CONTENTS

1. KNOTS, LINKS, 3-MANIFOLDS AND CHERN-SIMONS THEORY

1.1. **Knots, Links and Braids.** A *topological knot* is a continuous embedding of the circle $\mathbb{S}^1$ into $\mathbb{R}^3$, that is, a continuous, injective map

$$K : \mathbb{S}^1 \to \mathbb{R}^3$$

of $\mathbb{S}^1$ into $\mathbb{R}^3$. Since $\mathbb{S}^1$ is compact and $\mathbb{R}^3$ is Hausdorff, K is a homeomorphism of $\mathbb{S}^1$ onto its image $K(\mathbb{S}^1)$. It is often convenient to regard $\mathbb{R}^3$ as an open subspace of its 1-point compactification $\mathbb{S}^3$

$$\mathbb{R}^3 \subseteq \mathbb{R}^3 \cup \{\infty\} = \mathbb{S}^3$$

(see page 73 of [Nab2]) so that K can be thought of as a continuous embedding of $\mathbb{S}^1$ into $\mathbb{S}^3$.

Remark 1.1. We will generally use the same symbol K for the map $K : \mathbb{S}^1 \to \mathbb{R}^3$ and for its image $K(\mathbb{S}^1)$ and thereby think of a knot either as an embedding or as a subspace of $\mathbb{R}^3$ or $\mathbb{S}^3$.

Topologically, every knot K in $\mathbb{R}^3$ or $\mathbb{S}^3$ is just a copy of $\mathbb{S}^1$, but they are distinguished one from another by the manner in which they are "knotted" by the embedding. There are two notions of knot equivalence in common use for making this distinction mathematically.

(1) Two topological knots K_0 and K_1 in $\mathbb{R}^3$ are said to be *equivalent* if there is a homeomorphism $h : \mathbb{R}^3 \to \mathbb{R}^3$ of $\mathbb{R}^3$ onto itself for which $h(K_0) = K_1$.

   Remark 1.2. Such a homeomorphism h extends uniquely to a homeomorphism of $\mathbb{S}^3$ onto itself that carries K_0 onto K_1 so K_0 and K_1 are equivalent if and only if there is a homeomorphism $h : \mathbb{S}^3 \to \mathbb{S}^3$ of $\mathbb{S}^3$ onto itself for which $h(K_0) = K_1$.

This clearly defines an equivalence relation on the set of knots. The corresponding equivalence class of a knot K is called its *knot type*.

(2) Two topological knots K_0 and K_1 in $\mathbb{R}^3$ are said to be *isotopic* if there exists a homotopy $H : \mathbb{R}^3 \times [0,1] \to \mathbb{R}^3$ for which each $h_t(\,\cdot\,) = H(\,\cdot\,,t)$, $0 \le t \le 1$, is a homeomorphism of $\mathbb{R}^3$ onto itself, $h_0 = \mathrm{id}_{\mathbb{R}^3}$, and $h_1(K_0) = K_1$.

Remark 1.3. As in the previous Remark, one can replace $\mathbb{R}^3$ by $\mathbb{S}^3$ everywhere in this definition. Isotopic knots are clearly equivalent, but the converse is false as we will see shortly. We will refer to the equivalence class of a knot K under this equivalence relation as its *isotopy type*. To understand the difference between the two notions in geometric terms we recall that the notion of *orientability*, usually introduced in the category of *smooth* manifolds, can also be defined for *topological* manifolds in terms of their homology (see Section 22 of [Green]). Corollary 22.28 of [Green] implies that a compact, connected, topological n-manifold X is orientable if and only if $H_n(X;\mathbb{Z}) \cong \mathbb{Z}$. Furthermore, a connected, orientable manifold has exactly two distinct orientations (see Corollary 22.9 of [Green]). In particular, any space homeomorphic to a sphere, such as a knot, is orientable with two distinct orientations. In the case of a knot these are identified with the two directions in which the curve might be traversed and are indicated pictorially by attaching arrows to the curve. An *oriented knot* is a knot which has been equipped with one of its two possible orientations. Now, a homeomorphism $h : \mathbb{S}^3 \to \mathbb{S}^3$ of $\mathbb{S}^3$ onto itself induces an isomorphism h_* of $H_3(\mathbb{S}^3;\mathbb{Z}) \cong \mathbb{Z}$ onto itself. Consequently, h_* is either $\mathrm{id}_{\mathbb{Z}}$ or $-\mathrm{id}_{\mathbb{Z}}$. In the first case, h is said to be *orientation preserving* and, in the second, *orientation reversing*. The identity map $\mathrm{id}_{\mathbb{S}^3}$ is clearly orientation preserving and, for isotopic knots, the homotopy $H : \mathbb{S}^3 \times [0,1] \to \mathbb{S}^3$ is continuous in t. It follows that each homeomorphism h_t, $0 \le t \le 1$, is orientation preserving since it is homotopic to the identity and therefore induces the same map in homology. In particular, h_1 is orientation preserving. Consequently, if K_0 and K_1 are isotopic, there exists an orientation preserving homeomorphism of $\mathbb{S}^3$ onto itself that carries K_0 onto K_1. In fact, one can show that every orientation preserving homeomorphism of $\mathbb{S}^3$ onto itself arises in this way as h_1 from a homotopy $H : \mathbb{S}^3 \times [0,1] \to \mathbb{S}^3$ with each h_t, $0 \le t \le 1$, a homeomorphism of $\mathbb{S}^3$ onto itself and $h_0 = \mathrm{id}_{\mathbb{S}^3}$ (see [Fisher]). Consequently, *two topological knots K_0 and K_1 are isotopic if and only if there exists an orientation preserving homeomorphism $h : \mathbb{S}^3 \to \mathbb{S}^3$ of $\mathbb{S}^3$ onto itself with $h(K_0) = K_1$*. The difference between *equivalent* and *isotopic* is *orientation*.

A *knot invariant* is a mathematical object (a number, or a group, or a polynomial, etc.) that one can associate with each knot K and with the property that if two knots are equivalent, then their associated objects are the same. Given such a knot invariant one might hope to show that two given knots are *not* equivalent by showing that their invariants are not the same.

Remark 1.4. In the best of all possible worlds one might also hope to conclude that if the associated objects *are* the same, then the knots are equivalent. Unfortunately, such a *complete invariant* that one can actually compute is not known and there does not seem to be any reasonable prospect that one will be known in the foreseeable future (see Remark 1.5).

One begins the search for such knot invariants by considering the *knot complement* $\mathbb{R}^3 - K$. This is an open subspace of $\mathbb{R}^3$ since K is compact. Moreover, two equivalent knots clearly have homeomorphic knot complements so it would seem profitable to look at some of the invariants associated to $\mathbb{R}^3 - K$ by algebraic topology. The homology groups will not do because knot complements all have the same homology. Specifically, for any knot K in $\mathbb{R}^3$,

$$H_p(\mathbb{R}^3 - K; \mathbb{Z}) = \begin{cases} \mathbb{Z}, & \text{if } p = 1, 2 \\ 0, & \text{if } p = 3, 4, \ldots. \end{cases}$$

(see page 80 of [Green]). We therefore turn to the homotopy groups instead.

Any knot complement $\mathbb{R}^3 - K$ is pathwise connected, but this is not obvious. One can piece together a proof in the following way. The Borsuk Separation Theorem (see Theorem 2.1, Chapter XVII, of [Dug]) asserts that if A is a compact subset of $\mathbb{R}^n$, then $\mathbb{R}^n - A$ is connected if and only if every continuous map $f : A \to \mathbb{S}^{n-1}$ from A to $\mathbb{S}^{n-1}$ is nullhomotopic. Moreover, if $k < n$, then every continuous map of $\mathbb{S}^k$ into $\mathbb{S}^n$ is nullhomotopic (see Theorem 4.2, Chapter XVI, of [Dug]). A knot K is homeomorphic to $\mathbb{S}^1$ so any continuous map from K to $\mathbb{S}^2$ is nullhomotopic. It follows that $\mathbb{R}^3 - K$ is connected. But since $\mathbb{R}^3 - K$ is open in $\mathbb{R}^3$ every point in $\mathbb{R}^3 - K$ has a neighborhood that is pathwise connected and this, together with connectedness, implies that $\mathbb{R}^3 - K$ is pathwise connected (see Theorem 5.5, Chapter V, of [Dug]). Consequently, the fundamental group of $\mathbb{R}^3 - K$ is independent of the base point and we will therefore write this simply as $\pi_1(\mathbb{R}^3 - K)$. Moreover, it follows from van Kampen's Theorem that

$$\pi_1(\mathbb{R}^3 - K) \cong \pi_1(\mathbb{S}^3 - K)$$

so it does not matter whether we view K as a subspace of $\mathbb{R}^3$ or $\mathbb{S}^3$ (see the first Proposition, page 51, of [Rolf]). This is called the *knot group* of K. Equivalent knots have isomorphic knot groups because $h|_{\mathbb{R}^3 - K_0} : \mathbb{R}^3 - K_0 \to \mathbb{R}^3 - K_1$ is a homeomorphism, but the converse is false (see 3D10 and 8E15 of [Rolf]). Consequently, the knot group fails to be a complete invariant of knots (see Remark 1.4).

Remark 1.5. It has been shown by Gordon and Luecke [GL] that if one knows that $\mathbb{R}^3 - K_0$ and $\mathbb{R}^3 - K_1$ are homeomorphic, then the knots K_0 and K_1 are equivalent and therefore have isomorphic knot groups. Moreover, if there is an orientation preserving homeomorphism of $\mathbb{R}^3 - K_0$ onto $\mathbb{R}^3 - K_1$, then K_0 and K_1 are isotopic. In this sense the topological type of the knot complement is itself a complete invariant for knots (see Remark 1.4), but one can hardly think of it as "computable".

The most obvious example of a knot is the inclusion map $\iota : \mathbb{S}^1 \hookrightarrow \mathbb{R}^3$. This is called the *trivial knot*. The complement $\mathbb{R}^3 - \mathbb{S}^1$ has fundamental group $\mathbb{Z}$. This is (4.1) of Section 4, Chapter VI, of [CF], but the result can also be derived more directly from van Kampen's Theorem (see Example 1.23, page 46, of [Hatch]). Any knot equivalent to this one is called an *unknot* so the knot group of any unknot is $\mathbb{Z}$. We will see more examples of knot groups as we proceed.

A *polygonal knot* is a topological knot that is the union of a finite number of closed straight line segments in $\mathbb{R}^3$. A *smooth knot* is (the image of) a smooth embedding $K : \mathbb{S}^1 \to \mathbb{R}^3$ of $\mathbb{S}^1$ in $\mathbb{R}^3$ (or $\mathbb{S}^3$).

Remark 1.6. Recall that a smooth embedding $F : M \to N$ of one manifold into another is a smooth map with injective derivative F_{*m} at each point $m \in M$ that is a homeomorphism of M onto its image $F(M)$. It follows that $F(M)$ is a smooth submanifold of N diffeomorphic to M. Thus, a smooth knot is a submanifold of $\mathbb{R}^3$ (or $\mathbb{S}^3$) that is diffeomorphic to $\mathbb{S}^1$.

By polygonal approximation, any smooth knot is equivalent to a polygonal map (see Appendix 1 of [CF]) and, by rounding corners, any polygonal knot is equivalent to a smooth knot. A topological knot is *tame* if it is equivalent to a polygonal knot (or, equivalently, to a smooth knot); otherwise, it is *wild*. Wild knots have rather complicated topological structure (see Figure 1).

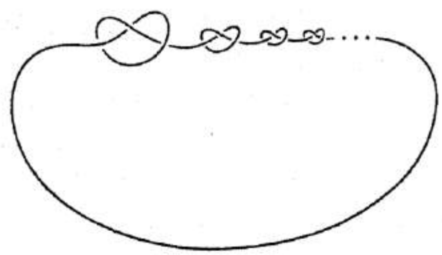


FIGURE 1. A Wild Knot

Henceforth we will consider only tame knots.

From this point on *knot* will always mean *tame knot*. Since we are concerned primarily with knot type we can therefore choose a polygonal or smooth representative as the need arises. There are algorithms for computing the knot group of any tame knot in terms of generators and relations (see Chapter VI of [CF] or Chapter 3, Section D, of [Rolf]). The problem is that it is difficult to determine whether or not two such presentations represent non-isomorphic groups. In particular, there is no algorithm for deciding this (see [Collins] for a simple example of an undecidable "Word Problem"). It is for this reason that one generally looks for other knot invariants with which to distinguish knots that are either not equivalent or not isotopic. We will eventually (Section 1.2) focus attention on one of these called the *Jones polynomial*.

A knot K in $\mathbb{R}^3$ is pictured geometrically by its orthogonal projection into a plane. Many such plane projections exist, some good and some not-so-good, and even the good ones can look quite different. We define a "good projection" of K as follows. Since K is assumed tame and since we are interested only in its knot type we can assume at the outset that K is a smooth knot. Let P be a plane in $\mathbb{R}^3$ and $\pi : \mathbb{R}^3 \to P$ the orthogonal projection of $\mathbb{R}^3$ onto P. A point $p \in \pi(K) \subseteq P$ is called a *multiple point* if $\pi^{-1}(p) \cap K$ consists of more that one point. It is a *double point* if $\pi^{-1}(p) \cap K$ consists of two points, a *triple point* if $\pi^{-1}(p) \cap K$ consists of three points, and so on. We will say that π is a *regular projection* for K if the following two conditions are satisfied.

(1) The set of multiple points in the image $\pi(K)$ consists of a finite number of double points.
(2) If $p \in \pi(K)$ is a double point with $\pi^{-1}(p) \cap K = \{k_1, k_2\}$, then locally the arcs of K through k_1 and k_2 project onto smooth curves that intersect transversally, that is, that are *not* tangent at p.

Condition (2) guarantees that every double point corresponds to a genuine "crossing" in the image $\pi(K)$ of K. Every knot K has a regular projection; in fact, there is a sense in which *almost every* projection of K is regular. More precisely, an arbitrarily small rotation of $\mathbb{R}^3$ carries any projection of K onto a regular projection (see (3.1) of Section 3, Chapter I, in [CF]).

From a regular projection of K one obtains a *knot diagram* of K in the following way. Each double point $p \in \pi(K)$ is the image under π of two *distinct* points k_1 and k_2 in K. Since π is the orthogonal projection onto P, the distance from k_1 to P is not equal to the distance from k_2 to P. Assuming, without loss of generality, that the former is less than the latter we redraw the crossing at p with the projected arc of K through k_1 *under* the projected arc of K through k_2. This is done by breaking the projected arc through k_1 at p. Repeating the process for each crossing one obtains something like Figure 2 which is a knot diagram for the so-called *left-handed trefoil knot*. Interchanging the over-crossings and under-crossings in Figure 2

FIGURE 2. Left-Handed Trefoil Knot

gives a knot diagram for what is called the *right-handed trefoil knot* shown in Figure 3.

FIGURE 3. Right-Handed Trefoil Knot

Remark 1.7. The "left-handed" and "right-handed" terminology is a matter of convention, chosen in the following way. Select an orientation (direction) for the trefoil knot. Choose a normal vector $\mathbf{n}$ to the plane containing its knot diagram. Now look at one of the three crossing points. Assuming the knot is smooth we can let $\mathbf{t}_o$ and $\mathbf{t}_u$ be the tangent vectors to the forward over-crossing and forward under-crossing arcs, respectively. Applying the right-hand rule from $\mathbf{t}_o$ to $\mathbf{t}_u$ yields a vector in the direction of either $\mathbf{n}$ or $-\mathbf{n}$. In the first case the crossing is "right-handed" and in the second it is "left-handed". Now note that this is independent of the orientation chosen because reversing the orientation reverses both $\mathbf{t}_o$ and $\mathbf{t}_u$. In the case of the trefoil, but not in general, one also notes that the "handedness" is the same for all of the crossings.

Thus, the “right-handed” and “left-handed” terminology can be applied to the knot itself. In Figure 2 and Figure 3 we have chosen the normal vector **n** *out of* the page. Whether or not one actually gets two different, that is, non-isotopic knots in this way is not so obvious and we will return to this question shortly.

These are called *mirror images* of each other and each is the image of the other under an *orientation reversing* homeomorphism of $\mathbb{R}^3$ or $\mathbb{S}^3$ onto itself. In particular, their complements are homeomorphic so, by the result of Gordon and Luecke [GL] referred to above in Remark1.5, the left- and right-handed trefoils are equivalent. They therefore have the same knot group which, for reasons that will become clear momentarily, is denoted $G_{2,3}$. One can show that $G_{2,3}$ has two generators x and y and one relation $x^2 = y^3$ (see page 52 of [Rolf]).

$$G_{2,3} = \langle\, x, y \,|\, x^2 = y^3 \,\rangle$$

This group is not Abelian (see the Lemma, page 52, of [Rolf]) so, in particular, the trefoil is not equivalent to the trivial knot; it is genuinely “knotted”. The left- and right-handed trefoils are, however, *not* isotopic, as we will see in a moment.

The trefoil is a particular example of an important class of knots that we will now briefly describe. For these we will identify $\mathbb{S}^1$ with the unit complex numbers $e^{\theta i}$, $0 \le \theta < 2\pi$. The torus $\mathbb{S}^1 \times \mathbb{S}^1$ lives naturally in $\mathbb{C} \times \mathbb{C} \cong \mathbb{R}^4$, but is diffeomorphic to the familiar surface of revolution in $\mathbb{R}^3$ and we will implicitly compose with such a diffeomorphism. Now let p and q be integers and define a map from $\mathbb{S}^1$ to $\mathbb{S}^1 \times \mathbb{S}^1$ by

$$e^{\theta i} \mapsto (\, e^{p\theta i}, e^{q\theta i}\,), \quad 0 \le \theta < 2\pi.$$

The map is easily seen to be injective if and only if p and q are relatively prime. In this case the map, which is clearly continuous, is a homeomorphism onto its image because $\mathbb{S}^1$ is compact and $\mathbb{S}^1 \times \mathbb{S}^1$ is Hausdorff. Thus, we can define, for any relatively prime integers p and q, the (p, q)*-torus knot*

$$T_{p,q} : \mathbb{S}^1 \to \mathbb{S}^1 \times \mathbb{S}^1$$

by

$$T_{p,q}(e^{\theta i}) = (\, e^{p\theta i}, e^{q\theta i}\,), \quad 0 \le \theta < 2\pi.$$

All of these are clearly smooth and therefore tame. Moving a few of them into $\mathbb{R}^3$ by a diffeomorphism of $\mathbb{S}^1 \times \mathbb{S}^1$ onto a torus of revolution in $\mathbb{R}^3$ gives pictures of the sort shown in Figure 4. We will record just a few facts about these torus knots for future reference (see pages 53-55 of [Rolf]).

(1) The knot group $G_{p,q}$ of $T_{p,q}$ has two generators x and y and one relation $x^p = y^q$.

$$G_{p,q} = \langle\, x, y \,|\, x^p = y^q \,\rangle$$

(2) The knot type of $T_{p,q}$ is unchanged by changing the sign of p or the sign of q or by interchanging p and q. Otherwise, all of the torus knots are inequivalent.

(3) $T_{\pm 1,q}$ and $T_{p,\pm 1}$ are equivalent to the trivial knot.

(4) $T_{2,3}$ and $T_{3,2}$ are both equivalent to the trefoil knot. $T_{2,3}$ is isotopic to the right-handed trefoil and $T_{3,2}$ is isotopic to the left-handed trefoil.

(5) $T_{p,q}$ and $T_{p,-q}$ are mirror images of each other. They are equivalent, but *not* isotopic.

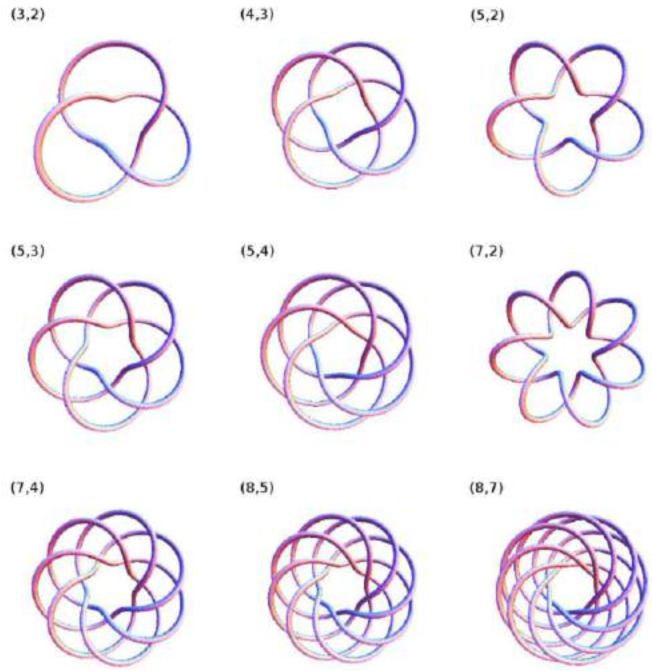


FIGURE 4. Some Torus Knots

Consider next an oriented knot K and a knot diagram D for it. We would like to use the orientation to associate with each crossing p in D a sign $\mathrm{sgn}(p) = \pm 1$. The procedure is indicated pictorially in Figure 5 and can be regarded as an application of the so-called "right-hand rule". Choose and fix a normal vector $\mathbf{n}$ to

FIGURE 5. Crossing Sign

the plane in which the diagram D lives and let $\mathbf{t}_o$ and $\mathbf{t}_u$ be the tangent vectors to the forward over-crossing and forward under-crossing at p, respectively. Then $\mathrm{sgn}(p)$ is $+1$ if curling the right hand from $\mathbf{t}_o$ to $\mathbf{t}_u$ leaves the thumb pointing in the direction of $\mathbf{n}$; otherwise the sign is -1. The sign depends, of course, on the choice of $\mathbf{n}$, but once this is fixed the signs are well-defined. Note that these signs are unchanged if the orientation of the knot is reversed. We define the *writhe* $w(D)$ of the knot diagram to be the sum of these signs over all the crossings in D; it is, by definition, zero if there are no crossings.

$$w(D) = \sum_{\text{crossings } p \in D} \mathrm{sgn}(p)$$

Although $w(D)$ is defined from an orientation of K it is, in fact, independent of the orientation and so is a property of the knot diagram itself. It fails to be a knot invariant, however, as one can see by comparing knot diagrams for the trivial knot and the unknot in Figure 6.

FIGURE 6. Unknot with $w(D) \neq 0$

Next we consider "linked" families of knots. A *link* in $\mathbb{R}^3$ or $\mathbb{S}^3$ is a disjoint union

$$L = K_1 \sqcup \cdots \sqcup K_m$$

of knots $K_1, \ldots, K_m$ each of which is called a *component* of the link. A knot is just a link with one component. The so-called *Whitehead link* is shown in Figure 7. We will always fix an ordering of the components

FIGURE 7. Whitehead Link

by $1, \ldots, m$. The link L is *topological*, *polygonal*, or *smooth* if each of its components $K_i, i = 1, \ldots, m$, is a topological, polygonal, or smooth knot, respectively. An *oriented link* is one for which each of its component knots has been assigned an orientation. Two links $L = K_1 \sqcup \cdots \sqcup K_m$ and $L' = K'_1 \sqcup \cdots \sqcup K'_m$ with the same number of components are *equivalent* (respectively, *isotopic*) if there is a homeomorphism (respectively, orientation preserving homeomorphism) h of $\mathbb{R}^3$ onto itself (or of $\mathbb{S}^3$ onto itself) with $h(K_i) = K'_i$ for each $i = 1, \ldots, m$. In particular, each of the knots K_i in L is equivalent (respectively, isotopic) to the corresponding knot K'_i in L'. The equivalence class of L is called its *link type* (respectively, *isotopy type*). A link is *tame* if it is equivalent to a polygonal (or, equivalently, smooth) link and we will consider only these so that, henceforth, *link* always means *tame link*. A link is *trivial*, or *unlinked*, if it is equivalent to a link that lies entirely in a plane in $\mathbb{R}^3$. We will make little use of the fundamental group of a link complement, but mention for those who are interested that two examples are computed in Example 1.23, page 46, of [Hatch].

Let P be a plane in $\mathbb{R}^3$ and $\pi : \mathbb{R}^3 \to P$ the orthogonal projection of $\mathbb{R}^3$ onto P. A point $p \in \pi(L) \subseteq P$ is called a *multiple point* if $\pi^{-1}(p) \cap L$ consists of more that one point. It is a *double point* if $\pi^{-1}(p) \cap L$ consists of two points, a *triple point* if $\pi^{-1}(p) \cap L$ consists of three points, and so on. As in the case of knots we will say that π is a *regular projection* for L if there are only finitely many multiple points, all of which are transverse double points. Every link has regular projections. Indeed, as for knots, there is a sense in which almost every projection is a regular projection for L. In particular, one can find a common regular projection for any two links. From any regular projection for L one obtains a *link diagram* from $\pi(L)$ by indicating the over-crossings and under-crossings just as for a knot. Interchanging the over-crossings and under-crossings gives a link diagram for the *mirror image* of L. One possible link diagram for the Whitehead link (Figure 7) is shown in Figure 8. For an oriented link one defines the sign of a crossing in a link diagram and the *writhe*

FIGURE 8. Whitehead Link Diagram

of the diagram in precisely the same way as for a knot diagram. A variant of this is obtained by ignoring all of the "internal" crossings that arise from the component knots themselves and looking only at those that correspond to one component crossing another (for example, we would ignore the crossing at the center of Figure 8). More precisely, we consider the oriented link $L = K_1 \sqcup \cdots \sqcup K_m$ with link diagram D and fix two distinct components, say, K_i and K_j with $1 \le i < j \le m$. Define the *linking number of K_i with K_j*, denoted $lk\,(K_i, K_j)$, to be one-half the sum of the signs of the crossings of K_i with K_j. Then the *linking number* of the diagram D, denoted $lk\,(D)$, is the sum of these over all distinct pairs of components.

$$lk\,(D) = \sum_{1 \le i < j \le m} lk\,(K_i, K_j)$$

The definition of the linking number can be formulated in a great variety of equivalent ways, including the famous integral formula of Gauss (see Chapter Five, Section D, of [Rolf] and [RN]).

Since drawing knots and links in space is rather a tricky business one would like to know if one can tell from their regular projections whether or not two links are "the same" (equivalent or isotopic). An answer to the question was obtained independently by Kurt Reidemeister [Reid] and Alexander and Briggs [AlBr]. The procedure involves applying a sequence of local "moves", now called *Reidemeister moves*, which alter a given link diagram in a small region without altering the link type of the link from which the diagram arose. The three basic Reidemeister moves are indicated in Figure 9 and labeled I, II, and III corresponding to the number of arcs of the projection that are effected by the move. The relevant result proved in [Reid] and [AlBr] is as follows.

Theorem 1.1. *Let L and L' be two links in $\mathbb{R}^3$ and let $\pi : \mathbb{R}^3 \to P$ be a regular projection for both L and L'. Let D be a link diagram of L in the plane P, and D' a link diagram of L' in P. Then L and L' are isotopic links if and only if D' can be obtained from D by a sequence of Reidemeister moves and orientation preserving homeomorphisms of P.*

The principal use one makes of the Reidemeister moves is to show that two links are isotopic or that something is a link invariant by proving that it is left unchanged by any of these moves. We include one pictorial application of Theorem 1.1 (Figure 10) which shows that the so-called *figure eight knot* is isotopic to its mirror image.

We turn next to a notion that is very closely related not only to knots and links, but to physics as well (see Remark 1.9). Fix an integer $n \ge 2$. A *braid* on *n strings* (or *n strands*), also called an *n-braid*, is the union ω of a family of smooth arcs $\{l_1, \ldots, l_n\}$ in $\mathbb{R}^3$ constructed in the following way (see Figure 11). Choose two parallel planes P_0 and P_1 in $\mathbb{R}^3$ and two parallel straight lines L_a on P_0 and L_b on P_1. Select an *ordered* set $\{a_1, \ldots, a_n\}$ of distinct points on L_a and then an ordered set $\{b_1, \ldots, b_n\}$ of points on L_b such that the distance

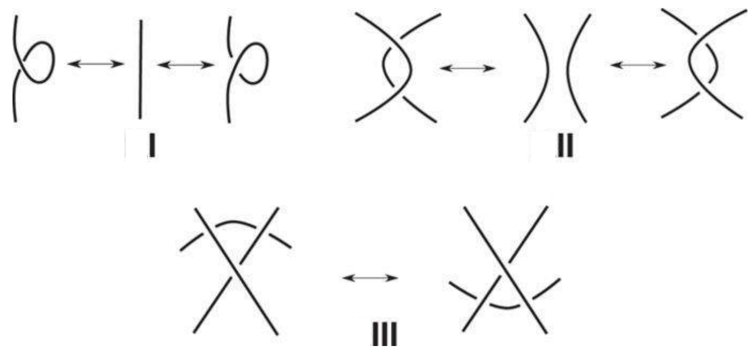


FIGURE 9. Reidemeister Moves

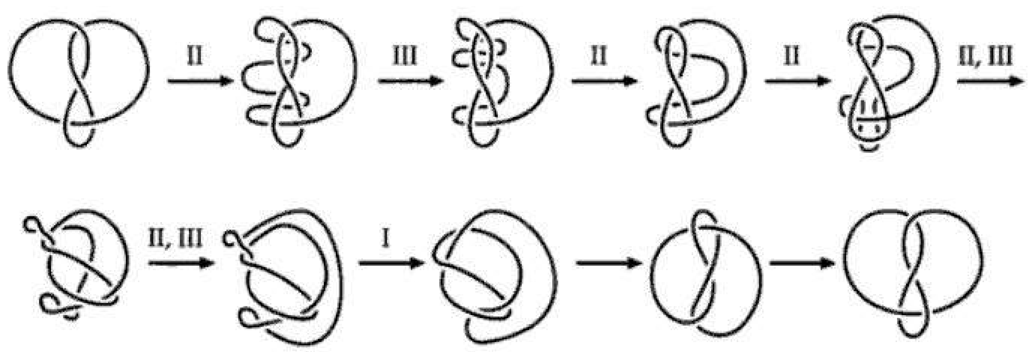


FIGURE 10. Figure Eight Knot and its Mirror Image

from a_i to b_i is the same for each $i = 1, \ldots, n$. Denote by $\{P_t\}_{0 \le t \le 1}$ the set of all parallel planes between P_0 and P_1. Finally, let π be some permutation of $\{1, \ldots, n\}$. Now take $\{l_1, \ldots, l_n\}$ to be a set of simple, pairwise non-intersecting smooth arcs satisfying each of the following conditions.

(1) l_i joins a_i and $b_{\pi(i)}$ for each $i = 1, \ldots, n$.
(2) Each l_i, $i = 1, \ldots, n$, intersects each P_t, $0 \le t \le 1$, exactly once.

If π is the identity permutation, then ω is called a *pure braid*. If, for some $i = 1, \ldots, n-1$, π is the transposition $i \leftrightarrow i+1$, then ω is called a *simple braid*.

With $P_0, P_1, \{a_1, \ldots, a_n\}$, and $\{b_1, \ldots, b_n\}$ all fixed we define an equivalence relation on the set of braids with n strings as follows. Let $P = \cup_{0 \le t \le 1} P_t$ denote the region between P_0 and P_1. Then the braids ω_1 and ω_2 are *equivalent* if there is a homeomorphism $h : P \to P$ of P onto itself satisfying each of the following.

