue that it is exactly soluble and has important implications for three dimensional geometry and two dimensional conformal field theory.

The first fundamental property of the Chern-Simons theory is the quantization law first discussed in [21]. It arises because the group $\hat{G}$ of continuous maps $M \to G$ is not connected. In the homotopy classification of such maps one meets at least the fact that $\pi_3(G) \simeq \mathbf{Z}$ for every compact simple group G. Though (1.3) is invariant under the component of the gauge group that contains the identity, it is not invariant under gauge transformations of non-zero "winding number", gauge transformations associated with non-zero elements of $\pi_3(G)$. Under a gauge transformation of winding number m, the transformation law of (1.3) is

$$\mathscr{L} \to \mathscr{L} + \mathrm{const} \cdot m . \tag{1.4}$$

As in Dirac's famous work on magnetic monopoles, consistency of quantum field theory does not quite require the single-valuedness of $\mathscr{L}$, but only of $\exp(i\mathscr{L})$. For this purpose, it is necessary and sufficient that the "constant" in (1.4) should be an integral multiple of 2π. This gives a quantization condition on the parameter called k in (1.3). If G is $SU(N)$ and "Tr" means a trace in the N dimensional representation, then the requirement is that k should be an integer. In general, for any G, we can uniquely fix the so far unspecified normalization of "Tr" so that the quantization condition is $k \in \mathbf{Z}$.

We will see later that k is very closely related to the central charge in the theory of highest weight representations of affine Lie algebras. It is no accident that the reasoning which shows that k must be quantized in (1.3) has a $1+1$ dimensional analogue [26] which leads to quantization of the central charge in the representation theory of affine algebras.

In quantum field theory, in addition to a Lagrangian, one also wishes to pick a suitable class of gauge invariant observables. In the present context, the usual gauge invariant local operators would not be appropriate, as they spoil general covariance. However, the "Wilson lines" so familiar in QCD give a natural class of

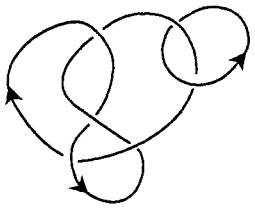

Fig. 2. Several linked but non-intersecting oriented knots in a three manifold M. Such a collection of knots is called a "link"

gauge invariant observables that do not require a choice of metric. Let C be an oriented closed curve in M. Intrinsically C is simply a circle, but the topological classification of embeddings of a circle in M is very complicated, as we observe in Fig. 1. Let R be an irreducible representation of G. One then defines the "Wilson line" $W_R(C)$ to be the following functional of the connection A_i. One computes the holonomy of A_i around C, getting an element of G that is well-defined up to conjugacy, and then one takes the trace of this element in the representation R. Thus, the definition is

$$W_R(C) = \mathrm{Tr}_R\, P \exp \int_C A_i\, dx^i\,. \tag{1.5}$$

The crucial property of this definition is that there is no need to introduce a metric, so general covariance is maintained.

We now can formulate the general problem of interest. In an oriented three manifold M, we take r oriented and non-intersecting knots C_i, $i = 1 \ldots r$, whose union is what knot theorists would call a "link" L. We assign a representation R_i to each C_i, and we propose to calculate the Feynman path integral

$$\int D\mathscr{A} \exp(i\mathscr{L}) \prod_{i=1}^{r} W_{R_i}(C_i)\,. \tag{1.6}$$

The symbol $D\mathscr{A}$ represents Feynman's integral over all gauge orbits, that is, an integral over all equivalence classes of connections modulo gauge transformations. Of course, (1.6) has exactly the formal structure of some familiar observables in QCD, the difference being that we are in three dimensions instead of four and we have chosen a somewhat exotic gauge theory action. We will call (1.6) the "partition function" of M with the given link, or the (unnormalized) "expectation value" of the given link; we will denote it as $Z(M; C_i, R_i)$ or simply as $Z(M; L)$ for short.

For the case of links in S^3, we will claim that the invariants (1.6) are exactly those that appear in the Jones theory and its generalizations. Simply replacing S^3 with a general oriented three manifold M gives a very intriguing (and as we will see, effectively computable) generalization of the known knot polynomials. Taking $r = 0$ (no knots), (1.6) gives invariants of the oriented three manifold M which also turn out to be effectively computable. Before getting into any details, let us note a few preliminary indications of a possible connection between (1.6) and the Jones theory:

(1) In (1.6) we see the right variables, namely a compact Lie group G, a choice of representation R_i for each component C_i of the link L, and an additional

variable k. [In knot theory one usually makes an analytic continuation and replaces k by a complex variable q, but it has been known since Jones' original work that there are special properties at special values of q. We claim that these properties reflect the fact that the three dimensional gauge theory with action (1.3) is well-defined only if k is an integer.] The two variable generalization of the Jones polynomial corresponds to the case that G is $SU(N)$, and the R_i are all the defining N dimensional representation of $SU(N)$. The two variables are N and k, analytically continued to complex values. The Kauffman polynomial similarly arises for $G = SO(N)$ and R the N dimensional representation.

(2) As a further check on the plausibility of a relation between (1.6) and the knot polynomials, let us note first of all that (1.6) depends on a choice of the orientation of M, as this enters in fixing the sign of the Chern-Simons form. Likewise, (1.6) depends on the orientations of the C_i, since these enter in defining the Wilson lines (in computing the holonomy around C_i, one must decide in which direction to integrate around C_i). If, however, one reverses the orientation of one of the C_i and simultaneously exchanges the representation R_i with its complex conjugate $\bar{R}_i$, then the definition of the Wilson lines is unchanged, so (1.6) is invariant under this process. And if (without changing the R_i) one reverses the orientations of *all* components C_i of the link L, then (1.6) is unchanged because of a symmetry that physicists would call "charge conjugation". This is an involution of the Lie algebra of G that exchanges all representations with their complex conjugates; applying this involution to all integration variables in (1.6) leaves (1.6) invariant while exchanging all R_i with their conjugates or equivalently reversing the orientation of all the C_i. These are important formal properties of the knot polynomials.

2. The Weak Coupling Limit

To begin with, since a non-abelian gauge theory with only a Chern-Simons action may seem unfamiliar, one might ask whether this Lagrangian really does lead to a sensible quantum theory, and really can be regulated to give topologically invariant results. In this section, we will briefly investigate this point by studying the theory in a weak coupling limit in which computations are comparatively straightforward. This is the limit of large k.[2] For large k, the path integral

$$Z = \int D\mathscr{A} \exp\left(\frac{ik}{4\pi} \int_M \mathrm{Tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right)\right) \tag{2.1}$$

(for the moment we omit knots) contains an integrand which is wildly oscillatory. The large k limit of such an integral is given by a sum of contributions from the points of stationary phase. The stationary points of the Chern-Simons action are precisely the "flat connections", that is, the gauge fields for which the curvature vanishes

$$F^a_{ij} = 0\,. \tag{2.2}$$

[2] The reader may wish to bear in mind that the discussion in this section and the next contains a number of technicalities which are part of the logical story but perhaps not essential on a first reading

Gauge equivalence classes of such flat connections correspond to homomorphisms

$$\phi\colon\ \pi_1(M) \to G, \tag{2.3}$$

or more exactly to equivalence classes of such homomorphisms, up to conjugation. If for simplicity we suppose that the topology of M is such that there are only finitely many classes of homomorphisms (2.3), then the large k behavior of (2.1) will be a sum

$$Z = \sum_\alpha \mu(A^{(\alpha)}), \tag{2.4}$$

where the $A^{(\alpha)}$ are a complete set of gauge equivalence classes of flat connections, and $\mu(A^{(\alpha)})$ is to be obtained by stationary phase evaluation of (2.1), expanding around $A^{(\alpha)}$. This reduction to a stationary phase evaluation means that the nonabelian theory, for large k, is closely related to the abelian theory. This in turn has been shown [18] to lead to Ray-Singer analytic torsion [27], which is closely related to the purely topological Reidemeister torsion. The $\mu(A^{(\alpha)})$ may be evaluated as follows. We make in (2.1) the change of variables $A_i = A_i^{(\alpha)} + B_i$, where B_i is the new integration variable. An important invariant of the flat connection $A^{(\alpha)}$ is its Chern-Simons invariant

$$I(A^{(\alpha)}) = \frac{1}{4\pi} \int_M \mathrm{Tr}(A^{(\alpha)} \wedge dA^{(\alpha)} + \tfrac{2}{3} A^{(\alpha)} \wedge A^{(\alpha)} \wedge A^{(\alpha)}). \tag{2.5}$$

When the Chern-Simons action is expanded in powers of B_i, the first terms are

$$\mathscr{L} = k \cdot I(A^{(\alpha)}) + \frac{k}{4\pi} \int_M \mathrm{Tr}(B \wedge DB). \tag{2.6}$$

Here it is understood that in (2.6), the expression DB denotes the covariant exterior derivative of B with respect to the background gauge field $A^{(\alpha)}$; it does not depend on a metric on M. A salient point is that in (2.6) there is no term linear in B, since $A^{(\alpha)}$ is a critical point of the action.

To carry out the Gaussian integral in (2.6), gauge fixing is needed. There is no way to carry out this gauge fixing without picking a metric on M (or in some other way breaking the symmetry of the problem). After picking such a metric, a convenient gauge choice is $D_i B^i = 0$ (with D_i the covariant derivative constructed from the metric and the background gauge field $A^{(\alpha)}$). The standard Faddeev-Popov construction then gives rise to a gauge fixing Lagrangian

$$\mathscr{L}_{\text{gauge}} = \int_M (\mathrm{Tr}\,\phi\, D_i B^i + \mathrm{Tr}\,\bar{c}\, D_i D^i c). \tag{2.7}$$

Here ϕ is a Lagrangian multiplier that enforces the gauge condition $D_i B^i = 0$, and c, $\bar{c}$ are anticommuting "ghosts" that are introduced to get the right measure on the space of gauge fields modulo gauge transformations. The quadratic terms in ϕ and B that can be found in (2.6), (2.7) have a natural geometric interpretation, described (in the abelian case) in [18]. Let D be the exterior derivative on M, twisted by the flat connection $A^{(\alpha)}$, and let $*$ be the Hodge operator that maps k forms to $3-k$ forms. On a three manifold one has a natural self-adjoint operator $L = *D + D*$ which maps differential forms of even order to forms of even order

and forms of odd order to forms of odd order. Let L_- denote its restriction to forms of odd order. With B and ϕ regarded as a one form and a three form, respectively, the boson kinetic operator in (2.6), (2.7) is precisely this operator L_-. The kinetic operator of the ghosts is also a natural geometrical operator, the Laplacian, which we will call Δ. We can now give a formula for the stationary point contributions $\mu(A^{(\alpha)})$ that appear in (2.4). This is

$$\mu(A^{(\alpha)}) = \exp(ikI(A^{(\alpha)})) \cdot \frac{\det(\Delta)}{\sqrt{\det(L_-)}}. \tag{2.8}$$

The phase factor in (2.8) is the value of the integrand in (2.1) at the point of stationary phase, and the determinants (whose absolute values can be defined by zeta functions) result from the Gaussian integral over B, ϕ, c, and $\bar{c}$.

Now we come to the crucial point. To regularize the path integral, we have had to pick a Riemannian metric on M. Therefore, it is not obvious a priori that the $\mu(A^{(\alpha)})$ computed this way will really be topological invariants. Perhaps the Chern-Simons theory suffers from anomalies, and cannot be regularized in a generally covariant fashion. Happily, we can now appeal to [18], where it was shown (in the context of the abelian theory, but this aspect of [18] generalizes) that the absolute value of the ratio of determinants appearing in (2.8) is precisely the Ray-Singer analytic torsion of the flat connection $A^{(\alpha)}$, and so in particular is a topological invariant. (The phase of this ratio of determinants is more delicate, and will be discussed later.) This is the first indication that topological invariants really can be obtained from the Chern-Simons theory.

The Phase of the Determinant

Though the absolute value of the ratio of determinants in (2.8) is the analytic torsion discussed long ago by Schwarz, the phase requires additional study. The ghost determinant $\det \Delta$ is real and positive, so the real issue is to study the phase of $\det L_-$. Because the operator L_- can be interpreted as a twisted Dirac operator, the phase of its determinant can be related to the study of the phase of odd dimensional fermion determinants, as studied by various authors [28]. However, I will here give a brief derivation of the relevant facts from the bosonic point of view, which is perhaps more natural in the present context. After an irrelevant rescaling of B and ϕ, the integral of interest is

$$\int DBD\phi \exp\left(i \int_M \operatorname{Tr}(B \wedge DB + \phi D * B)\right). \tag{2.9}$$

Upon changing variables to an orthonormal basis of eigenfunctions x_i of the operator L_-, with eigenvalues λ_i, (2.9) becomes

$$\prod_i \int_{-\infty}^{\infty} \frac{dx_i}{\sqrt{\pi}} e^{i\lambda_i x_i^2}. \tag{2.10}$$

Therefore the crucial integral to understand is

$$I = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{\pi}} e^{i\lambda x^2}, \tag{2.11}$$

for real λ. We consider this integral to be defined by taking the limit as $\varepsilon \to 0$ of the absolutely convergent integral

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{\pi}} e^{i\lambda x^2} \cdot e^{-\varepsilon x^2}. \tag{2.12}$$

With this or any other physically reasonable definition, the integral (2.11) is

$$I = \frac{1}{|\sqrt{\lambda}|} \cdot \exp\left(\frac{i\pi}{4} \operatorname{sign} \lambda\right). \tag{2.13}$$

The phase of the path integral is thus proportional to $\sum_i \operatorname{sign} \lambda_i$, or better, to its regularized version which is the "eta invariant" of Atiyah et al. [29]:

$$\eta(A^{(\alpha)}) = \frac{1}{2} \lim_{s \to 0} \sum_i \operatorname{sign} \lambda_i |\lambda_i|^{-s}. \tag{2.14}$$

Thus, the phase of the path integral may be expressed in the formula

$$\frac{1}{\sqrt{\det L_-}} = \frac{1}{|\sqrt{\det L_-}|} \cdot \exp\left(\frac{i\pi}{2} \eta(A^{(\alpha)})\right). \tag{2.15}$$

This can be made more explicit by using the Atiyah-Patodi-Singer theorem, which for our purposes can be regarded as a formula that expresses the dependence of η on the flat connection $A^{(\alpha)}$ about which we are expanding. In fact, in the case of the operator L_-, the formula is

$$\frac{1}{2}(\eta(A^{(\alpha)}) - \eta(0)) = \frac{c_2(G)}{2\pi} \cdot I(A^{(\alpha)}). \tag{2.16}$$

Here $I(A^{(\alpha)})$ is the Chern-Simons invariant of the flat connection $A^{(\alpha)}$, as defined in (2.5), $\eta(0)$ is the eta invariant of the trivial gauge field $A = 0$, and $c_2(G)$ is the value of the quadratic Casimir operator of the group G in the adjoint representation, normalized so that $c_2(SU(N)) = 2N$. The effect of this factor is to replace k in (2.8) by $k + c_2(G)/2$; in fact, the partition function (2.4) may now be written

$$Z = e^{i\pi\eta(0)/2} \cdot \sum_\alpha e^{i(k + c_2(G)/2) I(A^{(\alpha)})} \cdot T_\alpha \tag{2.17}$$

with T_α [the absolute value of the ratio of determinants in (2.8)] being the torsion invariant of the flat connection $A^{(\alpha)}$.

Unfortunately, although $I(A^{(\alpha)})$ and T_α are topological invariants, $\eta(0)$ is not; it depends on the choice of a metric on M in gauge fixing. Thus, to make sense of the phase of (2.17) requires further discussion, in the next subsection.

Before launching into that technical discussion, let us note that the computation just sketched actually has a very interesting spin-off. The fact that k in (2.8) has been replaced by $k + c_2(G)/2$ in (2.17) appears to be the beginning of an explanation of the fact that in many formulas of $1+1$ dimensional current algebra, quantum corrections have the effect of replacing k by $k + c_2(G)/2$. In turn, this is probably related to the fact that in various integrable models in $1+1$

dimensions, such as the sine-gordon model, the WKB approximation is exact if one makes suitable and seemingly *ad hoc* changes in the values of the parameters, analogous to replacing k by $k + c_2(G)/2$.

Trivialization of the Tangent Bundle

Now, let us discuss how the mysterious phase factor $e^{i\pi\eta(0)/2}$ in (2.17) should be interpreted.