(1) $h|_{P_0 \cup P_1} = id_{P_0 \cup P_1}$
(2) $h(P_t) = P_t$ for each $0 \le t \le 1$
(3) $h(\omega_1) = \omega_2$

It is customary to refer to the equivalence class of a braid ω also simply as a *braid* and even to use the same symbol ω for it rather than something more descriptive such as $[\omega]$. This is generally harmless and we will

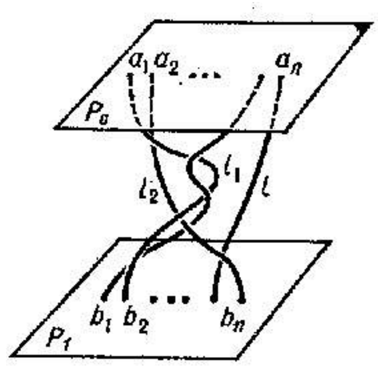


FIGURE 11. Braid

adhere to the custom, but when it seems advisable to make a distinction we will refer to any representative of a braid equivalence class as a *geometric braid*.

Remark 1.8. As was the case for knots and links, the theory of braids that we will discuss could equally well begin with polygonal, rather than smooth arcs. Furthermore, this theory, which is topological and algebraic, does not depend on any particular choice we might make for $P_0, P_1, \{a_1, \ldots, a_n\}$, and $\{b_1, \ldots, b_n\}$. Since it can simplify the exposition we will generally make the following choice for these (x, y, z are standard coordinates on $\mathbb{R}^3$).

$$P_0 = \{(x, y, 1) \in \mathbb{R}^3 : (x, y) \in \mathbb{R}^2\} \quad \text{and} \quad P_1 = \{(x, y, 0) \in \mathbb{R}^3 : (x, y) \in \mathbb{R}^2\}$$
$$a_i = (i, 0, 1) \quad \text{and} \quad b_i = (i, 0, 0) \quad \text{for} \quad i = 1, \ldots, n$$
$$P = \{(x, y, z) \in \mathbb{R}^3 : 0 \leq z \leq 1\}$$

Thus, each l_i is strictly *descending* from a_i to $b_{\pi(i)}$ and equivalence is defined in terms of *height preserving* homeomorphisms that leave the top and bottom fixed. One could, of course, turn the picture upside down and regard the l_i as strictly *ascending* from $z = 0$ to $z = 1$.

We will now introduce a multiplicative structure on the collection of all n-braids. Intuitively, one multiplies two geometric n-braids by simply putting them end-to-end. Let ω' and ω'' be two such geometric n-braids in copies P' and P'' of P, respectively. Place P'' below P' in such a way that P'_1 coincides with P''_0 and b'_i coincides with a''_i for each $i = 1, \ldots, n$. Then one can extend each l'_i by $l''_{\pi'(i)}$. This gives arcs in $P' \cup P''$ joining each a'_i to some b''_j, $i, j = 1, \ldots, n$. Specifically,

$$a'_i \to b'_{\pi'(i)} = a''_{\pi'(i)} \to b''_{\pi''(\pi'(i))}.$$

This would be an n-braid if $P' \cup P''$ were P, but it is not so we compress $P' \cup P''$ vertically upward to half its width, that is, we define a map from $P' \cup P''$ to P' that pushes every point to the point vertically above it and half the distance to P'_0. The resulting geometric braid (see Figure 12) is called the *product* of ω' and ω'' and written

$$\omega'\omega''.$$

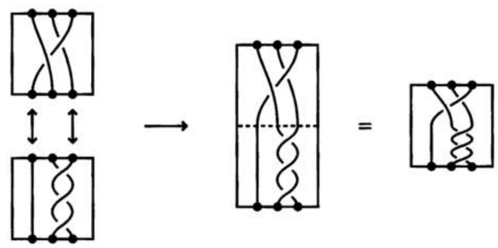


FIGURE 12. Braid Multiplication

The braid product induces a well-defined multiplication on equivalence classes of braids. The *identity braid e* is the equivalence class of the pure braid with straight line segments joining a_i and b_i for each $i = 1, \ldots, n$ and this clearly acts as an identity element for the multiplication of braid equivalence classes. For these equivalence classes one can show that inverses always exist. This means that for any geometric braid ω there exists a geometric braid ω' such that $\omega\omega'$ and $\omega'\omega$ are both in the equivalence class e (see Figure 13). If we adopt the convention that ω denotes the equivalence class and not any particular representative, then it one would write ω^{-1} rather than ω'. Furthermore, the multiplication of braid equivalence classes is associative so we have a group, called the *braid group* on *n strings* (or *n strands*) and denoted $\mathcal{B}_n$.

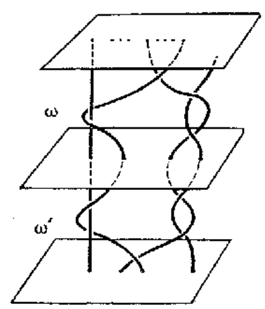


FIGURE 13. Braid Inverse

The structure of the group $\mathcal{B}_n$ can be explicitly described in terms of generators and relations. For each $i = 1, \ldots, n-1$ there are two equivalence classes of simple braids corresponding to the transposition $i \leftrightarrow i+1$, one with an overcrossing from i to $i+1$, denoted σ_i, and one with an undercrossing from i to $i+1$, which is σ_i^{-1}. Figure 12 represents $\sigma_1\sigma_2$. One can show that any n-braid can be written as a product of the σ_i and σ_i^{-1} for $i = 1, \ldots, n-1$. This is done by subdividing P into horizontal regions each of which contains at most one crossing. For each of these horizontal strips that contains a crossing one obtains a factor of either σ_i or σ_i^{-1} (see Figure 14). Thus, every element of $\mathcal{B}_n$ can be identified with a *word* in the generators $\sigma_1, \ldots, \sigma_{n-1}$. Emil Artin, who introduced the notions we have been discussing, showed that the following relations on these generators determine the group $\mathcal{B}_n$.

$$\sigma_i\,\sigma_j = \sigma_j\,\sigma_i \quad \text{whenever} \quad |\,i-j\,| \geq 2, \quad i, j = 1, \ldots, n-1 \tag{1}$$

$$\sigma_i\,\sigma_{i+1}\,\sigma_i = \sigma_{i+1}\,\sigma_i\,\sigma_{i+1} \quad \text{for all} \quad i = 1, \ldots, n-2 \tag{2}$$

The first of these relations is generally referred to as *far commutativity* since it asserts the commutativity

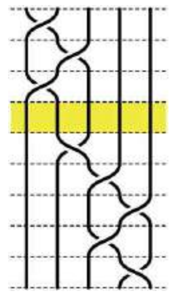
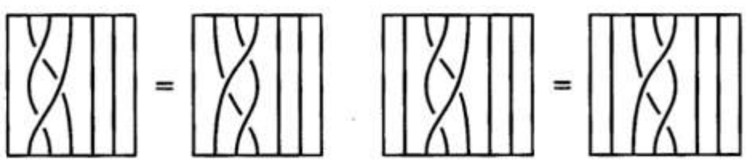


FIGURE 14. $\sigma_1^{-1}\sigma_2^{-1}\sigma_1^{-1}\sigma_2\sigma_3^{-1}\sigma_4^{-1}\sigma_3^{-1}\sigma_4$

of generators whose indices are sufficiently far apart, while the second is simply called the *braid relations*. These relations are illustrated in Figure 15 and Figure 16.

$\sigma_1\sigma_3 = \sigma_3\sigma_1$

FIGURE 15. Far Commutativity



$\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$ $\qquad$ $\sigma_2\sigma_3\sigma_2 = \sigma_3\sigma_2\sigma_3$

FIGURE 16. Braid Relations

The result proved by Artin is as follows (see Proposition 10.3 of [BZ] or Theorem 5.5 of [PraSos]).

Theorem 1.2. *(Artin's Theorem) The braid group $\mathcal{B}_n$ is isomorphic to the abstract group B_n with $n-1$ generators $g_1, \ldots, g_{n-1}$ subject to the relations*

$$g_i g_j = g_j g_i \quad \textit{whenever} \quad |i-j| \geq 2, \quad i, j = 1, \ldots, n-1$$
$$g_i g_{i+1} g_i = g_{i+1} g_i g_{i+1} \quad \textit{for all} \quad i = 1, \ldots, n-2.$$

It follows, for example, that $\mathcal{B}_2$ is isomorphic to $\mathbb{Z}$ and $\mathcal{B}_3$ has two generators x and y subject to only one relation $xyx = yxy$. There are many other ways to describe the braid group, but for these we will simply refer to [G-M].

The relationship between braids and links is expressed in a classic theorem of Alexander [Alex1]. Notice that if ω is any geometric braid (such as the one in Figure 17 for example) and if one chooses an arc joining a_i and b_i for each $i = 1, \dots, n$ in the manner shown in Figure 17, then the result is a link (perhaps a knot). This link is called the *closure* of the braid and any link that arises in this way is called a *closed braid*. Notice that any such link is naturally oriented by, for example, following the strand from a_i to b_i and then back to a_i along the inserted arc. The theorem of Alexander asserts that *every* link is a closed braid. For proofs of the following result we refer to [Alex1], Theorem 2 of [BB], or Theorem 6.5 of [PraSos].

Theorem 1.3. *(Alexander's Theorem) Every link (in particular, every knot) is isotopic to some closed braid.*

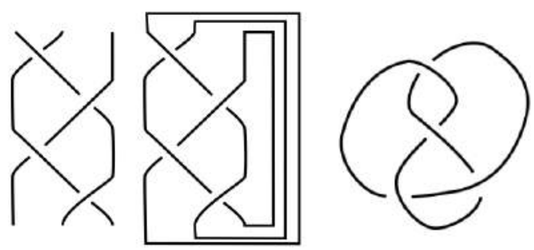


FIGURE 17. The Figure Eight Knot as a Closed Braid

Figure 17 exhibits a braid whose closure is the figure eight knot, but there are many other braids with the same closure. In general, the map that assigns to each braid its closure is many-to-one. For example, the unknot can be represented by $\sigma_1\sigma_2\cdots\sigma_{n-1} \in \mathcal{B}_n$ *for any* $n \geq 2$. Indeed, one can show that conjugate elements of $\mathcal{B}_n$ have isotopic closures. The question then arises as to how one can tell if two braids represent the same link. We will not pursue this here, but for those who are interested we mention that an answer was obtained by Markov [Mark] in 1935 who showed that two braids ω_1 and ω_2 in $\mathcal{B}_n$ have isotopic closures if ω_2 can be obtained from ω_1 by a sequence of the following two *Markov moves*.

$$\text{Replace } \omega_1 \text{ by } \omega\,\omega_1\,\omega^{-1} \text{ for some } \omega \in \mathcal{B}_n. \tag{3}$$

$$\text{Replace } \omega_1 \text{ by } \omega_1\,\sigma_n^{\pm 1} \in \mathcal{B}_{n+1}, \text{ or the inverse of this operation.} \tag{4}$$

Remark 1.9. We mentioned earlier that the theory of braids is closely related not only to knots and links, but to physics as well. The first suggestion of a connection with physics comes from the braid relations

$$\sigma_i\,\sigma_{i+1}\,\sigma_i = \sigma_{i+1}\,\sigma_i\,\sigma_{i+1} \quad \text{for all} \quad i = 1, \dots, n-2$$

for the braid group $\mathcal{B}_n$. As it happens these are formally identical to a special case of the so-called *Yang-Baxter equations* which, since the mid 1960s, have been the center of attention in statistical mechanics (see

Section 2.3 of [Jimbo]). Since this observation was made the interaction between statistical mechanics and knot theory has been pursued with a great deal of vigor and has yielded many deep and beautiful results. We will discuss one of these in Section 1.2. For those interested in pursuing this we suggest [Wu1], [Wu2], [Jones4], and [Tur1]. The canonical reference for the relevant statistical mechanical models is [Bax].

1.2. **The Jones Polynomial.** In 1928 J.W. Alexander [Alex2] introduced an invariant of oriented knots and links that, to a large degree, inaugurated *knot theory* as a branch of topology. The invariant associated to a link L is a Laurent polynomial $\Delta_L(t)$ with integer coefficients in the formal variable $\sqrt{t}$, that is, an integer polynomial in $t^{1/2}$ and $t^{-1/2}$. This is now universally known as the *Alexander polynomial* of L. It is determined only up to multiples of $\pm t^{k/2}$ for some positive integer k so one often finds $\Delta_L(t)$ written in a variety of forms.

There are a number of algebraic, combinatorial, and geometric ways to approach the definition of $\Delta_L(t)$. [Alex2] contains a detailed discussion of Alexander's original approach via the algebraic topology of the infinite cyclic cover of the knot complement and one can consult [Scher] for this and several other approaches. All of these, however, give rise to rather tedious calculations when applied to any particular link or knot. This changed with an observation of Conway (see [Con]) that we will briefly describe. Toward the end of Alexander's paper [Alex2] there is a section of "Miscellaneous Theorems" describing various properties of $\Delta_L(t)$. One of these is a so-called *skein relation* that apparently went unnoticed for quite some time, but has since become a focal point for the definition and application of polynomial invariants. This relation can be described as follows. Focus attention on one particular crossing in a diagram D for L. The skein relation compares the Alexander polynomial of three related links, denoted L_-, L_0, and L_+, that are identical to L except perhaps near the crossing in question where, in some neighborhood, their diagrams look like one of the pictures in Figure 18 or a rotation of one of these. Choosing the normal vector to the plane containing D appropriately (out of the page in Figure 18), L_- has sign -1 and L_+ has sign $+1$. As we observed earlier this is unaffected by a reversal of the orientation of the link.

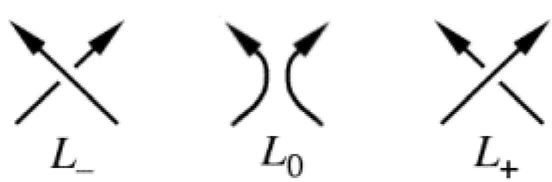


FIGURE 18. Skein Relation Diagrams

The skein relation satisfied by the Alexander polynomial is

$$\Delta_{L_+}(t) - \Delta_{L_-}(t) = (t^{1/2} - t^{-1/2})\,\Delta_{L_0}(t). \tag{5}$$

Conway showed that this skein relation, together with the normalization

$$\Delta_{\text{unknot}}(t) = 1, \tag{6}$$

completely determines one of the Alexander polynomials (recall that these are determined only up to multiples of $\pm t^{k/2}$). Although one must choose an orientation for L in order to define $\Delta_L(t)$, the polynomial itself is independent of how the orientation is chosen.

With this one can compute the Alexander polynomial of any link recursively. We will illustrate with the example of the right-handed trefoil (see Figure 3). First, however, a much simpler application of (5) and (6). The first row in Figure 19 shows three knot diagrams for the unknot. The knots in the second row of Figure 19 we denote L_+, L_-, and L_0 from left to right. L_0 is the 2-component unlink (two unlinked, unknotted circles). We conclude from (5) and (6) that $1 - 1 = (t^{1/2} - t^{-1/2})\Delta_{L_0}(t)$ so $\Delta_{L_0}(t) = 0$. The Alexander polynomial for the 2-component unlink is identically zero.

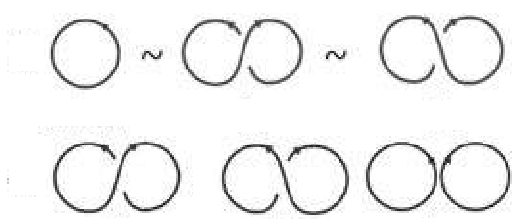


FIGURE 19. 2-Component Unlink

Now we turn to the right-handed trefoil. This is the diagram at the top of Figure 20 and we will denote it L^1_+, focusing our attention on the circled crossing. The second row in Figure 20 are, from left to right, L^1_- and L^1_0. The first is the unknot and the second is the so-called *Hopf link*. We conclude that

$$\Delta_{L^1_+}(t) - 1 = (t^{1/2} - t^{-1/2})\Delta_{L^1_0}(t).$$

To compute the Alexander polynomial for the Hopf link we focus our attention on the circled crossing in L^1_0 and rename that link L^2_+. The third row in Figure 20 is then, from left to right, L^2_- and L^2_0. The first is the 2-component unlink and the second is the unknot. We conclude that

$$\Delta_{L^2_+}(t) - 0 = (t^{1/2} - t^{-1/2})1.$$

so the Alexander polynomial of the Hopf link is

$$\Delta_{L^2_+}(t) = t^{1/2} - t^{-1/2}.$$

Combining these we find that the Alexander polynomial of the right-handed trefoil is $(t^{1/2} - t^{-1/2})(t^{1/2} - t^{-1/2})$, that is,

$$\Delta_{\text{right-handed trefoil}}(t) = t - 1 + t^{-1}.$$

Since the Alexander polynomial is a knot invariant we conclude from this that the right-handed trefoil is genuinely knotted, that is, not equivalent to the unknot. An analogous calculation for the left-handed trefoil gives the same result. Even so, we will see shortly that these two trefoils are not isotopic. The Alexander polynomial cannot, in general, distinguish a knot from its mirror image when these two fail to be isotopic. We will see that the Jones polynomial is more discerning.

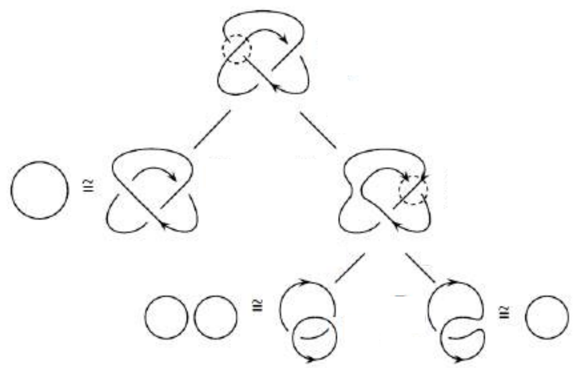


FIGURE 20. Skein Tree for the Right-Handed Trefoil

For over 50 years the Alexander polynomial remained the only known polynomial invariant for oriented knots and links. Then, in 1985, Vaughan Jones [Jones3] announced the discovery of another polynomial invariant that is now universally known as the *Jones polynomial* and denoted $V_L(t)$. For this discovery Jones was awarded the Fields Medal in 1990. We will briefly describe the path that led to this discovery as it is described in [Jones3] and then look at a few examples.

This path followed by Jones was somewhat unorthodox in that his research interests centered around von Neumann algebras rather than knot theory (see [Jones1] and [Jones2]). Recall that a von Neumann algebra is a *-algebra of bounded operators on a complex Hilbert space that contains the identity operator and is closed in the weak operator topology. Finite-dimensional von Neumann algebras are all products of matrix algebras for which the *-operation is conjugate transpose. Jones was led to consider a certain family of such finite-dimensional von Neumann algebras J_n (see Section 4 of [Jones1]). These were generated by the identity element 1 together with elements $e_1, \ldots, e_{n-1}$ that satisfy the following relations for some complex number t.

(1) $e_i^2 = e_i$ and $e_i^* = e_i, \quad i = 1, \ldots, n-1$
(2) $e_i e_j = e_j e_i, \quad i, j = 1, \ldots, n-1$ and $|i - j| \geq 2$
(3) $e_i e_{i+1} e_i = t/(1+t)^2 e_i, \quad i = 1, \ldots, n-2$
(4) $e_i e_{i-1} e_i = t/(1+t)^2 e_i, \quad i = 2, \ldots, n-1$

Relation (1) simply says that each e_i is a projection in the algebra J_n. Now notice that if we let

$$g_i = \sqrt{t}\,(\,(1+t)e_i - 1\,), \quad i = 1, 2, \ldots, n-1,$$

then relations (2), (3) and (4) together with $e_i^2 = e_i$ and a bit of algebra give

$$g_i g_j = g_j g_i, \quad i, j = 1, \ldots, n-1 \text{ and } |i - j| \geq 2,$$

and

$$g_i g_{i+1} g_i = g_{i+1} g_i g_{i+1}, \quad i = 1 \ldots, n-2$$

and that these are precisely the defining relations of the braid group $\mathcal{B}_n$ (see Theorem 1.2). Consequently, we have, for each t, a representation R_t of $\mathcal{B}_n$ defined by setting

$$R_t(\sigma_i) = g_i, \quad i = 1, 2, \ldots, n-1.$$

Thus we see braids, and therefore knots and links, begin to enter the picture.

Remark 1.10. One can see other things beginning to enter the picture as well. The *Jones algebra* J_n is essentially equivalent to what is called the *Temperley-Lieb algebra*. This was introduced in [TL] to prove the equivalence of two models in *statistical mechanics* (for more on this see [Jones4], [Jones5], [Jones6] and [Bax]). One should also be aware of the closely related *Iwahori Hecke algebra* which plays a prominent role in [Jones4].

Jones observes next that, if t is either a positive real number or one of the roots of unity $e^{2\pi i/k}, k = 3, 4, 5, \ldots$, then there is an arbitrarily large family $\{e_1, e_2, \ldots\}$ of projections such that the first $n-1$ of them determine such an algebra J_n. Then $J_n \subseteq J_{n+1}$, $n = 0, 1, \ldots$, where, by definition, we set $J_0 = \mathbb{C}$. Moreover, for such t, there is defined on each J_n a trace function

$$\text{tr} : J_n \to \mathbb{C}$$

that is uniquely characterized by the following properties.

(1) $\text{tr}\,(1) = 1$
(2) $\text{tr}\,(ab) = \text{tr}\,(ba)$ for all $a, b \in J_n$
(3) $\text{tr}\,(a^*a) > 0$ if $a \in J_n$ and $a \neq 0$
(4) $\text{tr}\,(ae_{n+1}) = t/(1+t)^2 \text{tr}\,(a)$ for all $a \in J_n$

The key result upon which the definition of the Jones polynomial is based is obtained by combining the representations R_t and the traces tr in the following way. Let ω be an element of the braid group $\mathcal{B}_n$. The closure of the braid ω is an oriented link (or knot) which we will denote ω^-. Furthermore, for each t, $R_t(\omega)$ is an element of J_n. Assuming that t is either a positive real number or a root of unity $R_t(\omega)$ has a trace $\text{tr}\,(R_t(\omega)) \in \mathbb{C}$ and Jones proves that the number

$$(-(1+t)/\sqrt{t})^{n-1} \text{tr}\,(R_t(\omega))$$

depends only on the isotopy class of the closed braid ω^-. Now, according to Alexander's Theorem 1.3, every link L is ω^- for some braid ω in some $\mathcal{B}_n$ and this gives rise to the following definition. Let L be an oriented link. Then there exists a positive integer n and an element ω in the braid group $\mathcal{B}_n$ such that L is isotopic to ω^-. If t is either a positive real number or one of the roots $e^{2\pi i/k}, k = 3, 4, 5, \ldots$, of unity we define the *Jones polynomial* of L by

$$V_L(t) = (-(1+t)/\sqrt{t})^{n-1} \text{tr}\,(R_t(\omega)).$$

This is an invariant of the isotopy type of L. Just as for the Alexander polynomial, L must be an oriented link in order to define the invariant $V_L(t)$, but the polynomial itself is independent of how the orientation is chosen.

The existence of the *-operation on J_n has played no role up to this point, but Jones points out that with it one can extend the definition of $V_L(t)$ to all complex values of t except 0. He then describes the structure of $V_L(t)$ as a function of t.

Theorem 1.4. *If the link L has an odd number of components (in particular, if L is a knot), then $V_L(t)$ is a Laurent polynomial in t with integer coefficients. If L has an even number of components, then $V_L(t)$ is $\sqrt{t}$ times a Laurent polynomial in t with integer coefficients.*

Jones proves a great many interesting and useful properties of $V_L(t)$, but we will focus on just a few that we will put to use in our examples. The first gives an explicit relationship between the Jones polynomial of an oriented link L and that of its mirror image which we will denote $\hat{L}$. Specifically, we have the following.

$$V_{\hat{L}}(t) = V_L(1/t)$$

Consequently, a link L that is isotopic to its mirror image must have a Jones polynomial that satisfies $V_L(t) = V_L(1/t)$. In particular, if $V_L(t)$ does not have this $t \leftrightarrow 1/t$ symmetry, then L cannot be isotopic to its mirror image. We will see that this is the case for the trefoil so that the right-handed trefoil and the left-handed trefoil are not the same despite the fact that the Alexander polynomial failed to distinguish them.

Just as in the case of the Alexander polynomial, the definition of the Jones polynomial is computable, but rather unwieldy. One would like to see a skein relation for the Jones polynomial analogous to (5) which reduces the problem of computing $V_L(t)$ to a recursive, combinatorial one. Jones provides this in Theorem 12 of [Jones3]. Referring again to Figure 18 the result is

$$(1/t)V_{L_+}(t) - tV_{L_-}(t) = (t^{1/2} - t^{-1/2})V_{L_0}(t). \tag{7}$$

The skein relation (7) by itself, of course, does not determine the Jones polynomial of a link completely since one must get the recursion started by knowing $V_{\text{unknot}}(t)$. The usual normalization is the same as in the case of the Alexander polynomial, that is,

$$V_{\text{unknot}}(t) = 1. \tag{8}$$

With this we can compute the Jones polynomial of the right-handed trefoil in precisely the same way we computed its Alexander polynomial earlier. First, referring to Figure 19, we obtain from the skein relation (7) that

$$(1/t)1 - t1 = (t^{1/2} - t^{-1/2})V_{\text{2-component unlink}}(t)$$

so

$$V_{\text{2-component unlink}}(t) = -(t^{1/2} + t^{-1/2}).$$

Remark 1.11. Note that one can continue inductively and show that the Jones polynomial for the *p-component unlink* (p unknotted, unlinked circles) is given by

$$V_{\text{p-component unlink}}(t) = [\,-(t^{1/2} + t^{-1/2})\,]^{p-1}. \tag{9}$$

Now we refer to Figure 20. The top two rows give

$$(1/t)V_{\text{right-handed trefoil}}(t) - tV_{\text{unknot}}(t) = (t^{1/2} - t^{-1/2})V_{\text{Hopf link}}(t)$$

so

$$V_{\text{right-handed trefoil}}(t) = t^2 + (t^{3/2} - t^{1/2})V_{\text{Hopf link}}(t).$$

From the bottom two rows we obtain

$$(1/t)V_{\text{Hopf link}}(t) - tV_{\text{2-component unlink}}(t) = (t^{1/2} - t^{-1/2})V_{\text{unknot}}(t)$$

so

$$V_{\text{Hopf link}}(t) = -(t^{5/2} + t^{1/2})$$

and therefore

$$V_{\text{right-handed trefoil}}(t) = -t^4 + t^3 + t.$$

Since this polynomial is not invariant under $t \to 1/t$, we conclude that the right-handed and left-handed trefoils are *not* isotopic.

Remark 1.12. More information about the Jones polynomial is available in a great many sources, but a good place to start is [Jones5] where one can find a calculation of $V_{T_{p,q}}(t)$ for all of the torus knots. We should also point out that, shortly after the introduction of the Jones polynomial, generalizations were discovered that contained both the Alexander and Jones polynomials as special cases (see [HOMFLY] and [Jones4]). These are referred to as the HOMFLY polynomials and are denoted P_L. The defining relations can be expressed in a variety of ways, but the normalization is always the same as in the case of the Alexander and Jones polynomials, that is, $P_{\text{unknot}} = 1$. One can think of P_L as a homogeneous Laurent polynomial of degree zero in three variables x, y and z satisfying the skein relation

$$xP_{L_+}(x,y,z) + yP_{L_-}(x,y,z) + zP_{L_0}(x,y,z) = 0. \tag{10}$$

For example, one can proceed exactly as we did for the Alexander and Jones polynomials to show that

$$P_{\text{2-component unlink}}(x,y,z) = -(x+y)/z$$

and

$$P_{\text{right-handed trefoil}}(x,y,z) = y^{-2}z^2 - 2xy^{-1} - x^2y^{-2}.$$

Alternatively, one can think of P_L as a nonhomogeneous Laurent polynomial in two variables. A common way to do this is to define $\mathcal{P}_L(l,m) = P_L(l, l^{-1}, m)$ in which case the skein relation becomes

$$l\,\mathcal{P}_{L_+}(l,m) + l^{-1}\mathcal{P}_{L_-}(l,m) + m\,\mathcal{P}_{L_0}(l,m) = 0. \tag{11}$$

The Alexander polynomial is then recovered by taking $l = i$ and $m = i(t^{1/2} - t^{-1/2})$ while the Jones polynomial corresponds to $l = it^{-1}$ and $m = i(t^{1/2} - t^{-1/2})$. We will make only a few comments about these generalizations in the sequel and so will simply refer to [HOMFLY] and [Jones4] for more details.

There is one somewhat unsatisfying aspect of the Alexander, Jones and HOMFLY polynomials. Knots and links are intrinsically 3-dimensional objects. There are no knots in $\mathbb{R}$ and only the unknot lives in $\mathbb{R}^2$. Moreover, in $\mathbb{R}^n$ for $n \geq 4$, one can use the extra dimensions to unknot any embedded circle so that every "knot" is an unknot (see [Gluck] and, for generalizations, [Zeeman] and [Lev]).

"*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.*"

- Edward Witten [Witt2]

Witten, formally at least, solved the problem of providing an intrinsically 3-dimensional definition of the Jones polynomial by identifying it with the expectation value of a certain observable called a *Wilson line* in a version of quantum field theory whose action is the so-called *Chern-Simons functional*. The solution is quite complex and draws heavily on many areas of mathematics and physics. In the next three sections we will very briefly describe some of the ideas involved.

1.3. **Classical Chern-Simons Theory.** Chern-Simons theory is a *gauge theory* so, as motivation, we will begin with a brief synopsis of the mathematical ingredients required to describe, at the semi-classical level, the interaction of a particle with a gauge field. One might keep in mind the interaction of a charged particle with an electromagnetic field (see Sections 2.3 and 2.4 of [Nab3]). Here is the semi-classical picture of the required ingredients.

(1) A smooth, oriented manifold M.

The particle "lives" in M. Depending on the problem at hand, M may or may not be equipped with a Riemannian or semi-Riemannian metric.

(2) A finite-dimensional vector space $\mathcal{V}$.

The particle has a wave function ψ that takes values in $\mathcal{V}$. The choice of $\mathcal{V}$ is dictated by the internal structure of the particle (for example, its spin) and so $\mathcal{V}$ is called the *internal space*. Typical examples are $\mathbb{C}$, $\mathbb{C}^2$, $\mathbb{C}^4$, or the Lie algebra of some Lie group (for example, $\mathfrak{u}(1)$, or $\mathfrak{su}(2)$). $\mathcal{V}$ is equipped with an inner product $\langle\ ,\ \rangle$ with which one computes squared norms and thereby the probabilities with which quantum theory deals.

(3) A matrix Lie group G and a representation

$$\rho : G \to \mathrm{GL}(\mathcal{V})$$

of G on $\mathcal{V}$ that is orthogonal with respect to the inner product on $\mathcal{V}$, that is,

$$\langle\, \rho(g)(v),\ \rho(g)(w)\,\rangle = \langle\, v, w\,\rangle$$

for all $g \in G$ and all $v, w \in \mathcal{V}$. This will generally be one of the classical groups (for example, U(1), SU(2), SL(2, $\mathbb{C}$) or a product of these) and plays a dual role.

(a) The inner product on $\mathcal{V}$ determines a class of orthonormal bases, or "frames" for $\mathcal{V}$ and these are related by the elements of G. More precisely, $g \in G$ acts, on the right, on a frame p to give a new frame $p \cdot g$. By fixing some "standard" frame at the outset one can therefore identify the elements of G with the frames.

(b) G also acts on $\mathcal{V}$, on the left, via the representation ρ

$$v \to g \cdot v = \rho(g)(v)$$

and so it acts on the wave function ψ at each point. This is essentially just a change of coordinates. If $\psi(p)$ is the value of the wave function, described relative to the frame p, then

$$\psi(p \cdot g) = g^{-1} \cdot \psi(p)$$

is its value when described relative to the frame $p \cdot g$.