First of all, $\eta(0)$ is the η invariant of the L_- operator coupled to (i) some metric g on M, and (ii) the trivial gauge field $A = 0$. Let $d = \dim G$ be the dimension of the gauge group G. Since the gauge field is trivial, the L_- operator consists of d copies of the purely gravitational L_- operator coupled to the metric only. Thus, as a preliminary, we write

$$\eta(0) = d \cdot \eta_{\text{grav}}, \tag{2.18}$$

where η_{grav} is the eta invariant of the purely gravitational operator. Our problematical phase factor is

$$\Lambda = \exp\left(\frac{id\pi}{2} \cdot \eta_{\text{grav}}\right). \tag{2.19}$$

Now, with a particular regularization of the Chern-Simons quantum field theory, we have obtained the formula (2.17) which contains the ambiguous phase factor Λ. The goal is to find a different regularization which will preserve general covariance. Two regularizations should differ by a local counterterm, and in this case, since the problem phase (2.19) depends on the background metric only, we want a counterterm that depends on the background metric only. It is easy to see that the counterterm with the right properties is a multiple of the gravitational Chern-Simons term, which is defined (by analogy with the Yang-Mills Chern-Simons term) as

$$I(g) = \frac{1}{4\pi} \int_M \operatorname{Tr}(\omega \wedge d\omega + \tfrac{2}{3}\omega \wedge \omega \wedge \omega). \tag{2.20}$$

Here ω is the Levi-Civita connection on the spin bundle of M.[3] $I(g)$ suffers from an ambiguity just similar to that of the Yang-Mills Chern-Simons action. To define $I(g)$ as a number, one requires a trivialization of the tangent bundle of M. Although the tangent bundle of a three manifold can be trivialized, there is no canonical way to do this. Any two trivializations differ by an invariantly defined integer, which is the number of relative "twists". The gravitational Chern-Simons functional has the property that if the trivialization of the tangent bundle of M is twisted by s units, $I(g)$ transforms by

$$I(g) \to I(g) + 2\pi s. \tag{2.21}$$

Now, the Atiyah-Patodi-Singer theorem says that the combination

$$\frac{1}{2}\eta_{\text{grav}} + \frac{1}{12} \cdot \frac{I(g)}{2\pi} \tag{2.22}$$

[3] (2.20) is not the integral of an intrinsic local functional, so it would not usually arise as a counterterm. Whether or not "counterterm" is the right word, we will have to view (2.20) as a correction that must be added to the action if one wishes to work in the gauge $D_i A^i = 0$

is a topological invariant, depending that is on the oriented three manifold M with a choice of trivialization of the tangent bundle, but not on the metric of M.[4] It is clear, therefore, what we must do. We replace $\eta(0)/2$ in (2.17) by d times the combination that appears in (2.22) [the factor of d is the one that entered in (2.19)], so (2.17) is replaced by

$$Z = \exp\left(i\pi d\left(\frac{\eta_{\text{grav}}}{2} + \frac{1}{12}\cdot\frac{I(g)}{2\pi}\right)\right)\cdot\sum_{\alpha} e^{i(k+c_2(G)/2)I(A^{(\alpha)})}\cdot T_\alpha. \tag{2.23}$$

So, finally, we can see that the Chern-Simons partition function, at least for large k, can be defined as a topological invariant of the oriented, framed three manifold M (a framed three manifold being one that is presented with a homotopy class of trivializations of the tangent bundle).

The fact that it is necessary to specify a framing of the three manifold may look like a nuisance, but there is no real loss of information. From (2.21) we see that if the framing is shifted by s units, the partition function is transformed by

$$Z \to Z\cdot\exp\left(2\pi i s\cdot\frac{d}{24}\right). \tag{2.24}$$

A topological invariant of framed, oriented three manifolds, together with a law for the behavior under change of framing, is more or less as good as a topological invariant of oreinted three manifolds without a choice of framing.

Of course, all of the discussion in this section, and in particular (2.24), has been limited to the behavior at large k. In Sect. (4.5), we will see that the generalization of (2.24) to finite k is

$$Z \to Z\cdot\exp\left(2\pi i s\cdot\frac{c}{24}\right), \tag{2.25}$$

with c being the central charge of two dimensional current algebra with symmetry group G at level k. It is well known that the large k limit of c is exactly d.

Moduli Spaces of Flat Connections

There is still an important gap in the above discussion of the large k behavior. The formula (2.8) is really only valid if the determinants that appear are all non-zero. In fact, the flat connection $A^{(\alpha)}$ determines a flat bundle E. The determinants in (2.8) are non-zero if and only if $A^{(\alpha)}$ is such that the de Rham cohomology of M, with values in E, is zero. If $H^1(M, E) \neq 0$, then the flat connection $A^{(\alpha)}$ is not isolated but lies on a moduli space $\mathscr{S}$ of gauge inequivalent flat connections; and the proper evaluation of the path integral (2.1) leads not to the discrete sum (2.4) but to an integral on $\mathscr{S}$. If $H^0(M, E)$ is not zero, then the fields ϕ, c and $\bar{c}$ in the above treatment have zero modes, and the gauge fixing requires more care. It is plausible that by more careful study of the path integral, the large k contribution of arbitrary flat connections can be extracted without assumptions about $H^*(M, E)$. But we will not attempt this.

---

[4] The crucial factor of 1/12 in (2.22) reflects the discrepancy between the Chern character $e^x = 1 + x^2/2 + \ldots$ that appears in gauge theory index theorems and the $\hat{A}$ genus $(x/2)/\sinh(x/2) = 1 - x^2/24 + \ldots$ that appears in gravitational index theorems

Some Examples

We will later on determine the partition functions of some simple three manifolds, giving results that can be compared to large k computations. For $S^2 \times S^1$, $Z = 1$, for any G and any k. For S^3 and $G = SU(2)$, we will obtain the formula [5]

$$Z(S^3) = \sqrt{\frac{2}{k+2}} \sin\left(\frac{\pi}{k+2}\right). \tag{2.26}$$

Of course, on S^3 the only flat connection is the trivial connection, for which (2.8) is not valid, since $H^0(M, E) \neq 0$ in this case. For $G = SU(2)$, the behavior $Z \sim k^{-3/2}$ in (2.26) is probably the general behavior of the contribution of the flat connection for homology spheres (on which the flat connection is isolated); it would be interesting to know how to obtain this behavior from path integrals. In Donaldson and Floer theory, the trivial connection, which has a negative formal dimension, is the cause of many subtleties. The vanishing of (2.26) in the classical limit of large k appears to be an interesting quantitative reflection of the "negative dimension" of the trivial connection.

2.1. *Incorporation of Knots*

We now wish to consider the large k behaviour in the presence of knots. For simplicity, we will limit ourselves to the case of S^3, and an abelian gauge group $G = U(1)$. Though the abelian gauge group is relatively trivial in the context of knot theory, it gives a quick and simple way to confirm the fact that the Chern-Simons action really does lead to topological invariants, and it also gives a simple context for explaining a technicality that is crucial in all that follows.

In the abelian theory, the gauge field is simply a one form A and the Lagrangian is

$$\mathscr{L} = \frac{k}{8\pi} \int_M e^{ijk} A_i \partial_j A_k . \tag{2.27}$$

We pick some circles C_a and some integers n_a [corresponding to representations of the gauge group $U(1)$]. As always in this paper, we assume C_a does not intersect C_b for $a \neq b$. We wish to calculate the expectation value of the product

$$W = \prod_{a=1}^{s} \exp\left(i n_a \int_{C_a} A\right) \tag{2.28}$$

with respect to the Gaussian measure determined by $e^{i\mathscr{L}}$. As was recently discussed by Polyakov (in a paper [23] in which he proposed to apply the Abelian Chern-Simons theory to high temperature superconductors), the result can be written in the form

$$\langle W \rangle = \exp\left(\frac{i}{2k} \sum_{a,b} n_a n_b \int_{C_a} dx^i \int_{C_b} dy^j \varepsilon_{ijk} \cdot \frac{(x-y)^k}{|x-y|^3}\right). \tag{2.29}$$

Here one has identified a region U of S^3 containing the knots with a region of three dimensional Euclidean space, and x^i, y^j are the Euclidean coordinates of U

[5] The appearance of $k+2$ in this formula is presumably an illustration of the $k + c_2(G)/2$ in (2.17)

evaluated along the knots. For $a \neq b$, the integral in (2.29) is essentially the Gauss linking number, which can be written as

$$\Phi(C_a, C_b) = \frac{1}{4\pi} \int_{C_a} dx^i \int_{C_b} dy^j \varepsilon_{ijk} \frac{(x-y)^k}{|x-y|^3}. \tag{2.30}$$

As long as C_a and C_b do not intersect, $\Phi(C_a, C_b)$ is a well defined integer; in fact, it is the most classic invariant in knot theory. Thus, if we could ignore the term $a = b$, we would have

$$\langle W \rangle = \exp\left(\frac{2i\pi}{k} \sum_{a,b} n_a n_b \Phi(C_a, C_b)\right). \tag{2.31}$$

The appearance of the Gauss linking number illustrates the fact that the Chern-Simons theory does lead to topological invariants as we hope. But we have to worry about the term with $a = b$. This integral is ill-defined near $x = y$; how do we wish to interpret it?

It is well known in knot theory that there is no natural and topologically invariant way to regularize the self-linking number of a knot. Polyakov in [23] used a regularization that is not generally covariant to get an answer that is interesting geometrically but not a topological invariant. We need a different approach for our present treatment in which general covariance is a primary goal. Though there is no completely invariant substitute for Polyakov's regularization, in the sense that there is no way to get a natural topological invariant from the integral in (2.29) or (2.30) with $a = b$, we cannot simply throw away the self-linking term and its non-abelian generalizations (which are sketched in Fig. 3a), since these terms are in fact not naturally zero. There is no reason to think that one could retain general covariance by dropping these terms. In the abelian theory, on a general three manifold M, on topological grounds the self-linking number can be a non-zero fraction, well-defined only modulo one. In such a case, it cannot be correct to set the self-linking number to zero, since it is definitely not zero. [Topologically, in such a situation, the self-linking number is well defined only modulo an integer, and this precision is definitely not good enough to evaluate (2.31).] In the non-abelian theory, we will get results later which amount to assigning definite, non-zero values to the non-abelian generalizations of the self-linking integral, so it would not be on the right track to try to throw these terms away.

Topologically, it is clear what data are needed to make sense of the self-linking of a knot C. One needs to give a "framing" of C; this is a normal vector field along C. The idea is that by displacing C slightly in the direction of this vector field one gets a new knot C', and it makes sense to calculate the linking number of C and C'. This can be defined as the self-linking number of the framed knot C. One can think of the framing as a thickening of the knot into a tiny ribbon bounded by C and C'; this is how it is drawn in Fig. 3b. It is clear that the self-linking number defined this way depends not on the actual vector field used to displace C to C' but only on the topological class of this vector field; and indeed by a "framing" we mean only the topological class. Though a choice of framing gives a definition of the self-linking number of a knot C, it is clear that by picking a convenient framing of C one can

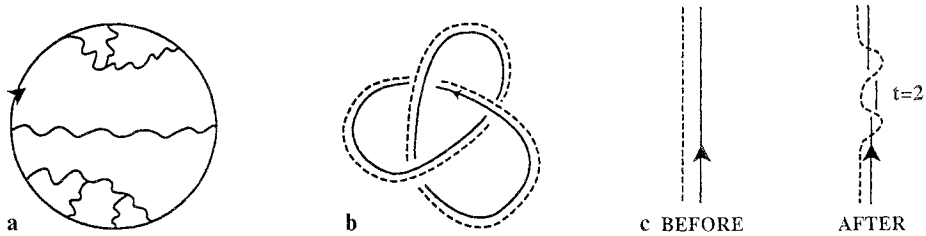

Fig. 3a–c. The self-linking integral is, in a non-abelian theory, the first in an infinite series of Feynman diagrams, with gauge fields emitted and absorbed by the same knot, as in **a**; these all pose similar problems. A topologically invariant but not uniquely determined regularization can be obtained by supposing that each knot is "framed", as in **b**. In **c**, the framing is shifted by 2 units by making a 2-fold twist

get any desired answer for its self-linking number; as illustrated in Fig. 3c, a t-fold twist in the framing of C will change its self-linking by t.[6]

Physically, the role of the framing is that it makes possible what physicists would call a point-splitting regularization. This is defined as follows: when one has to do the self-linking integral in (2.29), one lets x run on C and y on C'. This gives a well-defined integral, though of course it depends on the framing. In this paper, we will assume, without proof, that the framing gives sufficient information to make possible a consistent point-splitting regularization of all the non-abelian generalizations of the self-linking integral, without further arbitrary choices. This question is, perhaps, comparable to the question of whether the non-abelian Chern-Simons action defines a sensible quantum theory in the first place (even without introducing Wilson lines as observables); neither of these questions will be tackled here.

Of course, if it were always possible to pick a canonical framing of knots, then we could pick this framing and hide the question. On S^3, there is a canonical framing of every knot; it is determined by asking that the self-linking number should be zero. (This makes the abelian linking integral zero, but not its non-abelian generalizations.) On general three manifolds, this cannot be done since the self-linking number may be ill-defined or may differ from an integer by a definite fraction (so that it does not vanish with any choice of framing). Even when the canonical framing does exist, it is not convenient to be restricted to using it, since natural operations (like the surgery we study in Sect. 4) may not preserve it.

In general, therefore, we give up on finding a natural choice, and simply pick some framing and proceed. It would be rather unpleasing if the "physical" results depended uncontrollably on the framing of knots. What saves the day is that although we cannot in general make a natural choice of the framing, we can state a general rule for how expectation values of Wilson lines change under a change of the framing. First of all, let us note that while, in general, there is no canonical zero in the set of possible framings of a knot in a three manifold, if one compares two framings they always differ by a definite integer, which is the relative twist in going

[6] The discussion should make it clear that the need to frame knots is analogous to the need to frame three manifolds, as found in the last section. This hopefully justifies the use of the same word "framing" in each case

around the knot (Fig. 3c). (That is, in general there is no natural way to count how many times the ribbon in Fig. 3b is twisted, but there is a natural local operation of adding t extra twists to this ribbon.) In the abelian theory, it is clear from (2.29) and (2.30) how the partition function transforms under a change of framing. If we shift the framing of the link C_a by t units, its self-linking number is increased by t, and the partition function is shifted by a phase

$$\langle W \rangle \to \exp(2\pi i t \cdot (n_a^2/k)) \cdot \langle W \rangle . \tag{2.32}$$

The nonabelian analog of that result will be derived in Sect. 5.1; the transformation law in the non-abelian case is

$$\langle W \rangle \to \exp(2\pi i t \cdot h) \langle W \rangle , \tag{2.33}$$

where h is the conformal weight of a certain primary field in $1+1$ dimensional current algebra. This result, though it may seem rather technical, is a key ingredient enabling the Chern-Simons theory to work. It means that although we need to pick a framing for every link, because the self-linking integrals have no natural definition otherweise, there is no loss of information since we have a definite law for how the partition functions transform under change of framing.

Actually, it can be shown [13] that the structure of rational conformal field theory requires non-trivial monodromies. In the relationship that we will develop between the $2+1$ dimensional Chern-Simons theory and rational conformal field theory in $1+1$ dimensions, the need to frame all knots is the $2+1$ dimensional analog of the monodromies that arise in $1+1$ dimensions. [This will be clear in the derivation of (2.33).] Were it not for the seeming nuisance that knots must be framed to define the Wilson lines as quantum observables, one would end up proving that the Jones knot invariants were trivial.

An alternative description may make the physical interpretation of the framing of knots more transparent. A Wilson line can be regarded as the space-time trajectory of a charged particle. In $2+1$ dimensions, it is possible for a particle to have fractional statistics, meaning that the quantum wave function changes by a phase $e^{2\pi i\delta}$ under a 2π rotation. (See [30] for a discussion of these issues.) If one wishes to compute a quantum amplitude with propagation of a particle of fractional statistics, it is not enough to specify the orbit of the particle; it is necessary to also count the number of 2π rotations that the particle undergoes in the course of its motion. Equations (2.32) and (2.33) mean that the particles represented by Wilson lines in the Chern-Simons theory have fractional statistics with $\delta = n_a^2/2k$ in the abelian theory or $\delta = h$ in the non-abelian theory. This fractional statistics is the phenomenon claimed by Polyakov in [23], so in essence we agree with his substantive claim, though we prefer to exhibit this phenomenon in the context of a generally covariant regularization, where it appears in the behavior of Wilson lines under change of framing.

In this section, we have obtained some important evidence that the Chern-Simons theory can be regularized to give invariants of three manifolds and knots. We have also obtained the important insight that doing so requires picking a homotopy class of trivializations of the tangent bundle, and a "framing" of all knots. To actually solve the theory requires very different methods, to which we turn in the next section.

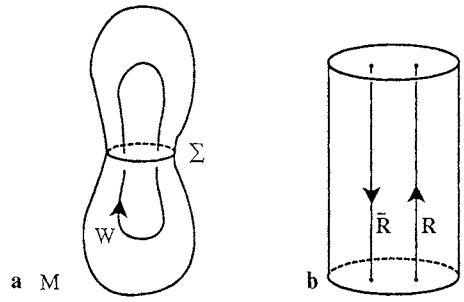

Fig. 4a and b. Cutting a three manifold M on an intermediate Riemann surface Σ is indicated in part **a**. Wilson lines W on M may pierce Σ and if so Σ comes with certain "marked points", with representations attached. Locally, near Σ, M looks like $\Sigma \times R^1$, indicated in part **b**

3. Canonical Quantization

The basic strategy for solving the Yang-Mills theory with Chern-Simons action on an arbitrary three manifold M is to develop a machinery for chopping M in pieces, solving the problem on the pieces, and gluing things back together. So to begin with we consider a three manifold M, perhaps with Wilson lines, as in Fig. 4a. We "cut" M along a Riemann surface Σ. Near the cut, M looks like $\Sigma \times R^1$, and our first step in learning to understand the theory on an arbitrary three manifold is to solve it on $\Sigma \times R^1$.