(4) A smooth principal G-bundle over M.

$$G \hookrightarrow P \xrightarrow{\pi} M$$

Typical examples include trivial bundles such as $\mathrm{SU}(2) \hookrightarrow \mathbb{S}^3 \times \mathrm{SU}(2) \to \mathbb{S}^3$ and Hopf bundles such as $\mathrm{U}(1) \hookrightarrow \mathbb{S}^3 \to \mathbb{S}^2$ (see Example 3, pages 24-25, of [Nab3]). At each $x \in M$ the fiber $\pi^{-1}(x)$ above x is a copy of G, thought of as the set of frames in the internal space $\mathcal{V}$ at x. A local cross section $s : U \to \pi^{-1}(U) \subseteq P$ is a smooth selection of a frame at each point in some open subset U of M relative to which the wave function ψ can be described on U. Such a local cross section is called a local *gauge* and is equivalent to a local trivialization of the bundle (see pages 220-221 of [Nab2]).

(5) A connection ω on $G \hookrightarrow P \xrightarrow{\pi} M$ with curvature $\Omega = d\omega + \omega \wedge \omega$.

The connection ω is what physicists would call a *gauge field*. If $s : U \to \pi^{-1}(U)$ is a local gauge, then the pullback $A = s^*\omega$ is called a local *gauge potential* and $F_A = s^*\Omega$ is a local *gauge field strength*. Whereas ω and Ω are globally defined $\mathfrak{g}$-valued forms on P, these gauge potentials and gauge field strengths are, in general, only locally defined on X since nontrivial principal bundles do not admit global cross sections and pullbacks by different local cross sections generally do not agree on the intersection of their domains.

Remark 1.13. Assuming, as we may, that the domain U of s is a local coordinate neighborhood for M with coordinates $x^1, \ldots, x^n$, we can write A and F_A on U as

$$A = A_\alpha dx^\alpha,$$

and

$$F_A = \frac{1}{2}(\,\partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta]\,)\, dx^\alpha \wedge dx^\beta,$$

where each A_α, $\alpha = 1, \ldots, n$, is a matrix of 1-forms on U and $\partial_\alpha = \partial/\partial x^\alpha$ is computed entrywise. Now let $s_j : U_j \to \pi^{-1}(U_j)$ and $s_i : U_i \to \pi^{-1}(U_i)$ be two local cross sections with $U_j \cap U_i \neq \emptyset$ and let $g_{ij} : U_j \cap U_i \to G$ be the corresponding transition function so that $s_j(x) = s_i(x) \cdot g_{ij}(x)$ for each $x \in U_j \cap U_i$. Then the corresponding gauge potentials $A_j = s_j^* \omega$ and $A_i = s_i^* \omega$ are related by

$$A_j = g_{ij}^{-1}\, A_i\, g_{ij} + g_{ij}^{-1}\, dg_{ij},$$

where dg_{ij} is the entrywise differential of g_{ij}. The gauge field strengths $F_{A_j} = s_j^* \Omega$ and $F_{A_i} = s_i^* \Omega$ are related by

$$F_{A_j} = g_{ij}^{-1}\, F_{A_i}\, g_{ij}.$$

Such a change of section is called a *local gauge transformation*.

(6) A global cross section of the vector bundle $P \times_\rho \mathcal{V}$ associated to $G \hookrightarrow P \xrightarrow{\pi} M$ by the representation ρ or, equivalently, a $\mathcal{V}$-valued map $\phi : P \to \mathcal{V}$ on P that is equivariant with respect to the right action of G on P and the left action of G on $\mathcal{V}$, that is,

$$\phi(p \cdot g) = g^{-1} \cdot \phi(p).$$

Particles coupled to (that is, experiencing the effects of) the field determined by ω have locally defined wave functions taking values in $\mathcal{V}$ that are obtained by solving appropriate field equations (see (8) below). A change of gauge ($s \to s \cdot g$) changes the local wave function by the representation ($\psi \to g^{-1} \cdot \psi$). Because of this these local wave functions piece together into a globally defined cross section of the the associated vector bundle $P \times_\rho \mathcal{V}$. This globally defined object $\psi : M \to P \times_\rho \mathcal{V}$ is the *wave function* of the particle coupled to the gauge field. That these cross sections are in one-to-one correspondence with equivariant $\mathcal{V}$-valued maps on P is Exercise 6.8.4 of [Nab2].

(7) A non-negative, smooth, real-valued function

$$V : \mathcal{V} \to \mathbb{R}$$

that is invariant under the action of G on $\mathcal{V}$, that is,

$$V(g \cdot v) = V(v)$$

for all $v \in \mathcal{V}$.

V is to be regarded as a potential function with $V \circ \phi$ describing the self-interaction of the particle field ϕ. Typically, these depend only on $\|\phi(x)\|$, for example, $\frac{1}{2} m \, \|\phi(x)\|^2$, or $\frac{\lambda}{8}(\,\|\phi(x)\|^2 - 1\,)^2$, where m and λ are non-negative constants.

(8) An *action functional* $S(\omega, \phi)$ that associates a real number with each pair (ω, ϕ) and whose stationary points are the physically significant configurations (ω, ϕ).

This is the so-called *Principle of Least Action*. The Calculus of Variations provides necessary conditions in the form of the Euler-Lagrange equations that must be satisfied by these stationary points. These are the appropriate *field equations*. The action $S(\omega, \phi)$ is often expressed as an integral over M of terms involving various squared norms of the gauge field strength, the covariant derivative of ϕ with respect to ω and the potential $V \circ \phi$. For physical reasons one is generally only interested in *finite action* solutions to the field equations, that is, those for which $S(\omega, \phi) < \infty$. When M is compact this is assured, but otherwise one must impose various asymptotic conditions on the terms in the integral. These asymptotic conditions generally have rather remarkable topological implications leading, for example, to the existence of *topological charges* (see Section 2.5 of [Nab3]).

Remark 1.14. Chapter 2 of [Nab3] contains four very detailed and physically significant examples in which all of the elements that we have just enumerated are spelled out carefully. Notice that it is a simple matter to extend the scenario we have just described to include the interaction of any finite number of particles with a gauge field. This is the context we will find ourselves in in Section 1.5 when we attempt to motivate Witten's interpretation of the Jones polynomial.

Our avowed intention in the preceding discussion was to describe the mathematical context in which one views the semi-classical interaction of a particle with a gauge field. Nevertheless, we should point out that the scenario we have described applies equally well to what are called *pure gauge theories* in which there is no particle at all and one is interested only in isolating a class of gauge fields appropriate to some particular problem. Somewhat more precisely, one would begin with a smooth, oriented manifold M, a matrix Lie group G and a smooth principal G-bundle $G \hookrightarrow P \xrightarrow{\pi} M$ over M. Let ω be a connection on the bundle. The physical or mathematical problem at hand will suggest an action functional $S(\omega)$ to be associated with ω and then the finite action stationary points of $S(\omega)$ will be the gauge fields appropriate to the problem. One such pure gauge theory has had a profound impact on the differential topology of 4-manifolds due to the work of Simon Donaldson. This is *SU(2)-Yang-Mills theory*. Here M is taken to be a compact, oriented, Riemannian 4-manifold, the Lie group G is SU(2) and $\mathrm{SU}(2) \hookrightarrow P \xrightarrow{\pi} M$ is the SU(2)-bundle over M with Chern number 1. For any connection ω on this bundle, $S(\omega)$ is taken to be the *Yang-Mills action*

$$\mathcal{YM}(\omega) = -\int_M \mathrm{tr}\,(F_A \wedge {}^*F_A),$$

where * is the Hodge dual determined by the given orientation and Riemannian metric. The Euler-Lagrange equations for this action are the *Yang-Mills equations*

$$d^A\, {}^*F_A = 0,$$

where d^A is the covariant exterior derivative with respect to the connection (for more on this see Section 6.3 of [Nab2]). We would like to turn now to another pure gauge theory, its quantization and its relation to the topological theory of knots, links and 3-manifolds. This is called *Chern-Simons theory* and we now proceed with its construction.

We will take M to be any compact, connected, oriented, smooth 3-manifold. Witten has done much more than exhibit an intrinsically 3-dimensional view of the classical Jones polynomial. The procedures described in [Witt2] apply equally well to embedded circles in any such 3-manifold and suggest a multitude of generalizations. Moreover, by taking the link to be empty the technique yields an invariant of the 3-manifolds themselves.

For G we could take any compact, simple Lie group, but we will generally restrict our attention to the case of most interest to us, that is,

$$G = \mathrm{SU}(2). \tag{12}$$

Notice that SU(2), being diffeomorphic to $\mathbb{S}^3$, is connected and simply connected so $\pi_0(\mathrm{SU}(2))$ and $\pi_1(\mathrm{SU}(2))$ are both trivial. Since $\pi_k(\mathbb{S}^n)$ is trivial whenever $k < n$ (see Corollary 4.9 of [Hatch]), $\pi_2(\mathrm{SU}(2))$ is also trivial. From this it follows from a classifying space argument that any principal SU(2)-bundle over the 3-manifold M is trivial. In the special case in which $M = \mathbb{S}^3$ we could also appeal to the following Classification Theorem for principal bundles over spheres (see Theorem 4.4.3 of [Nab2] for the C^0 case and Section 3.2 of [Nab3] for the extension to the smooth case).

Theorem 1.5. *(Classification Theorem) Let G be a connected Lie group. Then the set of equivalence classes of smooth principal G-bundles over $\mathbb{S}^n$, $n \geq 2$, is in one-to-one correspondence with the elements of the homotopy group $\pi_{n-1}(G)$.*

We will trivialize at the outset and take the principal SU(2)-bundle of our gauge theory to be

$$\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \xrightarrow{\pi} M, \tag{13}$$

where π is the projection onto the first factor and the right SU(2)-action is given by $(x, g) \cdot g' = (x, gg')$ for all $x \in M$ and all $g, g' \in \mathrm{SU}(2)$. Being trivial this bundle has global sections, for example,

$$s : M \to M \times \mathrm{SU}(2)$$
$$s(x) = (x, 1)$$

where, for convenience, we write 1 for the identity element of SU(2). Any other global section is of the form

$$s^g(x) = s(x) \cdot g(x) = (x, g(x))$$

for some smooth function $g : M \to \mathrm{SU}(2)$ so that global gauge transformations can be fully identified with such smooth functions and connections ω on the bundle are identified with globally defined gauge potentials $A = s^*\omega$ on M. The next item to be specified is an appropriate action functional $S(A)$. Since the definition we have in mind may look a bit strange at first glance we will begin with some motivation.

Recall first that if $\mathrm{SU}(2) \hookrightarrow P \xrightarrow{\pi} M$ is an SU(2)-bundle over a smooth manifold M, then the 2^{nd} *Chern class* of the bundle is a cohomology class $c_2(P) \in H^4(M; \mathbb{R})$ defined in the following way. Choose an arbitrary connection ω on the bundle with curvature Ω. A trivializing open cover of X gives rise to a family of local gauge potentials A and local gauge field strengths F_A on M. A change of gauge changes the F_A by conjugation and the trace is invariant under conjugation so the family $\mathrm{tr}\,(F_A \wedge F_A)$ determines a globally defined, closed 4-form on M that we will continue to denote $\operatorname{tr}(F_A \wedge F_A)$. The corresponding cohomology class is denoted $[\operatorname{tr}(F_A \wedge F_A)]$ and

$$c_2(P) = \frac{1}{8\pi^2}[\operatorname{tr}(F_A \wedge F_A)].$$

If M is a compact, oriented 4-manifold, then SU(2)-bundles over M are classified up to equivalence by $c_2(P)$. This is Theorem E.5 of [FU1] (a proof in the special case of $M = \mathbb{S}^4$ is available pages 328-334 of [Nab3]). The following is Lemma 6.4.1 of [Nab3].

Theorem 1.6. *Let $SU(2) \hookrightarrow P \xrightarrow{\pi} M$ be a principal SU(2)-bundle over the smooth manifold M, ω a connection on it with curvature Ω, $s : U \to \pi^{-1}(U) \subseteq P$ a local section and $A = s^*\omega$ and $F_A = s^*\Omega$ the corresponding local gauge potential and field strength. Then, on U,*

$$tr(F_A \wedge F_A) = d\,(\,tr\,(A \wedge dA + \frac{2}{3}A \wedge A \wedge A\,)).$$

Now suppose M is our 3-manifold so that the bundle is trivial. Let $s : M \to M \times \mathrm{SU}(2)$ be a *global* section. Since $\operatorname{tr}(F_A \wedge F_A)$ is a 4-form on a 3-manifold, it must be zero. Consequently, the globally defined 3-form

$$\operatorname{tr}(A \wedge dA + \frac{2}{3}A \wedge A \wedge A) \tag{14}$$

on M is closed. This is called the *Chern-Simons form* for the gauge potential A and with it we define the *Chern-Simons action* or *Chern-Simons functional* $S_{CS}(A)$

$$S_{CS}(A,k) = \frac{k}{4\pi}\int_M \operatorname{tr}(A \wedge dA + \frac{2}{3}A \wedge A \wedge A), \tag{15}$$

where k is a positive integer parameter called the *level*. The 3-manifold M, group (12), bundle (13), and action (15) specify all of the ingredients required for a pure gauge theory that we will refer to simply as *(classical) Chern-Simons theory*; more precisely one would call this *(classical) SU(2)-Chern-Simons theory on M*, but SU(2) and M will be fixed throughout our discussion. There is a great deal to be said about this gauge theory (see, for example, [Freed]) and particularly about its quantization. We will describe some of this in the sequel, but will close this section with just a few observations that we will need.

Remark 1.15. We should point out that the level k will play no particular role in the classical Chern-Simons theory, but will be crucial in the quantum theory that we will briefly sketch in Section 1.5. One finds that k plays the role of the reciprocal of a coupling constant (one might think of it as $1/\hbar$) so that the classical limit of quantum Chern-Simons corresponds to $k \to \infty$. It is also important to observe at this point that, unlike the Yang-Mills action $\mathcal{YM}$, the definition of S_{CS} *does not require the existence of a metric on M*. One final comment is that, whereas we have chosen to consider the non-Abelian gauge group SU(2) in order to describe Witten's theory, the Abelian case in which $G = \mathrm{U}(1)$ was considered much earlier by Schwarz [Schw1] and shown to be related to Ray-Singer torsion.

The first observation is that the Chern-Simon functional is invariant under orientation preserving diffeomorphisms of M. More precisely, what this means is the following. Let $\varphi : M \to M$ be an orientation preserving diffeomorphism of M and $A = s^*\omega$ a global gauge potential on M. Then $A' = \varphi^*A$ is an $\mathfrak{su}(2)$-valued 1-form on M. It is also a gauge potential for the connection induced by ω on the pullback bundle $\varphi^*(M \times \mathrm{SU}(2))$, which is also the trivial SU(2)-bundle over M. The claim is that

$$S_{CS}(\varphi^*A, k) = S_{CS}(A, k)$$

and this follows directly from the invariance of the integral under orientation preserving diffeomorphisms and the fact that pullback commutes with $\wedge$ and d.

Next we discuss gauge invariance or, rather, the lack thereof for the Chern-Simons action. In general, local sections of principal G-bundles are in one-to-one correspondence with local trivializations of the bundle (see pages 220-221 of [Nab2]). For this reason gauge potentials $A = s^*\omega$ and gauge field strengths $F_A = s^*\Omega$ are to be regarded as coordinate expressions for the gauge field ω and its curvature Ω, respectively. A local gauge transformation relates one such coordinatization to another ($A \to A^g = g^{-1}Ag + g^{-1}dg$ and $F_A \to F_{A^g} = g^{-1}F_A g$). In the Yang-Mills case, since F_A transforms by conjugation, so does $F_A \wedge {}^*F_A$ and therefore $\mathrm{tr}\,(F_A \wedge {}^*F_A)$ is gauge invariant. It follows that the Yang-Mills action $\mathcal{YM}(A)$ is gauge invariant and so, from the point of view of physics, represents a physically meaningful, coordinate independent characteristic of the gauge field. Regrettably, this is *not quite* true of the Chern-Simons action $S_{CS}(A, k)$ and we will need to understand precisely how $S_{CS}(A, k)$ responds to a gauge transformation.

For the trivial bundle $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \xrightarrow{\pi} M$ a global gauge transformation is identified with a smooth map $g : M \to \mathrm{SU}(2)$. Physicists distinguish two types of gauge transformations. A *small* gauge transformation is one for which g is smoothly homotopic to the map $g_0 : M \to \mathrm{SU}(2)$ that sends every point of M to the identity element 1 in SU(2), that is, for which there exists a smooth map $H : M \times [0, 1] \to \mathrm{SU}(2)$ with $H(x, 0) = g_0(x) = 1$ and $H(x, 1) = g(x)$ for all $x \in M$. Letting $g_t(x) = H(x, t)$ we have a smoothly varying 1-parameter family $\{g_t\}_{0 \le t \le 1}$ of gauge transformations beginning at g_0 and ending at $g_1 = g$. For this reason physicists would say that a small gauge transformation is one that is *connected to the identity*. Any other gauge transformation $g : M \to \mathrm{SU}(2)$ is deemed *large*.

Remark 1.16. To formulate the result we would like to state next we should recall that SU(2) is diffeomorphic to the 3-sphere $\mathbb{S}^3$ so that a gauge transformation can be regarded as a smooth map $g : M \to \mathbb{S}^3$. As such, g has a *Brouwer degree* $\deg(g)$ (see Section 5.6 of [Nab3]). This can be defined in a variety of ways, but the most relevant of these from our current point of view is as follows. For any smooth map $g : M \to \mathbb{S}^3$, $\deg(g)$ is the unique real number for which

$$\int_M g^*\alpha = \deg(g) \int_{\mathbb{S}^3} \alpha$$

for any closed 3-form α on $\mathbb{S}^3$. Although it is not apparent from this particular definition, $\deg(g)$ is necessarily an integer (see pages 297-298 of [Nab3]).

If $g : M \to \mathrm{SU}(2)$ is a small gauge transformation, each g_t determines a gauge potential

$$A^t = g_t^{-1}\, A\, g_t + g_t^{-1}\, dg_t$$

and each of these has a Chern-Simons action $S_{CS}(A^t, k)$. Computing the derivative of $S_{CS}(A^t, k)$ with respect to t one finds that

$$\frac{d}{dt} S_{CS}(A^t, k) = 0$$

(see pages 288-289 of [BM]). Consequently, the Chern-Simons action is preserved under gauge transformations that are smoothly homotopic to the identity.

$$S_{CS}(A^g, k) = S_{CS}(A, k), \quad (g \text{ small})$$

If $g : M \to \mathrm{SU}(2)$ is a general gauge transformation the analysis is more subtle (see pages 289-290 of [BM]). Roughly, one proceeds in the following way. Let A is a global gauge potential on $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \xrightarrow{\pi} M$ and A^g the gauge transformed potential. Since the set of connections forms an affine space one can define a family $\{A^s\}_{0 \leq s \leq 1}$ of gauge potentials by setting

$$A^s = A + s\,(A^g - A), \quad 0 \leq s \leq 1.$$

Then $A^0 = A$ and $A^1 = A^g$. Think of the family $\{A^s\}_{0 \leq s \leq 1}$ as defining an $\mathfrak{su}(2)$-valued 1-form on the cylinder $M \times [0, 1]$. This is a manifold with boundary. The boundary consists of two copies of M and the 1-form agrees with A on the "lower" copy and with A^g on the "upper" copy. Since A and A^g differ by a gauge transformation one can glue the two boundary components together to obtain a gauge potential on an SU(2)-bundle over the compact 4-manifold $M \times \mathbb{S}^1$. Writing out the 2^{nd} Chern class for this bundle and performing the integration with the aid of Stokes Theorem and Theorem 1.6 one finds that

$$S_{CS}(A^g, k) - S_{CS}(A, k) = 2\pi k \, \deg(g),$$

where k is the level of the Chern-Simons action. In particular, the Chern-Simons action is not gauge invariant unless $\deg(g) = 0$. Certainly, if g is homotopic to a constant map, then $\deg(g) = 0$. If $M = \mathbb{S}^3$, then a deep theorem of Hopf states that the converse is also true (see Chapter 7 of [Miln1] or Section 6, Chapter 3, of [GP]). For some examples of maps from $\mathbb{S}^3$ to SU(2) that *do not* have degree zero, see pages 332-334 of [Nab3].

Remark 1.17. The failure of the Chern-Simons action to be fully gauge invariant would seem to be rather unfortunate, but it's not so bad. We will see in Section 1.5 that Witten's approach to knot polynomials is based not on the classical Chern-Simons theory that we have been discussing, but rather on its quantization. In particular, the Jones polynomial arises as an expectation value for a certain observable in this quantum theory and this expectation value is expressed as a path integral which depends on $e^{iS_{CS}(A,k)}$ rather that just $S_{CS}(A, k)$. Since $S_{CS}(A, k)$ is gauge invariant up to an integer multiple of 2π, $e^{iS_{CS}(A,k)}$ is fully gauge invariant and all is well. Our eventual interest in this exponential phase factor also motivates the need to sort out the critical points of the Chern-Simons functional and this is our next objective.

We wish to determine conditions on the gauge potential A that must be satisfied if A is to be a critical point of the action functional S_{CS}. Since the space of connections on $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \xrightarrow{\pi} M$ is an affine space we can consider the line

$$A^t = A + tB$$

of gauge potentials, where B is an arbitrary gauge potential. If A is a critical point of the Chern-Simons action, then we must have

$$\frac{d}{dt} S_{CS}(A^t, k)\Big|_{t=0} = 0.$$

A simple computation gives

$$\begin{aligned} A^t \wedge dA^t + \frac{2}{3} A^t \wedge A^t \wedge A^t =& [\, A \wedge dA + \frac{2}{3} A \wedge A \wedge A \,] + t\, [\, A \wedge dB + B \wedge dA + \\ & \frac{2}{3} (\, A \wedge A \wedge B + A \wedge B \wedge A + B \wedge A \wedge A \,)\,] + t^2\, [\, B \wedge dB + \\ & \frac{2}{3} (\, A \wedge B \wedge B + B \wedge A \wedge B + B \wedge B \wedge A \,)\,] + \frac{2}{3} t^3\, [\, B \wedge B \wedge B \,]. \end{aligned}$$

Only the coefficient of t will survive when we compute the t-derivative at $t = 0$ so we will ignore the others. Taking the trace of this coefficient and using the cyclic property of the trace gives

$$\mathrm{tr}\,(A \wedge dB + dA \wedge B + 2\, A \wedge A \wedge B) = \mathrm{tr}\,(-d(A \wedge B) + 2\, dA \wedge B + 2\, A \wedge A \wedge B)$$

Integrating this over M then gives

$$2 \int_M \mathrm{tr}\,(dA \wedge B + A \wedge A \wedge B) = 2 \int_M \mathrm{tr}\,(\,(dA + A \wedge A) \wedge B\,) = 2 \int_M \mathrm{tr}\,(F_A \wedge B).$$

We conclude that

$$\frac{d}{dt} S_{CS}(A^t, k)\Big|_{t=0} = \frac{k}{2\pi} \int_M \mathrm{tr}\,(F_A \wedge B).$$

Since B was arbitrary we find that $\frac{d}{dt} S_{CS}(A^t, k)\big|_{t=0} = 0$ is only possible if A satisfies

$$F_A = 0.$$

Remark 1.18. Recall that a connection ω on a principal bundle $G \hookrightarrow P \xrightarrow{\pi} M$ is said to be *flat* if its curvature 2-form Ω is identically zero. If the bundle is trivial so that $P = M \times G$, then flat connections always exist. Indeed, if Θ is the Maurer-Cartan 1-form on G and $\pi_G : M \times G \to G$ is the projection onto the second factor, then the pullback $\omega = \pi_G^* \Theta$ is a flat connection whose horizontal subspace at any $(x, g) \in M \times G$ is the tangent space to the submanifold $M \times \{g\}$ of $M \times G$ (see Exercise 6.2.12 of [Nab3]). This is called the *canonical flat connection* on $M \times G$. If the base manifold M is simply connected, then every flat connection on $M \times G$ is equivalent to the canonical one (see Chapter II, Corollary 9.2, of [KN1]). Flat connections cannot exist on nontrivial bundles whose base space M is simply connected (see Chapter II, Corollary 9.2, of [KN1]).

Thus, critical points, or stationary points, of the Chern-Simons functional are gauge potentials corresponding to flat connections on $\mathrm{SU}(2) \hookrightarrow M{\times}\mathrm{SU}(2) \xrightarrow{\pi} M$. If M is simply connected there is, up to gauge equivalence, only one such flat connection. This is the case, for example, when $M = \mathbb{S}^3$.

1.4. **Wilson Lines.** The observables that give rise to polynomial invariants of knots and links in Witten's TQFT interpretation are called *Wilson lines* or *Wilson loops*. Roughly, each of these is the trace of the holonomy of a connection around a closed curve. To introduce them properly we will begin by reviewing the relevant ideas from differential geometry.

Remark 1.19. At first glance the seemingly abstract notion of *holonomy* that we are about to describe may appear rather far removed from the workaday world of physics, but this is by no means the case. Measurable physical effects directly associated with the holonomy of a connection on a bundle are ubiquitous. The best known of these are the *Aharonov-Bohm effect* (see [AB]) and, more generally, the *Berry phase* (see [Berry] and [Simon]). For an overview of the subject one might consult [CJ].

Let $G \hookrightarrow P \xrightarrow{\pi} M$ be a smooth principal G-bundle with connection ω and $\alpha : [0,1] \to M$ a smooth curve in M with $\alpha(0) = x_0$ and $\alpha(1) = x_1$. At each point $p \in P$, ω_p is a Lie algebra-valued linear map on the tangent space $T_p(P)$. The kernel of ω_p is the *horizontal space* $\mathrm{Hor}_{\omega,p}(P)$ at p corresponding to ω and a smooth curve in P is said to be *horizontal* if its tangent vector is in the horizontal space at each of its points. A fundamental fact is that if p_0 is any fixed point in the fiber $\pi^{-1}(x_0)$, then α has a unique *horizontal lift* to P starting at p_0. More precisely, we have the following result.

Theorem 1.7. *Let $G \hookrightarrow P \xrightarrow{\pi} M$ be a smooth principal G-bundle over M with connection ω. Let $\alpha : [0,1] \to M$ be a smooth curve in M with $\alpha(0) = x_0$ and $\alpha(1) = x_1$. Then, for any $p_0 \in \pi^{-1}(x_0)$, there exists a unique smooth curve $\tilde{\alpha}_{\omega,p_0} : [0,1] \to P$ in P such that*

(1) $\tilde{\alpha}_{\omega,p_0}(0) = p_0$,
(2) $\pi \circ \tilde{\alpha}_{\omega,p_0}(t) = \alpha(t)$ *for all* $t \in [0,1]$, *and*
(3) $\tilde{\alpha}'_{\omega,p_0}(t) \in Hor_{\omega,\tilde{\alpha}_{\omega,p_0}(t)}(P)$ *for all* $t \in [0,1]$.

Remark 1.20. The gist of the proof is simple enough to describe. One is looking for a curve $\tilde{\alpha}_{\omega,p_0} : [0,1] \to P$ in P that is horizontal with respect to ω and this requires

$$\omega_{\tilde{\alpha}_{\omega,p_0}(t)}(\,\tilde{\alpha}'_{\omega,p_0}(t)\,) = 0 \tag{16}$$

for each t. The remaining two conditions require that, locally, $\tilde{\alpha}_{\omega,p_0}(t) = s(\alpha(t)) \cdot g(t)$, where s is a local section and $g : [0,1] \to G$ is a smooth function with $g(0)$ equal to the identity element. Writing out (16) in these terms gives an initial value problem for an ordinary differential equation in the function $g(t)$ to which one can apply standard existence and uniqueness theorems (see Theorem 6.1.4 of [Nab2] or Proposition 3.1, Chapter II, of [KN1]).

Notice that $\tilde{\alpha}_{\omega,p_0}(1)$ is in $\pi^{-1}(x_1)$. $\tilde{\alpha}_{\omega,p_0}(1)$ is called the *parallel translation of p_0 along α* determined by ω. Since $p_0 \in \pi^{-1}(x_0)$ was arbitrary we can define a map

$$\tau_{\omega,\alpha} : \pi^{-1}(x_0) \to \pi^{-1}(x_1)$$

called *parallel translation* along α by

$$\tau_{\omega,\alpha}(p) = \tilde{\alpha}_{\omega,p}(1)$$

for every $p \in \pi^{-1}(x_0)$. We list some basic properties of the parallel translation map.

(1) $\tau_{\omega,\alpha}$ is independent of the parametrization of α (see page 71 of [KN1]).

(2) $\tau_{\omega,\alpha}$ commutes with the action of G on $\pi^{-1}(x_0) \subseteq P$, that is,

$$\tau_{\omega,\alpha}(p \cdot g) = \tau_{\omega,\alpha}(p) \cdot g \quad \text{for all } p \in \pi^{-1}(x_0) \text{ and all } g \in G$$

(see Exercise 6.1.22 of [Nab2] or Proposition 3.2, Chapter II, of [KN1]).

(3) If $\alpha^{\leftarrow} : [0,1] \to M$, defined by $\alpha^{\leftarrow}(t) = \alpha(1-t)$, is "$\alpha$ backwards", then

$$\tau_{\omega,\alpha^{\leftarrow}} = \tau_{\omega,\alpha}^{-1}$$

(see Exercise 6.1.23 of [Nab2] or Proposition 3.3 (a), Chapter II, of [KN1]).

(4) If $\alpha, \beta : [0,1] \to M$ are smooth curves in M with $\alpha(1) = \beta(0)$ and $\alpha\beta : [0,1] \to M$, defined by

$$(\alpha\beta)(t) = \begin{cases} \alpha(2t), & \text{if } 0 \le t \le \frac{1}{2} \\ \beta(2t-1), & \text{if } \frac{1}{2} \le t \le 1, \end{cases}$$

is "α followed by β", then

$$\tau_{\omega,\alpha\beta} = \tau_{\omega,\beta} \circ \tau_{\omega,\alpha}$$

(see Exercise 6.1.23 of [Nab2] or Proposition 3.3 (b), Chapter II, of [KN1]).

It follows from (2) that, when $\pi^{-1}(x_0)$ and $\pi^{-1}(x_1)$ are identified with G via a local trivialization, each $\tau_{\omega,\alpha}$ is an isomorphism. Moreover, by virtue of (4), the definition of $\tau_{\omega,\alpha}$ extends immediately to piecewise smooth curves α.