The special case of a three manifold of the form $\Sigma \times R^1$ is tractable by means of canonical quantization. Canonical quantization on $\Sigma \times R^1$ will produce a Hilbert space $\mathscr{H}_\Sigma$, "the physical Hilbert space of the Chern-Simons theory quantized on Σ". [7] These will turn out to be finite dimensional spaces, and moreover spaces that have already played a noted role in conformal field theory. In rational conformal field theories, one encounters the "conformal blocks" of Belavin, Polyakov, and Zamolodchikov. Segal has described these in terms of "modular functors" that canonically associate a Hilbert space to a Riemann surface, and has described in algebra-geometric terms a particular class of modular functors, which arise in current algebra of a compact group G at level k [16]. The key observation in the present work was really the observation that precisely those functors can be obtained by quantization of a three dimensional quantum field theory, and that this three dimensional aspect of conformal field theory gives the key to understanding the Jones polynomial.

[7] It is conventional in physics to call vector spaces obtained in this fashion "Hilbert spaces", and we will follow this terminology. In fact, the claim that comes most naturally from path integrals and that we will actually use is only that $\mathscr{H}_\Sigma$ is a vector space canonically associated with Σ, and exchanged with its dual when the orientation of Σ is reversed. However, a Hilbert space structure is natural in the Hamiltonian viewpoint, and in the particular problem we are considering here, an inner product on $\mathscr{H}_\Sigma$ is important in more delicate aspects of conformal field theory; such an inner product gives a "metric on the flat vector bundle" in the language of Friedan and Shenker [31]. According to Segal [16], $\mathscr{H}_\Sigma$ in fact has a canonical projective Hilbert space structure

Actually, the general situation that must be studied is that in which possible Wilson lines on M are "cut" by Σ, as in the figure. In this case Σ is presented with finitely many marked points $P_1, \ldots P_k$, with a G representation R_i assigned to each P_i (since each Wilson line has an associated representation). To this data – an oriented topological surface with marked points, and for each marked point a representation of G – we wish to associate a vector space. This is also the general situation that arises in conformal field theory – the marked points are points at which operators with non-vacuum quantum numbers have been inserted. If one reverses the orientation of Σ (and replaces the representations R_i associated with the marked points with their complex conjugates) the vector space $\mathscr{H}_\Sigma$ must be replaced with its dual.

The Canonical Formalism. At first sight, (1.3) might look like a typically intractable nonlinear quantum field theory, but this is far from being so. Working on $\Sigma \times R^1$, it is very natural to choose the gauge $A_0 = 0$ (with A_0 being the component of the connection in the R^1 direction). In this gauge we immediately see that the Lagrangian becomes quadratic. It reduces to

$$\mathscr{L} = \frac{k}{8\pi} \int dt \int_\Sigma \varepsilon^{ij} \operatorname{Tr} A_i \frac{d}{dt} A_j . \tag{3.1}$$

For the time being we will ignore extra complications due to Wilson lines that may be present on $\Sigma \times R^1$. From (3.1) we may deduce the Poisson brackets,[8]

$$\{A_i^a(x), A_j^b(y)\} = \frac{4\pi}{k} \cdot \varepsilon_{ij} \delta^{ab} \delta^2(x-y) . \tag{3.2}$$

Before rushing ahead to quantize these commutation relations, we should remember that the system is subject to a "Gauss law" constraint, which is $\delta \mathscr{L} / \delta A_0 = 0$, or (ignoring the Wilson lines)

$$\varepsilon^{ij} F_{ij}^a = 0 . \tag{3.3}$$

This constraint equation is nonlinear (since F contains a quadratic term), and – as (3.1) is certainly a free theory – this nonlinearity is what remains of the underlying nonlinearity of (1.3).

In quantum field theory, one very often quantizes first and then imposes the constraints. The situation that we are considering here is a situation in which it is far more illuminating to first impose the constraints and then quantize. For the phase space $\mathscr{M}_0$ of connections $A_i^a(x)$ without the constraints is an infinite dimensional phase space; imposing the constraints will reduce us to a rather subtle but eminently finite dimensional phase space $\mathscr{M}$. The problem that faces us here, of reducing from $\mathscr{M}_0$ to $\mathscr{M}$ by imposing the constraints (3.2), has been studied before – and has proved to have extremely rich properties – in the work of Atiyah and Bott on equivariant Morse theory, two dimensional Yang-Mills theory, and

[8] This is a typical problem in which it is not appropriate to "introduce canonical momenta". The purpose of introducing such variables is to reexpress a given Lagrangian in a form which is first order in time derivatives, but (3.1) is already first order in time derivatives. The variables in (3.1) are already canonically conjugate, as indicated in the following equation

the moduli space of holomorphic vector bundles [33]. In our present investigation, this familiar problem appears from a novel three dimensional vantage point.

It is necessary to recall the nature of constraint equations in classical physics. The constraints (3.2) are functions that should vanish, but they also generate gauge transformations via Poisson brackets. Imposing the constraints means two things classically: First, we restrict ourselves to values of the canonical variables for which the constraint functions vanish; and second, we identify two solutions of the constraint equations if they differ by a gauge transformation. In the case at hand, the first step means that we should consider only "flat connections", that is, connections for which $F_{ij}^a = 0$. The second step means that we identify two flat connections if they differ by a gauge transformation. Taking the two steps together, we see that the physical phase space, obtained by imposing the constraints (3.2), is none other than the moduli space of flat connections on Σ, modulo gauge transformations. Such flat connections are completely characterized by the "Wilson lines", that is, the holonomies around non-contractible loops on Σ. A simple count of parameters shows that on a Riemann surface of genus $g > 1$, the moduli space $\mathscr{M}$ of flat connections modulo gauge transformations has dimension $(2g-2) \cdot d$, where d is the dimension of the group G.

The topology of $\mathscr{M}$ is rather intricate (and this was in fact the main subject of interest in [33]). On general grounds $\mathscr{M}$ inherits a symplectic structure (that is, a structure of Poisson brackets) from the symplectic structure present on $\mathscr{M}_0$ before imposing the constraints. $\mathscr{M}$ is a compact space (with some singularities), and in particular its volume with the natural symplectic volume element is finite. Since in quantum mechanics there is one quantum state per unit volume in classical phase space, the finiteness of the volume of $\mathscr{M}$ means that the quantum Hilbert spaces will be finite dimensional. We would like to determine them.

3.1. The Holomorphic Viewpoint

Quantization of classical mechanics is usually carried out by separating the canonical variables into "coordinates", q^i, which are a maximal set of real commuting variables, and "momenta", p^j, which are conjugate to the q^i. The quantum Hilbert space is then the space $\mathscr{H}$ of square integrable functions of the q^i.

Such a scheme definitely requires a noncompact phase space of infinite volume, since – though the q^i may take values in a compact space – the p^j are definitely unbounded. Accordingly, the space $\mathscr{H}$ is infinite dimensional.

Quantizing a compact, finite volume phase space, such as the moduli space $\mathscr{M}$ of flat connections modulo gauge transformations, is quite a different kind of problem. It has no known general solution, but there is one important class of cases in which there is a natural notion of quantization. This arises in the case in which $\mathscr{M}$ is a Kähler manifold, and the symplectic structure on $\mathscr{M}$ is the curvature form that represents the first Chern class of a holomorphic line bundle L endowed with some metric. In this case, one carries out quantization not by separating the variables in phase space into "coordinates" and "momenta", q's and p's, but by separating them into holomorphic and anti-holomorphic degrees of freedom, essentially $z \sim q + ip$ and $\bar{z} \sim q - ip$. The quantum Hilbert space $\mathscr{H}$ is then a suitable space of holomorphic "functions". More exactly, $\mathscr{H}$ is the space of

holomorphic sections of the line bundle L. If $\mathcal{M}$ is compact, this latter space will be finite dimensional. In our problem, with $\mathcal{M}$ being the moduli space of flat connections modulo gauge transformations on an oriented smooth surface Σ, is there a natural Kähler structure on $\mathcal{M}$? The answer is crucial for all that follows. There is not quite a *natural* Kähler structure on $\mathcal{M}$, but there is a natural way to obtain such structures. Once one picks a complex structure J on Σ, the moduli space $\mathcal{M}$ of flat connections can be given a new interpretation - it is the moduli space of stable holomorphic G_C bundles on Σ which are topologically trivial (G_C is the complexification of the gauge group G). Let us refer to the latter space as $\mathcal{M}_J$. $\mathcal{M}_J$ is naturally a complex Kähler (and in fact projective algebraic) variety. Upon picking a linear representation of G (for our purposes it is convenient to pick a representation with the smallest value of the quadratic Casimir operator, e.g. the N dimensional representation of $SU(N)$ or the adjoint representation of E_8), and passing from a principal G_C bundle to the associated vector bundle, we can think of $\mathcal{M}_J$ as the moduli space of a certain family of holomorphic vector bundles. For $G = SU(N)$, $\mathcal{M}_J$ is simply the moduli space of all stable rank N holomorphic vector bundles of vanishing first Chern class.

The symplectic form on $\mathcal{M}$ that appears in (3.1) or (3.2) *without* picking a complex structure on Σ has a very special interpretation in holomorphic terms once we *do* pick such a complex structure. Let us recall the notion [34] of the determinant line bundle of the $\bar{\partial}$ operator. The $\bar{\partial}$ operator on Σ can be "twisted" by any holomorphic vector bundle. $\mathcal{M}_J$ parametrizes a family of holomorphic vector bundles on Σ, and thus it can be regarded as parametrizing a family of $\bar{\partial}$ operators. Taking the determinant line gives a line bundle L over the base space $\mathcal{M}_J$ of this family. Furthermore [34], the Dirac determinant gives a natural metric on L, and the first Chern class of L, computed with this metric, is precisely the symplectic form that appears in (3.1) or (3.2), provided $k = 1$. For general k, the symplectic form that appears in (3.1) or (3.2) represents the first Chern class of the kth power of the determinant line bundle.[9]

Thus, all of the conditions are met for a straightforward quantization of (3.1), taking into account the constraints (3.3). The constraints mean that the classical space to be quantized is the moduli space $\mathcal{M}$ of flat connections. Picking an arbitrary complex structure J on Σ, $\mathcal{M}$ becomes a complex manifold, and the symplectic form of interest represents the first Chern class of $L^{\otimes k}$, the kth tensor power of the determinant line bundle. The quantum Hilbert space $\mathcal{H}_\Sigma$ is thus the space of global holomorphic sections of $L^{\otimes k}$.

3.2. A Flat Vector Bundle on Moduli Space

This gives an answer to the problem of canonically quantizing the Chern-Simons theory on $\Sigma \times R^1$, but a crucial point now requires discussion.

[9] This description is valid for the gauge group $G = SU(N)$, but in general the following modification is needed. For groups other than $SU(N)$ the determinant line bundle L is not the fundamental line bundle on $\mathcal{M}$ but a tensor power thereof. For instance, for $G = E_8$, there is a line bundle L' with $(L')^{\otimes 30} \simeq L$. It is then L' whose first Chern class corresponds to (3.1) or (3.2) with $k = 1$

Quantizing (3.1), with the constraints (3.3), is a problem that can be naturally asked whenever one is given an oriented smooth surface Σ. Beginning with a generally covariant Lagrangian in three dimensions, we were led to this problem in a context in which it was not natural to assume any metric or complex structure on Σ. However, to solve the problem and construct $\mathscr{H}_\Sigma$, it was very natural to pick a complex structure J on Σ. Thus, our description of $\mathscr{H}_\Sigma$ depends on the choice of J, and what we have called $\mathscr{H}_\Sigma$ might perhaps be better called $\mathscr{H}_\Sigma^{(J)}$. As J varies, the $\mathscr{H}_\Sigma^{(J)}$ vary holomorphically with J, and thus we could interpret this object as a holomorphic vector bundle on the moduli space of complex Riemann surfaces. But since $\mathscr{H}_\Sigma^{(J)}$ is the answer to a question that depends on Σ and not on J, we would like to believe that likewise the $\mathscr{H}_\Sigma^{(J)}$ canonically depend only on Σ and not on J. The assertion that the $\mathscr{H}_\Sigma^{(J)}$ are canonically independent of J, and depend only on Σ, is the assertion that the vector bundle on moduli space given by the $\mathscr{H}_\Sigma^{(J)}$ has a canonical flat connection that permits one to identify the fibers. Such "flat vector bundles on moduli space" first entered in conformal field theory somewhat implicitly in the differential equations of Belavin, Polyakov, and Zamolodchikov [32]. They were discussed much more explicitly by Friedan and Shenker [31], who proposed that they would play a pivotal role in conformal field theory, and they have been prominent in subsequent work such as [12, 13]. At least in one important class of examples, we have just met a natural origin of "flat vector bundles on moduli space". The problem "quantize the Chern-Simons action" can be posed without picking a complex structure, so the answer is naturally independent of complex structure and thus gives a "flat bundle on moduli space". The particular flat bundles on moduli space that we get this way are those that Segal has described [16] in connection with conformal field theory; Segal also rigorously proved the flatness, which is explained somewhat heuristically by the physical argument sketched above. (Because of the conformal anomaly, this bundle has only a projectively flat connection, with the projective factor being canonically odd under reversal of orientation.)

The role of these flat bundles in conformal field theory is as follows. If one considers current algebra on a Riemann surface, with a symmetry group G, at "level" k, then one finds that in genus zero the Ward identities uniquely determine the correlation functions for descendants of the identity operator, but this is not so in genus $\geqq 1$. On a complex Riemann surface Σ of genus $\geqq 1$, the space of solutions of the Ward identities for descendants of the identity is a vector space $\hat{\mathscr{H}}_\Sigma$, which might be called the "space of conformal blocks". Segal calls the association $\Sigma \to \hat{\mathscr{H}}_\Sigma$ a "modular functor", and has given an algebra-geometric description of the modular functors that arise in current algebra. In quantizing the Chern-Simons theory we have exactly reproduced this description! This is then the secret of the relation between current algebra in $1+1$ dimensions and Yang-Mills theory in $2+1$ dimensions: the space of conformal blocks in $1+1$ dimensions are the quantum Hilbert spaces obtained by quantizing a $2+1$ dimensional theory. It would take us to far afield to explain here the algebra-geometric description of the space of conformal blocks. Suffice it to say that when one tries to use the Ward identities of current algebra to uniquely determine the correlation functions of descendants of the identity on a curve Σ of genus $\geqq 1$, one meets an obstruction which involves the existence of non-trivial holomorphic vector bundles on Σ; the

Ward identities reduce the determination of the correlation functions to the choice of a holomorphic section of $L^{\otimes k}$ over the moduli space of bundles.

It seems appropriate to conclude this discussion with some remarks on the formal properties of the association $\Sigma \to \mathscr{H}_\Sigma$. It is good to first think of the functor $\Sigma \to H^1(\Sigma, R)$ which to a Riemann surface Σ associates its first de Rham cohomology group. This functor is defined for every smooth surface Σ, independent of complex structure. A diffeomorphism of Σ induces a linear transformation on $H^1(\Sigma, R)$, so $H^1(\Sigma, R)$ furnishes in a natural way a representation of the mapping class group. The formal properties of the functors $\Sigma \to \mathscr{H}_\Sigma$ that come by quantizing the Chern-Simons theory are quite analogous. Though a complex structure J on Σ is introduced to construct $\mathscr{H}_\Sigma$, the existence of a natural projectively flat connection on the moduli space of complex structures permits one locally to (projectively) identify the various $\mathscr{H}_\Sigma^{(J)}$ and forget about the complex structure. One might think that the global monodromies of the flat connection on moduli space would mean that globally one could not forget the complex structure, but this is not so; these monodromies just correspond to an action of the purely topological mapping class group, so that the formal properties of $\mathscr{H}_\Sigma$ are just like those of $H^1(\Sigma, R)$.

3.3. *Inclusion of Wilson Lines*

So far, we have discussed the quantization of the Chern-Simons theory on a Riemann surface Σ *without* Wilson lines. Now we wish to include the Wilson lines, which, as in Fig. 4b, pierce Σ in some points P_i; associated with each such point is a representation R_i. Quantizing the Chern-Simons theory in the presence of the Wilson lines should give a Hilbert space $\mathscr{H}_{\Sigma; P_i, R_i}$ that is canonically associated with the oriented surface Σ together with the choice of P_i and R_i.

It is pretty clear what problem in conformal field theory this should correspond to. Instead of simply considering correlation functions of the descendants of the identity, we should consider in the conformal field theory primary fields transforming in the R_i representations of G. With these fields (or their descendants) inserted at points P_i on Σ, one gets in conformal field theory a more elaborate space $\hat{\mathscr{H}}_{\Sigma; P_i, R_i}$ of conformal blocks. Again, there is an algebra-geometric description of this space [16], and this is what we should expect to recover by quantizing the Chern-Simons theory in the presence of the Wilson lines.