Now we will restrict our attention to the case in which α is a smooth *loop* in M, that is, $x_0 = \alpha(0) = \alpha(1)$. Then $\tau_{\omega,\alpha}$ is an automorphism of $\pi^{-1}(x_0)$. Now, G acts transitively on $\pi^{-1}(x_0)$. Consequently, for each $p \in \pi^{-1}(x_0)$ there is a unique $g(\omega, \alpha, p) \in G$ such that $\tau_{\omega,\alpha}(p) = p \cdot g(\omega, \alpha, p)$. Holding p fixed and letting α vary over all smooth loops at x_0 in M we obtain a subset $\mathrm{Hol}_{\omega,p}$ of G consisting of all of those $g \in G$ such that p is parallel translated to $p \cdot g$ over some smooth loop at x_0 in M. By virtue of properties (3) and (4) above, $\mathrm{Hol}_{\omega,p}$ is a subgroup of G, called the *holonomy group of ω at p*. Moreover, if $p \cdot g$ is another point in $\pi^{-1}(x_0)$, then the holonomy groups at $p \cdot g$ and p are conjugate in G and therefore isomorphic (see Exercise 6.1.24 of [Nab2] or Proposition 4.1, Chapter II, of [KN1]). Specifically, as subgroups of G,

$$\mathrm{Hol}_{\omega,p\cdot g} = g\,\mathrm{Hol}_{\omega,p}\,g^{-1}.$$

More generally, suppose M is connected and let x_0 and x_1 be two points in M. For a manifold, connectedness is equivalent to pathwise connectedness so there exists a smooth curve $\gamma : [0,1] \to M$ in M from $x_0 = \gamma(0)$ to $x_1 = \gamma(1)$. Then γ and $\gamma^{\leftarrow}$ establish a one-to-one correspondence between loops at x_0 and loops at x_1. More precisely, if α is a loop at x_0, then $\gamma^{\leftarrow}\alpha\gamma$ is a loop at x_1, whereas if β is a loop at x_1, then $\gamma\beta\gamma^{\leftarrow}$ is a loop at x_0. From this one can argue that, for any $p_0 \in \pi^{-1}(x_0)$ and any $p_1 \in \pi^{-1}(x_1)$, the holonomy groups $\mathrm{Hol}_{\omega,p_0}$ and $\mathrm{Hol}_{\omega,p_1}$ are conjugate and therefore isomorphic (see page 73 of [KN1]). Consequently, if M is connected (as we will assume from now on), the holonomy group depends, up to conjugacy, only on the connection ω itself and one can denote the abstract group that all of these are isomorphic to simply

Hol_ω. This is called the *holonomy group of* ω and is determined up to conjugacy in the following way. Fix an arbitrary point $x_0 \in M$ and an arbitrary point $p \in \pi^{-1}(x_0)$. For every smooth loop α at x_0 in M write $\tau_{\omega,\alpha}(p) = p \cdot g(\omega, \alpha, p)$, where $g(\omega, \alpha, p) \in G$. Then

$$\mathrm{Hol}_\omega \cong \{g(\omega, \alpha, p) : \alpha \text{ is a smooth loop at } x_0 \text{ in } M\}.$$

If one considers only smooth loops α that are nullhomotopic in M, the result is a subgroup of Hol_ω called the *restricted holonomy group* and denoted Hol^0_ω. If M is simply connected, then the holonomy group and the restricted holonomy group coincide. In general, the structures of Hol_ω and Hol^0_ω are described in the following result (see Theorem 4.2, Chapter II, of [KN1]).

Theorem 1.8. *Let $G \hookrightarrow P \xrightarrow{\pi} M$ be a smooth principal G-bundle with connection ω over the connected manifold M. Then Hol_ω is a Lie subgroup of G and Hol^0_ω is the connected component containing the identity in Hol_ω.*

To define Wilson lines we would like to use this information about holonomy to associate a *number* with every pair consisting of a connection ω on $G \hookrightarrow P \xrightarrow{\pi} M$ and a smooth, oriented, closed curve C in M. This will require some additional input due to the fact that essentially everything we have introduced to this point is defined only up to conjugacy in G. To remedy this we select some finite-dimensional representation ρ of G. The case that will be of most interest to us will be $G = \mathrm{SU}(2)$ with ρ the fundamental (spin 1/2) representation of SU(2) on $\mathbb{C}^2$. Notice that, for any $g_0 \in G$, the trace of $\rho(g_0)$ is a number that is invariant under conjugation in G, that is,

$$\mathrm{tr}_\rho\left(\rho(g\, g_0\, g^{-1})\right) = \mathrm{tr}_\rho\left(\rho(g)\,\rho(g_0)\,\rho(g)^{-1}\right) = \mathrm{tr}_\rho\left(\rho(g_0)\right) \tag{17}$$

for any $g \in G$. Now proceed as follows. Choose a parametrization $\alpha : [0,1] \to M$ of C consistent with its orientation and let $x_0 = \alpha(0)$. Choose some point $p \in \pi^{-1}(x_0)$ and write $\tau_{\omega,\alpha}(p) = p \cdot g(\omega, \alpha, p)$, where $g(\omega, \alpha, p)$ is in G. Any other choices for the parametrization, for x_0, and for p will conjugate $g(\omega, \alpha, p)$ by some element of G. Consequently, $\mathrm{tr}_\rho\left(\rho(g(\omega, \alpha, p))\right)$ is a number that is independent of these choices. Now we define the *Wilson line* or *Wilson loop* $W_{C,\rho}(\omega)$ corresponding to the connection ω and the closed curve C with respect to the representation ρ by

$$W_{C,\rho}(\omega) = \mathrm{tr}_\rho\left(\rho\left(g(\omega, \alpha, p)\right)\right). \tag{18}$$

Remark 1.21. The Wilson line depends only on the connection, the closed curve, and the representation. The notation $W_{C,\rho}(\omega)$ is chosen because we will eventually want to think of it as a function of the connection for each fixed choice of the curve and the representation. In the physics literature the definition of a Wilson line will look nothing like this. Rather, one will see something of the following sort.

$$W_{C,\rho}(\omega) = \mathrm{tr}_\rho\left(\mathcal{P} \exp i \int_C A_\mu dx^\mu\right)$$

Here A is a gauge potential for ω, $\mathcal{P}$ is, as they say, "the familiar path ordering symbol", and

$$\mathcal{P} \exp i \int_C A_\mu dx^\mu$$

is called a *path ordered exponential*. Since we have no intention of utilizing this notation we will not attempt to either motivate or define it, but will simply refer those interested to the Introduction and Section 2.4 of [CMV].

Next we will prove that, for a fixed curve C and representation ρ, Wilson lines are *gauge invariant*. In preparation for the proof we introduce the following notation. Let $S : P \to P$ be a global gauge transformation of $G \hookrightarrow P \xrightarrow{\pi} M$, that is, S is a diffeomorphism of P onto itself that preserves the fibers ($\pi \circ S = \pi$) and respects the action of G on P ($S(p \cdot g) = S(p) \cdot g$). Now define

$$g_S : P \to G$$

by

$$S(p) = p \cdot g_S(p)$$

for each $p \in P$. Note that $g_S(p)$ is uniquely defined because the action of G on P is free and it is transitive on each fiber. Then our main result is the following.

Theorem 1.9. *(Gauge Invariance of Wilson Lines) Let $G \hookrightarrow P \xrightarrow{\pi} M$ be a smooth principal G-bundle over the connected manifold M and $S : P \to P$ a global gauge transformation. Let ω be a connection on the bundle and $\omega^S = S^*\omega$ the (gauge equivalent) connection induced by S. Let ρ be a finite-dimensional representation of G. Finally, let C be a smooth, oriented, closed curve in M. Then*

$$W_{C,\rho}(\omega^S) = W_{C,\rho}(\omega).$$

Proof. Using the notation established in (18) and appealing to (17), it will be enough to prove that

$$g(\omega^S, \alpha, p) = g_S(p)^{-1} g(\omega, \alpha, p)\, g_S(p). \tag{19}$$

We begin by proving that, for any $g \in G$,

$$g(\omega, \alpha, p \cdot g) = g^{-1} g(\omega, \alpha, p)\, g. \tag{20}$$

First note that, by definition,

$$\tilde{\alpha}_{\omega, p\cdot g}(1) = p \cdot (\, g\, g(\omega, \alpha, p \cdot g)\,) \tag{21}$$

Now we will obtain another expression for $\tilde{\alpha}_{\omega,p\cdot g}(1)$. Consider the curve $\sigma_g \circ \tilde{\alpha}_{\omega,p}(t) = \tilde{\alpha}_{\omega,p}(t) \cdot g$. Note that

$$\sigma_g \circ \tilde{\alpha}_{\omega,p}(0) = p \cdot g$$

and

$$\pi(\sigma_g \circ \tilde{\alpha}_{\omega,p}(t)) = \pi(\tilde{\alpha}_{\omega,p}(t) \cdot g) = \pi(\tilde{\alpha}_{\omega,p}(t)) = \alpha(t)$$

so $\sigma_g \circ \tilde{\alpha}_{\omega,p}$ begins at $p \cdot g$ and lifts α. We claim that it is also horizontal so that, by the uniqueness assertion in Theorem 1.7, it must be $\tilde{\alpha}_{\omega,p\cdot g}$. To see this we compute

$$\begin{aligned}\omega_{\sigma_g(\tilde{\alpha}_{\omega,p}(t))}\left(\frac{d}{dt}\left(\sigma_g(\tilde{\alpha}_{\omega,p}(t)\right)\right) &= \omega_{\sigma_g(\tilde{\alpha}_{\omega,p}(t))}\left(\left(\sigma_g\right)_{*\tilde{\alpha}_{\omega,p}(t)}\left(\tilde{\alpha}'_{\omega,p}(t)\right)\right)\\ &= (\sigma_g^*\omega)_{\sigma_g(\tilde{\alpha}_{\omega,p}(t))}\left(\tilde{\alpha}'_{\omega,p}(t)\right)\\ &= g^{-1}\,\omega_{\tilde{\alpha}_{\omega,p}(t)}\left(\tilde{\alpha}'_{\omega,p}(t)\right)g\\ &= 0\end{aligned}$$

where we have used a defining property of a connection and the fact that $\tilde{\alpha}'_{\omega,p}(t)$ is horizontal. We conclude then that $\sigma_g \circ \tilde{\alpha}_{\omega,p}(t) = \tilde{\alpha}_{\omega,p\cdot g}(t)$. Since $\sigma_g \circ \tilde{\alpha}_{\omega,p}(1) = p \cdot (\, g(\omega,\alpha,p)\, g\,)$ we conclude from (21) that

$$p \cdot (\, g\, g(\omega,\alpha,p\cdot g)\,) = p \cdot (\, g(\,\omega,\alpha,p)\, g\,).$$

Since the action of G on P is free this implies $g\, g(\omega,\alpha,p\cdot g) = (g(\omega,\alpha,p)\, g$ and this is (20).

Next we consider

$$\tilde{\alpha}_{\omega^S,p}(1) = p \cdot g(\omega^S,\alpha,p).$$

We claim that

$$S \circ \tilde{\alpha}_{\omega^S,p} = \tilde{\alpha}_{\omega,S(p)}. \tag{22}$$

Both of the curves in (22) begin at $S(p)$ and, since S is a gauge transformation, both curves lift α. It will therefore suffice to show that $S \circ \tilde{\alpha}_{\omega^S,p}$ is ω-horizontal and appeal again to the uniqueness statement in Theorem 1.7. For this we compute

$$\omega_{S(\tilde{\alpha}_{\omega^S,p}(t))}\left(\frac{d}{dt}\,S(\tilde{\alpha}_{\omega^S,p}(t))\right) = \omega^S_{\tilde{\alpha}_{\omega^S,p}(t)}\left(\frac{d}{dt}\,\tilde{\alpha}_{\omega^S,p}(t)\right) = 0$$

since $\omega^S = S^*\omega$ and $\tilde{\alpha}_{\omega^S,p}(t)$ is ω^S-horizontal. Thus, (22) is satisfied and therefore

$$S(p)\cdot g(\omega^S,\alpha,p) = S(p)\cdot g(\omega,\alpha,S(p)).$$

Since the action of G on P is free this implies

$$g(\omega^S,\alpha,p) = g(\omega,\alpha,S(p)) = g(\omega,\alpha,p\cdot g_S(p)) = g_S(p)^{-1} g(\omega,\alpha,p)\, g_S(p),$$

where the last equality is by (20). This proves (19) and therefore the Theorem. □

Remark 1.22. When it comes time to describe the quantization of Chern-Simons theory we will want to consider Wilson lines $W_{C,\rho}(\omega)$ with the curve C and the representation ρ fixed, but ω varying over the space of connections on $G \hookrightarrow P \xrightarrow{\pi} M$. In this context $W_{C,\rho}$ is a function on the space of connections and is referred to as a *Wilson operator*. According to Theorem 1.9 $W_{C,\rho}$ descends to a function on the space of gauge equivalence classes of connections, called the *moduli space of connections* on $G \hookrightarrow P \xrightarrow{\pi} M$. We will have more to say about this in Section 1.5.

There are two identities analogous to (19) and (20) that come in handy and that we will record without proof. For this we let α and β be two smooth loops at $x_0 \in M$ and $\gamma : [0,1] \to M$ a smooth path from x_0 to $x_1 \in M$. Let p be an arbitrary point in $\pi^{-1}(x_0)$. Then

$$g(\omega, \alpha, p) = g(\omega, \gamma\alpha\gamma^{\leftarrow}, \tau_{\omega,\gamma}(p))$$

and

$$g(\omega, \alpha\beta, p) = g(\omega, \beta, p)g(\omega, \alpha, p).$$

In particular, the last of these implies that the map $\mu_{\omega,p}$ that carries a loop α at x_0 to $g(\omega, \alpha, p)^{-1}$ in G preserves products.

$$\mu_{\omega,p}(\alpha\beta) = \mu_{\omega,p}(\alpha)\,\mu_{\omega,p}(\beta)$$

Now, it is *not* true that homotopic loops at x_0 have the same holonomy. Indeed, even a nullhomotopic loop need not have trivial holonomy, that is, the restricted holonomy group need not be trivial. This issue is governed by the curvature Ω of the connection ω (see the *Ambrose-Singer Theorem*, Chapter II, Section 8, of [KN1]). In particular, if the curvature is zero (see Remark 1.18), then nullhomotopic loops do, indeed, have trivial holonomy and it follows that homotopic loops have the same holonomy. In this case, $\mu_{\omega,p}$ descends to a homomorphism from the fundamental group $\pi_1(M, x_0)$ of M into G, the image being a subgroup of G isomorphic to Hol_ω.

Theorem 1.10. *Let $G \hookrightarrow P \xrightarrow{\pi} M$ be a smooth principal G-bundle over the connected manifold M and let ω be a flat connection on the bundle. For any $x_0 \in M$ and any $p \in \pi^{-1}(x_0)$ the map*

$$\mu_{\omega,p} : \pi_1(M, x_0) \to G,$$

defined by

$$\mu_{\omega,p}(\,[\alpha]\,) = g(\omega, \alpha, p)^{-1}$$

for every $[\alpha] \in \pi(M, x_0)$, is a homomorphism whose image is a subgroup of G isomorphic to the holonomy group Hol_ω of ω. A different choice of $x_0 \in M$ or of $p \in \pi^{-1}(x_0)$ conjugates the homomorphism by some element of G and therefore gives a conjugate subgroup of G.

A principal G-bundle over M and a flat connection on it therefore determine a conjugacy class of homomorphisms from the fundamental group of M into G. Stated otherwise, we have a map from flat connections on principal G-bundles over M to conjugacy classes of homomorphisms from $\pi_1(M)$ to G. One can then show the following.

(1) The map is surjective, that is, given a conjugacy class of homomorphisms $\pi_1(M) \to G$ there is a principal G-bundle over M and a flat connection ω on it for which the conjugacy class of homomorphisms described in the previous theorem is precisely the given class.

(2) If two flat connections ω_1 and ω_2 on two principal G-bundles $G \hookrightarrow P_1 \stackrel{\pi_1}{\to} M$ and $G \hookrightarrow P_2 \stackrel{\pi_2}{\to} M$, respectively, over M give rise to the same conjugacy class of homomorphisms $\pi_1(M) \to G$, then ω_1 and ω_2 are gauge equivalent, that is, there exists a diffeomorphism $S : P_1 \to P_2$ of P_1 onto P_2 satisfying $\pi_1 = \pi_2 \circ S$, and $S \circ (\sigma_1)_g = (\sigma_2)_g \circ S$ with $\omega_1 = S^*\omega_2$.

These culminate in the following result (see Proposition 2.9 of [Morita]).

Theorem 1.11. *Let M be a connected, smooth manifold and G a Lie group. Then there is a bijective correspondence between the set of conjugacy classes of homomorphisms $\mu : \pi_1(M) \to G$ from the fundamental group of M into G and the set of gauge equivalence classes of flat connections ω on principal G-bundles over M.*

1.5. **Witten's TQFT Interpretation of the Invariants.**

1.5.1. *Introduction.* In this section we will attempt to sketch the ideas behind the construction by Edward Witten [Witt2] of a *topological quantum field theory (TQFT)* in which the Jones polynomials are realized as expectation values for observables arising from Wilson lines. This construction is quite complex and draws on an enormous variety of very deep ideas from both theoretical physics and mathematics. As such we can offer here only a brief aerial view of the terrain, but we will provide rather detailed references for those who would like explore in earnest. Aside from [Witt2] and [Witt3] our primary sources are [Atiy3], [Atiy4], [Atiy5], [ADPW], [BM], [Hu], and [Kiri].

The intricacy of the construction makes it desirable to reverse the historical order and begin with a rather abstract mathematical context in which to fit the various pieces. This is provided by a set of "axioms" for TQFT devised by Michael Atiyah [Atiy3] and motivated by Witten's ideas in [Witt2] and an analogous axiomatization of conformal field theory due to Graeme Segal [Segal2]. There are a number of variants of these axioms and we will begin with a version best suited to our purposes. This is the appropriate context in which to discuss the 3-manifold invariants introduced by Witten in [Witt2]. These axioms require some modification when the objective is to define invariants for knots and links in M and we will say more about this in due course.

1.5.2. *The Context: Atiyah's Axioms for TQFT.* The most efficient formulation of Atiyah's definition of a TQFT is rather abstract and somewhat off-putting, but we will record it anyway and then try to spell out in detail precisely what it means.

Atiyah's Axioms for TQFT: Categorical Formulation

An $(n+1)$-dimensional topological quantum field theory (TQFT) is a symmetric, monoidal functor from the category whose objects are compact, oriented, smooth n-manifolds and whose morphisms are equivalence classes of oriented cobordisms between these manifolds to the category whose objects are finite-dimensional, complex vector spaces and whose morphisms are linear maps between these vector spaces.

For the categorically-challenged we will now unravel the mathematical jargon to arrive at a statement of the axioms in a form which we will actually use. We will then say a few words about the physical interpretation.

Remark 1.23. We will require some basic information about cobordisms so we will pause momentarily to record the items we will need. Let Σ_0 and Σ_1 be compact, oriented, smooth n-manifolds. An *oriented cobordism* from Σ_0 to Σ_1 consists of a compact, oriented, smooth $(n+1)$-manifold M with boundary (with its induced orientation) together with smooth maps

$$\Sigma_0 \stackrel{i_0}{\rightarrow} M \stackrel{i_1}{\leftarrow} \Sigma_1,$$

where i_0 is an *orientation reversing* diffeomorphism of Σ_0 onto $i_0(\Sigma_0) \subseteq \partial M$ and i_1 is an *orientation preserving* diffeomorphism of Σ_1 onto $i_1(\Sigma_1) \subseteq \partial M$ such that $i_0(\Sigma_0)$ and $i_1(\Sigma_1)$ are disjoint and

$$\partial M = i_0(\Sigma_0) \sqcup i_1(\Sigma_1).$$

It is customary to suppress the embeddings i_0 and i_1 and write this as

$$\partial M = -\Sigma_0 \sqcup \Sigma_1,$$

where $-\Sigma_0$ is Σ_0 with its orientation reversed. Thus, one can think of the cobordism as an $(n+1)$-manifold "connecting" Σ_0 to Σ_1. We will write this symbolically as

$$M : \Sigma_0 \to \Sigma_1.$$

For example, the boundary of the cylinder $\Sigma \times [0,1]$ is $-\Sigma \sqcup \Sigma$ so $\Sigma \times [0,1]$ is an oriented cobordism from Σ to Σ. Figure 21 represents a cobordism from the circle $\Sigma_0 = \mathbb{S}^1$ to a disjoint union $\Sigma_1 = \mathbb{S}^1 \sqcup \mathbb{S}^1$ of two circles. If $\mathbb{S}^1$ is given some fixed orientation, then the sense of rotation on the left-hand side is opposite to that on the right-hand side.

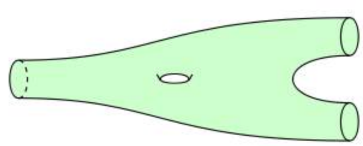


FIGURE 21. Cobordism

Two oriented cobordisms $\Sigma_0 \stackrel{i_0}{\rightarrow} M \stackrel{i_1}{\leftarrow} \Sigma_1$ and $\Sigma_0 \stackrel{i_0'}{\rightarrow} M' \stackrel{i_1'}{\leftarrow} \Sigma_1$ from Σ_0 to Σ_1 are *equivalent* if there exists an orientation preserving diffeomorphism $\varphi : M \to M'$ such that $\varphi \circ i_0 = i_0'$ and $\varphi \circ i_1 = i_1'$. The axioms for TQFT that we will describe do not distinguish between equivalent cobordisms. In particular, anything equivalent to $\Sigma \times [0,1]$ will be referred to as the *identity cobordism*.

Two oriented cobordisms $M : \Sigma_0 \to \Sigma_1$ and $M' : \Sigma_1 \to \Sigma_2$ can be *composed* by "gluing" the copies of Σ_1 together to obtain an oriented cobordism $M' \circ M : \Sigma_0 \to \Sigma_2$. More precisely, if $\Sigma_0 \stackrel{i_0}{\rightarrow} M \stackrel{i_1}{\leftarrow} \Sigma_1$ and $\Sigma_1 \stackrel{i_1'}{\rightarrow} M' \stackrel{i_2}{\leftarrow} \Sigma_2$ are the cobordisms, then $i_1' \circ i_1^{-1}$ is an orientation reversing diffeomorphism of $i_1(\Sigma_1)$ onto $i_1'(\Sigma_1)$. Denote by $M \cup_{\Sigma_1} M'$ the topological space obtained from the disjoint union $M \sqcup M'$ by identifying x with $(i_1' \circ i_1^{-1})(x)$ for every $x \in i_1(\Sigma_1)$. It is not obvious, but true nonetheless that $M \cup_{\Sigma_1} M'$

admits a smooth structure, unique up to diffeomorphism, for which M and M' are smoothly embedded submanifolds with boundary (see Theorem 1.4 of [Miln2]). Furthermore, $M \cup_{\Sigma_1} M'$ is naturally oriented and $\partial(M \cup_{\Sigma_1} M') = -\Sigma_0 \sqcup \Sigma_2$. Thus, we can define

$$M' \circ M = M \cup_{\Sigma_1} M' : \Sigma_0 \to \Sigma_2$$

(see Figure 22).

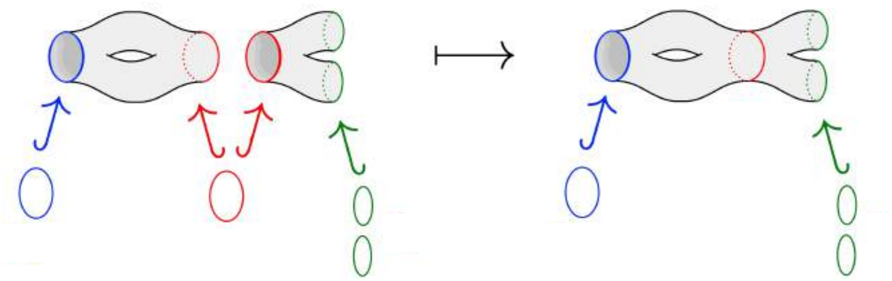


FIGURE 22. Gluing Cobordisms

With this information we can return to the task at hand. An $(n+1)$*-dimensional TQFT* is an assignment Z of

(Z1) a finite-dimensional, complex vector space $Z(\Sigma)$ to each compact, oriented, smooth, n-dimensional manifold Σ without boundary, and

(Z2) a linear transformation $Z(M) : Z(\Sigma_0) \to Z(\Sigma_1)$ to each oriented cobordism $M : \Sigma_0 \to \Sigma_1$.

Remark 1.24. Although generally not included in the basic axioms for TQFT we will want to assume as well that each of the complex vector spaces $Z(\Sigma)$ is equipped with some specific non-degenerate Hermitian inner product and so, being finite-dimensional, is a Hilbert space (see **(A3)** below). One should be aware of the fact that, however, that in the physics literature it is common to refer to any vector space that plays the role of a "state space" for some system (as we will see $Z(\Sigma)$ does) as a "Hilbert space" whether or not such a structure has been introduced, or even exists.

These assignments are assumed to satisfy each of the following.

(A1) Equivalent oriented cobordisms $\Sigma_0 \stackrel{i_0}{\to} M \stackrel{i_1}{\leftarrow} \Sigma_1$ and $\Sigma_0 \stackrel{i'_0}{\to} M' \stackrel{i'_1}{\leftarrow} \Sigma_1$ have the same image, that is,

$$Z(M) = Z(M') : Z(\Sigma_0) \to Z(\Sigma_1). \tag{23}$$

(A2) If $-\Sigma$ denotes Σ with the opposite orientation and $Z(\Sigma)^*$ is the vector space dual of $Z(\Sigma)$, then

$$Z(-\Sigma) \cong Z(\Sigma)^*. \tag{24}$$

Remark 1.25. Axiom **(A2)** describes the effect on Z of reversing the orientation of the n-manifold Σ, but there is no mention of the effect of orientation reversal for the $(n+1)$-manifold M. For this we would need the Hermitian structures of the vector spaces $Z(\Sigma)$ mentioned in Remark 1.24. Suppose that $\partial M = -\Sigma_0 \sqcup \Sigma_1$ so that $Z(M)$ is a linear transformation $Z(M) : Z(\Sigma_0) \to Z(\Sigma_1)$. Then $\partial(-M) = -\Sigma_1 \sqcup \Sigma_0$ so $Z(-M)$ is a linear transformation $Z(-M) : Z(\Sigma_1) \to Z(\Sigma_0)$ in the opposite direction.

(A3) If $\partial M = -\Sigma_0 \sqcup \Sigma_1$, then $Z(-M) : Z(\Sigma_1) \to Z(\Sigma_0)$ is the adjoint of $Z(M) : Z(\Sigma_0) \to Z(\Sigma_1)$ with respect to the Hermitian structures of $Z(\Sigma_0)$ and $Z(\Sigma_1)$.

$$Z(-M) = Z(M)^* \tag{25}$$

(A4) Z is multiplicative in the sense that, if $\Sigma = \Sigma_1 \sqcup \Sigma_2$ is the disjoint union of Σ_1 and Σ_2, then

$$Z(\Sigma) = Z(\Sigma_1 \sqcup \Sigma_2) \cong Z(\Sigma_1) \otimes Z(\Sigma_2) \tag{26}$$

Remark 1.26. We point out that (26) is what makes the functor we are in the process of defining *monoidal*. Notice that **(A2)** and **(A4)** provide another way of thinking about $Z(M)$ when $M : \Sigma_0 \to \Sigma_1$. Indeed, since $\partial M = -\Sigma_0 \sqcup \Sigma_1$,

$$Z(\partial M) = Z(-\Sigma_0 \sqcup \Sigma_1) = Z(\Sigma_0)^* \otimes Z(\Sigma_1) \cong \mathrm{Hom}\,(Z(\Sigma_0), Z(\Sigma_1))$$

so we can identify $Z(M) : Z(\Sigma_0) \to Z(\Sigma_1)$ with an element of $Z(\partial M)$.

$$Z(M) \in Z(\partial M) \tag{27}$$

Conversely, every element of $Z(\partial M)$ is canonically identified with a linear transformation from $Z(\Sigma_0)$ to $Z(\Sigma_1)$. The formulation of Atiyah's axioms in [Atiy3] did, in fact, adopt the point of view that $Z(M)$ is to be regarded as an element of $Z(\partial M)$. We have chosen the equivalent view of $Z(M)$ as a linear transformation primarily because the "functorial" nature of TQFT, as expressed in the following axiom, seems a bit more natural.

(A5) If $M : \Sigma_0 \to \Sigma_1$ and $M' : \Sigma_1 \to \Sigma_2$ are oriented cobordisms and $M' \circ M : \Sigma_0 \to \Sigma_2$ is their composition, then

$$Z(M' \circ M) = Z(M') \circ Z(M). \tag{28}$$

Remark 1.27. Viewing $Z(M)$ as an element of $Z(\partial M)$ this functoriality axiom is expressed by Atiyah in the following way. If $\partial M_1 = \Sigma_1 \sqcup \Sigma_3$ and $\partial M_2 = -\Sigma_3 \sqcup \Sigma_2$ and if $M = M_1 \cup_{\Sigma_3} M_2$ is the manifold obtained by gluing together the common Σ_3 boundary of M_1 and M_2, then

$$Z(M) = \langle\, Z(M_1),\, Z(M_2)\,\rangle \tag{29}$$

where $\langle\ ,\ \rangle$ is the natural pairing

$$Z(\Sigma_1) \otimes Z(\Sigma_3) \otimes Z(\Sigma_3)^* \otimes Z(\Sigma_2) \to Z(\Sigma_1) \otimes Z(\Sigma_2).$$

Note that, when $\Sigma_3 = \emptyset$, $M = M_1 \sqcup M_2$ and this reduces to

$$Z(M) = Z(M_1 \sqcup M_2) = Z(M_1) \otimes Z(M_2).$$

To motivate the next two axioms we will think of the empty set $\emptyset$ as either the empty n-manifold or the empty $(n+1)$-manifold and will, at least temporarily, denote these $\emptyset_n$ and $\emptyset_{n+1}$, respectively. Thus, an $(n+1)$-manifold M without boundary can be thought of as a cobordism from $\emptyset_n$ to $\emptyset_n$. Applying **(A4)** to $\partial M = -\emptyset_n \sqcup \emptyset_n = \emptyset_n$ we find that $Z(\emptyset_n) \cong Z(\emptyset_n) \otimes Z(\emptyset_n)$ so $\dim Z(\emptyset_n) = (\dim Z(\emptyset_n))^2$. Consequently, $\dim Z(\emptyset_n)$ is either 0 or 1. To exclude the trivial case in which $\dim Z(\emptyset_n) = 0$ we make the following non-triviality assumption.

(A6) $Z(\emptyset_n) = \mathbb{C}$

Similarly, we regard $\emptyset_{n+1}$ as a cobordism from $\emptyset_n$ to $\emptyset_n$. Taking both $M : \Sigma_0 \to \Sigma_1$ and $M' : \Sigma_1 \to \Sigma_2$ in **(A5)** to be $\emptyset_{n+1} : \emptyset_n \to \emptyset_n$ one finds that $M' \circ M : \Sigma_0 \to \Sigma_2$ is the same cobordism. Consequently, $Z(\emptyset_{n+1})$ is an idempotent linear transformation on the 1-dimensional vector space $Z(\emptyset_n) = \mathbb{C}$ so, when identified with the complex number that determines it, $Z(\emptyset_{n+1})$ is either 0 or 1. Again we evade the trivial case with the following.

(A7) $Z(\emptyset_{n+1}) = 1$

Analogous arguments (to be found on page 179 of [Atiy3]) for $Z(\Sigma \times [0,1]) \in \mathrm{End}(Z(\Sigma))$ lead to the final non-triviality axiom.

(A8) $Z(\Sigma \times [0,1]) = \mathrm{id}_{Z(\Sigma)}$

Before turning to a brief description of how all of this is to be thought of physically we would like to elaborate just a bit on one particular aspect of the formalism we have just introduced. First note that, if M is a compact, oriented, smooth $(n+1)$-manifold *without boundary*, then we can regard M as an oriented cobordism from $\emptyset_n$ to $\emptyset_n$. Thus, $Z(M)$ is a linear transformation from $\mathbb{C}$ to $\mathbb{C}$ and is therefore uniquely determined by a single complex number which we will also denote $Z(M)$. An orientation preserving diffeomorphism of M gives another, equivalent oriented cobordism from $\emptyset_n$ to $\emptyset_n$ and therefore, by **(A1)**, preserves $Z(M)$. Consequently, $Z(M)$ is a numerical invariant of compact, oriented, smooth $(n+1)$-manifolds. In Witten's theory $n = 2$ and $Z(M)$ is Witten's 3-manifold invariant.

Now, let us assume for the moment that it is possible to construct a meaningful example that satisfies Atiyah's rather elaborate set of axioms (we will sketch a bit of Witten's construction of such an example when $n = 2$ in Sections 1.5.3 and 1.5.4. One might then ask if it is actually possible to *compute* this invariant $Z(M)$ for at least some compact, oriented, smooth $(n+1)$-manifolds M. We are certainly in no position to answer such a question at this point, but we will say a few words about Witten's strategy.

The basic strategy ... is to develop a machinery for chopping M in pieces, solving the problem on the pieces, and gluing things back together.