I will now briefly sketch how this works out, deferring a fuller treatment for another occasion. First of all, the Wilson lines correspond to static non-abelian charges which show up as extra terms in the constraint equations. So (3.3) is replaced by

$$\frac{k}{8\pi}\,\varepsilon^{ij} F^a_{ij}(x) = \sum_{s=1}^{r} \delta^2(x - P_s)\, T^a_{(s)}, \tag{3.4}$$

where P_s, $s = 1 \ldots r$ are the points at which static external charges have been placed, and $T^a_{(s)}$, $a = 1 \ldots \dim G$ are the group generators associated with the external charges. Now, a naive attempt to quantize (3.1) with the generalized constraints (3.4) would run into extremely unpleasant difficulties. One could try to quantize first and then impose the constraints, but this is difficult to see through

even in the absence of the external charges. Alternatively, one can try to impose the constraints at the classical level and then quantize, as we did above. But it is hard to make sense of (3.4) as constraints in the classical theory; the solution A_i^a of (3.4) cannot be an ordinary c-number connection, since non-commuting operators appear on the right-hand side. It is clear that to solve (3.4), A_i^a would have to be some sort of "q-number connection", whose holonomy would presumably be an element of a "quantum group", not an ordinary classical group. Indeed, it seems likely that the theory of quantum groups [35] can be considered to arise in this way.

However, there is a much better way to quantize the Chern-Simons theory with static charges. We certainly wish to impose (3.4) at the classical level. This cannot be done directly, since on the right-hand side there appear quantum operators. A useful point of view is the following. A representation R_i of a group G should be seen as a quantum object. This representation should be obtained by quantizing a classical theory. The Borel-Weil-Bott theorem gives a canonical way to exhibit for every irreducible representation R of a compact group G a problem in classical physics, with G symmetry, such that the quantization of this classical problem gives back R as the quantum Hilbert space. One introduces the "flag manifold" G/T, with T being a maximal torus in G, and for each representation R one introduces a symplectic structure ω_R on G/T, such that the quantization of the classical phase space G/T, with the symplectic structure ω_R, gives back the representation R. Many aspects of representation theory find natural explanations by thus regarding representations of groups as quantum objects that are obtained by quantization of classical phase spaces.

In the problem at hand, this point of view can be used to good effect. We extend the phase space $\mathscr{M}_0$ of G connections on Σ by including at each marked point P_i a copy of G/T, with the symplectic structure appropriate to the R_i representation. The quantum operators $T_{(i)}^a$ that appear on the right of (3.4) can then be replaced by the classical functions on G/T whose quantization would give back the $T_{(i)}^a$. The constraints (3.4) then make sense as classical equations, and the analysis can be carried out just as we did without marked points, though the details are a bit longer. Suffice it to say that after imposing the classical constraints, one gets a finite dimensional phase space $\mathscr{M}_{P_i, R_i}$ that incorporates the static charges; a point on this space is a flat G connection on Σ with a reduction of structure group to T at the points P_i. Upon picking an arbitrary conformal structure on Σ, this phase space can be quantized. In this way one gets exactly Segal's description of the space of conformal blocks in current algebra in a general situation with primary fields in the R_i representation inserted at the points P_i. (In current algebra at level k, one only permits certain representations, the "integrable ones". If one formally tries to include other representations, the Ward identities show that they decouple [36]. According to Segal, the analogous statement in algebraic geometry is that the appropriate line bundle over $\mathscr{M}_{P_i, R_i}$ has no non-zero holomorphic sections unless the R_i all correspond to integrable representations. For the Chern-Simons theory, this means that unless the representations R_i are all integrable, the zero vector is the only vector in the physical Hilbert space.)

Finally, let us note that the Borel-Weil-Bott theorem should not be used simply as a tool in quantization. It should be built into the three dimensional description.

One should use the theorem to replace the Wilson lines (1.5) that appear in (1.6) with a functional integral over maps of the circle S into G/T (or actually an integral over sections of a G/T bundle, twisted by the restriction to S of the G-bundle E). This gives a much more unified formalism.

3.4. The Riemann Sphere with Marked Points

The above description may seem a little bit dense, and we will supplement it by giving a simple intuitive description of the physical Hilbert space $\mathscr{H}_{\Sigma;R_i,P_i}$ in the important case of genus zero. Let Σ be an oriented surface of genus zero, with static charges in the R_i representation at points P_i. Let us consider the case of very large k. Now, the gauge coupling in (1.3) is of order $1/k$, so for large k we are dealing with very weak coupling. Rather naively, one might believe that for extremely weak coupling the physical Hilbert space is the same as it would be if the charges were not coupled to gauge fields. If so, the physical Hilbert space would be simply the tensor product $\mathscr{H}_0 = \otimes_i R_i$ of the Hilbert spaces R_i of the individual charges. However, there is a key error here. No matter how weak the gauge coupling may be, we must remember that in a closed universe the total charge must be zero (since the electric flux has nowhere to go). The total charge being zero means in a nonabelian theory that all of the charges together must be coupled to the trivial representation of G. So the physical Hilbert space, for large k, is precisely the G-invariant subspace of $\mathscr{H}_0$, or

$$\mathscr{H} = \mathrm{Inv}\,(\otimes_i R_i)\,. \tag{3.5}$$

This is a familiar answer in conformal field theory for the space of conformal blocks obtained, in the large k limit, in coupling representations R_i. Considerations of conformal field theory also show that for finite k the correct answer is always a subspace of (3.5). The most important modification of (3.5) that arises for finite k (and is explained algebra-geometrically in [16]) is that $\mathscr{H}$ is zero unless the R_i correspond to integrable representations of the loop group; in what follows a restriction to such representations is always understood.

Now we consider some important special cases.

(i) For the Riemann sphere with no marked points, the Hilbert space is one dimensional. This is well known in conformal field theory – for descendants of the identity on the Riemann sphere, there is only one conformal block.

(ii) For the Riemann sphere with one marked point in a representation R_i, the Hilbert space is one dimensional if R_i is trivial, and zero dimensional otherwise.

(iii) For the Riemann sphere with two marked points with representations R_i and R_j, the Hilbert space is one dimensional if R_j is the dual of R_i (so that there is an invariant in $R_i \otimes R_j$) and zero dimensional otherwise. Again, this is well known in conformal field theory.

(iv) For the Riemann sphere with three marked points in representations R_i, R_j, and R_k, the dimension of $\mathscr{H}$ is the number N_{ijk} for which Verlinde has proposed [12] and Moore and Seiberg have proved [13] rather striking properties. Here, N_{ijk} may in general be less than its large k limit which is the dimension of (3.5).

(v) From the results of Verlinde, the dimensions of the physical Hilbert spaces for an arbitrary collection of marked points on S^2 can be determined from a knowledge of the N_{ijk}. But let us consider a particularly important special case.

Suppose that there are four external charges, and that the representations are R, R, $\bar{R}$, and $\bar{R}$. If the decomposition of $R \otimes R$ is

$$R \otimes R = \bigoplus_{i=1}^{s} E_i, \tag{3.6}$$

with the E_i being distinct irreducible representations of G, then the physical Hilbert space $\mathscr{H}$ at large k will be s dimensional, since the possible invariants in $R \otimes R \otimes \bar{R} \otimes \bar{R}$ are uniquely fixed by giving the representation to which $R \otimes R$ is coupled. (For small k the dimension of $\mathscr{H}$ might be less than s.) In understanding the knot polynomials, an important special case is that in which G is $SU(N)$ and R is the defining N dimensional representation. In that case, $s = 2$ and the physical Hilbert space is two dimensional (except for $k = 1$ where it is one dimensional).

4. Calculability

Our considerations so far may have seemed somewhat abstract, and we would now like to show that in fact these considerations can actually be used to calculate things. As an introduction to the requisite ideas, we will first deduce a certain theoretical principle that is of great importance in its own right.

Consider, as in Fig. 5a, a three manifold M which is the connected sum of two three manifolds M_1 and M_2, joined along a two sphere S^2. There may be knots in M_1 or M_2, but if so they do not pass through the joining two sphere. If for every three manifold X we denote the partition function or Feynman path integral (1.6) as $Z(X)$, then we wish to deduce the formula

$$Z(M) \cdot Z(S^3) = Z(M_1) \cdot Z(M_2) \tag{4.1}$$

[it being understood that $Z(S^3)$ denotes the partition function of a three sphere that contains no knots]. This can be rewritten

$$\frac{Z(M)}{Z(S^3)} = \frac{Z(M_1)}{Z(S^3)} \cdot \frac{Z(M_2)}{Z(S^3)}. \tag{4.2}$$

In some special cases, (4.2) is equivalent or closely related to known formulas. If M_1 and M_2 are copies of S^3 with knots in them, then the ratios appearing in (4.2) turn out to be the knot invariants that appear in the Jones theory, and (4.2) expresses the fact that these invariants are multiplicative when one takes the disjoint sum of knots. If M_1 and M_2 are arbitrary three manifolds without knots, then (in view of our discussion in Sect. 2) (4.2) is closesly related to the multiplicativity of Reidemeister and Ray-Singer torsion under connected sums.

So let us study Fig. 5a using the general ideas of quantum field theory. On the left of this figure, we see a three manifold M_1 with boundary S^2. According to the general ideas of quantum field theory, one associates a "physical Hilbert space" $\mathscr{H}$ with this S^2; as we have seen in the last section, it is one dimensional. The Feynman path integral on M_1 determines a vector χ in $\mathscr{H}$. Likewise, on the right of

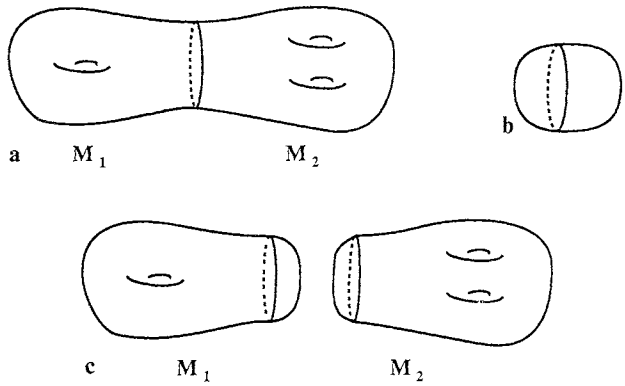

Fig. 5a–c. In **a** is sketched a three manifold M which is the connected sum of two pieces M_1 and M_2, joined along a sphere S^2. Similarly, a three sphere S^3 can be cut along its equator, as in **b**. Cutting both M and S^3 as indicated in **a** and **b**, the pieces can be rearranged into the *disconnected* sum of M_1 and M_2, as in **c**

Fig. 5a we see a three manifold M_2 whose boundary is the same S^2 with opposite orientation; its Hilbert space $\mathscr{H}'$ is canonically the dual of $\mathscr{H}$. The path integral on M_2 determines a vector ψ in $\mathscr{H}'$, and according to the general ideas of quantum field theory, the partition function of the connected sum M is

$$Z(M) = (\chi, \psi). \tag{4.3}$$

The symbol (χ, ψ) denotes the natural pairing of vectors $\chi \in \mathscr{H}$, $\psi \in \mathscr{H}'$. We cannot evaluate (4.3), since we do not know χ or ψ. Instead, let us consider some variations on this theme. The two sphere S^2 that separates the two parts of Fig. 5a could be embedded in S^3 in such a way as to separate S^3 into two three balls B_L and B_R. The path integrals on B_L and B_R would give vectors v and v' in $\mathscr{H}$ and $\mathscr{H}'$, and the same reasoning as led to (4.3) gives

$$Z(S^3) = (v, v'). \tag{4.4}$$

Again, we do not know v or v' and cannot evaluate (4.4). But we can say the following. As $\mathscr{H}$ is one dimensional, v is a multiple of χ; likewise, since $\mathscr{H}'$ is one dimensional, v' is a multiple of ψ. It is then a fact of one dimensional linear algebra that

$$(\chi, \psi) \cdot (v, v') = (\chi, v') \cdot (v, \psi). \tag{4.5}$$

The two terms on the right-hand side of (4.5) are respectively $Z(M_1)$ and $Z(M_2)$, as we see in Fig. 5c. So (4.5) is equivalent to the desired result (4.1).

One may wonder what is the mysterious object $Z(S^3)$ that is so prominent in (4.1). Can it be set to one? Actually, the axioms of quantum field theory are strong enough so that the value of $Z(S^3)$ is uniquely determined and cannot be postulated arbitrarily; as we will see later it can be calculated from the theory of affine Lie algebras. For $G = SU(2)$ the formula has been given in (2.26).

As a special case of (4.1), pick s irreducible representations of G, say $R_1, \ldots R_s$, and consider a link in S^3 that consists of s unlinked and unknotted circles C_i, with

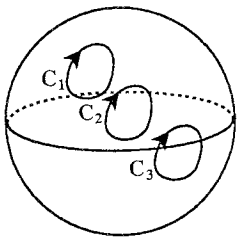

Fig. 6. A three sphere with 3 unlinked and unknotted circles C_i, associated with representations $R_1 \dots R_3$. The figure can be cut in various ways to separate the circles

one of the R_i associated with each circle. This is indicated in Fig. 6. Denote the partition function of S^3 with this collection of Wilson lines as $Z(S^3; C_1, \dots C_s)$ (the representations R_i being understood). Then by cutting the figure to separate the circles, and repeatedly using (4.1), we learn that

$$\frac{Z(S^3; C_1, \dots C_s)}{Z(S^3)} = \prod_{k=1}^{s} \frac{Z(S^3; C_k)}{Z(S^3)}. \tag{4.6}$$

If we introduce the normalized expectation value of a link L, defined by $\langle L \rangle = Z(S^3; L)/Z(S^3)$, then (4.6) becomes

$$\langle C_1 \dots C_s \rangle = \prod_k \langle C_k \rangle \tag{4.7}$$

for an arbitrary collection of unlinked, unknotted Wilson lines on S^3.

In knot theory there is another notion of connected sum, the "connected sum of links". The Jones invariants also have a simple multiplicative behavior under this operation, as we will sketch briefly at the end of Sect. 4.5.

4.1. Knots in S^3

We will now describe the origin of the "skein relation" which can be taken as the definition of the knot polynomials for knots on S^3. (A special case of the skein relation was first used by Conway in connection with the Alexander polynomial.)

Consider a link L on a general three manifold M, as indicated in Fig. 7a. The components of the link are associated with certain representations of G, and we wish to calculate the Feynman path integral (1.6), which we will denote as $Z(L)$ (with the representations understood). We will evaluate it by deducing an algorithm for unknotting knots. If the lines in Fig. 7 could pass through each other unimpeded, all knots could be unknotted. As it is, this is prevented by some unfortuitous crossings, such as the one circled in the figure. Let us draw a small sphere about this crossing, cut it out, and study it more closely. This cuts M into two pieces, which after rearrangement are shown in Fig. 7b as a complicated piece M_L shown on the left of the figure and a simple piece M_R shown on the right. M_R consists of a three ball with boundary S^2; on this boundary there are four marked points that are connected by two lines in the interior of the ball.

To make the discussion concrete, let us suppose that the gauge group is $G = SU(N)$ and that the Wilson lines are all in the defining N dimensional representation of $SU(N)$, which we will call R. Then, as we saw at the end of the

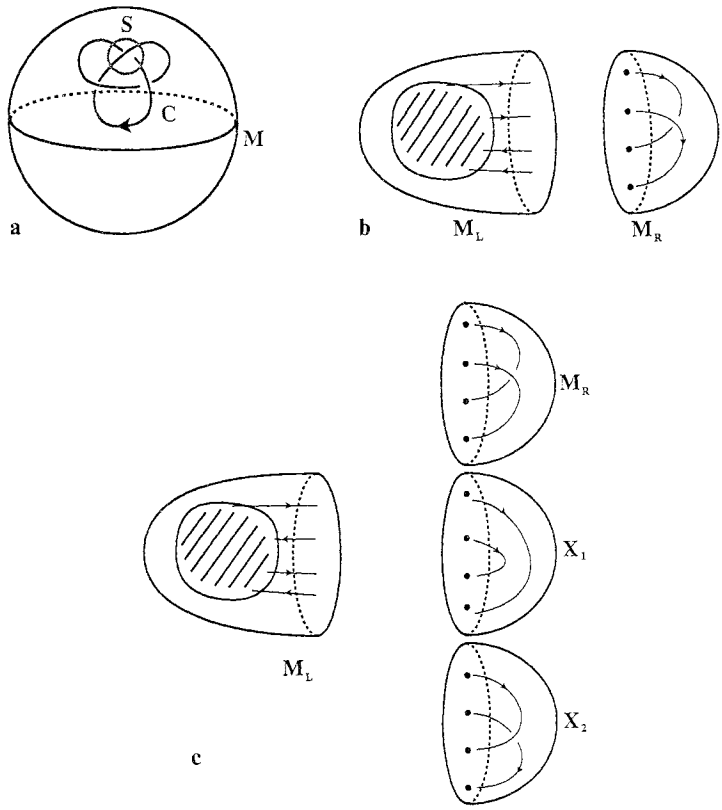

Fig. 7a–c. A link C on a general three manifold M is sketched in **a**. A small sphere S has been drawn about an inconvenient crossing; it cuts M into a simple piece (the interior of S) and a complicated piece. In **b**, the picture is rearranged to exhibit the cutting of M more explicitly; the two pieces now appear on the left and right as M_L (the complicated piece whose details are not drawn) and M_R (the interior of S). The key to the skein relation is to consider replacing M_R with some substitutes, as shown in **c**

last section, the physical Hilbert spaces $\mathscr{H}_L$ and $\mathscr{H}_R$ associated with the boundaries of M_L and M_R are two dimensional.