-Edward Witten [Witt2]

Remark 1.28. To describe the machinery Witten has in mind we should review a few items from topology. The Jordan-Brouwer Separation Theorem (Theorem 7.10 of [MT] or Theorem 2.4, Chapter XVII, of [Dug]) implies that a subset Σ_0 of $\mathbb{R}^3$ that is homeomorphic to the 2-sphere $\mathbb{S}^2$ separates $\mathbb{R}^3$ into two subsets $\mathbb{R}^3 = M_1 \cup M_2$, each of which is a 3-dimensional manifold with boundary and which intersect only on their common boundary $\partial M_1 = \partial M_2 = \Sigma_0$. The boundaries ∂M_1 and ∂M_2 inherit opposite orientations from the orientation of $\mathbb{R}^3$. The same is true for embedded 2-spheres in $\mathbb{S}^3$ and also for embeddings of compact Riemann surfaces Σ_g of higher genus $g \geq 1$ in $\mathbb{R}^3$ or $\mathbb{S}^3$. On the other hand, an embedding of a compact surface in a more general 3-manifold M need not separate M at all. There is, for example, an embedding of the 2-sphere in the 3-torus $\mathbb{T}^3 = \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$ that does not separate the torus; Figure 23 shows the analogous situation in one less dimension where the complement of the embedded circle has only one component.

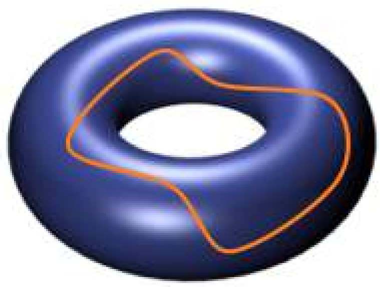


FIGURE 23. Smooth Embedding of $\mathbb{S}^1$ in $\mathbb{T}^2$ that does not separate $\mathbb{T}^2$

Nevertheless, there are very useful procedures for embedding certain surfaces Σ in a 3-manifold M that separate M into simpler pieces M_1 and M_2 on each of which a given problem may be more tractable. For example, if K is a knot in M one can choose a closed tubular neighborhood $\overline{N(K)}$ of K in M (see Chapter 4, Sections 5 and 6 of [Hirsch]). The boundary of $\overline{N(K)}$ is a torus $\partial\,\overline{N(K)} = \Sigma_1$ and the interior $N(K)$ is an open solid torus (see Figure 24). The complement $M - N(K)$ is a 3-manifold with boundary Σ_1. Taking $M_1 = \overline{N(K)}$ and $M_2 = M - N(K)$ we have "chopped" M into two pieces with the required properties, at least one of which is quite simple. One can now continue the process inside the 3-manifold $M - \overline{N(K)}$.

FIGURE 24. Tubular Neighborhood of a Knot

In general, if we embed a smooth surface Σ in a 3-manifold M which separates M into two 3-manifolds M_1 and M_2 with common boundary $\partial M_1 = \partial M_2 = \Sigma$, then this process is called *cutting M along* Σ (see [Haken] for more on this).

Now suppose that we have in hand a 3-dimensional TQFT and M is a compact, oriented, smooth 3-manifold without boundary for which we would like to calculate the invariant $Z(M)$. Suppose also that Σ is a compact, oriented, smoothly embedded surface without boundary in M along which we can cut M into two pieces M_1 and M_2 as described above. We can regard M, M_1 and M_2 as cobordisms. Specifically, $M : \emptyset_2 \to \emptyset_2$, $M_1 : \emptyset_2 \to \Sigma$ and $M_2 : \Sigma \to \emptyset_2$ so $Z(M)$, $Z(M_1)$ and $Z(M_2)$ are linear maps from $\mathbb{C}$ to $\mathbb{C}$, from $\mathbb{C}$ to $Z(\Sigma)$, and from $Z(\Sigma)$ to $\mathbb{C}$, respectively. According to **(A5)**,

$$Z(M) = Z(M_2 \circ M_1) = Z(M_2) \circ Z(M_1)$$

which says that we can obtain the value of $Z(M)$ if we can compute Z on each of the pieces M_1 and M_2. We can, in other words, chop M into pieces, solve the problem on the pieces, and glue things back together, as per Witten's strategy. One should observe that there will, in general, be a great many ways in which to cut M into two pieces along some surface and that, according to **(A4)**, $Z(M)$ is computable from *any* of these. The idea would be to do the cutting in such a way that at least one of the pieces is simpler than M so that we have a better chance of being able to solve the problem there.

Finally, let's write this in a somewhat different form that is more common in the physics literature. Notice that $Z(M_1)$ and $Z(M_2)$ lie in dual vector spaces so that the natural pairing $\langle\, Z(M_1), Z(M_2)\,\rangle$ is well-defined. Identifying the linear map $Z(M) : \mathbb{C} \to \mathbb{C}$ with the complex number, still denoted $Z(M)$, by which one multiplies to get its values we find that

$$Z(M) = \langle\, Z(M_1), Z(M_2)\,\rangle$$

(see Remark 1.27). It will be useful in comparing our formalism with ordinary quantum mechanics to think of this natural pairing as an inner product.

Now we should say a few words about how this scenario is to be thought of physically. To obtain something in the way of motivation we will begin by recalling the familiar *Schrödinger picture* of ordinary quantum mechanics. Here one begins with a *quantum system* such as a free particle or a harmonic oscillator. The system resides in some physical space Σ which, in the simplest case, one can take to be to be just a Euclidean space $\mathbb{R}^n$. One then associates with Σ a *Hilbert space* $Z(\Sigma)$, the nonzero elements of which are identified with the (perhaps unnormalized) *states* of the system. If Σ is $\mathbb{R}^n$ one takes $Z(\Sigma)$ to be $L^2(\mathbb{R}^n)$ with respect to Lebesgue measure on $\mathbb{R}^n$. The *observables* of the system are then certain unbounded, self-adjoint operators on $Z(\Sigma)$. One then introduces unitary *time evolution operators* $U(t_0, t_1)$, $0 \leq t_0 \leq t_1 < \infty$, which carry the state $\psi(t_0)$ of the system at time t_0 onto the state $\psi(t_1)$ of the system at time t_1. These operators arise in the following way. The physics of the system dictates the selection of a particular observable H, called the *Hamiltonian* or *energy*, which, in turn, determines the Schrödinger equation $i\hbar\frac{d\psi(t)}{dt} = H(\psi(t))$. The formal solution with initial condition $\psi(t_0)$ is $\psi(t) = U(t_0, t)\psi(t_0)$, where

$$U(t_0, t) = e^{-i(t-t_0)H/\hbar}.$$

Evaluating at $t = t_1$ gives the time evolution operator $U(t_0, t_1)$. These time evolution operators are linear and satisfy

$$U(t_0, t_2) = U(t_1, t_2) \circ U(t_0, t_1), \quad t_0 \leq t_1 \leq t_2,$$

and

$$U(t,t) = \mathrm{id}_{Z(\Sigma)}$$

for all t.

Now let's rephrase this slightly. Although Schrödinger quantum mechanics is non-relativistic one can construct a picture in the Minkowski space $\mathbb{R}^{n,1}$ by fixing an inertial frame of reference, that is, an admissible basis. At each t one then has an n-dimensional spacelike hypersurface $\Sigma_t \cong \mathbb{R}^n$ in $\mathbb{R}^{n,1}$ representing all of space at the instant t. Identifying Σ_t with $\mathbb{R}^n$ also identifies $L^2(\Sigma_t)$ with $L^2(\mathbb{R}^n)$ so one can regard

$$U(t_0, t_1) : Z(\Sigma_{t_0}) \to Z(\Sigma_{t_1})$$

as a unitary map of $Z(\Sigma_{t_0})$ onto $Z(\Sigma_{t_1})$. Furthermore, the region between Σ_{t_0} and Σ_{t_1} in $\mathbb{R}^{n,1}$, which is just $M_{t_0,t_1} = \cup_{t_0 \le t \le t_1} \Sigma_t$, is a cobordism $M_{t_0,t_1} : \Sigma_{t_0} \to \Sigma_{t_1}$ from Σ_{t_0} to Σ_{t_1}. If we define

$$Z(M_{t_0,t_1}) = U(t_0, t_1), \quad 0 \le t_0 \le t_1 < \infty,$$

then

$$Z(M_{t_0,t_2}) = Z(M_{t_1,t_2}) \circ Z(M_{t_0,t_1})$$

and M_{t_0,t_2} is just M_{t_1,t_2} glued to M_{t_0,t_1}.

Thus, the axioms of Atiyah have the same general structure as this geometric view of the Schrödinger picture of quantum mechanics. The compact, oriented, smooth, n-dimensional manifolds Σ are to be thought of as physical space, but without any mathematical structure beyond that which we have just enumerated. In particular, they are not assumed to carry any Riemannian metric. The Hilbert space $Z(\Sigma)$ associated to Σ is to be regarded as a Hilbert space of quantum states of the system.

Remark 1.29. It may not be entirely clear at this point what the elements of $Z(\Sigma)$ are intended to be the states of, that is, to what quantum system we are referring. This will become clear when we discuss Witten's example, but for the moment we will content ourselves with the following remark. The $(n+1)$-manifold will have defined on it a space of *fields*. In Witten's theory these are connections (gauge potentials) on the trivial SU(2)-bundle over a 3-manifold. Each of these can be restricted to the trivial sub-bundle over a tubular neighborhood of any Σ. Then $Z(\Sigma)$ will result from the canonical quantization of this space of fields on Σ. In Witten's words, $Z(\Sigma)$ is "the physical Hilbert space of the Chern-Simons theory quantized on Σ" (page 366 of [Witt2]). We will discuss all of this in more detail shortly.

An oriented cobordism $M : \Sigma_0 \to \Sigma_1$ is to be thought of as representing an evolution of Σ_0 into Σ_1. Notice, however, that M is assumed only to be a compact, oriented, smooth, $(n+1)$-dimensional manifold with boundary $\partial M = -\Sigma_0 \sqcup \Sigma_1$. In particular, there is no Lorentz metric and no physically meaningful notion time. Nevertheless, M is the analogue in this abstract context of "spacetime" and one often finds the cobordism referred to in the literature as the "time evolution" of Σ_0 into Σ_1. The essential feature of a TQFT is that it does not presume any fixed topology for either space Σ or spacetime M and that changes in the topology of space are mediated by nontrivial topologies for spacetime (the topology of Σ is unaltered by any cobordism equivalent to $\Sigma \times [0,1]$).

Remark 1.30. It is worthwhile to contrast this with the situation in classical general relativity where results of Geroch and Tipler imply that spacetimes in which the topology of space changes can be constructed only at the expense of introducing causality violations (see [Borde] for references and some more recent results).

Axiom (**A4**) is a reflection of the quantum mechanical principle that the Hilbert space of a compound system is the tensor product of the Hilbert spaces of the independent subsystems. The operator $Z(M)$ describes the evolution of the quantum states in $Z(\Sigma_0)$ to states in $Z(\Sigma_1)$. This is often referred to as the *propagator*. If $\partial M = \emptyset$, then $Z(M)$ is a complex number and is identified with the *vacuum-vacuum expectation value*, also called the *partition function* (see page 182 of [Atiy3]). For the particular 3-dimensional TQFT constructed by Witten, which we will describe in the next three subsections, it is this partition function $Z(M)$ that defines an orientation preserving diffeomorphism invariant of the 3-manifold M.

Remark 1.31. It may seem disconcerting that our definition of a topological quantum *field* theory contains no elements that are identified with *fields* of any sort. One should keep in mind, however, that Atiyah's axioms are intended to specify a very general set of properties that a quantum field theory should possess in order to be deemed *topological*. The fields enter at the earlier stage of constructing the, hopefully topological, quantum field theory. Witten's example, for instance, begins with a 3-manifold and considers the space of SU(2)-connections on the trivial SU(2)-bundle over it. These are the fields. Witten's quantization of these classical fields to produce the invariants is an intricate and very delicate interplay of geometric quantization, path integrals and surgery on 3-manifolds which we will try to outline, ever so briefly, in Sections 1.5.3 and 1.5.4. Although it inverts the actual historical order, one might say that Witten shows that this quantum field theory is, indeed, topological by showing that it satisfies the Atiyah axioms.

It will not have escaped the reader's attention that we have yet to say anything about the presence of knots or links in the 3-manifold M. To determine the proper context in which to place Witten's approach to knot and link invariants the axioms of Atiyah require some modification. Indeed, there are a great many refinements of Atiyah's axioms, each geared specifically to the treatment of particular problems (see page 181 of [Atiy3] and pages 15-17 of [Atiy4]). Before moving on to the details in the next few sections we will briefly describe the sort of adjustments that are required for Witten's knot and link invariants. Since it is the only relevant case we will restrict our comments to 3-dimensional TQFTs. We begin then with a compact, oriented, smooth 3-manifold M and a smooth, oriented link $L = K_1 \sqcup \cdots \sqcup K_m$ in M.

Remark 1.32. The details of Witten's arguments in [Witt2] require that both M and L be assigned the additional structure of a *framing*. This means that one must fix, for M, a trivialization of its tangent bundle and, for L, a trivialization of its normal bundle in M. To be more precise then Witten's theory begins with a *framed 3-manifold M* and a *framed link L* in M. We will have a bit more to say about this somewhat later.

The physics enters in the following way. Select a compact, simply connected, simple Lie group G (we will consider only SU(2)). A principal G-bundle on the 3-manifold M is necessarily trivial and a connection on the trivial G-bundle over M can therefore be identified with a globally defined $\mathfrak{g}$-valued 1-form A on M. We will consider the affine space $\mathcal{A}$ of all such gauge potentials. These are the classical fields that will give rise to Witten's TQFT. Each of the knots K_i is a loop in M and the observables Witten has in mind for his quantum field theory are the Wilson lines (Section 1.4) of these loops with respect to a connection A. To define these one must select, for each knot K_i, a finite-dimensional representation ρ_i of G. Consequently, the objective is now to assign an invariant $Z(M, L, \rho_1, \ldots, \rho_m)$ to each collection $(M, L, \rho_1, \ldots, \rho_m)$, where M is a compact, oriented, smooth 3-manifold, $L = K_1 \sqcup \cdots \sqcup K_m$ is a smooth link in M, and ρ_i, $i = 1, \ldots, m$, is a finite-dimensional representation of G associated with the knot K_i.

Remark 1.33. One might think of this intuitively in the following way. Each knot K_i is the trajectory of a particle. Each A determines a gauge field to which the particle is coupled. The representation ρ_i characterizes the type of the particle, that is, the number of components in its wave function and how these components transform under a gauge transformation. We will see that Witten defines $Z(M, L, \rho_1, \ldots, \rho_m)$ as an expectation value, that is as a path integral over the space of gauge equivalence classes of connections.

As was the case above when no link was present, Witten's strategy for calculating $Z(M, L, \rho_1, \ldots, \rho_m)$ is to cut M along a surface Σ, solve the problem on each piece and then glue the solutions together. In our present context cutting M along Σ will produce submanifolds with boundary Σ that will generally cut the link L as well. The Transversality Theorem allows us to assume that any intersection of L with Σ is tranversal so that $\partial L \subseteq \Sigma$ is an oriented, 0-dimensional manifold, that is, a finite set $p_1, \ldots, p_r \in \Sigma$ of signed points, where the signs are determined by the orientations of Σ and the knots. As a result the object of interest here is no longer simply a 2-dimensional surface Σ, but rather a surface $\Sigma_{p_1 \ldots, p_r}$ with a finite number of *punctures* or *marked points* $p_1, \ldots, p_r \in \Sigma$, each of which is labeled with a sign ± 1 and some representation of G. In this context a *morphism* is not simply an oriented cobordism, but consists of the following data: a compact, oriented, smooth 3-manifold M_1 with boundary $\partial M_1 = -\Sigma_0 \sqcup \Sigma_1$; a finite set of marked points $p_1, \ldots, p_r$ on ∂M_1; a disjoint union of oriented, embedded circles and arcs, each labeled with a representation of G, which intersect ∂M_1 in the points $p_1, \ldots, p_r$ in such a way that the sign of each intersection point agrees with the sign with which the point is marked (see Figure 25). We will return to the problem of including the presence of links in M in Section 1.5.4.

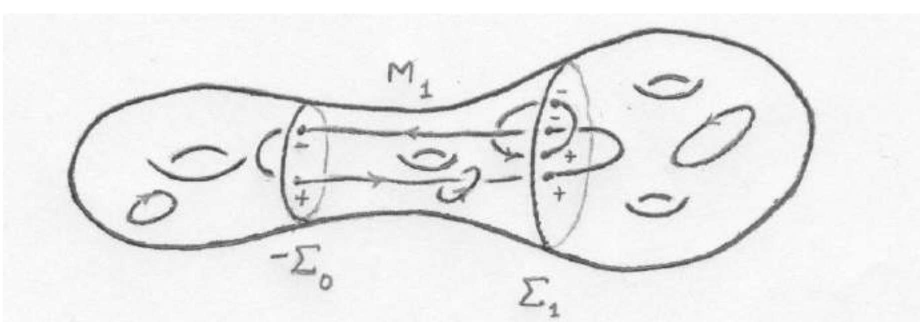


FIGURE 25. Morphism for Manifolds with Links

1.5.3. *The Hilbert Spaces* $Z(\Sigma)$*: Geometric Quantization.* In the next two sections we will look somewhat more closely at the construction of the 3-dimensional TQFT in [Witt2]. The construction is deep and difficult and we cannot pretend to offer more than a very broad sketch of how it is done and, we hope, adequate references for those who wish to know the whole story. We will, moreover, eventully focus most of our attention on the 3-manifold invariants $Z(M)$ and be content with somewhat more abbreviated remarks on $Z(M, L, \rho_1, \ldots, \rho_m)$. General references for this material are [AHS], [Atiy3], [Atiy4], [Atiy5], [ADPW] and, of course, [Witt2].

We begin by setting the stage as it was seen by Witten. As usual, M is a compact, oriented, smooth 3-manifold and $L = K_1 \sqcup \cdots \sqcup K_m$ is a smooth, oriented link in M. G is a compact, simply connected, simple Lie group with Lie algebra $\mathfrak{g}$ and, for each $i = 1, \ldots, m$, ρ_i is a finite-dimensional representation of G associated with the knot K_i. $G \hookrightarrow M \times G \xrightarrow{\pi} M$ is the trivial G-bundle over M and $\mathcal{G}$ is the group of gauge transformations of $G \hookrightarrow M \times G \xrightarrow{\pi} M$ identified with smooth maps $g : M \to G$ of M into G. $\mathcal{A}$ is the affine space of connections on $G \hookrightarrow M \times G \xrightarrow{\pi} M$ identified with globally defined gauge potentials $A \in \Omega^1(M, \mathfrak{g})$. For each $i = 1, \ldots, m$,

$$W_{K_i,\rho_i}(A) = \mathrm{tr}_{\rho_i}(\,\rho_i(\,g(A, \alpha_i, \rho_i)\,)\,)$$

is the Wilson line corresponding to the connection A, the loop K_i and the representation ρ_i (see (18)) and we will write

$$\prod_{i=1}^{m} W_{K_i,\rho_i}(A)$$

for the product of these over all of the knots in L. These are gauge invariant by Theorem 1.9. The Chern-Simons action functional on $\mathcal{A}$ at level k is given by

$$S_{CS}(A, k) = \frac{k}{4\pi} \int_M \mathrm{tr}\,(\,A \wedge dA + \frac{2}{3} A \wedge A \wedge A\,).$$

It is worth pointing out once again that the definition of the action $S_{CS}(A, k)$ does not require the presence of a metric on the 3-manifold M and so has a chance of leading to a quantum theory that is “topological” in the sense of the physicists, that is, dependent only on the differential topology of M.

All of this is perfectly rigorous mathematically, but now we must descend into the murky world of Feynman integrals. The *partition function* of Chern-Simons theory at level k is represented by the Feynman integral

$$Z_k(M) = \int_{\mathcal{A}/\mathcal{G}} e^{\,i\,S_{CS}(A,k)} \mathcal{D}A, \tag{30}$$

where the “integral” is intended to be over the moduli space $\mathcal{A}/\mathcal{G}$ of gauge equivalence classes of connections on $G \hookrightarrow M \times G \xrightarrow{\pi} M$ and $\mathcal{D}A$ is a non-existent “measure” on $\mathcal{A}/\mathcal{G}$. More generally, the (unnormalized) *expectation value* of the Wilson line observable $\prod_{i=1}^{m} W_{K_i,\rho_i}(A)$ is given by

$$Z_k(M, L, \rho_1, \ldots, \rho_m) = \int_{\mathcal{A}/\mathcal{G}} e^{\,i\,S_{CS}(A,k)} \prod_{i=1}^{m} W_{K_i,\rho_i}(A)\, \mathcal{D}A. \tag{31}$$

Witten sets himself the task of “evaluating” these Feynman integrals. The calculations in [Witt2] are rather far removed from the realm of rigorous mathematics. Nevertheless, they lead to quite explicit formulas for the invariants that can be validated by independent methods.

Remark 1.34. When $M = \mathbb{S}^3$, $G = \mathrm{SU}(2)$, and $\rho_1, \ldots, \rho_m$ are all taken to be the standard 2-dimensional (spin 1/2) representation ρ of SU(2) on $\mathbb{C}^2$, then the result of Witten's calculation is

$$Z_k(\mathbb{S}^3, L, \rho, \overset{m}{\ldots}, \rho) = V_L\left(\frac{2\pi i}{k+2}\right),$$

where V_L is the Jones polynomial of L. Known generalizations of the Jones polynomial are obtained when $M = \mathbb{S}^3$, $G = \mathrm{SU}(n)$, and $\rho_1, \ldots, \rho_m$ are all taken to be the standard n-dimensional representation ρ of SU(n). Other choices lead to a wealth of new invariants. We will have more to say about Witten's interpretation of the Jones polynomial in Section 1.5.4.

We will make no attempt to reproduce Witten's calculations here. Rather we will simply try to sketch the sense in which the constructions in [Witt2] give rise to an example of a topological quantum field theory in the sense of Atiyah when no link is present. The inclusion of links will be discussed later. In this section we will be concerned with **(Z1)**, that is, the assignment of a Hilbert space $Z_k(\Sigma)$ to each compact, oriented, smooth 2-manifold Σ without boundary. This corresponds to Section 3 (Canonical Quantization) in [Witt2], where $\Sigma \to Z_k(\Sigma)$ is discussed from the physical point of view.

The assignment $\Sigma \mapsto Z_k(\Sigma)$ can be approached rigorously in a variety of ways. We will sketch a technique based on an infinite-dimensional generalization of a procedure known as *geometric quantization*. This is a rigorous mathematical construction, due to Kostant and Souriau, of a natural canonical quantization of *certain* finite-dimensional classical mechanical systems. A standard reference for this material is [Wood] and one can also consult Chapters 22 and 23 of [Hall]; nice synopses are available in [Kiri], [Blau] and Section 1 of [ADPW]. The infinite-dimensional generalization appropriate to the Witten theory is carried out rigorously in [ADPW]. We will begin with a brief description, in the finite-dimensional case, of that part of the geometric quantization procedure that we will need here (adopting units in which the Planck constant $\hbar$ is equal to 1).

Remark 1.35. Since our objective here is simply to associate a Hilbert space to any Σ we will require only that part of the geometric quantization program that produces the quantum Hilbert space. For this reason we will describe the process only up to that point. This still involves rather considerable technical difficulties, however, and we must be content with a cursory sketch. For the rest of the story we will simply refer to [Wood] and [Kiri].

One begins with a classical mechanical system whose phase space is represented by a $2n$-dimensional manifold X with a symplectic form, that is, a non-degenerate, closed 2-form $\omega \in \Omega^2(X; \mathbb{R})$. The pair (X, ω) is called a *symplectic manifold*. Non-degeneracy implies that $\frac{1}{n!}\omega^n = \frac{1}{n!}\omega \wedge \overset{n}{\cdots} \wedge \omega$ is a volume form on X called the *Liouville measure* or *Liouville volume form*. In order for geometric quantization to succeed one must assume that $\frac{\omega}{2\pi}$ represents an *integral* cohomology class.

$$\left[\frac{\omega}{2\pi}\right] \in H^2(X; \mathbb{Z}) \tag{32}$$

This means simply that the integral of the 2-form $\omega/2\pi$ over any generator of $H_2(X; \mathbb{Z})$ is an integer. Physicists call (32) the *Bohr-Sommerfeld quantization condition*.

Example 1.1. Any oriented, 2-dimensional, smooth manifold Σ admits a volume (that is, area) form ω (see Theorem 4.3.1 of [Nab3]). It follows that ω is non-degenerate and it is closed since any 2-form on a 2-manifold is closed. Thus, ω is a symplectic form on Σ. When $\Sigma = \mathbb{S}^2$ and x^1, x^2 and x^3 are standard coordinates in $\mathbb{R}^3$ a volume form on $\mathbb{S}^2$ is given by

$$\omega = \iota^*(\, x^1 dx^2 \wedge dx^3 - x^2 dx^1 \wedge dx^3 + x^3 dx^1 \wedge dx^2 \,),$$

where $\iota : \mathbb{S}^2 \hookrightarrow \mathbb{R}^3$ is the inclusion map (see page 211 of [Nab3]). Integrating ω over $\mathbb{S}^2$ gives

$$\int_{\mathbb{S}^2} \omega = 4\pi$$

(see page 248 of [Nab3]). Consequently,

$$\int_{\mathbb{S}^2} \frac{\omega}{2\pi} = 2.$$

Since $\mathbb{S}^2$ generates $H_2(\mathbb{S}^2; \mathbb{Z}) \cong \mathbb{Z}$ we conclude that $\frac{\omega}{2\pi}$ represents an integral cohomology class.

If (32) is satisfied, then one can show that there exists a Hermitian line bundle L over X with a connection ∇ that is compatible with the Hermitian metric $\langle \; , \; \rangle_L$ of L and whose curvature is $-i\,\omega$. This is called *Weil's Theorem* (see Proposition 8.3.1 of [Wood] or Proposition 1.3 of [Kohno]).

Remark 1.36. Recall that if ∇ is any connection on L, then $\frac{i}{2\pi}$ times the curvature of ∇, which is $\frac{\omega}{2\pi}$ above, represents the 1^{st} Chern class of L and that this is an integral cohomology class that determines the bundle up to isomorphism (see Theorem E.5 of [FU1]). Recall also that ∇ is said to be compatible with $\langle \; , \; \rangle_L$ if, for every smooth vector field V on X and any two smooth sections s_1 and s_2 of L,

$$V(\langle\, s_1, s_2 \,\rangle_L) = \langle\, \nabla_V s_1, s_2 \,\rangle_L \,+\, \langle\, s_1, \nabla_V s_2 \,\rangle_L.$$

The curvature R^∇ is given, for any two smooth vector fields V and W on X and any smooth section s of L, by

$$R^\nabla(V, W)(s) = \nabla_V \nabla_W(s) - \nabla_W \nabla_V(s) - \nabla_{[V,W]}(s).$$

Since L is a Hermitian line bundle $R^\nabla(V, W)$ acts on s pointwise by multiplication so $R^\nabla(V, W) = -i\,\alpha(V, W)$ for some real-valued 2-form α. If (32) is satisfied, then $R^\nabla = -i\,\omega$ as a multiplication operator on sections.

L is called the *pre-quantum line bundle* of the symplectic manifold (X, ω). With respect to the Liouville measure one can then consider the Hilbert space $\mathcal{H}_0$ of L^2-sections of the line bundle L. This is called the *pre-quantum Hilbert space* of (X, ω).

Remark 1.37. Somewhat more precisely, we consider first the space $\Gamma_c(L)$ of smooth sections s of L with compact support. Then $\Gamma_c(L)$ has a natural L^2-Hermitian inner product $\langle\!\langle \; , \; \rangle\!\rangle$ defined by

$$\langle\!\langle s_1, s_2 \rangle\!\rangle = \int_X \langle s_1, s_2 \rangle_L \,(\frac{1}{n!}\,\omega^n).$$

$\mathcal{H}_0$ is the completion of $\Gamma_c(L)$ with respect to the corresponding norm.

$\mathcal{H}_0$ is *not* the Hilbert space of the quantization that one constructs in geometric quatization since it is generally regarded as "too big". For example, if X is the cotangent bundle $T^*\mathbb{R}^n$ of $\mathbb{R}^n$ with its standard symplectic form $\omega = dq^i \wedge dp_i$, then $\mathcal{H}_0$ is $L^2(\mathbb{R}^{2n})$, whereas the "correct" Hilbert space in this context is $L^2(\mathbb{R}^n)$. Intuitively, the Hilbert space should be L^2-functions on the configuration space, not the phase space, that is, functions of $q^1, \ldots, q^n$ and not $q^1, \ldots, q^n, p_1, \ldots, p_n$.

Nevertheless, $\mathcal{H}_0$ has one very nice property. The observables of the classical theory are the elements of the Poisson-Lie algebra $C^\infty(X)$ of smooth real-valued functions on X under pointwise multiplication and the Poisson bracket. At very least, a quantization procedure should contain a mechanism for mapping (at least some) classical observables to self-adjoint operators in such a way that a Poisson bracket of two classical observables is mapped to i times the commutator of the corresponding operators. As it happens one can write down an explicit mapping that sends *every* element of $C^\infty(X)$ to a self-adjoint operator on $\mathcal{H}_0$ and has this property (see (2.5) of [Kiri]). $\mathcal{H}_0$ together with this mapping is called the *pre-quantization* of (X, ω) (see Section 2.3 of [Kiri] for explicit calculations of the pre-quantizations for the 2-sphere $\mathbb{S}^2$ and the torus $\mathbb{S}^1 \times \mathbb{S}^1$ when ω is taken to be the standard area form). The next objective then is to cut down the size of the pre-quantum Hilbert space $\mathcal{H}_0$ and yet preserve as much of this nice property as possible.

Remark 1.38. A few remarks are in order before we proceed. First, one should keep in mind that it is generally too much to ask of any quantization procedure that it represent *all* of the classical observables by self-adjoint operators in this way (see, for example, Theorem 7.2.9 of [Nab4]). As a result one can only hope to preserve "as much as possible" of the nice property of $\mathcal{H}_0$ that we described above. Second, we should point out that there is a subtlety in the preceding discussion that we have and will continue to gloss over here. The self-adjoint operators that represent observables in quantum mechanics are invariably *unbounded* and therefore only *densely defined*. As a result one must deal with all of the usual domain issues that arise, in particular, when one tries to define the commutator of two such operators. We feel that contending with these issues here would only be a distraction so, for those who would like to see a more careful discussion, we will simply refer to Section 7.2 of [Nab4].

Cutting down the size of $\mathcal{H}_0$ is accomplished by choosing what is called a *polarization*. This should be thought of intuitively as an abstract device for cutting the number of variables in half, analogous to replacing functions of $q^1, \ldots, q^n, p_1, \ldots, p_n$ by functions of $q^1, \ldots, q^n$ in the case of $X = T^*\mathbb{R}^n$.

Remark 1.39. To define polarizations precisely we will need to recall a few items from symplectic geometry (for more details see Section 5.3 of [AM]). First suppose that V is a finite-dimensional, real vector space and ω a *symplectic form* on V, that is, a nondegenerate, skew-symmetric, bilinear form on V. The pair (V, ω) is called a *symplectic vector space*. The obvious example that we have in mind is the following. Let (X, ω) be a symplectic manifold and TX its tangent bundle. For any $x \in X$ the fiber of TX above x is the tangent space $T_x(X)$ to X at x and ω_x is a non-degenerate, skew-symmetric, bilinear form on $T_x(X)$. Thus, each $(T_x(X), \omega_x)$ is a symplectic vector space. Now let W be a linear subspace of V and define its *symplectic complement* $W^\perp$ by

$$W^\perp = \{v \in V : \omega(v, w) = 0 \ \forall w \in W\}.$$

Then W is said to be *isotropic* if $W \subseteq W^{\perp}$. This is the case if and only if $\omega|_{W\times W}$ is identically zero. W is *Lagrangian* if $W = W^{\perp}$. It will be convenient to note that this last definition can be rephrased in several ways (see Proposition 5.3.3 of [AM]).

$$\begin{aligned} W \text{ is Lagrangian } &\iff W \ \text{ is isotropic and } \dim W = \frac{1}{2}\dim V \\ &\iff W \ \text{ is a maximal isotropic subspace of } V \end{aligned}$$

Now suppose (X, ω) is a symplectic manifold of dimension $2n$. Then each fiber $T_x(X)$ of the tangent bundle TX is a $2n$-dimensional vector space on which is defined a symplectic form ω_x. Thus, we can speak of Lagrangian subspaces of $T_x(X)$ with respect to ω_x. A submanifold Y of X is said to be *Lagrangian* if, for every $y \in Y$, $T_y(Y)$ is a Lagrangian subspace of $T_x(X)$. Notice that, if $\iota : Y \hookrightarrow X$ is the inclusion map, then Y is a Lagrangian submanifold of X if and only if $\iota^*\omega = 0$ and $\dim Y = \frac{1}{2}\dim X$.

Example 1.2. Suppose X is taken to be the cotangent bundle T^*M of a smooth n-manifold M with its standard symplectic form ω. If U is a coordinate neighborhood in M with coordinates $q^1, \ldots, q^n$ and $p_1, \ldots, p_n$ are the corresponding canonical coordinates, then $q^1, \ldots, q^n, p_1, \ldots, p_n$ are coordinates on $T^*U \subseteq T^*M$ relative to which the standard symplectic form ω on T^*M is given by $\omega = -d\theta = dq^i \wedge dp_i$, where $\theta = p_i dq^i$. The image $s_0(M)$ of the 0-section s_0 of T^*M is an n-dimensional submanifold of T^*M diffeomorphic to M. The intersection $s_0(M) \cap T^*U$ of $s_0(M)$ with any T^*U is given by the equations $p_1 = \cdots = p_n = 0$ so θ vanishes on this intersection. Thus, if $\iota_0 : s_0(M) \hookrightarrow T^*M$ is the inclusion map, then $\iota_0^*\omega = -\iota_0^*(d\theta) = 0$. We conclude that $s_0(M)$ is a Lagrangian submanifold of T^*M.

Next consider a fiber $T^*_{m_0}(M)$ of T^*M. This is also an n-dimensional submanifold of T^*M. Choose a coordinate neighborhood U on M with coordinates $q^1, \ldots, q^n$ and with $m_0 \in U$. Denote the coordinates of m_0 by $q_0^1, \ldots, q_0^n$. Then $T^*_{m_0}(M)$ is given by the equations $q^1 = q_0^1, \ldots, q^n = q_0^n$ with $p_1, \ldots, p_n$ arbitrary. Since the q^i are all constant on $T^*_{m_0}(M)$, the restriction of θ to it is identically zero so once again $T^*_{m_0}(M)$ is a Lagrangian submanifold of T^*M. Notice that the fibers $T^*_m(M)$, $m \in M$, are pairwise disjoint and their union if all of T^*M. One refers to this collection of fibers as the *vertical Lagrangian foliation* of T^*M and the fibers themselves are called the *leaves* of the foliation. Next we introduce analogous notions on a general symplectic manifold.

In general, an l-dimensional *foliation* of a smooth, k-dimensional manifold X, where $0 \le l \le k$, is a family $\mathcal{F} = \{L_\alpha : \alpha \in \mathcal{A}\}$ of connected subspaces of X, called the *leaves* of the foliation, which satisfy the following properties:

(1) $L_\alpha \cap L_{\alpha'} = \emptyset$ for all $\alpha, \alpha' \in \mathcal{A}$ with $\alpha \neq \alpha'$;
(2) $\cup_{\alpha\in\mathcal{A}} L_\alpha = X$;
(3) Every point in X is contained in a coordinate neighborhood U with coordinates $x^1, \ldots, x^k$ such that, for each leaf L_α, the connected components of $U \cap L_\alpha$ are described by equations of the form $x^{l+1} = a_1, \ldots, x^k = a_{k-l}$, where $a_1, \ldots, a_{k-l}$ are constants.

In particular, it follows from (3) that each leaf L_α is an immersed l-dimensional submanifold of X. One says that X is *foliated* by the submanifolds L_α. If (X, ω) is a symplectic manifold of dimension $2n$, then an

n-dimensional foliation of X for which each leaf is a Lagrangian submanifold of X is called a *Lagrangian foliation* of X.

We recall also the classical theorem of Frobenius that relates foliations of X to smooth distributions on X (see Chapter 6 of [Spivak]). An l-dimensional *(real) distribution* on a smooth, k-dimensional manifold X is a function Δ that assigns to every $x \in X$ an l-dimensional subspace Δ_x of the tangent space $T_x(X)$. A distribution on a symplectic manifold (X, ω) is *Lagrangian* if Δ_x is a Lagrangian subspace of $(T_x(X), \omega_x)$ for every $x \in X$. The distribution is *smooth* if, for any $x_0 \in X$, there exists an open neighborhood U of x_0 and l smooth vector fields $V_1, \ldots, V_l$ on U such that $V_1(x), \ldots, V_l(x)$ form a basis for Δ_x for every $x \in U$.

An l-dimensional submanifold Y of X is called an *integral manifold* for Δ if, for each $y \in Y$, $\iota_*(T_y(Y)) = \Delta_y$, where $\iota : Y \hookrightarrow X$ is the inclusion map and ι_* is its derivative. A *maximal integral manifold* for Δ is an integral manifold for Δ that is not properly contained in any other integral manifold for Δ. When $l = 1$ an integral manifold is just an integral curve of a smooth vector field and the existence of a unique maximal integral curve through any point is assured by the basic existence and uniqueness theorems for solutions to ordinary differential equations (see Theorem 5.7.2 of [Nab2]). When $l > 1$ integral manifolds need not exist, even locally (see the example on page 6-4 of [Spivak]). To ensure the existence of a unique maximal integral manifold through any point one appeals to a theorem of Frobenius that we will now describe.

Consider a smooth, l-dimensional distribution Δ on the smooth, k-dimensional manifold X. A smooth vector field V on X is said to *belong to* Δ if $V(x) \in \Delta_x$ for every x. We will say that the distribution Δ is *integrable*, if, whenever the vector fields V and W belong to Δ, then their Lie bracket $[V, W]$ also belongs to Δ. The following is Theorem 6, page 6-22, of [Spivak], but be sure to look at Problem 5 at the end of Chapter 6 as well; alternatively, one can consult Chapter III, Section VIII, of [Chev].

Theorem 1.12. *(Frobenius) Let Δ be a smooth, integrable distribution on the smooth manifold X. Then X is foliated by maximal integral manifolds of Δ. Conversely, if $\mathcal{F}$ is a foliation of X, then the collection of tangent spaces to the leaves of $\mathcal{F}$ gives a smooth, integrable distribution on X the maximal integral manifolds of which are the leaves of $\mathcal{F}$.*

Example 1.3. Consider the vertical Lagrangian foliation of $X = T^*M$ described in Example 1.2. The leaves are just the cotangent spaces $T^*_m(M)$ for $m \in M$ and every tangent space to such a leaf can be canonically identified $T^*_m(M)$. The corresponding distribution is therefore just the assignment $m \mapsto T^*_m(M) \subseteq T(T^*M)$. Now let $\pi : T^*M \to M$ be the natural projection. This is a surjective submersion so its derivative $\pi_* : T(T^*M) \to TM$ has a kernel $\ker(\pi_*)$ that is a sub-bundle of $T(T^*M)$, called the *vertical tangent bundle* of T^*M. It consists precisely of the vectors tangent to the fibers of $\pi : T^*M \to M$ so the foliation and the distribution can be identified with the sub-bundle $\ker(\pi_*)$ of $T(T^*M)$.

Polarizations of symplectic manifolds come in two varieties. A *real polarization* $\mathcal{P}^{\mathbb{R}}$ of a symplectic manifold (X, ω) of dimension $2n$ is a foliation of X by Lagrangian submanifolds or, equivalently, it is a smooth distribution Δ on X that is integrable and Lagrangian. One can also identify a real polarization of (X, ω) with a smooth, Lagrangian sub-bundle of TX that is integrable in the sense that its sections are closed under Lie bracket.

Remark 1.40. Notice that symplectic manifolds need not have real polarizations. Consider, for example, the 2-sphere $\mathbb{S}^2$ with the symplectic (area) form ω described in Example 1.1. The existence of a smooth, Lagrangian distribution on $\mathbb{S}^2$ would imply, in particular, the existence of a smooth vector field V on $\mathbb{S}^2$ whose value at any $x \in \mathbb{S}^2$ spans the tangent space to the Lagrangian curves of the corresponding foliation of $\mathbb{S}^2$. In particular, V would be a smooth, *non-vanishing* vector field on $\mathbb{S}^2$, but it is a classical theorem in topology that $\mathbb{S}^2$ admits no continuous, non-vanishing vector fields (see Theorem 7.3 of [MT] or Theorem 3.3, Chapter XVI, of [Dug]).

Assuming that the Bohr-Sommerfeld quantization condition (32) is satisfied we have a complex line bundle L with a connection ∇ that gives rise to the pre-quantum Hilbert space $\mathcal{H}_0$ of sections of L that are square integrable with respect to the Liouville measure. If (X, ω) is supplied with a real polarization, then we can consider just those sections s of L that satisfy $\nabla_V s = 0$ for every smooth vector field V that belongs to the distribution Δ. These are constant on the leaves of the corresponding foliation which is Lagrangian and therefore n-dimensional. Consequently, s depends only on the remaining n directions in X, that is, only on n of the $2n$ coordinates in X. These are called *polarized sections* of L.

Remark 1.41. This argument should make clear why we would be interested in distributions Δ on X that are integrable and n-dimensional, but, aside from the fact that they are, indeed, n-dimensional, it may not be so clear why we insist that they be Lagrangian. We will try to provide some motivation. Suppose Δ is simply integrable and n-dimensional. As above we consider only sections s of the pre-quantum line bundle L that satisfy $\nabla_V s = 0$ for every smooth vector field V on M that belongs to the distribution Δ. These depend on just n of the $2n$ coordinates in X and this was our goal. Notice, however, that it is not obvious that nontrivial sections of this sort exist. We will look for a necessary condition for their existence. Consider the connection ∇ on L whose curvature R^∇ is $-i\,\omega$. By definition

$$R^\nabla(V, W)(s) = \nabla_V \nabla_W(s) - \nabla_W \nabla_V(s) - \nabla_{[V,W]}(s).$$

Suppose that s is a section of the required type. If V and W belong to Δ the first two terms are zero and, since Δ is integrable, $[V, W]$ also belongs to Δ so the last term is zero as well. We conclude that $R^\nabla(V, W)(s) = -i\omega(V, W)(s) = 0$ so $\omega(V, W)(s) = 0$. This condition must be satisfied by any section of the type we are looking for and it will certainly be satisfied if the restriction of ω to any Δ_x is identically zero, that is, if each Δ_x is a Lagrangian subspace of $T_x(X)$. Assuming that Δ is a Lagrangian distribution therefore removes at least the integrability condition $\omega(V, W)(s) = 0$ for the existence of sections s of the required type.

For the vertical Lagrangian foliation of T^*M, for example, the polarized sections of L are just the sections that depend only on $q^1, \ldots, q^n$. In this sense we have cut the number of variables in half. One might guess then that the quantum Hilbert space we are looking for consists of the L^2 polarized sections. There are, however, quite a few problems with this idea. Perhaps the most obvious of these is that polarized sections are constant on the leaves of the foliation so, if these leaves are not compact, the sections cannot be square integrable (see pages 158-159 of [Kiri] for additional remarks on the difficulties). We will return to the quantum Hilbert space shortly, but first must describe polarizations of the second variety. These are complex rather than real and their advantage is that they make available the powerful techniques of complex analysis and geometry.

Remark 1.42. (*Complex Structures*) To understand the complex analogue of a real polarization, and for other purposes as well, we will take a few moments to review some notions from linear algebra and differential geometry. If V is a k-dimensional real vector space, then we will denote its *complexification* by $V^{\mathbb{C}} = V \otimes \mathbb{C}$. The elements of $V^{\mathbb{C}}$ can be thought of as $v_1 + iv_2$, where $v_1, v_2 \in V$, and multiplication by $\alpha_1 + i\alpha_2 \in \mathbb{C}$ is defined in the obvious way by $(\alpha_1 + i\alpha_2)(v_1 + iv_2) = (\alpha_1 v_1 - \alpha_2 v_2) + i(\alpha_1 v_2 + \alpha_2 v_1)$. Then $V^{\mathbb{C}}$ is a complex vector space and $\dim_{\mathbb{C}} V^{\mathbb{C}} = \dim_{\mathbb{R}} V = k$. Thus, $V^{\mathbb{C}}$ is the direct sum of two copies of V called its *real* and *imaginary parts* and written

$$V^{\mathbb{C}} = V \oplus iV.$$

If $k = 2n$, then a symplectic form ω on V extends to a complex form, also denoted ω, on $V^{\mathbb{C}}$ by complex bilinearity. A complex subspace W of $V^{\mathbb{C}}$ is then *(complex) Lagrangian* if the restriction of ω to $W \times W$ is identically zero and $\dim_{\mathbb{C}} W = \frac{1}{2} \dim_{\mathbb{C}} V^{\mathbb{C}} = n$.

On the other hand, a *complex structure* on the $2n$-dimensional real vector space V is a linear transformation $J : V \to V$ of V onto itself that satisfies $J^2 = J \circ J = -id_V$. Notice that the determinant of J is ± 1 so J is an isomorphism. Given such a complex structure one can define scalar multiplication by $\alpha = \alpha_1 + i\alpha_2 \in \mathbb{C}$ on V by $\alpha v = (\alpha_1 + i\alpha_2)v = \alpha_1 v + \alpha_2 Jv$ for every $v \in V$. With this V becomes a complex vector space that we will denote V^J. If $\{e_1, \ldots, e_k\}$ is a basis for V^J over $\mathbb{C}$, then $\{e_1, Je_1, \ldots, e_k, Je_k\}$ is a basis for V over $\mathbb{R}$ so we must have $k = n$ and therefore, unlike $V^{\mathbb{C}}$, $\dim_{\mathbb{C}} V^J = \frac{1}{2} \dim_{\mathbb{R}} V = n$. We note that a real linear subspace W of V is a complex linear subspace of V^J if and only if W is invariant under J, that is, $JW \subseteq W$ and a real linear transformation $A : V \to V$ of V is a complex linear transformation of V^J if and only if A commutes with J, that is, $A \circ J = J \circ A$. If V has a symplectic form ω, then J is said to be *compatible with* ω or simply ω*-compatible* if it preserves ω in the sense that $\omega(Ju, Jv) = \omega(u, v)$ for all $u, v \in V$. This is the case if and only if J is skew-adjoint with respect to ω, that is, $\omega(Ju, v) = -\omega(u, Jv)$ for all $u, v \in V$. Notice that if one is given both a symplectic form ω and a complex structure J on V, then one can define a non-degenerate bilinear form g on V by $g(u, v) = \omega(u, Jv)$ for all $u, v \in V$ and that g is symmetric if and only if J is compatible with ω. If, in addition, $\omega(v, Jv) > 0$ for all nonzero $v \in V$, then g is positive definite and we will say the J is a *positive* ω*-compatible* complex structure on V. Thus, if V is supplied with a symplectic form ω and a complex structure J that is positive ω-compatible, then one obtains a *positive definite inner product* g on V. Notice that $g(Ju, Jv) = \omega(Ju, J^2 v) = \omega(Ju, -v) = -\omega(Ju, v) = \omega(u, Jv) = g(u, v)$ so J is an orthogonal transformation with respect to g.

Now extend $J : V \to V$ to $V^{\mathbb{C}}$ by complex linearity and continue to denote the extension by $J : V^{\mathbb{C}} \to V^{\mathbb{C}}$. Then, for $w \in V^{\mathbb{C}}$, $Jw = \lambda w \Rightarrow J(Jw) = \lambda Jw \Rightarrow -w = \lambda^2 w \Rightarrow \lambda = \pm i$ so J has eigenvalues $\pm i$. The corresponding eigenspaces are

$$V_+^{\mathbb{C}} = \{v - iJv : v \in V\} \qquad (\lambda = i)$$

and

$$\overline{V}_+^{\mathbb{C}} = \{v + iJv : v \in V\} \qquad (\lambda = -i)$$

so

$$V^{\mathbb{C}} \cong V_+^{\mathbb{C}} \oplus \overline{V}_+^{\mathbb{C}}.$$

The map $v \mapsto v - iJv$ is an isomorphism of the complex vector space V^J onto $V_+^{\mathbb{C}}$ so V^J can be identified with a subspace of $V^{\mathbb{C}}$. Then $\overline{V}_+^{\mathbb{C}}$ is a complex vector space called the *conjugate* of V^J and denoted $\overline{V}^J$. $\overline{V}^J$ has the same elements and the same additive structure as V^J, but with complex scalar multiplication $*$ defined by $\alpha * v = \overline{\alpha} \cdot v$, where $\alpha \in \mathbb{C}$, $v \in V$ and $\cdot$ is the scalar multiplication on V (we will generally not bother with the $*$ and the $\cdot$). Thus, any complex structure J on V gives rise to a decomposition of the complexification $V^{\mathbb{C}}$ of V.

$$V^{\mathbb{C}} = V^J \oplus \overline{V}^J$$

If J is compatible with the symplectic form ω on V and satisfies $\omega(v, Jv) > 0$ for all nonzero $v \in V$, then the corresponding positive definite inner product g on V induces a Hermitian inner product, also denoted g, on $V^{\mathbb{C}}$ by defining $g(u, iv) = ig(u, v) = -g(iu, v)$ for all $u, v \in V$ (our convention is that a Hermitian form is linear in the *second* slot and conjugate linear in the first). This, in turn, induces a Hermitian inner product on V^J.

Unlike V^J, a general complex vector space does not come equipped with a canonically given notion of conjugation. This is an additional element of structure that must be introduced independently. A *real structure*, or *conjugation*, on a complex vector space W is a conjugate linear map $\sigma : W \to W$ satisfying $\sigma^2 = \sigma \circ \sigma = id_W$. Thus, $\sigma(\alpha v + \beta w) = \overline{\alpha}\sigma(v) + \overline{\beta}\sigma(w)$ and $\sigma(\sigma(w)) = w$ for all $\alpha, \beta \in \mathbb{C}$ and all $v, w \in W$. For example, one real structure on $\mathbb{C}$ is just the usual conjugation map $\sigma(z) = \overline{z}$. The set of fixed points of σ is denoted

$$W^{\mathbb{R}} = \{w \in W : \sigma(w) = w\}$$

and is a *real* linear subspace of W since, for $\alpha, \beta \in \mathbb{R}$ and $v, w \in W^{\mathbb{R}}$, $\sigma(\alpha v + \beta w) = \overline{\alpha}\sigma(v) + \overline{\beta}\sigma(w) = \alpha v + \beta w$. Moreover, the complexification of $W^{\mathbb{R}}$ is precisely W.

$$W^{\mathbb{R}} \otimes \mathbb{C} = W$$

Thus, any complex vector space that admits a real structure is the complexification of a real vector space. Conversely, if W is the complexification $V^{\mathbb{C}} = V + iV$ of a real vector space V, then $\sigma(v_1 + iv_2) = v_1 - iv_2$ defines a natural real structure on $V^{\mathbb{C}}$. Consequently, choosing a real structure on W is precisely the same as choosing a real vector space whose complexification is W and this, in turn, is the same as choosing a notion of conjugation on W.

Now suppose X is a $2n$-dimensional smooth manifold. Then an assignment ω of a symplectic form ω_x to each tangent space $T_x(X)$ that varies smoothly with x in the sense that $x \mapsto \omega_x(V(x), W(x))$ is a smooth real-valued function on X for any smooth vector fields V and W on X is nothing other than a symplectic form on X. Similarly, an assignment g of a positive definite inner product g_x to each $T_x(X)$ that varies smoothly with x in the same sense is just a Riemannian metric on X. But now consider an assignment J of a complex structure $J_x : T_x(X) \to T_x(X)$ to each tangent space. Then J can be regarded as a (1,1)-tensor field on X. If this tensor field is smooth, then J is called an *almost complex structure* on X and (X, J) is an *almost complex manifold*. An almost complex structure can be thought of as a vector bundle isomorphism $J : TX \to TX$ satisfying $J^2 = -id_{TX}$. Any complex manifold X has a natural almost complex structure. This is clear locally since X has holomorphic coordinates $z^j = x^j + iy^j$ and the map J defined by $J(\partial/\partial x^j) = \partial/\partial y^j$ and $J(\partial/\partial y^j) = -\partial/\partial x^j$ is smooth and squares to minus the identity. Since the transition maps between such local coordinates are holomorphic and therefore satisfy the Cauchy-Riemann equations, these local almost complex structures piece together into a global one. The existence of an almost complex structure is therefore necessary for X to admit the structure of a complex manifold. It is *not* sufficient, however. An almost complex structure that is induced in this way from the structure of a complex manifold is said to be *integrable*. There is a well-know and quite deep result of Newlander and Nirenberg that characterizes the integrable almost complex structures on an even dimensional smooth manifold. There are a number of ways to state the result and we will describe one that is reminiscent of the Frobenius Theorem 1.12. Complexifying each fiber $T_x(X)$ of the tangent bundle TX gives the *complexified tangent bundle* $TX^{\mathbb{C}} = TX \otimes \mathbb{C}$, where $\mathbb{C}$ here means the trivial complex line bundle over X. A smooth, *complex vector field* on X is a smooth section of $TX^{\mathbb{C}}$. Extending J to $TX^{\mathbb{C}}$ as above gives a decomposition of $TX^{\mathbb{C}}$ into $\pm i$-eigenbundles of J. These are generally written $T^{1,0}M$ ($\lambda = i$) and $T^{0,1}M$ ($\lambda = -i$) and called the *holomorphic* and *anti-holomorphic tangent bundles*, respectively. In this way one obtains a Whitney sum decomposition

$$TX^{\mathbb{C}} = T^{1,0}X \oplus T^{0,1}X.$$

The Newlander-Nirenberg Theorem then states that the almost complex structure J on X is integrable if and only if

$$[\,T^{0,1}X,\, T^{0,1}X\,] \subseteq T^{0,1}X, \tag{33}$$

meaning that the Lie bracket of two smooth sections of $T^{0,1}X$ is also a section of $T^{0,1}X$ (see Proposition 2.6.17 and Theorem 2.6.19 of [Huy]). There is also a uniqueness statement. To formulate this precisely we recall that a *complex structure* on the $2n$-dimensional real manifold X is a maximal holomorphic atlas, that is, a maximal open covering of X by charts $\varphi : U \to \varphi(U) \subseteq \mathbb{R}^{2n} = \mathbb{C}^n$ for which the transition maps are holomorphic. The Newlander-Nirenberg Theorem then asserts that, if J is an integrable almost complex structure on X, then J is induced by a *unique* complex structure on X. For this reason we will generally identify a complex structure on X with the integrable almost complex structure J that it induces.

Remark 1.43. Shortly we will need to provide a very brief introduction to some of the basic notions of "conformal field theory". Here one is interested in *Riemann surfaces*, that is, smooth, orientable, 2-dimensional manifolds Σ with a complex structure. In this case it turns out that *every* almost complex structure is integrable (this is Exercise 2.6.2 of [Huy]) so that one can identify complex structures with almost complex structures.

Now suppose that (X, ω) is a symplectic manifold and J is an integrable almost complex structure on X. X therefore has the structure of a complex manifold. We say that J is *ω-compatible* if, for each $x \in X$, J_x is compatible with ω_x, that is, $\omega_x(J_x u, J_x v) = \omega_x(u, v)$ for all $u, v \in T_x(X)$. If, in addition, $\omega_x(v, J_x v) > 0$ for all $x \in X$ and all $v \in T_x(X)$, then J is *positive ω-compatible* and one defines an associated Riemanian metric g on X by $g_x(u, v) = \omega_x(u, J_x v)$ for all $x \in X$ and all $u, v \in T_x(X)$. An even dimensional smooth manifold X with all of this structure, that is, a symplectic form ω, an integrable, positive ω-compatible almost complex structure J, and the corresponding Riemannian metric g, is called a *Kähler manifold*. Although we will see in a moment that it is redundant we would like to emphasize this entire ensemble of structures by writing

$$(X, \omega, J, g)$$

to specify a Kähler manifold. The symplectic form ω is then called the *Kähler form* and g is the *Kähler metric*. Note that g is also referred to as a *Hermitian metric* on X since, as we have seen, it extends pointwise to a Hermitian metric on the complexified tangent bundle $TX^{\mathbb{C}}$. Being closed, ω determines a 2-dimensional de Rham cohomology class $[\omega] \in H^2(X; \mathbb{R})$ called the *Kähler class*.

The three items of structure in a Kähler manifold (ω, J and g) are closely related by the following compatibility conditions (V and W are smooth vector fields on X, while JV and JW are their pushforwards by the isomorphism J).

$$\omega(JV, JW) = \omega(V, W) \tag{34}$$

$$g(V, W) = \omega(V, JW) \tag{35}$$

$$\omega(V, W) = g(JV, W) \tag{36}$$

Indeed, each of these items is uniquely determined by the other two and so there are a variety of alternative definitions of a Kähler manifold. We have chosen one that begins with a symplectic manifold because these are the phase spaces of physics, but we should record at least the definition that is most prominent in differential geometry. Here one defines a Kähler manifold X in the following way. X is a complex manifold (which therefore possesses an integrable, almost complex structure J) on which is defined a Riemannian metric g satisfying $g(JV, JW) = g(V, W)$ and for which the non-degenerate 2-form ω defined by $\omega(V, W) = g(JV, W)$ is closed (and therefore a symplectic form). Thus, in geometry one generally sees a Kähler manifold specified by beginning with a complex manifold (for example, complex projective space $\mathbb{CP}^n$) and defining a Kähler metric g on it (for example, the Fubini-Study metric on $\mathbb{CP}^n$). Many classical examples of Kähler manifolds are described in Section 6, Chapter IX, of [KN2].

With this rather lengthy detour behind us we can now return to the task at hand. A *(complex) distribution* on X is a function Δ that assigns to every $x \in X$ a complex subspace Δ_x of $T_x(X)^{\mathbb{C}}$ and it is *(complex) Lagrangian* if each Δ_x is a (complex) Lagrangian subspace of $(T_x(X)^{\mathbb{C}}, \omega_x)$. The distribution is *smooth* if, for any $x_0 \in X$, there is an open neighborhood U of x_0 and n smooth, complex vector fields $V_1, \ldots, V_n$ on U such that $V_1(x), \ldots, V_n(x)$ form a basis for Δ_x for every $x \in X$. A smooth, complex vector field V on X is said to *belong to* Δ if $V(x) \in \Delta_x$ for every x. The distribution Δ is *integrable* if, whenever the complex vector fields V and W belong to Δ, then their Lie bracket $[V, W]$ also belongs to Δ. A *(complex) polarization* $\mathcal{P}^{\mathbb{C}}$ of (X, ω) is identified with a smooth, (complex) distribution Δ on X that is integrable and Lagrangian. One can identify a (complex) polarization of (X, ω) with a smooth, Lagrangian sub-bundle of $TX^{\mathbb{C}}$ that is integrable in the sense that its sections are closed under Lie bracket.

Remark 1.44. We will be interested almost exclusively in complex polarizations and so will henceforth omit the parenthetical references "(complex)" as well as the superscript "$\mathbb{C}$". Thus, $\mathcal{P}$ will always denote a complex polarization. Should we wish to consider a polarization that is real we will say so explicitly.

We should also point out that Gotay [Gotay] has found examples of symplectic manifolds that admit no polarizations at all, either real or complex, and for these the geometric quantization procedure is of no use. We also mention in passing that that there are complex versions of the Frobenius Theorem 1.12 (see, for example, [Nir]), but we will have so occasion to make use of them.

We will now begin to narrow our view and focus attention on the special case whose infinite-dimensional generalization is relevant to Witten's 3-manifold invariants. The first step is to assume that the symplectic manifold (X, ω) is a Kähler manifold (X, ω, J, g). In particular, X is a complex manifold with integrable almost complex structure J. Then J provides a decomposition of the complexified tangent bundle $TX^{\mathbb{C}} = T^{1,0}X \oplus T^{0,1}X$ into holomorphic and anti-holomorphic parts. One can then define a polarization $\mathcal{P}_J$ of X to be the anti-holomorphic sub-bundle $\mathcal{P}_J = T^{0,1}X$ of $TX^{\mathbb{C}}$. By (33), the integrability of the distribution $\mathcal{P}_J$ is equivalent to the integrability of the almost complex structure J. The Lagrangian condition follows from

$$\omega_x(u, v) = \omega_x(J_x u, J_x v) = \omega_x(-iu, -iv) = -\omega_x(u, v)$$

for all $u, v \in T_x^{0,1}X$. $\mathcal{P}_J$ is called the *Kähler polarization* of (X, ω, J, g) and it is this polarization that we will choose for the remainder of our discussion.. Note, however, that $\overline{\mathcal{P}}_J = T^{1,0}X$ is also a polarization of X. We are, of course, still assuming the Bohr-Sommerfeld quantization condition

$$\left[\frac{\omega}{2\pi}\right] \in H^2(X; \mathbb{Z})$$

so there exists a Hermitian line bundle L over X with a connection ∇ that is compatible with the Hermitian metric $\langle\ ,\ \rangle_L$ of L and whose curvature is $-i\,\omega$. L is what we have called the *pre-quantum line bundle* of (X, ω). The space $\mathcal{H}_0$ of those sections of L that are square integrable with respect to the Liouville measure is the *pre-quantum Hilbert space* of (X, ω). A section s of L is said to be *polarized* with respect to $\mathcal{P}_J$ if $\nabla_V s = 0$ for every complex vector field V with $V(x)$ in the anti-holomorphic part $\mathcal{P}_J$ of $TX^{\mathbb{C}}$ for every $x \in X$. Using only these polarized sections to define the local trivializations of L one can show that the transition functions are holomorphic and so L admits the structure of a holomorphic line bundle (see Section 9.2 of [Wood] or Proposition 1.4.17 of [Kob]). The $\mathcal{P}_J$-polarized sections of L are therefore precisely the holomorphic sections of L. The set $\mathcal{H}_J$ of elements in $\mathcal{H}_0$ that are $\mathcal{P}_J$-polarized, that is, holomorphic, is a closed subspace of $\mathcal{H}_0$ and is therefore a Hilbert space. $\mathcal{H}_J$ is the *quantum Hilbert space* of the Kähler manifold (X, ω, J, g). This construction of $\mathcal{H}_J$ is the first step in the special case of geometric quantization known as *Kähler quantization*. The next step would be the construction of a representation of the Heisenberg group in $\mathcal{H}_J$, but as we mentioned it will not be necessary for us to take this step.

The difficulty with Kähler quantization in this generality is that we would eventually like to construct 3-manifold invariants and, in general, $\mathcal{H}_J$ depends on the choice of the complex structure J. Fortunately, we do not require this much generality. Roughly, the reason is as follows (we will expand upon these remarks shortly). The *classical fields* in [Witt2] are connections on the trivial SU(2)-bundle over a 3-manifold X and these form an *affine space* $\mathcal{A}$, albeit an infinite-dimensional one. The moduli space $\mathcal{A}/\mathcal{G}$ of gauge equivalence classes of such fields is a rather complicated object and still infinite-dimensional. However, the Principle of Least Action suggests that the *physical fields* in Witten's theory are the gauge equivalence classes of *stationary points* of the Chern-Simons action and, as we have seen in Section 1.3, these are in the moduli space $\mathcal{A}_{Flat}/\mathcal{G}$ of *flat* connections. We will refer to this as the *physical phase space* of Witten's theory. This moduli space is, as Witten points out, “rather subtle, but eminently finite-dimensional” (page 367 of [Witt2]). Moreover, we will soon see that the moduli space of flat connections can be viewed as the *symplectic quotient* of $\mathcal{A}$ by $\mathcal{G}$ (defined momentarily). Consequently, the objects one really needs to study are finite-dimensional symplectic quotients of infinite-dimensional affine spaces. This is the study that is undertaken in considerable detail in [ADPW]. The analysis carried out in [ADPW] is quite lengthy and technical and we will not pretend to offer anything more that a crude outline. We begin with a brief discussion of symplectic quotients.