The strategy is now the same as the strategy which led to the multiplicativity relation (4.1). The Feynman path integral on M_L determines a vector χ in $\mathscr{H}_L$. The Feynman path integral on M_R determines a vector ψ in $\mathscr{H}_R$. The vector spaces $\mathscr{H}_L$ and $\mathscr{H}_R$ (which are associated with the same Riemann surface S^2 with opposite orientation) are canonically dual, and the partition function or Feynman path integral $Z(L)$ is equal to the natural pairing

$$Z(L) = (\chi, \psi). \tag{4.8}$$

We cannot evaluate (4.8), since we know neither χ nor ψ. The one thing that we do know, at present, is that (for the groups and representations we are considering) this pairing is occurring in a two dimensional vector space. A two dimensional vector space has the marvelous property that any three vectors obey a relation of

linear dependence. Thus, given any two other vectors ψ_1 and ψ_2 in $\mathscr{H}_R$, there would be a linear relation

$$\alpha\psi + \beta\psi_1 + \gamma\psi_2 = 0, \tag{4.9}$$

where α, β, and γ are complex numbers. Physically, there is a very natural way to get additional vectors in $\mathscr{H}_R$. If one replaces M_R in Fig. 7b by any other three manifold X with the same boundary (and with suitable strings in X connecting the marked points on the boundary of M_R), then the Feynman path integral on X gives rise to a new vector in $\mathscr{H}_R$. Picking any two convenient three manifolds X_1 and X_2 for this computation gives vectors ψ_1 and ψ_2 that can be used in (4.9). We will consider the case in which X_1 and X_2 are the same manifolds as M_R but with different "braids" connecting the points on the boundary; this is indicated in Fig. 7c.

Once ψ_1 and ψ_2 are obtained in this way, (4.9) has the obvious consequence that

$$\alpha(\chi, \psi) + \beta(\chi, \psi_1) + \gamma(\chi, \psi_2) = 0. \tag{4.10}$$

The three terms in (4.10) have a "physical" interpretation, evident in Fig. 7c. By gluing M_L back together with M_R or one of its substitutes X_1 and X_2, one gets back the original three manifold M, but with the original link L replaced by some new links L_1 and L_2. Thus, (4.10) amounts to a relation among the link expectation values of interest, namely

$$\alpha Z(L) + \beta Z(L_1) + \gamma Z(L_2) = 0. \tag{4.11}$$

This recursion relation is often drawn as in Fig. 8. The meaning of this figure is as follows. If one considers three links whose plane projections are identical outside a disc, and look inside this disc like the three drawings in the figure, then the expectation values of those links, weighted with coefficients α, β, and γ, add to zero.

It is well known in knot theory that (4.11) uniquely determines the expectation values of all knots in S^3. For convenience we include a brief explanation of this. One starts with a plane projection of a knot, indicated in Fig. 9. The number p of crossings is finite. Inductively, suppose that all knot expectation values for knots with at most $p-1$ crossings have already been computed. One wishes to study knots with p crossings. If one had $\beta = 0$ in (4.11), one could at each crossing pass the two strands through each other with a factor of $-\gamma/\alpha$ in replacing an over-crossing by an under-crossing. If this were possible, the lines would be effectively transparent, and one could untie all knots. As it is, $\beta \neq 0$, but the term proportional to β reduces the number of crossings, giving rise to a new link whose expectation value is already known by the induction hypothesis.

This process reduces the discussion to the case $p = 0$ where there are no crossings, and therefore we are dealing only with a certain number of unlinked and unknotted circles. For practice with (4.11), let us discuss this case explicitly. In Fig. 10, we sketch a useful special case of Fig. 8. The first and third links in links in Fig. 10 consist of a single unknotted circle, and the second consists of two unlinked and unknotted circles. If we denote the partition function for s unlinked

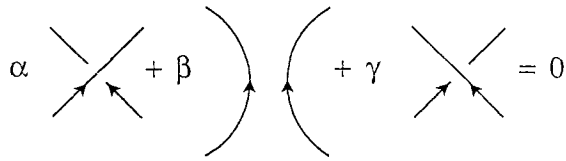

Fig. 8. A recursion relation for links

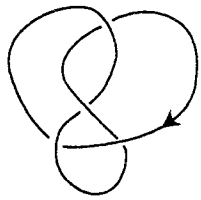

Fig. 9. A plane projection of a knot, with four crossings

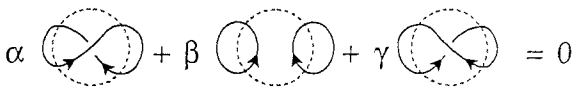

Fig. 10. A special case of the use of Fig. 8. The idea is that the three pictures are identical outside of the dotted lines, and look like figure (4.11) inside them

and unknotted circles in the N dimensional representation of $SU(N)$ as $Z(S^3; Cs)$, then (4.11) amounts in this case to the assertion that

$$(\alpha+\gamma)\, Z(S^3;\, C)+\beta Z(S^3; C^2)=0. \tag{4.12}$$

Together with (4.7),[10] this implies that the expectation value of an unknotted Wilson line in the N dimensional representation of $SU(N)$ is

$$\langle C\rangle = -\frac{\alpha+\gamma}{\beta}. \tag{4.13}$$

Presently we will make this formula completely explicit by computing α, β, and γ in terms of the fundamental quantum field theory parameters N and k.

The induction sketched above expresses any knot expectation value as a rational function of α, β and γ (a ratio of polynomials), after finitely many steps. It is in this sense that the Jones knot invariants and their generalizations are "polynomials". While it is, as we have seen, comparatively elementary to prove that (4.11) uniquely determines the knot invariants, the converse is far less obvious. Equations (4.11) can be used in many different ways to obtain the expectation value of a given link, and one must show that one does not run into

[10] This is the only point at which (4.7) has to be used. The induction sketched in the previous paragraph reduces all computations for knots in S^3 to this special case without using (4.7)

any inconsistency. While this has been proved in a variety of ways, the proofs have not been intrinsically three dimensional – (4.11) has not previously been derived from a manifestly invariant three dimensional framework. This is the novelty of the present discussion.

Change of Framing. We want to compute α, β, and γ, but as a prelude we must discuss a certain technical point. At the end of Sect. 2, we learned that choosing a circle C and a representation R is not enough to give a well defined quantum holonomy operator $W_R(C) = \mathrm{Tr}_R\, P \exp \int_C \mathrm{Ad}\, x$. It is also necessary to pick a "framing" of the circle C, which enters when one has to calculate the self-linking number of C and its non-abelian and quantum generalizations. At the end of Sect. 2, we promised to derive a formula (2.33) showing how any partition function with an insertion of $W_R(C)$ transforms under a change of framing. Now it is time to deliver on this promise.

As in Fig. 7b, let us cut the three manifold M on a Riemann surface Σ that intersects C in a point P (and perhaps in some other points that will not be material). In our previous argument, we used the fact that associated with the boundaries M_L or M_R are Hilbert spaces $\mathscr{H}_L$ and $\mathscr{H}_R$. Moreover, $\mathscr{H}_R$ (for example) is "a flat bundle on moduli space" so the mapping class group of the boundary Σ acts naturally on $\mathscr{H}_R$. We wish to act on the boundary of M_R with a very particular diffeomorphism before gluing the pieces of Fig. 7b back together again. The diffeomorphism that we want to pick is a t-fold "Dehn twist" about the point P on Σ. Making this diffeomorphism and then gluing the pieces of Fig. 7b back together again, one gets an identical looking picture, but the framing of the circle C has been shifted by t units. On the other hand, one knows in conformal field theory how the Dehn twist acts on $\mathscr{H}_R$. Associated with the representation R is a number h_R, the "conformal weight of the primary field in the R representation". The t-fold Dehn twist acts on $\mathscr{H}_R$ as multiplication by $e^{2\pi i t h_R}$. So we have obtained (2.33) with $h = h_R$.

Explicit Evaluation. We will now determine the parameters α, β, and γ that appear in the crucial equation (4.10). We need to determine the explicit relation among the three vectors ψ, ψ_1 and ψ_2 that appear in Fig. 7c. This requires a further study of the two dimensional Hilbert space which arises as the space of conformal blocks for the R, R, $\bar{R}$, $\bar{R}$ four point function on S^2 [R being in this case the defining N dimensional representation of $SU(N)$ and $\bar{R}$ its dual]. The three configurations in Fig. 7c can be regarded as differing from each other by a certain diffeomorphism of S^2; the diffeomorphism in question is the "half-monodromy" under which the two copies of R change places by taking a half-step around one another, as indicated in Fig. 11. Moore and Seiberg call this operation B and study it extensively. The states ψ_1 and ψ_2 are none other than

$$\psi_1 = B\psi, \qquad \psi_2 = B^2\psi\,. \tag{4.14}$$

The matrix B, since it acts in a two dimensional space, obeys a characteristic equation

$$B^2 - yB + z = 0\,, \tag{4.15}$$

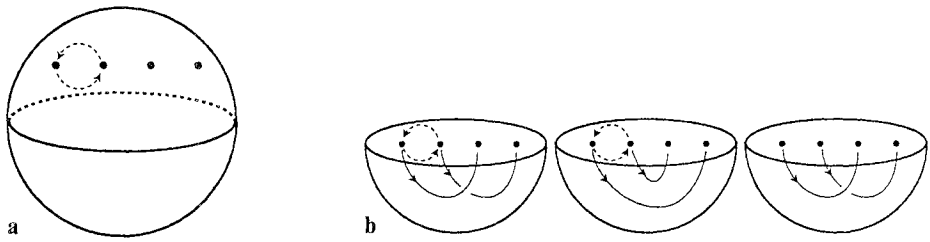

Fig. 11a and b. The half-monodromy operation exchanging two equivalent points on S^2 is sketched in **a**; the arrows are meant to suggest a process in which the first two points change places by executing a half-twist about one another. The idea in **b** is that if the two points on the left in the first picture undergo a half-twist about one another, the first picture becomes the second, and if this is done again, the second picture becomes the third. In this way the three pictures on the right of Fig. 7c differ by a succession of half-monodromies

where
$$y = \operatorname{Tr} B, \qquad z = \det B. \tag{4.16}$$
In view of (4.14), the linear relation among ψ, ψ_1, and ψ_2 is (up to an irrelevant common factor) just
$$z \cdot \psi - y \cdot \psi_1 + \psi_2 = 0, \tag{4.17}$$
and according to (4.16), to make this explicit we need only to know the eigenvalues (and thus the determinant and trace) of B.

These can be obtained from [13], but before describing the formulas, I would like to point out an important subtlety. As we have discussed in the last subsection, all concrete results such as the values of α, β, and γ depend on the framing of knots. The convention that is most natural in working on an arbitrary three manifold is not the convention usually used in discussing knots on S^3.

In studying Fig. 7c, to describe the relative framings, the task is to specify the relative framing of the three pictures on the right, since the picture on the left is being held fixed. If one just looks at these three pictures and ignores the fact that the lines cannot pass through each other, there is an obvious sense in which one would like to pick "the same" framing for each picture; for instance, a unit vector coming out of the page defines a normal vector field on each link in the picture.

This is equivalent to the convention of Moore and Seiberg in defining the eigenvalues of B, so we can now quote their results. Let h_R be the conformal weight of a primary conformal field transforming as R, let E_i be the irreducible representations of $SU(N)$ appearing in the decomposition of $R \otimes R$, and let h_{E_i} be the weights of the corresponding primary fields. Then the eigenvalues of B are
$$\lambda_i = \pm \exp\left(i\pi\left(2h_R - h_{E_i}\right)\right), \tag{4.18}$$
where the $+$ or $-$ sign corresponds to whether E_i appears symmetrically or antisymmetrically in $R \otimes R$. If R is the N dimensional representation of $SU(N)$, then one finds [11] that the eigenvalues of B are
$$\lambda_1 = \exp\left(\frac{i\pi(-N+1)}{N(N+k)}\right), \qquad \lambda_2 = -\exp\left(\frac{i\pi(N+1)}{N(N+k)}\right). \tag{4.19}$$

[11] For this representation, $h_R = (N^2-1)/(2N(N+k))$. In the decomposition of $R \otimes R$, the symmetric piece is an irreducible representation with $h_{E_1} = (N^2+N-2)/N(N+k)$, and the antisymmetric piece is an irreducible representation with $h_{E_2} = (N^2-N-2)/N(N+k)$

It is straightforward to put these formulas in (4.16), (4.17) and thus make our previous results completely explicit.

Before comparing to the knot theory literature, it is necessary to make a correction in these results. For a link in S^3, there is always a standard framing in which the self-intersection number of each component of the link is zero. Values of the knot polynomials for knots in S^3 are usually quoted without specifying a framing; these are the values for the link with standard framing. However, if on the right of Fig. 7c we use the "same" framing for each picture, then when the right of Fig. 7c is glued to the left, one does not have the canonical framing for each link. If the first knot is framed in the standard fashion, then the second is in error by one unit and the third by two units. So after using (4.19) to compute α, β, γ, we must, if we wish to agree with the knot theory literature, multiply β by $\exp(-2\pi i h_R)$ and γ by $\exp(-4\pi i h_R)$. After these corrections, one gets

$$\begin{aligned}
\alpha &= -\exp\left(\frac{2\pi i}{N(N+k)}\right),\\
\beta &= -\exp\left(\frac{i\pi(2-N-N^2)}{N(N+k)}\right)+\exp\left(\frac{i\pi(2+N-N^2)}{N(N+k)}\right),\\
\gamma &= \exp\left(\frac{2\pi i(1-N^2)}{N(N+k)}\right).
\end{aligned} \tag{4.20}$$

If one multiplies α, β, γ by an irrelevant common factor $\exp(i\pi(N^2-2)/N(N+k))$ and introduces the variable

$$q=\exp(2\pi i/(N+k)), \tag{4.21}$$

then the skein relation can be written more elegantly as

$$-q^{N/2}L_+ + (q^{1/2}-q^{-1/2})\,L_0 + q^{-N/2}L_- = 0. \tag{4.22}$$

Here L_+, L_0, and L_- [equivalent to L, L_1, and L_2 in (4.11)] are standard notation for overcrossing, zero crossing, and undercrossing; and for $i=+,0,-$, we now write simply L_i, instead of $Z(L_i)$. Equation (4.22) is correctly normalized to give the right answers for knots on S^3 with their standard framing, and if one is only interested in knots on S^3 one can use it without ever thinking about the framings. Finally, comparing (4.13) and (4.22), we see that the expectation value of an unknotted Wilson line on S^3, with its standard framing, is

$$\langle C\rangle = \frac{q^{N/2}-q^{-N/2}}{q^{1/2}-q^{-1/2}}. \tag{4.23}$$

This formula can be subjected to several interesting checks. First of all, the right-hand side of (4.23) is positive for all values of the positive integers N and k. This is required by reflection positivity of the Chern-Simons gauge theory in three dimensions. Second, in the weak coupling limit of $k\to\infty$, we have $\langle C\rangle\to N$. This is easily interpreted; in the weak coupling limit, the fluctuations in the connection A_i on S^3 are irrelevant, and the expectation value of the Wilson line approaches its value for $A_i=0$, which is the dimension of the representation, or in this case N.

4.2. Surgery on Links

We have seen that it is possible to effectively calculate the expectation value of an arbitrary link in S^3. We would now like to generalize this to computations on an arbitrary three manifold. The basic idea is that by the operation of "surgery on links" any three manifold can be reduced to S^3, so it is enough to understand how the invariants that we are studying transform under surgery. The operation of surgery can be described as follows. One begins with a three manifold M and an arbitrarily selected embedded circle C. Note that there is, to begin with, no Wilson line associated with C; C is simply a mathematical line on which we are going to carry out "surgery". To do so we first thicken C to a "tubular neighborhood", a solid torus centered on C. Removing this solid torus, M is split into two pieces; the solid torus is called M_R in Fig. 12b, and the remainder is called M_L. One then makes a diffeomorphism on the boundary of M_R and glues M_L and M_R back together to get a new three manifold $\tilde{M}$.

It is a not too deep result that every three manifold can be obtained from or reduced to S^3 (or any other desired three manifold) by repeated surgeries on knots. However, such a description is far from unique and it is often difficult to use a description of a three manifold in terms of surgery to compute the invariants of interest. We will now see that the invariants studied in this paper can be effectively computed from a surgery presentation.

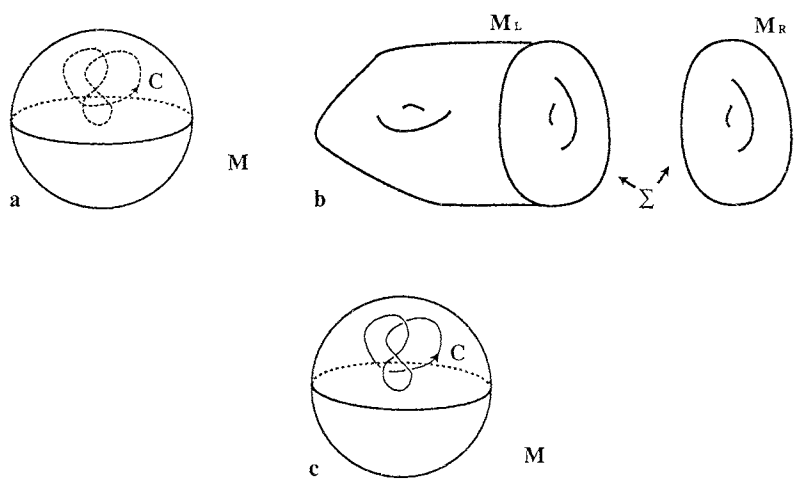

Fig. 12a–c. Surgery on a circle C in a three manifold M is carried out by removing a tubular neighborhood of C, depicted in **a**. At this point M has been separated into two pieces, M_L and M_R, with a torus Σ for their boundaries, as sketched in **b**. M_R is simply a solid torus. Surgery is completed by gluing the two pieces back together after making a diffeomorphism of the boundary of M_R. At the end of this process, M has been replaced by a new three manifold $\tilde{M}$. As we will eventually see, computations on $\tilde{M}$ are equivalent to computations on M with a physical Wilson line where the surgery was made, as in **c**. The difference between **a** and **c** is that in **a** the circle C is just a locale for surgery, but in **c** it is a Wilson line

We study Fig. 12b by the standard arguments. Hilbert spaces $\mathscr{H}_L$ and $\mathscr{H}_R$, canonically dual to one another, are associated with the boundaries of M_L and M_R. The path integrals on M_L and M_R give vectors ψ and χ in $\mathscr{H}_L$ and $\mathscr{H}_R$, and the partition function on M is just the natural pairing (ψ, χ). If we act on the boundary of M_R with a diffeomorphism K before gluing M_L and M_R back together, then χ is replaced by $K\chi$ so (ψ, χ) is replaced by $(\psi, K\chi)$.

This potentially gives a way to determine how the partition function of the quantum field theory transforms under surgery. Upon gaining a suitable understanding of $K\chi$, we will be able to reduce calculations on $\tilde{M}$ to calculations on M.

4.3. The Physical Hilbert Space in Genus One

At this point we need a description of the physical Hilbert space in genus one. A beautiful description, perfectly adapted for our needs, appears in the work of Verlinde [12].

First of all, the loop group LG has at level k finitely many integrable highest weight representations. Let t be the number of these. For each such highest weight representation of the loop group, the highest weight space is an irreducible representation of the finite dimensional group G. In this way there appear t distinguished representations of G; we label these as $R_0, R_1 \ldots R_{t-1}$, with R_0 denoting the trivial representation (which is always one of those on this list). Verlinde showed that if Σ is a Riemann surface of genus one, then the dimension of the physical Hilbert space $\mathscr{H}_\Sigma$ is t. Moreover, though there is no canonical basis for $\mathscr{H}_\Sigma$, Verlinde showed that every choice of a homology basis for $H^1(\Sigma, \mathbf{Z})$, consisting of two cycles a and b, gives a canonical choice of basis in $\mathscr{H}_\Sigma$. For our purposes, this can be described as follows. Topologically, there are many inequivalent ways to identify a torus Σ as the boundary of a solid torus U. The choice of U can be fixed by requiring that the cycle a is contractible in U. This is indicated in Fig. 13a. Next, for every $i = 0 \ldots t-1$, one defines a state v_i in $\mathscr{H}_\Sigma$ as follows. One places a Wilson line in the R_i representation in the interior of U,

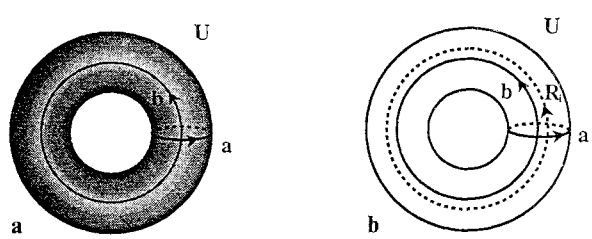

Fig. 13a and b. A Riemann surface Σ of genus one is shown in **a** as the boundary of a solid torus U; the indicated a-cycle is contractible in U. In **b**, a basis of the physical Hilbert space is indicated consisting of states obtained by placing a Wilson line, in the R_i representation, in the interior of U, parallel to the cycle b, and performing the path integral to get a vector v_i in $\mathscr{H}_\Sigma$

running in the b direction,[12] and one performs the Feynman path integral in U to define a vector v_i in $\mathscr{H}_\Sigma$. The v_i make up the Verlinde basis in $\mathscr{H}_\Sigma$. It must be understood that a Wilson line in the trivial representation is equal to 1, so the vector v_0 obtained by this definition is the same as the vector χ which in the last subsection was obtained by a path integral on U with no Wilson lines:

$$\chi = v_0 . \tag{4.24}$$

A diffeomorphism K of Σ is represented in the Verlinde basis by an explicit matrix K_i^j, defined by the formula

$$K \cdot v_i = \sum_j K_i^j v_j . \tag{4.25}$$

In the space spanned by the v_i, there is a natural inner product, defined by the tensor g_{ij} which is one if R_i is the dual of R_j and zero otherwise. We may sometimes use this metric to raise and lower indices, letting $K_{ij} = \sum_m g_{mj} K_i^m$.

We can now get a much more concrete description of the behavior of the quantum field theory partition function under surgery. In discussing surgery, we began with a three manifold M and a knot C. Cutting out a tubular neighborhood of this knot, whose boundary we call Σ, we separated M into M_L and M_R, with M_R being a solid torus. The path integrals on M_L and M_R gave vectors ψ and χ in the Hilbert spaces $\mathscr{H}_L$ and $\mathscr{H}_R$. As we have just noted, χ is the same as v_0, so the partition function on M is (ψ, v_0). Now we want to make a diffeomorphism K on the boundary of M_R, and then glue together M_L and M_R to make a new three manifold $\tilde{M}$. The partition function of $\tilde{M}$ is $Z(\tilde{M}) = (\psi, Kv_0)$, as we saw in the last section. To say that $Z(\tilde{M})$ is computable from a surgery presentation means that the evaluation of this invariant of $\tilde{M}$ can be reduced to tractable calculation on M. We will now show this. From (4.25), we can write

$$Z(\tilde{M}) = \sum_j K_0{}^j (\psi, v_j) . \tag{4.26}$$

But each term (ψ, v_j) has an interpretation in terms of path integrals on M! Indeed, it is the very definition of the v_j that v_j differs from v_0 just by an insertion of an extra Wilson line in the R_j representation at the center of M_R. So just as (ψ, v_0) represents the original partition function of M, (ψ, v_j) represents a modified partition function with an extra Wilson line in the R_j representation placed on C. So we rewrite (4.26) in the form

$$Z(\tilde{M}) = \sum_j K_0{}^j \cdot Z(M;\, R_j) , \tag{4.27}$$

where $Z(M;\, R_j)$ is the partition function of M with an extra Wilson line in the R_j representation included on the circle C (in addition to whatever Wilson lines are already present on M). This is indicated in Fig. 12c. To use (4.27), one needs to know the matrix K_i^j, which is precisely the matrix by which the diffeomorphism K of the torus is represented on the characters of the irreducible level k representations of LG; these matrices appear in [37] and have remarkable properties

[12] The b cycle on the boundary of U gives a framing of this Wilson line

recently investigated in [12, 13]. Given a knowledge of the K_i^j, (4.27) is a completely explicit formula expressing computations on $\tilde{M}$ in terms of computations on M. By repeated use of this formula, computations on any three manifold can be reduced to computations on S^3, [13] with appropriate Wilson lines. Of course, the surgery will generate Wilson lines on S^3 in representations of G corresponding to arbitrary integrable representations of LG.

Generalized Surgery. The surgery law (4.27) has a useful generalization. While so far we have only considered surgery on a purely imaginary circle C, as in Fig. 12, there is no reason not to generalize this to a situation in which before the surgery a Wilson line in the R_i representation was already present on C. Surgery amounts to cutting out a neighborhood of C and then gluing it back in, and after this process the R_i Wilson line will still be present in $\tilde{M}$. So the left-hand side of (4.27) is replaced by $Z(\tilde{M}; R_i)$, where the notation schematically indicates the presence of the R_i Wilson line. What about the right-hand side of (4.27)? Before surgery, with a Wilson line R_i on C, the path integral on a tubular neighborhood U of C gives on the boundary a state $\chi' = v_i$; this is the generalization of (4.24). If we cut out C and glue it back in with a diffeomorphism K of the boundary, then v_i is replaced according to (4.25) with $K_i^j v_j$. If we remember that v_j could have been obtained by putting a Wilson line on C in the R_j representation, we see that the right-hand side of (4.27) becomes $\sum_j K_i^j Z(M; R_j)$, so we get the generalized surgery formula

$$Z(\tilde{M}; R_j) = \sum_j K_i^j Z(M; R_j). \tag{4.28}$$

This formula will be used later in a new proof of Verlinde's conjecture.

Cabling of Knots; Satellites. Finally, let us note that similar methods can be used to determine the behavior of the knot invariants under "cabling", and more generally to relate the invariants of the "satellites" of a knot to invariants of the original knot. Any knot C in any three manifold M has a neighborhood that looks like a solid torus U. If we replace C by an arbitrary satellite $\tilde{C}$ of itself (an arbitrary knot that can be placed in U), with representations R_σ associated with the connected components of $\tilde{C}$, then the path integral on U will define a vector ψ in the physical Hilbert space $\mathscr{H}_\Sigma$ associated with the boundary Σ of U. Like any vector in $\mathscr{H}_\Sigma$, ψ can be expanded in the Verlinde basis,

$$\psi = \sum_i \alpha_i v_i. \tag{4.29}$$

The α_i are complex numbers that depend on the choice of satellite $\tilde{C}$ and on the choice of representations R_σ, but they do not depend on what three manifold M the solid torus U has been extracted from or on what other knots are present on M. The vectors $v_i \in \mathscr{H}_\Sigma$ are the vectors that would be produced by the path integral on U with a Wilson line in the R_i representation placed on the original knot C (and not a satellite of C). Thus, a knowledge of the invariants of C in arbitrary representations together with a knowledge of the universal coefficients α_i is enough to determine the invariants of arbitrary satellites of C.

[13] Or on any desired three manifold; we will see that $S^2 \times S^1$ is more tractable than S^3

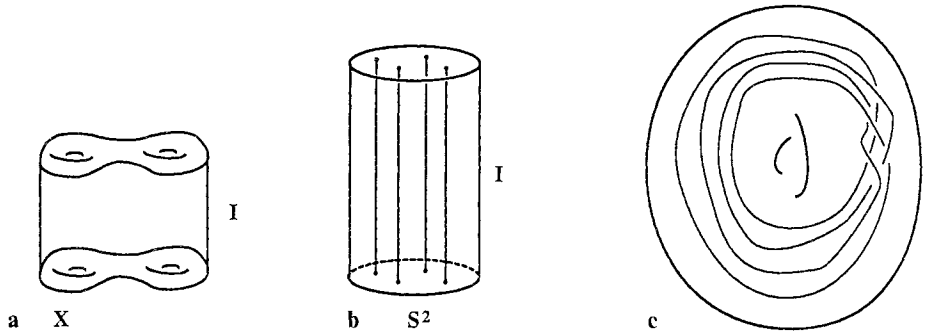

Fig. 14a–c. Beginning with $X \times I$, shown in **a**, one makes $X \times S^1$ by identifying $X \times \{0\}$ with $X \times \{1\}$. If X is S^2 with some marked points P_i, then this construction gives the picture of **b**. If $S^2 \times \{0\}$ is joined to $S^2 \times \{1\}$ via a non-trivial diffeomorphism B, one makes in this way a braid, as in **c**

4.4. Path Integrals on $S^1 \times X$

In this subsection we will describe a few facts which are useful in their own right and will enable us to carry out some concrete surgeries.

First of all, we have not so far determined the partition function of S^3 without Wilson lines. It may come as a surprise to topologists that we cannot trivially assert that this is 1. In quantum field theory there is no particularly strong axiom governing the partition function of S^3. The three manifolds whose partition functions can be computed in a particularly simple way, from the axioms of quantum field theory, are those of the form $X \times S^1$, for various X. $X \times S^1$ can conveniently be studied in a "Hamiltonian" formalism, as indicated in Fig. 14a. One constructs the Hilbert space $\mathscr{H}_X$ of X. Then one introduces a "time" direction, represented by a unit interval $I = [0, 1]$, and one propagates the vectors in $\mathscr{H}_X$ from "time" 0 to "time" 1. This operation is trivial, since the Chern-Simons theory, like any generally covariant theory, has a vanishing Hamiltonian. Finally, one forms $X \times S^1$ by gluing $X \times \{0\}$ to $X \times \{1\}$; this identifies the initial and final states, giving a trace:

$$Z(X \times S^1) = \mathrm{Tr}_{\mathscr{H}_X}(1) = \dim \mathscr{H}_X . \tag{4.30}$$

For example, the physical Hilbert space of S^2 is one dimensional, for any G and k, so one has

$$Z(S^2 \times S^1) = 1 . \tag{4.31}$$

It is possible to generalize (4.30) as follows. If we are given a diffeomorphism K: $X \to X$, then one can form the mapping cylinder $X \times_K S^1$ by identifying $x \times \{1\}$ with $K(x) \times \{0\}$ for every $x \in X$. At the level of quantum field theory, when one goes from $X \times I$ to $X \times_K S^1$, the initial and final states are identified via K, so the generalization of (4.30) is

$$Z(X \times_K S^1) = \mathrm{Tr}_{\mathscr{H}_X}(K) . \tag{4.32}$$

The situation that we actually wish to apply this to is the case in which X is S^2 with some marked points P_a, $a = 1 \ldots s$ to which representations $R_{i(a)}$ are assigned.

[For $a = 1 \ldots s$, $i(a)$ is one of the values $0 \ldots t-1$ corresponding to integrable level k representations of the loop group.] In this case, the simple product $X \times S^1$ is just $S^2 \times S^1$ with some Wilson lines which are unknotted, parallel circles of the form $\{P_a\} \times S^1$, as sketched in Fig. 14b. To determine the path integral on $S^2 \times S^1$ in the presence of these Wilson lines, which we will denote as $Z(S^2 \times S^1; \langle R \rangle)$, one needs to study the Hilbert space of S^2 with charges in the representations R_{a_i}; we will denote this as $\mathscr{H}_{S^2;\langle R \rangle}$. The analog of (4.31) is then

$$Z(S^2 \times S^1; \langle R \rangle) = \dim \mathscr{H}_{S^2;\langle R \rangle}. \tag{4.33}$$

The dimensions of these spaces were discussed at the end of Sect. 3. Thus, if the collection of representations $\langle R \rangle$ consists of a single representation R_a, we get

$$Z(S^2 \times S^1; R_a) = \delta_{a,0}, \tag{4.34}$$

since the physical Hilbert space with a single charge in the R_a representation is one dimensional if R_a is the trivial representation ($a = 0$) and zero dimensional otherwise.[14] For two charges in the representations R_a and R_b, we get

$$Z(S^2 \times S^1; R_a, R_b) = g_{ab}, \tag{4.35}$$

where g_{ab}, introduced earlier, is 1 if R_b is the dual of R_a and zero otherwise. The formula (4.35) follows from the result of Sect. (4.4) for the Hilbert space on S^2 with two charges. Finally, if there are three charges in the representations R_a, R_b, R_c, we get

$$Z(S^2 \times S^1; R_a, R_b, R_c) = N_{abc}, \tag{4.36}$$

with N_{abc} the trilinear "coupling" of Verlinde, since this is the dimension of the physical Hilbert space.