A *symplectic quotient*, also called a *Marsden-Weinstein reduction*, is a device for reducing the size of the phase space of a physical system by exploiting symmetries of the system and the conservation laws to which these symmetries give rise. We will briefly record the relevant definitions and refer the reader to Chapter 4 of [AM] or Chapter II of [GS] for detailed introductions to the topic.

Suppose then that (X, ω) is a symplectic manifold and G is a matrix Lie group with Lie algebra $\mathfrak{g}$. We assume that G acts smoothly on X on the left

$$\sigma : G \times X \to X$$
$$\sigma(g, x) = g \cdot x = \sigma_g(x)$$

and that this G-action respects the symplectic structure in the sense that the diffeomorphisms σ_g, $g \in G$, all preserve ω, that is,

$$\sigma_g^* \omega = \omega.$$

In this case we say that the action of G on X is *symplectic* and each σ_g is a *symplectomorphism*. G is then called a *symmetry group* of the symplectic manifold (X, ω). The first order of business is to define what is called a “moment map” of this symplectic G-action on X. We will denote by $\mathfrak{g}^*$ the vector space dual of $\mathfrak{g}$ and by $\langle \, , \, \rangle : \mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R}$ the natural pairing of $\mathfrak{g}^*$ and $\mathfrak{g}$ that evaluates the covector in the first slot on the vector in the second slot. Both $\mathfrak{g}$ and $\mathfrak{g}^*$ are finite-dimensional real vector spaces so they have natural smooth structures. Every $\xi \in \mathfrak{g}$ determines a vector field ξ_X on X called the *infinitesimal action* of ξ on X and defined, at each $x \in X$, by

$$\xi_X(x) = \frac{d}{dt} exp\,(t\xi) \cdot x \Big|_{t=0}.$$

Now consider a smooth function

$$\mu : X \to \mathfrak{g}^*.$$

Then $\mu(x) \in \mathfrak{g}^*$ for every $x \in X$ and so $\langle \mu(x), \xi \rangle$ is defined for every $\xi \in \mathfrak{g}$. Thus, we can define, for each $\xi \in \mathfrak{g}$, a smooth map

$$\hat{\mu}(\xi) : X \to \mathbb{R}$$

by

$$\hat{\mu}(\xi)(x) = \langle \mu(x), \xi \rangle$$

for every $x \in X$. The differential $d\hat{\mu}(\xi)$ is a 1-form on X for every $\xi \in \mathfrak{g}$. The contraction $\iota_{\xi_X}\omega$ of ω with ξ_X is also a 1-form on X for every $\xi \in \mathfrak{g}$. We will say that $\mu : X \to \mathfrak{g}^*$ is a *moment map* for the symplectic

G-action on X if these two are equal

$$d\hat{\mu}(\xi) = \iota_{\xi_X}\omega$$

for every $\xi \in \mathfrak{g}$. One can alleviate some of the notational clutter by rephrasing this in the following way. Let $V_{\hat{\mu}(\xi)}$ be the Hamiltonian vector field on X determined by $\hat{\mu}(\xi) \in C^\infty(X;\mathbb{R})$ so that $\iota_{V_{\hat{\mu}(\xi)}}\omega = d\hat{\mu}(\xi)$. Then μ is a moment map if and only if $\iota_{V_{\hat{\mu}(\xi)}}\omega = \iota_{\xi_X}\omega$ and, since ω is nondegenerate, this is the case if and only if

$$V_{\hat{\mu}(\xi)} = \xi_X.$$

Moment maps need not exist, but when they do their task is to provide a Hamiltonian function $\hat{\mu}(\xi)$ for each infinitesimal generator ξ_X.

This definition may appear rather abstract so we should point out two examples (Examples 4.2.15 (i) and (ii) of [AM]) and a general result (Theorem 4.2.2 of [AM]) that may help to clarify the significance of a moment map. The two examples are both from classical mechanics so we will take the configuration space to be $Q = \mathbb{R}^3$. The phase space is $T^*Q = T^*\mathbb{R}^3 = \mathbb{R}^3 \times (\mathbb{R}^3)^*$ with its standard symplectic structure and on which we will denote the canonical position and momentum coordinates by $(q,p) = (q^1,q^2,q^3,p_1,p_2,p_3)$. For the first example, we let $G = \mathbb{R}^3$ act on $Q = \mathbb{R}^3$ by translation, that is,

$$\sigma : G \times Q = \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$$
$$\sigma(a,q) = \sigma_a(q) = a + q.$$

It follows from Corollary 4.2.11 of [AM] that a moment map μ is given by

$$\mu(q,p) = p$$

and is therefore simply the classical momentum. Recall that if a classical mechanical system is invariant under translations, then the classical momentum is conserved (see page 26 of [Nab4]). This hints at a relationship between the moment map and conserved quantities that we will make more explicit in a moment. For the second example, we let the rotation group $G = \mathrm{SO}(3)$ act on the (column) vectors in $Q = \mathbb{R}^3$ by matrix multiplication, that is,

$$\sigma : G \times Q = \mathrm{SO}(3) \times \mathbb{R}^3 \to \mathbb{R}^3$$
$$\sigma(A,q) = \sigma_A(q) = Aq.$$

Identifying the Lie algebra $\mathfrak{so}(3)$ with $\mathbb{R}^3$ and $\mathbb{R}^3$ with $(\mathbb{R}^3)^*$, it follows from Corollary 4.2.13 of [AM] that a moment map μ is given by the cross product

$$\mu(q,v) = q \times v$$

which is just the classical angular momentum (when the mass $m = 1$). Again we recall that if a classical mechanical system is invariant under rotation, then the classical angular momentum is conserved (see page 40 of [Nab4]). We will make this connection between moment maps and conservation laws a bit more explicit by quoting Theorem 4.2.2 of [AM].

Theorem 1.13. *Let (X, ω) be a symplectic manifold, G a Lie group with Lie algebra $\mathfrak{g}$ and $\sigma : G \times X \to X$, $\sigma(g, x) = g \cdot x = \sigma_g(x)$, a symplectic action of G on X. Assume that the G-action has a moment map $\mu : X \to \mathfrak{g}^*$. Suppose that the smooth map $H : X \to \mathbb{R}$ is invariant under the G-action, that is,*

$$H(\sigma_g(x)) = H(x)$$

for all $x \in X$ and all $g \in G$. Then H is constant along each integral curve of the Hamiltonian vector field V_H of H.

We will need to make one more assumption about the moment maps used to construct symplectic quotients. Recall the the adjoint action of a matrix Lie group G on its Lie algebra $\mathfrak{g}$ is defined, for every $g \in G$, by

$$\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$$
$$\mathrm{Ad}_g \xi = g \xi g^{-1}$$

for every $\xi \in \mathfrak{g}$ and that the co-adjoint action of G on $\mathfrak{g}^*$ is defined, for every $g \in G$, by

$$\mathrm{Ad}_g^* : \mathfrak{g}^* \to \mathfrak{g}^*$$
$$\langle\, \mathrm{Ad}_g^* \eta,\ \xi \,\rangle = \langle\, \eta,\ \mathrm{Ad}_g \xi \,\rangle$$

for every $\eta \in \mathfrak{g}^*$ and every $\xi \in \mathfrak{g}$. We will say that a moment map $\mu : X \to \mathfrak{g}^*$ is *Ad^*-equivariant* if

$$\mu(\sigma_g(x)) = \mathrm{Ad}^*_{g^{-1}}(\mu(x))$$

for every $x \in X$ and every $g \in G$ or, equivalently,

$$\hat{\mu}(\mathrm{Ad}_g \xi)(\sigma_g(x)) = \hat{\mu}(\xi)(x)$$

for every $x \in X$, every $g \in G$ and every $\xi \in \mathfrak{g}$.

Now we will see how the moment map gives rise to symplectic quotients. We let (X, ω) be a symplectic manifold, G a matrix Lie group with Lie algebra $\mathfrak{g}$ and $\sigma : G \times X \to X$, $\sigma(g, x) = g \cdot x = \sigma_g(x)$, a symplectic G-action on X. We assume that the G-action has an Ad^*-equivariant moment map $\mu : X \to \mathfrak{g}^*$. Fix an $\eta \in \mathfrak{g}^*$ and let

$$G_\eta = \{\, g \in G : \mathrm{Ad}^*_{g^{-1}} \eta = \eta \,\}$$

be the isotropy subgroup of η under the co-adjoint action and

$$\mu^{-1}(\eta) = \{\, x \in X : \mu(x) = \eta \,\}$$

the level set of μ at η. Notice that G_η is a closed subgroup of the Lie group G and so G_η is also a Lie group (see Proposition 4.1.12 of [AM]). Now, if $g \in G_\eta$ and $x \in \mu^{-1}(\eta)$, then

$$\mu(\sigma_g(x)) = \mathrm{Ad}^*_{g^{-1}}(\mu(x)) = \mathrm{Ad}^*_{g^{-1}}(\eta) = \eta$$

so $\sigma_g(x)$ is also in $\mu^{-1}(\eta)$. Thus, G_η acts on $\mu^{-1}(\eta)$ so the orbit space

$$X_\eta = \mu^{-1}(\eta)/G_\eta$$

is a well-defined topological space with the quotient topology determined by the natural projection

$$\pi_\eta : \mu^{-1}(\eta) \to X_\eta = \mu^{-1}(\eta)/G_\eta.$$

The quotient space X_η is generally not a manifold, but one can impose additional conditions that ensure this. First one chooses η to be a *regular value* of μ so that the level set $\mu^{-1}(\eta)$ is a submanifold of X. By Sard's Theorem this is the case for "almost all" $\eta \in \mathfrak{g}^*$ (see Theorem 1.3, Chapter 3, of [Hirsch]). Next assume that the action of G_η on $\mu^{-1}(\eta)$ is *free* (only the identity element of G_η fixes any element of $\mu^{-1}(\eta)$) and *proper* (the mapping $(g, x) \mapsto (x, \sigma_g(x))$ of $G_\eta \times \mu^{-1}(\eta)$ to $\mu^{-1}(\eta) \times \mu^{-1}(\eta)$ is a proper map, that is, has the property that the inverse image of any compact set is compact). Then, according to Theorem 4.3.1 of [AM], X_η has a unique smooth structure for which the natural projection π_η is a submersion and, moreover, X_η has a unique symplectic form ω_η with the property that the pullback of ω_η to $\mu^{-1}(\eta)$ by π_η is the restriction of ω to $\mu^{-1}(\eta)$. In more detail, ω_η is defined in the following way. Let $\tilde{x} \in X_\eta$ and $\tilde{v}, \tilde{w} \in T_{\tilde{x}}(X_\eta)$. Select $x \in \mu^{-1}(\eta)$ and $v, w \in T_x(\mu^{-1}(\eta))$ with $\pi_\eta(x) = \tilde{x}$, $(\pi_\eta)_{*x}v = \tilde{v}$ and $(\pi_\eta)_{*x}w = \tilde{w}$. Then

$$(\omega_\eta)_{\tilde{x}}\,(\tilde{v}, \tilde{w}) = \omega_x(v, w).$$

G-invariance implies that this is independent of the choices of x, v and w. Thus, (X_η, ω_η) is a symplectic manifold, called a *symplectic quotient*, or *Marsden-Weinstein reduction*, of (X, ω) by the symplectic action of G on X. When $\eta = 0$ one often sees the symplectic quotient written $X//G$.

The idea behind this construction originates in classical theorems in dynamics according to which the existence of symmetries or, equivalently, conserved quantities in a mechanical system permits one to integrate out some of the variables upon which the system depends by restricting to level sets of the conserved quantities and thereby reducing the size, that is, the dimension of the phase space. Many concrete examples from physics are to be found in Chapter 4 of [AM].

Finally, we return to the gauge theory problem of interest in the construction of Witten's 3-manifold invariants. Here we have a compact, oriented, smooth 3-manifold M and a compact, oriented, smooth 2-manifold Σ embedded in M. We will assume that Σ is connected and so is characterized topologically by its genus g. Locally, near Σ, M looks like a tubular neighborhood $\Sigma \times \mathbb{R}$. The objective is to assign to Σ a finite-dimensional, complex Hilbert space $Z_k(\Sigma)$ at each level $k = 1, 2, \ldots$. We recall once again Witten's description of how we are to think of this Hilbert space. It is to be regarded as

"... the physical Hilbert space of the Chern-Simons theory quantized on Σ*."*
-Edward Witten [Witt2]

We will therefore focus, for the time being, on $\Sigma \times \mathbb{R}$ and the affine space $\mathcal{A}(\Sigma \times \mathbb{R})$ of connections on the trivial SU(2)-bundle over $\Sigma \times \mathbb{R}$. Σ need not be embedded in a 3-manifold for this discussion. We would like to apply the techniques of geometric quantization to the physical phase space

$$\mathcal{A}_{Flat}(\Sigma \times \mathbb{R})/\mathcal{G}(\Sigma \times \mathbb{R})$$

of gauge equivalence classes of flat connections on $\Sigma \times \mathbb{R}$ to produce a quantum Hilbert space.

Remark 1.45. Both the affine space $\mathcal{A}_{Flat}(\Sigma\times\mathbb{R})$ and the gauge group $\mathcal{G}(\Sigma\times\mathbb{R})$ are infinite-dimensional and so carrying out the geometric quantization program rigorously involves a nontrivial extension of the finite-dimensional techniques we have discussed and a very considerable amount of technical labor. This is done quite carefully in [ADPW] which also contains a synopsis of much of the required background material. Here we can, at best, offer a crude schematic of the game plan which eschews all of the aforementioned technical labor. In particular, much of the analytical work cannot be carried out in the smooth category, but requires appropriate Sobolev completions of the objects under consideration (see [Schm]). We will generally pass over these issues with scarcely a word. Moreover, we will not scruple to use terminology and concepts in this infinite-dimensional category that we have tried to make sense of only in a finite-dimensional one. This is inexcusable, of course, but unavoidable given our very limited objectives. As always one can rationalize: the object of real interest $\mathcal{A}_{Flat}(\Sigma\times\mathbb{R})/\mathcal{G}(\Sigma\times\mathbb{R})$ to which we will be led by these shenanigans turns out to be finite-dimensional and can be studied rigorously, taking no heed of whence it came. A nice treatment in this same spirit, but with considerably more detail than our brief sketch, is available in [Mich].

The first order of business is to reduce the problem at hand to another problem that has been well-studied by Atiyah and Bott in [AtBott]. Witten [Witt2] approaches our problem by first choosing an "axial gauge" and then imposing what he calls a "Gauss law constraint" on the fields on $\Sigma\times\mathbb{R}$. The effect is to reduce the expected phase space $\mathcal{A}(\Sigma\times\mathbb{R})/\mathcal{G}(\Sigma\times\mathbb{R})$ of gauge equivalence classes of connections on $\Sigma\times\mathbb{R}$ to the space $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ of gauge equivalence classes of flat connections on the trivial SU(2)-bundle over Σ. This reduction can be established mathematically as well. In Section 4 of [Bas] it is shown that our physical phase space, that is, the moduli space $\mathcal{A}_{Flat}(\Sigma\times\mathbb{R})/\mathcal{G}(\Sigma\times\mathbb{R})$ of flat connections on the trivial SU(2)-bundle over $\Sigma\times\mathbb{R}$ can be identified with the moduli space $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ of flat connections on the trivial SU(2)-bundle over Σ so that, for our purposes, we can restrict our attention to the latter. The problem then is to quantize this physical phase space. For this we would like to suggest that there is yet another way to view the physical phase space, namely, as a symplectic quotient of $\mathcal{A}(\Sigma)$ by $\mathcal{G}(\Sigma)$, so that one might hope to apply the machinery of geometric quantization described above.

We begin then by considering the space $\mathcal{A}(\Sigma)$ of connections A on the trivial SU(2)-bundle over Σ, identified with $\mathfrak{su}(2)$-valued 1-forms, that is, gauge potentials, on Σ. $\mathcal{A}(\Sigma)$ is an affine space modeled on the vector space $\Omega^1(\Sigma;\mathfrak{su}(2))$ of Lie algebra-valued 1-forms on Σ, that is,

$$\mathcal{A}(\Sigma)=A_0+\Omega^1(\Sigma;\mathfrak{su}(2)),$$

where A_0 is any fixed element of $\mathcal{A}(\Sigma)$. The gauge group is the set $\mathcal{G}(\Sigma)=C^\infty(\Sigma,\mathrm{SU}(2))$ of smooth maps $g:\Sigma\to\mathrm{SU}(2)$ under pointwise multiplication. $\mathcal{G}(\Sigma)$ acts on $A\in\mathcal{A}(\Sigma)$ on the right by

$$A\mapsto A^g=g^{-1}Ag+g^{-1}dg$$

and on the curvature F_A of A by

$$F_A\mapsto F_A^g=g^{-1}\,F_A\,g.$$

$\mathcal{A}(\Sigma)$ is an affine space, albeit an infinite-dimensional one, and so it can be regarded as an infinite-dimensional smooth manifold whose tangent space $T_A(\mathcal{A}(\Sigma))$ at any $A\in\mathcal{A}(\Sigma)$ can be identified with the vector space underlying the affine space, that is, the space $\Omega^1(\Sigma;\mathfrak{su}(2))$ of Lie algebra-valued 1-forms on Σ. To define a symplectic form on $\mathcal{A}(\Sigma)$ one must assign to each $T_A(\mathcal{A}(\Sigma))$ a non-degenerate, real-valued, skew-symmetric, bilinear form $(\,,\,)_A$. To do this we choose a suitably normalized positive definite, Ad-invariant inner product $\langle\,,\,\rangle$ on the Lie algebra $\mathfrak{su}(2)$.

Remark 1.46. The normalizations, which depend on the group SU(2) and the level k, are intended to ensure that the symplectic form we eventually arrive at for the physical phase space satisfies the Bohr-Sommerfeld condition (32). A general procedure for arriving at the appropriate normalization for any compact, simple Lie group G is described in [ADPW] (pages 815 and 895).

Now let $\alpha, \beta \in T_A(\mathcal{A}(\Sigma)) = \Omega^1(\Sigma; \mathfrak{su}(2))$ and define a real-valued 2-form $\langle \alpha \wedge \beta \rangle$ on Σ by

$$\langle \alpha \wedge \beta \rangle_x(u, v) = \langle \alpha_x(u), \beta_x(v) \rangle - \langle \alpha_x(v), \beta_x(u) \rangle$$

for all $x \in \Sigma$ and all $u, v \in T_x(\Sigma)$. Integrate this 2-form over Σ to obtain a skew-symmetric bilinear form

$$(\alpha, \beta)_A = \int_\Sigma \langle \alpha \wedge \beta \rangle = -(\beta, \alpha)_A$$

on $T_A(\mathcal{A}(\Sigma))$. The corresponding 2-form ω on $\mathcal{A}(\Sigma)$ whose value at any $A \in \mathcal{A}(\Sigma)$ is defined by

$$\omega_A(\alpha, \beta) = (\alpha, \beta)_A = \int_\Sigma \langle \alpha \wedge \beta \rangle = -\omega_A(\beta, \alpha)$$

is closed because it is constant in A and non-degenerate (see Lemma 88 of [Mich]). Taking for granted that all of these finite-dimensional notions can actually be made rigorous sense of on the infinite-dimensional manifold $\mathcal{A}(\Sigma)$, we conclude that ω provides $\mathcal{A}(\Sigma)$ with a symplectic structure.

Remark 1.47. Notice also that one can define, in exactly the same way, a non-degenerate, real-valued, bilinear form on $\Omega^2(\Sigma; \mathfrak{su}(2)) \times \Omega^0(\Sigma; \mathfrak{su}(2))$ and thereby identify the dual of $\Omega^0(\Sigma; \mathfrak{su}(2))$ with $\Omega^2(\Sigma; \mathfrak{su}(2))$.

$$\Omega^0(\Sigma; \mathfrak{su}(2))^* \cong \Omega^2(\Sigma; \mathfrak{su}(2))$$

We will put this to use in just a moment.

The symmetry group of the symplectic manifold $(\mathcal{A}(\Sigma), \omega)$ is the group $\mathcal{G}(\Sigma)$ of gauge transformations of the trivial SU(2)-bundle over Σ, which we identify with the group of smooth maps from Σ into SU(2) under pointwise multiplication. This, or at least an appropriate Sobolev completion of it (see [Schm]), is a Hilbert Lie group. Elements of its Lie algebra Lie $(\mathcal{G}(\Sigma))$ are smooth, $\mathfrak{su}(2)$-valued functions on Σ since these can be exponentiated pointwise to give elements of $\mathcal{G}(\Sigma)$. Thus,

$$\mathrm{Lie}\,(\mathcal{G}(\Sigma)) = \Omega^0(\Sigma; \mathfrak{su}(2)),$$

with the bracket given by the pointwise $\mathfrak{su}(2)$ bracket. One then shows that the corresponding Ad*-equivariant moment map

$$\mu : \mathcal{A}(\Sigma) \to \Omega^0(\Sigma; \mathfrak{su}(2))^* \cong \Omega^2(\Sigma; \mathfrak{su}(2))$$

is just (minus) the curvature map, that is,

$$\mu(A) = -F_A$$

(see pages 63-66 of [Mich]). Consequently, $\mu^{-1}(0)$ is precisely the set of flat connections in $\mathcal{A}(\Sigma)$ and the symplectic quotient $\mu^{-1}(0)/\mathcal{G}(\Sigma)$ is the moduli space of gauge equivalence classes of flat connections, that is, Witten's physical phase space. Depending on the context we will write this moduli space in one of the following ways.

$$\mu^{-1}(0)/\mathcal{G}(\Sigma) = \mathcal{A}(\Sigma)//\mathcal{G}(\Sigma) = \mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$$

Since 0 is generally *not* a regular value of μ, the moduli space $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ does not have a manifold structure everywhere, but rather has singularities at points corresponding the *reducible* connections in $\mathcal{A}_{Flat}(\Sigma)$.

Remark 1.48. There are a number of ways to define what it means for a connection on a smooth principal G-bundle $G \hookrightarrow P \to M$ over M to be *reducible*. All of these are simply equivalent ways of saying that the connection "comes from" a connection on a reduced sub-bundle of $G \hookrightarrow P \to M$ so we will content ourselves with the following (see Chapter II, Sections 6 and 7, of [KN1] for more details). Let H be a Lie subgroup of G with Lie algebra $\mathfrak{h} \subseteq \mathfrak{g}$. Then a *reduction of the structure group* of $G \hookrightarrow P \to M$ to H is a principal H-bundle $H \hookrightarrow Q \to M$ over M together with an H-equivariant bundle map $\iota : Q \to P$ such that $\iota(Q)$ is an H-invariant submanifold of P. A connection on $G \hookrightarrow P \to M$ is said to be *reducible* if there is such a reduction of the structure group for which the pullback of the connection by ι takes values in $\mathfrak{h}$ and is a connection on $H \hookrightarrow Q \to M$. Reducible connections generally signal the appearance of singularities in the moduli space. If the connection is not reducible it is said to be *irreducible* and these generally correspond to points in the moduli space at which a manifold structure exists.

The moduli space $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ therefore splits into two disjoint pieces. The *reducible* $\mathcal{A}^{red}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$, respectively, *irreducible* $\mathcal{A}^{irred}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ *part* of $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ is the image under the canonical projection of the reducible, respectively, irreducible connections in $\mathcal{A}_{Flat}(\Sigma)$. The moduli space $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ has the structure of a compact orbifold in which the irreducible part has a finite-dimensional manifold structure, while the reducible part is the singular set. More precisely, one has the following combination of Theorems 95 and 96 of [Mich]. Note that when $g = 0$, Σ is the 2-sphere $\mathbb{S}^2$ and the moduli space turns out to be a single point and therefore uninteresting so we will henceforth assume that $g \geq 1$.

Theorem 1.14. *Let Σ be a compact, connected, oriented, smooth surface of genus $g \geq 1$ and $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ the moduli space of flat connections on the trivial SU(2)-bundle over Σ.*

(1) *The irreducible part $\mathcal{A}^{irred}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ of the moduli space $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ admits the structure of a smooth manifold of dimension $6g-6$.*
(2) *The reducible part $\mathcal{A}^{red}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ of $\mathcal{A}_{Flat}(\Sigma)/\mathcal{G}(\Sigma)$ is homeomorphic to*

$$\mathbb{S}^1 \times \overset{2g}{\cdots} \times \mathbb{S}^1/\mathbb{Z}_2,$$

where $\mathbb{Z}_2$ acts simultaneously on each circle $\mathbb{S}^1$ by complex conjugation.

The symplectic form on $\mathcal{A}(\Sigma)$, because it is $\mathcal{G}(\Sigma)$-invariant, descends to $\mathcal{A}_{Flat}^{irred}(\Sigma)/\mathcal{G}(\Sigma)$, so this is a finite-dimensional symplectic manifold. This is the object of interest to us here so, to ease the typography a bit, we will simply refer to it as $\mathcal{M}(\Sigma)$ for the remainder of our discussion.

$$\mathcal{M}(\Sigma) = \mathcal{A}_{Flat}^{irred}(\Sigma)/\mathcal{G}(\Sigma)$$

We will write $\omega_{\mathcal{M}}$ for the symplectic form on $\mathcal{M}(\Sigma)$ induced by symplectic reduction from $(\mathcal{A}(\Sigma), \omega)$.

The problem then is to apply the geometric quantization scheme to the symplectic manifold $(\mathcal{M}(\Sigma), \omega_{\mathcal{M}})$ in order to associate with it a quantum Hilbert space that we will take to be $Z_k(\Sigma)$ in **(Z1)** of Atiyah's axioms. The first step is to construct for $(\mathcal{M}(\Sigma), \omega_{\mathcal{M}})$ a pre-quantum line bundle L. The normalizations to which we referred earlier imply that $\omega_{\mathcal{M}}/2\pi$ represents an integral cohomology class of $\mathcal{M}(\Sigma)$ so such a line bundle is guaranteed to exist. There is an explicit construction of L as a complex line bundle $\mathcal{A}_{Flat}^{irred}(\Sigma) \times_{\Theta} \mathbb{C}$ associated to $\mathcal{A}(\Sigma) \times \mathcal{G}(\Sigma)$ by a certain U(1)-valued function Θ on $\mathcal{A}(\Sigma) \times \mathcal{G}(\Sigma)$ in [RSW]. We record Theorem 1 of [RSW], incorporating into it Remark 1 of the same paper.

Theorem 1.15. *Let Σ be a compact, connected, oriented, smooth surface of genus $g \geq 1$ and let $\mathcal{A}(\Sigma)/\mathcal{G}(\Sigma)$ and $\mathcal{M}(\Sigma) = \mathcal{A}_{Flat}^{irred}(\Sigma)/\mathcal{G}(\Sigma)$ be, respectively, the moduli spaces of gauge equivalence classes of all connections and all irreducible flat connections on the trivial SU(2)-bundle over Σ. Then there is a continuous complex line bundle over $\mathcal{A}(\Sigma)/\mathcal{G}(\Sigma)$ whose restriction L to $\mathcal{M}(\Sigma)$ is a smooth Hermitian line bundle carrying a compatible connection ∇ whose curvature is $-i\,\omega_{\mathcal{M}}$.*

With the pre-quantum line bundle L in hand the next order of business is to show that $(\mathcal{M}(\Sigma), \omega_{\mathcal{M}})$ admits a Kähler structure, that is, an integrable almost complex structure that is positive and $\omega_{\mathcal{M}}$-compatible. In order to do this one must choose a complex structure J on the Riemann surface Σ.

Remark 1.49. This presents something of a problem since we are trying to define differential topological invariants and there should be no dependence on the choice of a complex structure for Σ. Although the 2-sphere $\mathbb{S}^2$ has a unique complex structure by the Uniformization Theorem (see [Abik]), when $g \geq 1$ we will see momentarily that there are a great many complex structures available. In order to describe the proof that the constructions we intend to sketch are (projectively) independent of this choice we will need to briefly recall the definition of Teichmüller space (see [IT] for more on this). We let Σ denote a compact, connected, oriented, smooth surface of genus $g \geq 1$. Two complex structures J_1 and J_2 on Σ are said to be *equivalent* if there is a diffeomorphism $F : \Sigma \to \Sigma$ of Σ onto itself such that

(1) $F : (\Sigma, J_1) \to (\Sigma, J_2)$ is holomorphic, and
(2) F is isotopic to the identity map id_{Σ} on Σ, that is, there exists a continuous homotopy $H : \Sigma \times [0,1] \to \Sigma$ for which $H_t : \Sigma \to \Sigma$, defined by $H_t(x) = H(x,t)$ for each $t \in [0,1]$, is a diffeomorphism, $H_0 = F$, and $H_1 = id_{\Sigma}$.

The set of all equivalence classes $[J]$ of complex structures J on Σ is denoted $\mathcal{T}(\Sigma)$ and called the *Teichmüller space* of Σ. The Teichmüller space of the torus, for example, is the upper half-plane $\mathcal{T}(\mathbb{S}^1 \times \mathbb{S}^1) = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ (see Section 1.2.2 of [IT]) so there is one equivalence class of complex structures for $\mathbb{S}^1 \times \mathbb{S}^1$ for every complex number with positive imaginary part. That's quite a lot of complex structures. Teichmüller spaces have a very rich topological and geometric structure. In particular, every $\mathcal{T}(\Sigma)$ admits the structure of a simply connected Kähler manifold (see [IT]).

Now we fix some choice of a complex structure J on Σ. J determines an orientation for Σ and we will assume that J is chosen in such a way that this agrees with the given orientation of Σ. J induces a complex structure $J_{\mathcal{A}(\Sigma)} : T\mathcal{A}(\Sigma) \to T\mathcal{A}(\Sigma)$ on $\mathcal{A}(\Sigma)$ as follows. Any $\alpha \in \Omega^1(\Sigma; \mathfrak{su}(2))$ can be identified with a point in $T_A(\mathcal{A}(\Sigma))$ for any $A \in \mathcal{A}(\Sigma)$ and we define $J_{\mathcal{A}(\Sigma)}(\alpha)$ by

$$J_{\mathcal{A}(\Sigma)}(\alpha) = -J\alpha,$$

where $J\alpha$ is the pushforward of the 1-form α by the isomorphism $J : T\Sigma \to T\Sigma$ (the rationale behind the minus sign is explained on page 817 of [ADPW]).

The complex structure $J_{\mathcal{A}(\Sigma)}$, in turn, induces a Kähler structure on the moduli space $\mathcal{M}(\Sigma)$, but this is not at all obvious. The usual route is by way of algebraic geometry and a theorem of Narasimhan and Seshadri [NS] that identifies $\mathcal{M}(\Sigma)$ with a certain moduli space of stable, holomorphic $SL(2,\mathbb{C})$-bundles over (Σ, J) which are known to be Kähler (see pages 617-619 of [ADPW]). We will not attempt to describe this, but will simply denote by $\mathcal{M}_J(\Sigma)$ the moduli space $\mathcal{M}(\Sigma)$ with the Kähler structure induced by J. It is, however, important to note that $\mathcal{M}_J(\Sigma)$ depends on J only up to isotopy (see page 819 of [ADPW]) so that one can parametrize the moduli spaces $\mathcal{M}_J(\Sigma)$ by the points $[J]$ in the Teichmüller space $\mathcal{T}(\Sigma)$.