4.5. Some Concrete Surgeries

Now we would like to describe some useful results that can be obtained from concrete surgeries. The first goal is compute the partition function of S^3. Since we already know the partition function of $S^2 \times S^1$, we will try to interpret S^3 as a manifold obtained by surgery on $S^2 \times S^1$. This is readily done. We consider the circle C in $S^2 \times S^1$ indicated in Fig. 15a. A tubular neighborhood of C is a torus Σ; we pick a basis of $H^1(\Sigma; \mathbf{Z})$ consisting of cycles a and b indicated in the figure. Now we wish to make a particular surgery associated with a very special diffeomorphism S: $\Sigma \to \Sigma$. We pick S to map a to b and b to $-a$.[15] This surgery – removing the interior of Σ from $S^2 \times S^1$ and gluing it back after acting with S – produces a three manifold that is none other than S^3 (Fig. 16). Since this point is crucial in what follows, we pause to explain it. We regard S^3 as R^3 plus a point at infinity. In Fig. 16a a torus Σ has been embedded in R^3. Obviously, Σ with its

[14] Implicit in (4.34) and subsequent formulas is the use of the standard framing of the Wilson line which is invariant under rotations of S^1; it is for this choice that the path integral on $S^2 \times S^1$ computes the trace of the identity operator in the physical Hilbert space

[15] This transformation, which acts on the upper half plane as $\tau \to -1/\tau$, is indeed usually called S in the theory of the modular group $SL(2, \mathbf{Z})$ (which can be identified, via the basis a, b, with the mapping class group of Σ)

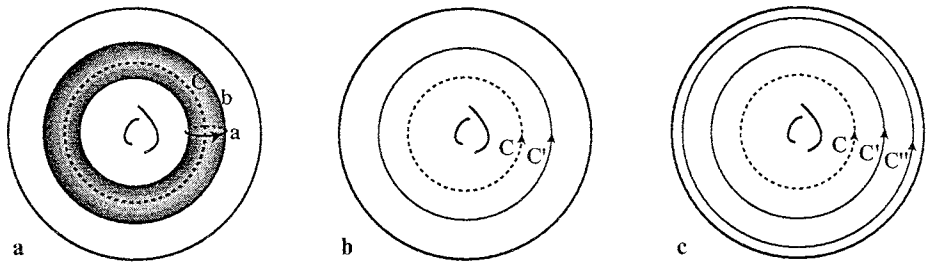

Fig. 15a–c. In part **a**, we consider surgery on a circle C in $S^2 \times S^1$. A tubular neighborhood of this circle is a torus Σ; a useful basis of $H^1(\Sigma, \mathbf{Z})$ is indicated. In **b**, in addition to the circle C on which we perform surgery, there is a parallel circle C' on which we place a Wilson line in the R_j representation. In **c**, there are two parallel circles C' and C'' with Wilson lines in representations R_j and R_k

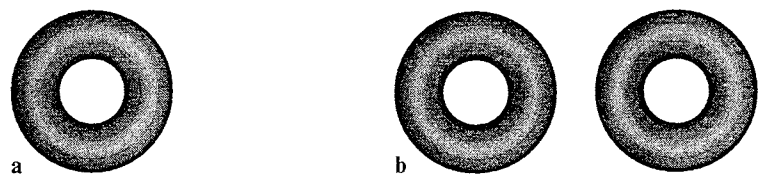

Fig. 16a and b. The purpose of this figure is to indicate how S^3 can be made by surgery starting with $S^2 \times S^1$. In **a** we show a torus Σ, sitting in R^3 and in **b** a pair of identical solid tori

interior make up a solid torus T. It is also relatively easy to see that the figure (a) is invariant under inversion, so that the exterior of Σ (including the point at infinity) is a second solid torus T'. Thus, S^3 can be made by gluing two solid tori along their boundaries. Now in (b) we sketch two identical solid tori T and T'; T' has been obtained by simply translating T in Euclidean space. If one glues together the boundaries pointwise with the identification that is indicated by saying that "T' is a translate of T", one gets $S^2 \times S^1$. (In fact, the solid torus T is $D \times S^1$, with D a two dimensional disc, and T' is $D' \times S^1$ with D' a second disc. Just as two discs D and D' glued on their boundary make S^2, $D \times S^1$ naturally glues to $D' \times S^1$ to make $S^2 \times S^1$.) On the other hand, we know from part (a) that S^3 can be obtained by gluing two solid tori. A little mental gymnastics, comparing the argument we gave in connection with (a) to that in (b), shows that to make S^3 we must glue together T and T' after making the modular transformation S on the boundary of T'.

Now we can use (4.27), with S^3 playing the role of $\tilde{M}$, $S^2 \times S^1$ playing the role of M, and the arbitrary diffeomorphism K replaced by S. So we learn

$$Z(S^3) = \sum_j S_0{}^j \, Z(S^2 \times S^1;\, R_j)\,. \tag{4.37}$$

We have learned in the last section that $Z(S^2 \times S^1;\, R_j)$ is 1 for $j = 0$ and 0 otherwise, so

$$Z(S^3) = S_{0,0}\,. \tag{4.38}$$

Here $S_{0,0}$ can be determined from the theory of affine Lie algebras; for $G = SU(2)$ one gets the formula stated earlier in (2.26). In fact, the whole matrix S_{ij} can be written very explicitly for $G = SU(2)$. The integrable representations of level k are those of spin $n/2$ for $n = 0 \ldots k$, and the matrix elements of S are

$$S_{mn} = \sqrt{\frac{2}{k+2}} \sin\left(\frac{(m+1)(n+1)\pi}{k+2}\right). \tag{4.39}$$

The Phase of the Partition Function. Now let us re-examine in the light of these methods a thorny question that appeared in Sect. 2 – the framing of three manifolds, and the phase of the partition function.

We have obtained S^3 from $S^2 \times S^1$, by performing surgery on a certain circle C, using the modular transformation S: $\tau \to -1/\tau$. Apart from S, there are other modular transformations that could be used to build S^3 by surgery on the same knot C in $S^2 \times S^1$. The general choice would be $T^n S T^m$, with n and m being arbitrary integers, and T being the modular transformation T: $\tau \to \tau + 1$. [S and T are the standard generators of the modular group, obeying $S^2 = (ST)^3 = 1$.] Had we used $T^n S T^m$, we would have gotten not (4.38) but

$$Z(S^3) = (T^n S T^m)_{0,0}. \tag{4.40}$$

This may readily be evaluated. In the Verlinde basis, T is a diagonal matrix with $T \cdot v_i = e^{2\pi i(h_i - c/24)} \cdot v_i$; h_i is the conformal weight of the primary field in the representation R_i and c is the central charge for current algebra with symmetry group G at level k. Since $h_0 = 0$, if we replace (4.38) by (4.40) the partition function transforms as

$$Z \to Z \cdot \exp\left(2\pi i(n-m) \cdot \frac{c}{24}\right). \tag{4.41}$$

Though we have obtained this formula in the example of a particular surgery (giving S^3 from $S^2 \times S^1$), the same ambiguity arises in any process of surgery. Whenever one makes surgery on a circle C, in a three manifold M, with the surgery being determined by an $SL(2, Z)$ element u, one could instead consider surgery on the same circle C, using the $SL(2, Z)$ element $u \cdot T^m$. This would have the same effect topologically, but our surgery law would give a partition function containing an extra phase $\exp(-2\pi i m \cdot c/24)$.

This phase ambiguity was already encountered, in the large k limit, in formula (2.24). What is more, from the discussion in Sect. 2, we know what topological structure on three manifolds must be considered in order to keep track of the factors of $\exp(2\pi i \cdot c/24)$. One must consider "framed" three manifolds. Two surgeries that have the same effect on the topology of a three manifold may have different effects on the framing. I will discuss elsewhere how to systematically keep track of the factors of $\exp(2\pi i \cdot c/24)$ under surgery. In the simple applications in this paper, this will not be necessary. All of our applications will involve considering the standard surgery (by the modular transformation S) that was used in the last subsection to obtain S^3 from $S^2 \times S^1$.

Some Expectation Values. Now let us see if we can go farther and determine the path integral $Z(S^3; R_j)$ on S^3 with an unknotted Wilson line on S^3 in an arbitrary

representation R_j.[16] To do this, we start on $S^2 \times S^1$ with a Wilson line in the R_j representation running parallel to the circle C on which we are doing surgery, as in Fig. 15b. Carrying out the same surgery as before turns $S^2 \times S^1$ into S^3, with a Wilson line in the R_j representation on S^3. Application of (4.27) now gives

$$Z(S^3; R_j) = \sum_i S_0{}^i Z(S^2 \times S^1; R_i, R_j). \tag{4.42}$$

Using (4.35), we can evaluate this and determine the partition function for a Wilson line in an arbitrary representation R_j; it is

$$Z(S^3; R_j) = \sum_i S_0{}^i g_{ij} = S_{0,j}. \tag{4.43}$$

Let us compare this to our previous evaluation (4.23) of the expectation value of an unknotted Wilson line in S^3. We must recall that the symbol $\langle C \rangle$ in (4.23) represented a ratio $\langle C \rangle = Z(S^3; R)/Z(S^3)$. Let us take $G = SU(2)$, so that we can use the explicit formulas (4.39), and take R to be the two dimensional representation of $SU(2)$, so that we can compare to (4.23). Using (4.38), (4.43), and (4.39), we get

$$\langle C \rangle = \frac{S_{0,1}}{S_{0,0}} = \frac{\sin(2\pi/(k+2))}{\sin(\pi/(k+2))}. \tag{4.44}$$

It is easy to see that setting $N = 2$ in (4.23) gives the same formula. Let us take this one step further and try to calculate by these methods the partition function $Z(S^3; R_j, R_k)$ for S^3 with two unknotted, unlinked Wilson lines in representations R_j and R_k. In Fig. 15c, we start on $S^2 \times S^1$ with *two* Wilson lines, in representations R_j and R_k, parallel to the circle C on which surgery is to be performed. Carrying out the surgery, we get to S^3 with the desired unlinked, unknotted circles. In this case, the surgery formula (4.27) tells us that

$$Z(S^3; R_j, R_k) = \sum_i S_0^i Z(S^2 \times S^1; R_i, R_j, R_k). \tag{4.45}$$

The right-hand side can be evaluated with (4.36), while the left-hand side can be reduced to (4.43) using (4.6). We get

$$\frac{S_{0,j} S_{0,k}}{S_{0,0}} = \sum_i S_0^i N_{ijk}. \tag{4.46}$$

Proof of Verlinde's Conjecture. The last equation is a special case of a celebrated conjecture by Verlinde, which has been proved by Moore and Seiberg [13]. We can use these methods to give a new proof of Verlinde's conjecture, in the case of current algebra. We will have to use the generalized surgery relation (4.28). We return to Fig. 15b but now instead of treating C as a purely imaginary contour on which surgery is to be performed, we suppose that there is a Wilson line on C in the R_i representation. In this case, the standard surgery on C will still turn $S^2 \times S^1$ into S^3, but now on S^3 we will have two Wilson lines, in the R_i and R_j representations. Some mental gymnastics shows that they are linked, as in Fig. 17a; schematically,

[16] We give this Wilson line the framing described in the footnote after (4.34); after surgery this turns into the standard framing on S^3

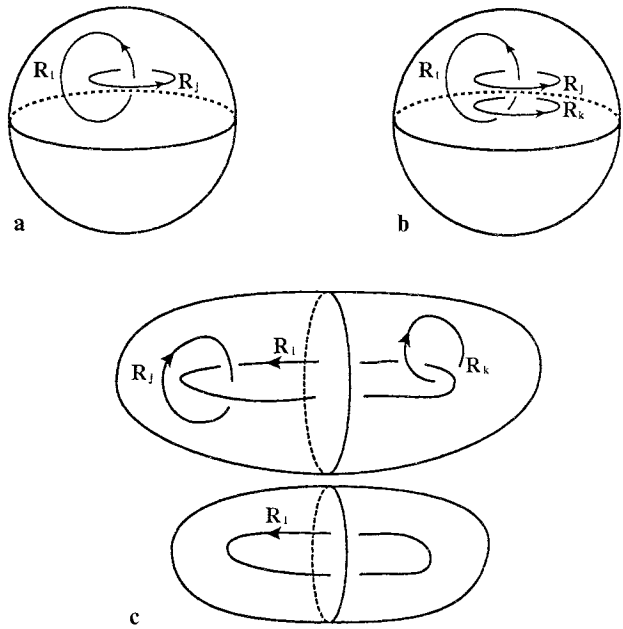

Fig. 17a–c. In **a**, we sketch linked but unknotted Wilson lines on S^3, in the R_i and R_j representations. In **b**, a Wilson line R_i is linked with two Wilson lines R_j and R_k on S^3. In **c**, we sketch how two crucial amplitudes can be factored through the same one dimensional space

we refer to this linked pair of Wilson lines as $L(R_i; R_j)$. The use of (4.28) therefore determines the partition function of S^3 with a pair of linked Wilson lines:

$$Z(S^3;\, L(R_i; R_j)) = \sum_k S_i^k Z(S^2 \times S^1;\, R_k,\, R_j) = S_{ij}\,. \tag{4.47}$$

In the second step, we have used the fact that the partition function of $S^2 \times S^1$ with static charges R_k and R_j is the metric that we have called g_{kj} – one if R_k is dual to R_j and otherwise zero.

Now let us go back to Fig. 15c, and again on what was previously the purely imaginary circle C we put a Wilson line in the R_i representation. The standard surgery on this link will now produce a picture sketched in Fig. 17b, with a Wilson line on S^3 in the R_i representation that links a pair of Wilson lines R_j, R_k that are themselves unlinked and unknotted. We call this configuration $L(R_i; R_j, R_k)$. The evaluation of (4.28) now gives

$$Z(S^3; L(R_i;\, R_j,\, R_k)) = \sum_m S_i^m Z(S^2 \times S^1;\, R_m, R_j, R_k) = \sum_m S_i^m N_{mjk}\,. \tag{4.48}$$

To obtain Verlinde's formula, it is now necessary to find an independent way to evaluate the left-hand side.

Such a method is provided by the following generalization of the multiplicativity formula (4.1). The key point in the derivation of (4.1) was that the physical Hilbert space for S^2 with no charges was one dimensional. It is likewise true that the physical Hilbert space $\mathscr{H}$ for S^2 with a pair of charges in the dual

representations R_i and $R_{\bar{i}}$ is one dimensional. Using this and otherwise repeating the derivation of (4.1) gives the following formula:

$$Z(S^3; L(R_i; R_j, R_k)) \cdot Z(S^3; R_i) = Z(S^3; L(R_i; R_j)) \cdot Z(S^3; L(R_i; R_k)). \quad (4.49)$$

The idea in (4.49) is that, as in Fig. 17c, the evaluation of $Z(S^3; L(R_i; R_j, R_k))$ can be expressed as a pairing (ψ, χ) where ψ and χ are certain vectors in $\mathscr{H}$ and its dual. Likewise the evaluation of $Z(S^3; R_i)$ is a pairing (v', v), where v' and v are vectors in $\mathscr{H}$ and its dual. Using the wonderful fact of one dimensional linear algebra $(\psi, \chi) \cdot (v', v) = (\psi, v) \cdot (v', \chi)$, we arrive at (4.49). Since all factors in (4.49) are known except the first, we arrive at the result

$$Z(S^3; L(R_i; R_j, R_k)) = S_{ij} S_{ik} / S_{0,i}. \quad (4.50)$$

Combining this with (4.48), we have

$$S_{ij} S_{ik} / S_{0,i} = \sum_m S_i^m N_{mjk}. \quad (4.51)$$

This is equivalent to Verlinde's statement that "the matrix S diagonalizes the fusion rules". In other words, in the basis v_i indicated in Fig. 13, the structure constants of the Verlinde algebra are by definition $v_i v_j = \sum_k N_{ij}{}^k v_k$, where $N_{ij}{}^k = \sum_r N_{ijr} g^{rk}$. If we introduce a new basis $w_i = S_{0,i} \cdot \sum S_i^m v_m$, then the Verlinde algebra reduces to $w_i w_j = \delta_{ij} w_j$. To verify this, we compute

$$w_i w_j = \sum_{k,l} S_i^k S_j^l v_k v_l \cdot S_{0,i} S_{0,j}. \quad (4.52)$$

Using $v_k v_l = \sum_m N_{kl}{}^m v_m$ and (4.51), this becomes

$$w_i w_j = S_j^l v^m \cdot S_{il} S_i^m \cdot S_{0,j}. \quad (4.53)$$

Using the unitarity of S, in the form $S_j^l S_{il} = \delta_{ij}$, we see that

$$w_i w_j = \delta_{ij} \sum_m S_j^m v_m \cdot S_{0,j} = \delta_{ij} \cdot w_j, \quad (4.54)$$

showing that the Verlinde algebra has been diagonalized and that the w_i are idempotents.

Connected Sum of Links. At the beginning of Sect. 4, we have seen that the quantum partition function has a multiplicative behavior under connected sum of three manifolds. From (4.1), if $M = M_1 + M_2$, then $Z(M) \cdot Z(S^3) = Z(M_1) \cdot Z(M_2)$. In the special case that M_1 and M_2 are copies of S^3 with links in them, the connected sum of M_1 and M_2 is a copy of S^3 containing the *disconnected* sum of the two links.

In knot theory there is also an operation of taking the *connected* sum of two links. This operation has appeared in the above discussion. The link that we have called $L(R_i; R_j, R_k)$ is the connected sum of the two links that we have called $L(R_i, R_j)$ and $L(R_i, R_k)$. In fact, in Fig. 17c, $L(R_i; R_j, R_k)$ is "cut" into two pieces, which are respectively $L(R_i, R_j)$ and $L(R_i, R_k)$ with in each case a connected segment removed. This is the defining configuration for the "connected sum of links". Accordingly, the reasoning that led to (4.49) has the following more

general consequence. If L_1 and L_2 are two links, and $L_1 + L_2$ is their connected sum, then

$$Z(S^3; L_1 + L_2) \cdot Z(S^3; C) = Z(S^3; L_1) \cdot Z(S^3; L_2). \tag{4.55}$$

Here it is understood that [as in (4.49)] representations have been assigned to the connected components of L_1, L_2, and $L_1 + L_2$ in a compatible fashion; C is an unknot placed in whatever representation is carried by the strand "cut" in the generalization of Fig. 17c. Equation (4.55) has a generalization in which L_1 and L_2 are links in arbitrary three manifolds M_1 and M_2; then the connected sum of links $L_1 + L_2$ is a link in the connected sum of manifolds $M = M_1 + M_2$, and (4.55) is replaced by

$$Z(M_1 + M_2; L_1 + L_2) \cdot Z(S^3; C) = Z(M_1; L_1) \cdot Z(M_2; L_2). \tag{4.56}$$

4.6. The Knot Polynomials and the Braid Group

The results in the last subsection are nice enough so that one may wonder if the partition function for an arbitrary link on S^3 can be evaluated in this way. This can indeed be done, and in a way that is closely related to the original route by which the Jones polynomial was discovered, though we cannot expect such explicit formulas as in the simple cases treated above. An arbitrary link L on S^3, whose partition function we will call $Z(S^3; L)$, can be arranged in the form of a braid, as indicated in Fig. 18a. One can imagine "lifting" this braid B out of S^3 and putting it on $S^2 \times S^1$. To get back to S^3 one would have to do surgery on a circle running parallel to the braid, as suggested in Fig. 18b. The general surgery formula (4.27) then tells us

$$Z(S^3; L) = \sum_j S_0{}^j Z(S^2 \times S^1; R_j, B), \tag{4.57}$$

where $Z(S^2 \times S^1; R_j, B)$ is the partition function on $S^2 \times S^1$ in the presence of both the braid B and a parallel Wilson line in the R_j representation. We want to rewrite this in the spirit of (4.32). Suppose that the braid B contains n strands making up a collection of representations $\langle R \rangle$. Then B can be regarded as defining an element of the Artin braid group on n letters. The braid group is closely related to the mapping class group for S^2 with marked points. The reason for that is that if in Fig. 18b we "cut" $S^2 \times S^1$, to get back to $S^2 \times I$ (this amounts to undoing what was done in Fig. 14, then the braid can be unbraided. Thus, the complete information about the braid is in the choice of a diffeomorphism of S^2 (constrained to preserve the marking of the points) by which the top and bottom of Fig. 18c are to be identified. This, however, does not quite mean that the braid group is the same as the mapping class group. In Fig. 18b, there are $n + 1$ strands, one of which arose from the surgery and does not participate in the braid, while the other n strands make up the braid. The braid group on n letters is the subgroup of the mapping class group on $n + 1$ letters which fixes one of the (framed) strands. There are a number of invariant traces on the braid group that can be naturally defined with the data at our disposal. They are

$$\tau_i(B) = Z(S^2 \times S^1; R_i, B). \tag{4.58}$$

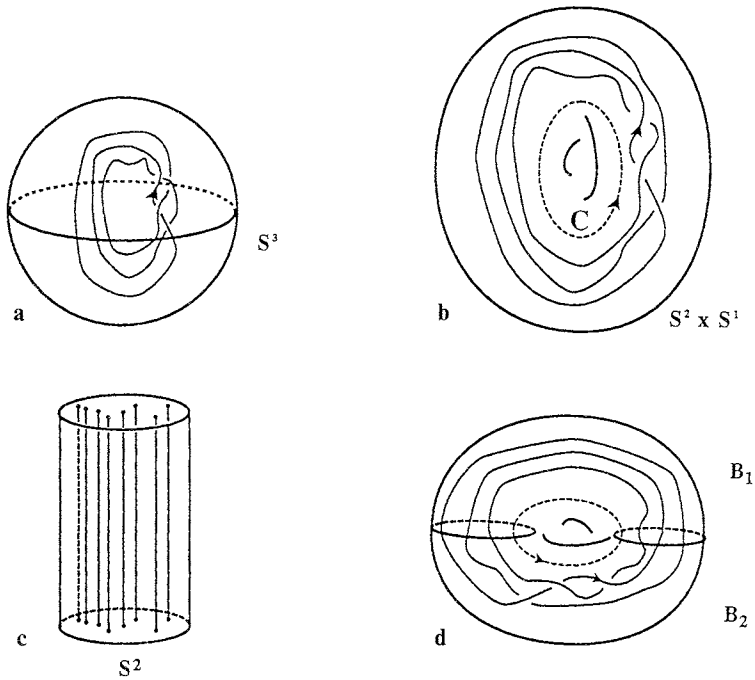

Fig. 18a–d. Any link on S^3 can be shaped into the form of a braid, as in **a**; putting the same braid on $S^2 \times S^1$, and doing surgery on the circle C indicated with the dotted line in part **b**, one gets back to the original link on S^3. If one "cuts" $S^2 \times S^1$ to get to $S^2 \times I$, the braid can be "unbraided"; the braid is recovered by prescribing a diffeomorphism of S^2, via which the top and bottom of part **c** are to be identified. In **d**, we sketch the origin of the key property of the braid traces. In the presence of an arbitrary Wilson line on the dotted contour (reflecting the results of surgery), two braids B_1 and B_2 are joined end to end in $S^2 \times S^1$. There is no way to tell which comes first, so the partition function is invariant under exchange of B_1 and B_2

This has the key property of a trace,

$$\tau_i(B_1 B_2) = \tau_i(B_2 B_1), \tag{4.59}$$

since the two sides of (4.59) have the same path integral representation, in which the two braids are glued end to end in $S^2 \times S^1$, as in Fig. 18d. Not only does (4.58) obey (4.59); it is actually equal to the trace of the operator B in a certain representation of the braid group, namely the representation furnished by the physical Hilbert space $\mathscr{H}$ for S^2 with the $n+1$ charges, it being understood that the braid group is acting on the first n charges, and the $(n+1)^{st}$ is fixed in the R_i representation. That $\tau_i(B)$ is the trace of B in this Hilbert space is a statement just along the lines of (4.32).

So we can rewrite (4.57) in the form

$$Z(S^3; L) = \sum_j S_0{}^j \tau_j(B). \tag{4.60}$$

This shows that the link invariants on S^3 may be written as linear combinations of braid traces. This is very close to how the knot polynomials were originally

discovered. It is clear from the work of Tsuchiya and Kanie [11] and Segal [16] along with what has been said above that the braid traces that arise from the Chern-Simons theory are precisely those that first appeared in the work of Jones.

5. Applications To Physics

Finally, I would like to comment on the likely implications of these results for physics. We have been exploring a three dimensional viewpoint about conformal field theory, at least for the important special case of current algebra on Riemann surfaces. Many aspects of rational conformal field theory have emerged as natural consequences of general covariance in three dimensions. It seems likely that the marvelous hexagons and pentagons of [13], and the other consistency conditions of rational conformal field theories, can be synthesized by saying that such theories come from generally covariant theories in three dimensions. If so, general covariance in three dimensions may well emerge as one of the main unifying themes governing two dimensional conformal field theory. Such considerations have motivated a study of $2+1$ dimensional gravity which will appear elsewhere [38].

The basic connection that we have so far stated between general covariance in $2+1$ dimensions and conformally invariant theories in $1+1$ dimensions is that the physical Hilbert spaces obtained by quantization in $2+1$ dimensions can be interpreted as the spaces of conformal blocks in $1+1$ dimensions. This connection may seem rather abstract, and I will now make a few remarks aimed at making a more concrete connection. Starting from three dimensions we were led to the problem of quantizing the Lagrangian (3.1), which we repeat for convenience:

$$\mathscr{L} = \frac{k}{8\pi} \int dt \int_{\Sigma} \varepsilon^{ij} \operatorname{Tr} A_i \frac{d}{dt} A_j . \tag{5.1}$$

This is to be quantized with constraints $\varepsilon^{ij} F^a_{ij} = 0$, these constraints being the generators of the infinitesimal gauge transformations

$$A_i \to A_i - D_i \varepsilon . \tag{5.2}$$

So far we have quantized (5.1) on Riemann surfaces without boundary, but now let us relax this requirement. Quantization of (5.2) in this more general case amounts to studying the three dimensional Chern-Simons theory on a three manifold with boundary, namely $\Sigma \times R^1$, where Σ is a Riemann surface with boundary. For instance, let Σ be a disc D. To quantize (5.1) on the disc, we must impose the constraints, which generate the gauge transformations (5.2). As D has a boundary, we must choose boundary conditions on A_i and ε (for closed surfaces this question did not arise). We will adopt free boundary conditions for A_i, but require $\varepsilon = 0$ on the boundary of S. (A rationale for this choice is that the Chern-Simons action is not invariant under gauge transformations that do not vanish on the boundary.) With this condition, ε generates in (5.2) the group $\hat{G}_1$ of gauge transformations which are the identity on the boundary of the disc. An element of this group is an arbitrary continuous map V: $D \to G$ whose restriction to S is the

identity. Now we impose the constraints. The first step is to require that $\varepsilon^{ij} F^a_{ij} = 0$. Since the disc is simply connected, this implies that $A_i = -\partial_i U \cdot U^{-1}$ for some map U: $D \to G$. U is uniquely defined up to

$$U \to U \cdot W, \tag{5.3}$$

for a constant element $W \in G$. Then we must identify two U's that differ by an element of the restricted gauge group $\hat{G}_1$. This means that we must impose the equivalence relation $U \simeq VU$ for any V such that $V = 1$ on the boundary S. The equivalence relation means that only the restriction of U to S is relevant. This restriction defines an element of the loop group LG, but because of the freedom (5.3), we should actually regard U as an element of LG/G. Geometrically, we have learned that the homogeneous space LG/G can be regarded as the symplectic quotient of the space of G connections on the disc D, by the group $\hat{G}_1$ of symplectic diffeomorphisms. Now we wish to quantize the theory, which means doing quantum mechanics on LG/G, which inherits a natural symplectic structure from (5.1). Clearly, the group LG of gauge transformations on the boundary of Σ acts on LG/G, so the quantum Hilbert space will be at least a projective representation of the loop group. In fact, according to Segal [39], the quantization of LG/G, with this symplectic structure, gives rise to the basic irreducible highest weight representation of the loop group. This makes the connection between $2+1$ dimensions and $1+1$ dimensions far more direct, since the irreducible representations of the loop group are a basic ingredient in the $1+1$ dimensional theory. It is obvious at this point that by considering more complicated Riemann surfaces with various boundary components, we can generate the whole $1+1$ dimensional conformal field theory, essentially by studying the generally covariant $2+1$ dimensional theory on various three manifolds with boundary.

Acknowledgements. This work originated with the realization that some results about conformal field theory described by G. Segal could be given a three dimensional interpretation by considering a gauge theory with Chern-Simons action. I am grateful to Segal for explaining his results, and to M. Atiyah for interesting me in and educating me about the Jones polynomial. V.F.R. Jones and L. Kauffman, and other participants at the IAMP Congress, raised many relevant questions. Finally, I must thank S. Deser and D.J. Gross for pointing out Polyakov's paper, G. Moore and N. Seiberg for explanations of their work, and the organizers of the IAMP Congress for their hospitality.

References

1. Atiyah, M.F.: New invariants of three and four dimensional manifolds. In: The mathematical heritage of Hermann Weyl. Proc. Symp. Pure Math., vol. **48**. Wells R. (ed.). Providence, RI: American Mathematical Society 1988, pp. 285–299
2. Donaldson, S.: An application of gauge theory to the topology of four manifolds. J. Diff. Geom. **18**, 269 (1983), Polynomial invariants for smooth four-manifolds. Oxford preprint
3. Floer, A.: An instanton invariant for three manifolds. Courant Institute preprint (1987). Morse theory for fixed points of symplectic diffeomorphisms. Bull. AMS **16**, 279 (1987)
4. Witten, E.: Topological quantum field theory. Commun. Math. Phys. **117**, 353 (1988)

5. Jones, V.F.R.: Index for subfactors. Invent. Math. **72**, 1 (1983). A polynomial invariant for links via von Neumann algebras. Bull. AMS **12**, 103 (1985), Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**, 335 (1987)
6. Freyd, P., Yetter, D., Hoste, J., Lickorish, W.B.R., Millett, K., Ocneanu, A.: A new polynomial invariant of knots and links. Bull. AMS **12**, 239 (1985)
7. Kauffman, L.: State models and the Jones polynomial. Topology **26**, 395 (1987). Statistical mechanisms and the Jones polynomial, to appear in the Proceedings of the July, 1986, conference on Artin's braid group, Santa Cruz, California; An invariant of regular isotopy preprint
8. Turaev, V.G.: The Yang-Baxter equation and invariants of links. LOMI preprint E-3-87, Inv. Math. **92**, 527 (1988)
9. Przytycki, J.H., Traczyk, P.: Invariants of links of conway type. Kobe J. Math., **4**, 115 (1988)
10. Birman, J.: On the Jones polynomial of closed 3-braids. Invent. Math. **81**, 287 (1985). Birman, J., Wenzl, H. Link polynomials and a new algebra, preprint
11. Tsuchiya, A., Kanie, Y.: In: Conformal field theory and solvable lattice models. Adv. Stud. Pure math. **16**, 297 (1988); Lett. Math. Phys. **13**, 303 (1987)
12. Verlinde, E.: Fusion rules and modular transformations in 2d conformal field theory. Nucl. Phys. **B 300**, 360 (1988)
13. Moore, G., Seiberg, N.: Polynomial equations for rational conformal field theories. To appear in Phys. Lett. B, Naturality in conformal field theory. To appear in Nucl. Phys. B, Classical and quantum conformal field theory. IAS preprint HEP-88/35
14. Schroer, B.: Nucl. Phys. **295**, 4 (1988) K.-H. Rehren, Schroer, B.: Einstein causality and Artin braids. FU preprint (1987)
15. Fröhlich, J.: Statistics of fields, the Yang-Baxter equation, and the theory of links and knots. 1987 Cargese lectures, to appear In Nonperturbative quantum field theory. New York: Plenum Press
16. Segal, G.: Conformal field theory. Oxford preprint; and lecture at the IAMP Congress, Swansea, July, 1988
17. Gromov, M.: Pseudo-holomorphic curves in symplectic manifolds. Invent. Math. **82**, 307 (1985)
18. Schwarz, A.: The partition function of degenerate quadratic functional and Ray-Singer invariants. Lett. Math. Phys. **2**, 247 (1978)
19. Schonfeld, J.: A mass term for three dimensional gauge fields. Nucl. Phys. **B 185**, 157 (1981)
20. Jackiw, R., Templeton, S.: How superrenormalizable theories cure their infrared divergences. Phys. Rev. **D 23**, 2291 (1981)
21. Deser, S., Jackiw, R., Templeton, S.: Three dimensional massive gauge theories. Phys. Rev. Lett. **48**, 975 (1983). Topologically massive gauge theory. Ann. Phys. NY **140**, 372 (1984)
22. Zuckerman, G.: Action principles and global geometry. In: The proceedings of the 1986 San Diego Summer Workshop, Yau, S.-T. (ed.)
23. Polyakov, A.M.: Fermi-bose transmutations induced by gauge fields. Mod. Phys. Lett. **A 3**, 325 (1988)
24. Hagen, C.R.: Ann. Phys. **157**, 342 (1984)
25. Arovas, D., Schrieffer, R., Wilczek, F., Zee, A.: Statistical mechanics of anyons. Nucl. Phys. **B 251 [FS 13]**, 117 (1985)
26. Witten, E.: Non-Abelian bosonization in two dimensions. Commun. Math. Phys. **92**, 455 (1984)
27. Ray, D., Singer, I.: Adv. Math. **7**, 145 (1971), Ann. Math. **98** 154 (1973)
28. Deser, S., Jackiw, R., Templeton, S.: In [21]; Affleck, I., Harvey, J., Witten, E.: Nucl. Phys. **B 206**, 413 (1982)
Redlich, A.N.: Gauge invariance and parity conservation of three-dimensional fermions. Phys. Rev. Lett. **52**, 18 (1984)
Alvarez-Gaumé, L., Witten, E.: Gravitational anomalies. Nucl. Phys. **B 234**, 269 (1983)
Alvarez-Gaumé, Della Pietra, S., Moore, G.: Anomalies and odd dimensions. Ann. Phys. (NY) **163**, 288 (1985)

Atiyah, M.F.: A note on the eta invariant (unpublished)
Singer, I.M.: Families of Dirac operators with applications to Physics. Astérisque, **1985**, p. 323
29. Atiyah, M.F., Patodi, V., Singer, I.: Math. Proc. Camb. Phil. Soc. **77**, 43 (1975), **78**, 405 (1975), **79**, 71 (1976)
30. Wilczek, F., Zee, A.: Linking numbers, spin, and statistics of solitons. Phys. Rev. Lett. **51**, 2250 (1983)
31. Friedan, D., Shenker, S.: Nucl. Phys. **B 281**, 509 (1987)
32. Belavin, A., Polyakov, A.M., Zamolodchikov, A.: Nucl. Phys. **B** (1984)
33. Atiyah, M.F., Bott, R.: The Yang-Mills equations over Riemann surfaces. Phil. Trans. R. Soc. Lond. **A 308**, 523 (1982)
34. Quillen, D.: Determinants of Cauchy-Riemann operators over a Riemann surface. Funct. Anal. Appl. **19**, 31 (1986)
35. Drinfeld, V.: Quantum groups. In: The Proceedings of the International Congress of Mathematicians, Berkeley 1986, Vol. 1, pp. 798–820
36. Gepner, D., Witten, E.: String theory on group manifolds. Nucl. Phys. B **278**, 493 (1986)
37. Kac, V.G.: Infinite dimensional Lie algebras. Cambridge: Cambridge University Press (1985)
Kac, V.G., Peterson, D.H.: Adv. Math. **53**, 125 (1984)
Kac, V.G., Wakimoto, M.: Adv. Math. **70**, 156 (1988)
38. Witten, E.: 2+1 dimensional gravity as an exactly soluble system. IAS preprint HEP-88/32
39. Segal, G.: Unitary representation of some infinite dimensional groups. Commun. Math. Phys. **80**, 301 (1981)

Communicated by A. Jaffe

Received September 12, 1988; in revised form September 27, 1988