Remark 1.50. $SL(2,\mathbb{C})$ is the complexification of the Lie group $SU(2)$ and this accounts for its appearance above. The complexification $\mathcal{G}(\Sigma)^{\mathbb{C}}$ of the gauge group $\mathcal{G}(\Sigma)$ is identified with the group of smooth maps from Σ into $SL(2,\mathbb{C})$. One can show that the $\mathcal{G}(\Sigma)$-action on $\mathcal{A}(\Sigma)$ can be analytically continued to a $\mathcal{G}(\Sigma)^{\mathbb{C}}$-action and that the symplectic quotient $\mathcal{A}(\Sigma)//\mathcal{G}(\Sigma)$ can be identified with the ordinary quotient $\mathcal{A}(\Sigma)/\mathcal{G}(\Sigma)^{\mathbb{C}}$ (see pages 817-819 of [ADPW]).

As in the finite-dimensional case, any pre-quantum line bundle for $\mathcal{M}_J(\Sigma)$ admits the structure of a holomorphic line bundle so that one is free to consider the space of holomorphic L^2-sections of L, or of any tensor power $L^{\otimes_k} = L \otimes \overset{k}{\cdots} \otimes L$ of L. Next one would like to associate to $\mathcal{M}_J(\Sigma)$ a Hilbert space $\mathcal{H}_J(\Sigma)$ of holomorphic sections. The appropriate way to do this, however, depends on the *level* k.

Fix a complex structure J on Σ and a level $k = 1, 2, \ldots$ and denote by

$$\mathcal{H}_J^k(\Sigma)$$

the Hilbert space of all holomorphic L^2-sections of the k-fold tensor power of the pre-quantum line bundle L over $\mathcal{M}_J(\Sigma)$. We will refer to $\mathcal{H}_J^k(\Sigma)$ as the *quantum Hilbert space of* $\mathcal{M}_J(\Sigma)$ *at level* k. The compactness of $\mathcal{M}(\Sigma)$ implies that each of these Hilbert spaces is finite-dimensional.

Remark 1.51. We precede the remainder of our discussion with a few definitions. A connection ∇ on a complex vector bundle $E \to X$ of rank n is *projectively flat* if and only if its curvature R^∇ takes values in the scalar multiples of the identity endomorphism, that is, if and only if there exists a complex 2-form α such that $R^\nabla = \alpha\, id_E$ (see Proposition 1.2.8 of [Kob]). The *projectivization* of $E \to X$ is constructed in the following way. Let $P_E \to X$ be the principal $GL(n,\mathbb{C})$-bundle of linear frames corresponding to $E \to X$.

The connection ∇ on $E \to X$ is determined by a connection on $P_E \to X$. Since $\mathrm{GL}(n, \mathbb{C})$ acts naturally on the projective space $\mathbb{CP}^{n-1}$ there is an associated fiber bundle $P_E \times_{\mathrm{GL}(n,\mathbb{C})} \mathbb{CP}^{n-1}$. This is the projectivization of $E \to X$ which one generally writes as $\mathcal{P}(E) \to X$. Now, a connection on a principal bundle determines an Ehresmann connection and therefore a parallel translation on any associated fiber bundle. Thus, the connection ∇ on $E \to X$ determines a connection on $P_E \to X$ which, in turn, determines a connection on $\mathcal{P}(E) \to X$. If ∇ is projectively flat, then the induced connection on the projectivization $\mathcal{P}(E) \to X$ is flat.

Now, let's fix the surface Σ and the level k. Then the Hilbert spaces $\mathcal{H}_J^k(\Sigma)$ depend on J only up to isotopy so one can parametrize the $\mathcal{H}_J^k(\Sigma)$ by the points $[J]$ in the Teichmüller space $\mathcal{T}(\Sigma)$ of Σ. Since our objective is to associate with Σ and k a single Hilbert space $Z_k(\Sigma)$, one would like to eliminate any dependence on J. In other words, one would like to somehow identify $\mathcal{H}_{J_1}^k(\Sigma)$ and $\mathcal{H}_{J_2}^k(\Sigma)$ when $[J_1] \neq [J_2]$, but, unfortunately, these Hilbert spaces need not be isomorphic. What we will see in just a moment, however, is that the corresponding *projectivizations* $\mathcal{P}(\,\mathcal{H}_{J_1}^k(\Sigma)\,)$ and $\mathcal{P}(\,\mathcal{H}_{J_2}^k(\Sigma)\,)$ *can* be identified. Notice that from the point of view of physics this is entirely satisfactory since quantum states are represented by *rays* in the quantum Hilbert space, that is, by points in the projectivization. We will refer to two Hilbert spaces whose corresponding projectivizations can be identified as *physically equivalent*.

Proving that $\mathcal{P}(\,\mathcal{H}_{J_1}^k(\Sigma)\,)$ and $\mathcal{P}(\,\mathcal{H}_{J_2}^k(\Sigma)\,)$ can be identified is by far the most difficult part of the argument. We will offer only a few words on the plan of attack and the source of the difficulties and then refer to [ADPW] for the very considerable technical work required to implement the plan (see [Hitchin] for another approach). One begins by regarding the Hilbert spaces $\mathcal{H}_J^k(\Sigma)$, for fixed Σ and fixed k, as the fibers of a Hilbert space bundle over Teichmüller space $\mathcal{T}(\Sigma)$.

$$\mathcal{H}_J^k(\Sigma) \longrightarrow [J] \in \mathcal{T}(\Sigma)$$

This is a complex vector bundle over $\mathcal{T}(\Sigma)$ and the goal is to construct a projectively flat connection on it (see Remark 1.51). Such a projectively flat connection would then induce a *flat* connection on the projectivization of the Hilbert space bundle, that is, on the fiber bundle obtained by replacing each $\mathcal{H}_J^k(\Sigma)$ by $\mathcal{P}(\,\mathcal{H}_J^k(\Sigma)\,)$.

$$\mathcal{P}(\,\mathcal{H}_J^k(\Sigma)\,) \longrightarrow [J] \in \mathcal{T}(\Sigma)$$

Now, a flat connection determines a parallel translation that depends only on the homotopy type of the curve over which the translation takes place. Moreover, since Teichmüller space is simply connected, any two paths joining $[J_1]$ and $[J_2]$ in $\mathcal{T}(\Sigma)$ are homotopic and so have the same parallel translation maps. This uniquely determined parallel translation map then gives a well-defined identification of the fiber above $[J_1]$ and the fiber above $[J_2]$, that is, of $\mathcal{P}(\,\mathcal{H}_{J_1}^k(\Sigma)\,)$ and $\mathcal{P}(\,\mathcal{H}_{J_2}^k(\Sigma)\,)$.

Assuming that one has managed to construct a projectively flat connection on $\mathcal{H}_J^k(\Sigma) \longrightarrow [J]$ and assuming also that one is content with identifying two Hilbert spaces when they are physically equivalent even if not isomorphic, then one can define the Hilbert space $Z_k(\Sigma)$ to be $\mathcal{H}_J^k(\Sigma)$ for any choice of the complex structure J on Σ.

$$Z_k(\Sigma) = \mathcal{H}_J^k(\Sigma)$$

Remark 1.52. If, on the other hand, one is not comfortable with identifying two Hilbert spaces that are merely physically equivalent, then comfort can be restored by refining Atiyah's TQFT axioms slightly. In their categorical formulation this amounts to replacing *functor* with *projective functor*. This same situation will recur when we discuss Segal's Axioms for Conformal Field Theory shortly.

How then does one construct such a projectively flat connection? The procedure adopted in [ADPW] is to begin with a careful examination of the analogous problem of the Kähler quantization of a finite-dimensional affine space and procedures for "pushing down" the appropriate structures to a symplectic quotient. Here everything can be done quite explicitly (see Section 1 of [ADPW]). Such a direct approach fails for the gauge theory problem because $\mathcal{A}(\Sigma)$ is "too infinite-dimensional" to have a rigorous quantization. One can, nevertheless, proceed *formally* to construct what "should" be a quantization of $\mathcal{A}(\Sigma)$ and push the information down to the finite-dimensional symplectic quotient $\mathcal{A}(\Sigma)//\mathcal{G}(\Sigma)$. The result is, at very least, a candidate for the quantization of $\mathcal{A}(\Sigma)//\mathcal{G}(\Sigma)$ (see Sections 2 and 3 of [ADPW]). The real work of showing that this candidate actually has all of the required properties takes up the remainder of the paper [ADPW] and we will have nothing further to say about it. Modulo all of the technical details concerning which we have been conspicuously silent we now have in hand the first item $Z_k(\Sigma)$ required in **(Z1)** of Atiyah's schematic of Witten's procedure in [Witt2]. The next order of business would be to turn to **(Z2)**, that is, to the partition function $Z_k(M)$.

Remark 1.53. We should also point out that there are formulas for computing the complex dimension of the quantum Hilbert spaces $\mathcal{H}_J^k(\Sigma)$ (see Chapter 11 of [Schott]). If Σ has genus g, then

$$\dim_{\mathbb{C}} \mathcal{H}_J^k(\Sigma) = \left(\frac{k+2}{2}\right)^{g-1} \sum_{j=1}^{k+1} \left(\sin^2 \frac{j\pi}{k+2}\right)^{1-g}.$$

For example, when $\Sigma = \mathbb{S}^2$ is the Riemann sphere ($g = 0$) the dimension is 1 for any k and any J. These formulas are derived from conformal field theory about which we will have more to say shortly.

As we have seen, the *partition function* of Chern-Simons theory at level k is represented in physics by the Feynman path integral

$$Z_k(M) = \int_{\mathcal{A}/\mathcal{G}} e^{i S_{CS}(A,k)}\, \mathcal{D}A, \tag{37}$$

where the integral is intended to be over the moduli space $\mathcal{A}/\mathcal{G}$ of gauge equivalence classes of connections on $G \hookrightarrow M \times G \xrightarrow{\pi} M$ and $\mathcal{D}A$ is a (non-existent) measure on $\mathcal{A}/\mathcal{G}$. Similarly, the *unnormalized expectation value* of the Wilson line observable $\prod_{i=1}^m W_{K_i,\rho_i}(A)$ is given by

$$Z_k(M, L, \rho_1, \ldots, \rho_m) = \int_{\mathcal{A}/\mathcal{G}} e^{i S_{CS}(A,k)} \prod_{i=1}^m W_{K_i,\rho_i}(A)\, \mathcal{D}A. \tag{38}$$

The *normalized expectation value* is defined by

$$\frac{Z_k(M, L, \rho_1, \ldots, \rho_m)}{Z_k(M)}.$$

It is by now a cliché to point out that these "integrals" are nothing of the sort. There are circumstances in which one can make some rigorous mathematical sense of these objects, but generally they must be regarded as heuristic devices which, when handled with care and discretion, can lead one to results that can be verified by independent means, either in the laboratory or by mathematical arguments that as a rule have nothing to do with path integrals. Physicists have developed intricate and subtle techniques for uncovering these results and their successes have been spectacular, but this terrain is vast and we will not venture into it (a somewhat dated, but classic introduction is [Schul]). Shortly we will have a few words to say about those few circumstances in which the path integral for $Z_k(M)$ can be represented by a rigorously defined mathematical object, but first we will relax, let down our guard and go through one essentially meaningless computation just for fun. Throughout this brief interlude we will assume that anything that looks like an integral really is an integral, that Fubini's Theorem applies whenever you want it to apply and, basically, that anything you would like to be true of an integral is true. The objective here is not to convince anyone of anything, but only to suggest the source of those properties of the non-existent path integrals that are being taken as axioms by Atiyah.

Let $\mathcal{A}_0(\Sigma)$ denote the space of connections on the trivial sub-bundle of $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \to M$ above the tubular neighborhood $\Sigma \times \mathbb{R} \subseteq M$ of Σ in M obtained by restricting the fields in $\mathcal{A}(M)$ to $\Sigma \times \mathbb{R}$. These are the classical field configurations on Σ. Now in quantum mechanics the Hilbert space of a quantum system is a space of functions on the configuration space (for example, $L^2(\mathbb{R}^n)$). Intuitively, one can think of $Z_k(\Sigma)$ as a "space of functions ψ of the classical fields on Σ". For each $A_0 \in \mathcal{A}_0(\Sigma)$ we set $\mathcal{A}(M, \Sigma, A_0) = \{A \in \mathcal{A}(M) : A|_\Sigma = A_0\}$. For each fixed level k we formally obtain such a function ψ_k on $\mathcal{A}_0(\Sigma)$ by taking its value at each $A_0 \in \mathcal{A}_0(\Sigma)$ to be the mythical beast

$$\psi_k(A_0) = \int_{\mathcal{A}(M,\Sigma,A_0)/\mathcal{G}} e^{i S_{CS}(A,k)} \, \mathcal{D}A. \tag{39}$$

An oriented cobordism $M_1 : \Sigma_0 \to \Sigma_1$ is thought of as the analogue of a spacetime in which the physical space Σ_0 evolves into Σ_1 (see Figure 26). The linear map $Z_k(M_1) : Z_k(\Sigma_0) \to Z_k(\Sigma_1)$ is to be identified with the evolution operator that carries quantum states ψ in $Z_k(\Sigma_0)$ to states $Z_k(M_1)\psi$ in $Z_k(\Sigma_1)$, that is, carries functions on $\mathcal{A}_0(\Sigma_0)$ to functions on $\mathcal{A}_0(\Sigma_1)$. By analogy with ordinary quantum mechanics this is taken to be the formal integral operator

$$(Z_k(M_1)\psi)(A_1) = \int_{\mathcal{A}_0(\Sigma_0)/\mathcal{G}} K_{M_1}(A_0, A_1) \, \psi(A_0) \, \mathcal{D}A_0 \tag{40}$$

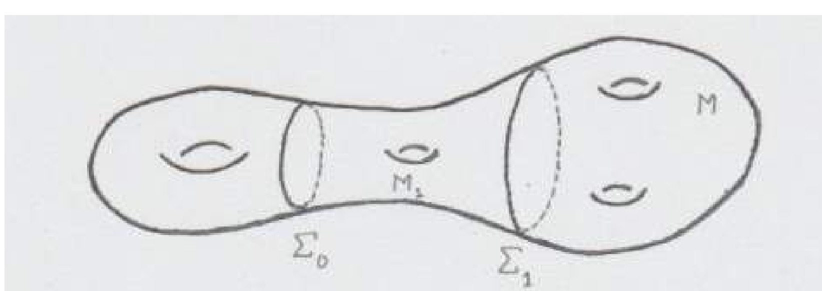


FIGURE 26. $M_1 : \Sigma_0 \to \Sigma_1$

with kernel

$$K_{M_1}(A_0, A_1) = \int_{\mathcal{A}(M_1,\Sigma_0,A_0)\,\cap\,\mathcal{A}(M_1,\Sigma_1,A_1)/\mathcal{G}} e^{\,i\,S_{CS}(A,k)}\,\mathcal{D}A. \tag{41}$$

Now let $M_2 : \Sigma_1 \to \Sigma_2$ be another oriented cobordism and $M_2 \circ M_1 : \Sigma_0 \to \Sigma_2$ the composition of M_1 and M_2. We will "verify" Atiyah's Axiom **(A5)**, that is,

$$Z_k(M_2 \circ M_1) = Z_k(M_2) \circ Z_k(M_1).$$

Now $(\,Z_k(M_2) \circ Z_k(M_1)\,)\psi = Z_k(M_2)\,(\,Z_k(M_1)\psi\,)$ is a function on $\mathcal{A}_0(\Sigma_2)$ given by

$$\begin{aligned}
Z_k(M_2)\,(\,Z_k(M_1)\psi\,)(A_2) &= \int_{\mathcal{A}_0(\Sigma_1)/\mathcal{G}} K_{M_2}(A_1, A_2)\,(\,Z_k(M_1)\psi\,)(A_1)\,\mathcal{D}A_1 \\
&= \int_{\mathcal{A}_0(\Sigma_1)/\mathcal{G}} K_{M_2}(A_1, A_2) \int_{\mathcal{A}_0(\Sigma_0)/\mathcal{G}} K_{M_1}(A_0, A_1)\,\psi(A_0)\,\mathcal{D}A_0\,\mathcal{D}A_1 \\
&= \int_{\mathcal{A}_0(\Sigma_0)/\mathcal{G}} \Big[\int_{\mathcal{A}_0(\Sigma_1)/\mathcal{G}} K_{M_1}(A_0, A_1) K_{M_2}(A_1, A_2)\,\mathcal{D}A_1\Big]\,\psi(A_0)\,\mathcal{D}A_0.
\end{aligned}$$

From this we conclude that it will suffice to show that

$$\int_{\mathcal{A}_0(\Sigma_1)/\mathcal{G}} K_{M_1}(A_0, A_1) K_{M_2}(A_1, A_2)\,\mathcal{D}A_1 = K_{M_2 \circ M_1}(A_0, A_2).$$

But

$$\begin{aligned}
\int_{\mathcal{A}_0(\Sigma_1)/\mathcal{G}} K_{M_1}(A_0, A_1) K_{M_2}(A_1, A_2)\,\mathcal{D}A_1 &= \int_{\mathcal{A}_0(\Sigma_1)/\mathcal{G}} \Big[\int_{\mathcal{A}(M_1,\Sigma_0,A_0)/\mathcal{G}} e^{i\,S_{CS}(A|_{M_1},k)}\mathcal{D}(A|_{M_1})\Big]\cdot \\
&\qquad \Big[\int_{\mathcal{A}(M_2,\Sigma_1,A_1)/\mathcal{G}} e^{i\,S_{CS}(A|_{M_2},k)}\mathcal{D}(A|_{M_2})\Big]\,\mathcal{D}A_1 \\
&= \int_{\mathcal{A}(M_1,\Sigma_0,A_0)\,\cap\,\mathcal{A}(M_2,\Sigma_2,A_2)/\mathcal{G}} e^{i\,(S_{CS}(A|_{M_1},k)+S_{CS}(A|_{M_2},k))}\mathcal{D}A \\
&= \int_{\mathcal{A}(M_1,\Sigma_0,A_0)\,\cap\,\mathcal{A}(M_2,\Sigma_2,A_2)/\mathcal{G}} e^{i\,S_{CS}(A,k)}\mathcal{D}A \\
&= K_{M_2 \circ M_1}(A_0, A_2)
\end{aligned}$$

so the "proof" is complete.

The preceding calculations are, from the mathematical point of view at least, sheer fantasy and we offer them only in deference to the well-known dictum from physics

"Shut up and calculate"

(see [Merm]). The problem of finding some solid mathematical ground upon which to build a rigorous definition of the path integral has been around ever since the notion was introduced into quantum mechanics by Feynman in his Ph.D. thesis (see [Brown]). Feynman viewed his "integral" as the limit, as $n \to \infty$, of certain n-fold multiple integrals (for an introduction to Feynman's motivation, his definition, explicit calculations for the free particle and the harmonic oscillator, and some discussion of various attempts at rigorous formulations, see Chapter 8 of [Nab4]). Feynman was not much concerned with rigorous convergence issues, but mathematicians were, of course (see, for example [Fuj1], [Fuj2], and [Fuj3]).

The issue of rigor becomes substantially more difficult when one strays outside of quantum mechanics into quantum field theory and, more particularly, gauge field theory. Indeed, rigorous definitions of the Chern-Simons path integrals (37) and (38) have been achieved only in very special cases and the general problem seems quite out of reach at present. These special cases are those, such as $M = \mathbb{R}^3$ and $M = \Sigma \times \mathbb{S}^1$, in which one can isolate a reasonable notion of Witten's "axial gauge". There is a very brief and readable account of one approach to the $M = \mathbb{R}^3$ case in [Seng]; more details are available in [AHS] and [AHKM]. For a look at the the situation when $M = \mathbb{S}^2 \times \mathbb{S}^1$ one can consult [Hahn]. Since even these special cases are quite involved and technical and since the ideas in [Witt2] certainly did not presume any rigorous notion of the path integral, we will not pursue this any further. Rather, we would like record a few preliminary remarks on the sort of arguments employed by Witten in [Witt2]. These arguments very often amount to statements of the sort, "It is well known in conformal field theory that ...".

The following quotations from [Witt2] suggest that the Hilbert spaces $Z_k(\Sigma)$ we constructed in this section appear also in *Conformal Field Theory (CFT)* in the guise of *conformal blocks* and *modular functors*, as does the projectively flat connection on the bundle of these Hilbert spaces by which one proves independence of the choice of complex structure on Σ.

"Canonical quantization on $\Sigma \times \mathbb{R}^1$ will produce a Hilbert space $\mathcal{H}_\Sigma$, 'the physical Hilbert space of the Chern-Simons theory quantized on Σ'. 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. 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."*

-*Witten* [Witt2], *page 366*

"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 in connection with conformal field theory; Segal also rigorously proved the flatness ... Because of the conformal anomaly, this bundle has only a projectively flat connection ..."*

-*Witten* [Witt2], *page 370*

Witten is referring specifically to *2-dimensional Conformal Field Theory (CFT)* and, even more specifically, to an axiomatization of 2-dimensional CFT proposed by Segal [Segal2] that is analogous to and, in fact, preceded Atiyah's axiomatization of topological quantum field theory (Section 1.5.2). It could be argued that the most profound insight in [Witt2] was not simply that knot, link and 3-manifold invariants can be viewed as arising from considerations in quantum field theory, but rather that there is a deep and unexpected connection between topology in dimension 3, where knots and links live, and in dimension 2, where they are studied and often defined via plane projections. This insight came about because Witten discovered an analogous relationship in physics between 3-dimensional Chern-Simons theory and 2-dimensional conformal field theory. The subject of 2-dimensional CFT is vast and technically difficult and we will devote an entire chapter to some of the basic ideas a bit later. For the moment we will be content to provide a bit of context by recording a very simplified version of the categorical definition of CFT proposed by Segal [Segal2] and a few rudimentary remarks on conformal blocks and modular functors.

Remark 1.54. We should say that Segal's paper [Segal2] is more of a program than a rigorous axiomatization and that much work has been done to put the program on an entirely rigorous footing. The categorical underpinnings, for example, are discussed in detail in [Fiore]. The underlying mathematical structure of CFT is described in [Schott], while those aspects related directly to the topological questions of interest to us here are discussed rigorously in [Kohno]. A rigorous study of conformal blocks and modular functors is available in [BK]. A very useful and quite influential, although unpublished source of information is [Walk]. A comprehensive introduction to CFT from the point of view of physics is to be found in [DFMS]. One should also consult the relatively brief paper [Segal4] of Segal. The discussion that follows should be regarded as a mere snapshot of the program without any of the hard work required to do it properly.

As motivation for Segal's Axioms let us consider again Atiyah's axioms for topological quantum field theory (Section 1.5.2). These were devised to meet the demands of *topological* field theory and can be thought of as an axiomatization of those properties that path integrals arising in TQFT *should* have if only they existed. The restriction to topological theories can be seen in the assumptions made of Σ and M, which are simply oriented manifolds with no additional structure, and in the morphisms, which are defined in terms of the purely topological notion of an oriented cobordism. Nevertheless, we have seen that even Schrödinger quantum mechanics can be phrased in very similar terms and one can easily imagine decorating the basic ingredients in the axioms with additional hypotheses in order to accommodate other situations. Indeed, at the end of Section 1.5.2 we indicated the alterations required for 3-manifolds containing a link if one is hoping to produce a link invariant. Viewed in this way the axioms of Atiyah and Segal appear as two instances of a general, geometrical framework for quantum field theory (see Chapter 3 of [Dijk]).

One should think of a conformal field theory as a particular type of quantum field theory that is invariant under local conformal transformations.

Remark 1.55. Needless to say, it is no mean feat to write down explicit examples of such things. In a subsequent chapter on 2-Dimensional Conformal Field Theory we will describe the *Wess-Zumino-Witten (WZW) models* which are the examples relevant to Witten's paper [Witt2]. For those who would like to see the details at this point we mention that the WZW-models are discussed in Chapter 1 of [Kohno] from the mathematical point of view and in Chapter 15 of [DFMS] from the perspective of physics.

Just as for Atiyah's TQFT axioms, Segal's definition of a 2-dimensional conformal field theory is intended to describe a very general framework within which to construct certain quantum field theories of this type. As was the case for Atiyah's axioms, the bare foundation for this framework can be encapsulated in a categorical definition (see page 424 and pages 454-456 of [Segal2] and Section 3 of [Segal4]). We will record a simplified version of this and then try to explain in a bit more detail what some of the words mean and what the definition is intended to convey.

Remark 1.56. At this point it will be useful to know a bit about the structure of smooth mapping spaces so we will pause momentarily to record a few items that we will need. These are generally Fréchet manifolds. The standard reference for this is [Ham], but a quick discussion is available on pages 74-77 of [GG]. Specifically, suppose M and N are finite-dimensional smooth manifolds with M compact and consider the set $C^\infty(M,N)$ of all smooth maps from M into N. Then $C^\infty(M,N)$ has the structure of a Fréchet manifold (Theorem 1.11, Chapter III, of [GG]). The set $\mathrm{Diff}\,(M)$ of diffeomorphisms of M onto itself is an open subset of $C^\infty(M,M)$ and is therefore also a Fréchet manifold (Proposition 1.10, Chapter III, of [GG]). The tangent space $T_{id_M}(\mathrm{Diff}\,(M))$ to $\mathrm{Diff}\,(M)$ at the identity is identified with the Fréchet space $\mathrm{Vect}\,(M)$ of smooth vector fields on M (Proposition 1.13, Chapter III, of [GG]). In addition, one can check that composition and inversion are smooth maps on $\mathrm{Diff}\,(M)$ so that $\mathrm{Diff}\,(M)$ is a Fréchet Lie group with Lie algebra $\mathrm{Vect}\,(M)$, the bracket being *minus* the usual Lie bracket of vector fields. Moreover, if M is oriented, then the set $\mathrm{Diff}_+(M)$ of orientation preserving diffeomorphisms of M is the connected component containing the identity in $\mathrm{Diff}\,(M)$ and is therefore also a Fréchet Lie group with the same Lie algebra. The case of most interest is $\mathrm{Diff}_+(\mathbb{S}^1)$. Although we will make no use of them we should point out two rather remarkable facts about the Fréchet Lie group $\mathrm{Diff}_+(\mathbb{S}^1)$. It was proved by Herman [Herm] that $\mathrm{Diff}_+(\mathbb{S}^1)$ is a simple group, that is, it has no nontrivial normal subgroups. In particular, there are no nontrivial homomorphisms of $\mathrm{Diff}_+(\mathbb{S}^1)$ into a connected, complex Lie group. Another consequence of this is that $\mathrm{Diff}_+(\mathbb{S}^1)$ does not admit a complexification, that is, there is no complex Lie group whose Lie algebra is $\mathrm{Vect}^{\mathbb{C}}(\mathbb{S}^1)$.

Another example of particular interest in conformal field theory can be described as follows. Suppose that $M = \mathbb{S}^1$ and $N = G$ is a compact, connected, matrix Lie group ($G = \mathrm{SU}(2)$ is the case of most interest to us). Denote by LG the Fréchet space $C^\infty(\mathbb{S}^1, G)$. An element γ of LG is called a *loop* in G and can be identified with a smooth map of $\mathbb{R}$ into G that is periodic with period 2π. We define a group structure on LG by pointwise multiplication, that is, for $\gamma_1, \gamma_2 \in LG$ we define $\gamma_1\gamma_2$ at each $z \in \mathbb{S}^1$ by

$$(\gamma_1\gamma_2)(z) = \gamma_1(z)\gamma_2(z).$$

If $e : \mathbb{S}^1 \to G$ is the constant map that sends every $z \in \mathbb{S}^1$ to the identity in G, then $\gamma e = e\gamma = \gamma$ for any $\gamma \in LG$ and $\gamma^{-1} : \mathbb{S}^1 \to G$ defined by $\gamma^{-1}(z) = \gamma(z)^{-1}$ is the inverse of γ in LG. Multiplication and inversion are both smooth in the Fréchet manifold structure so LG is a Fréchet Lie group, called the *loop group* of G. The standard reference for loop groups is [PreSeg]. The loop group LG has a Lie algebra that is identified with the algebra $L\mathfrak{g} = C^\infty(\mathbb{S}^1, \mathfrak{g})$ of loops in the Lie algebra $\mathfrak{g}$ of G under pointwise linear operations and bracket. One can therefore exponentiate elements of $L\mathfrak{g}$ pointwise to obtain elements of LG. However, this exponential map has certain deficiencies. For instance, the exponential map of a compact Lie group G maps *onto* the connected component of the identity in G, but this is not the case for LG (see the Example, page 27, of [PreSeg]).

With this we return to our brief preview of Segal's approach of conformal field theory.

Segal's Axioms for CFT: Categorical Formulation

A 2-dimensional conformal field theory (CFT) is a symmetric, monoidal functor from the category $\mathcal{C}$ whose objects are compact, oriented, smooth, 1-dimensional manifolds and whose morphisms are oriented, conformal cobordisms to the category whose objects are complex Hilbert spaces and whose morphisms are bounded linear maps.

Remark 1.57. As was the case for Atiyah's Axioms (see Remark 1.52) one should, strictly speaking, replace *functor* with *projective functor* since quantum theory cares only about projective Hilbert spaces. What this means in practice is that the functor will assign to each object in $\mathcal{C}$ a *projective* Hilbert space, that is, a Hilbert space uniquely defined only up to projective (physical) equivalence.

We will try to flesh out the definition with a few remarks. Much of this should be quite reminiscent of Atiyah's axioms for TQFT in Section 1.5.2. An object in the category $\mathcal{C}$ on which the functor is defined is a finite and possibly empty disjoint union of oriented, smooth, circles (every connected, compact 1-manifold is diffeomorphic to $\mathbb{S}^1$; see the Appendix of [Miln1], Appendix 2 of [GP], or Theorem 5.30 of [Nab1]). The CFT maps a circle S to a Hilbert space $\mathcal{H}_S$ and a diffeomorphism between circles to an isomorphism of the corresponding Hilbert spaces. $\mathcal{H}_S$ is the state space of the circle (*string*, if you like). The assumption that the CFT is monoidal means that a disjoint union of circles is mapped to the tensor product of the corresponding Hilbert spaces and the empty 1-manifold is sent to the Hilbert space $\mathbb{C}$. In particular, if $\mathbb{S}^1$ is the standard unit circle with its usual counterclockwise orientation and if we write $\mathcal{H}$ for $\mathcal{H}_{\mathbb{S}^1}$, then every $\mathcal{H}_S$ is isomorphic to $\mathcal{H}$ and an element of $\mathcal{C}$ consisting of n circles is mapped to the n-fold tensor product $\mathcal{H}^{\otimes n} = \mathcal{H} \otimes \overset{n}{\cdots} \otimes \mathcal{H}$.

Remark 1.58. To define the morphisms of $\mathcal{C}$, that is, the *oriented, conformal cobordisms*, we recall that a *Riemann surface with boundary* is a smooth, oriented, 2-dimensional manifold Σ with boundary $\partial\Sigma$ for which there exists a covering by charts $\varphi_\alpha : U_\alpha \to \varphi_\alpha(U_\alpha)$, where each U_α is an open set in Σ and $\varphi_\alpha(U_\alpha)$ is an open set in the closed upper half-plane $\mathrm{Im}\,(z) \geq 0$ in $\mathbb{C}$, for which the overlap mappings $\varphi_\alpha \circ \varphi_\beta^{-1}$: $\varphi_\beta(U_\alpha \cap U_\beta) \to \varphi_\alpha(U_\alpha \cap U_\beta)$ are holomorphic whenever $U_\alpha \cap U_\beta \neq \emptyset$. Recall also that a function f is said to be holomorphic at a boundary point of $\mathrm{Im}(z) \geq 0$ if f is extendible to a holomorphic function on an open neighborhood of that point in $\mathbb{C}$. In particular, $\Sigma - \partial\Sigma$ is a complex manifold. If Σ is compact and the boundary $\partial\Sigma$ is nonempty, then this boundary must be a disjoint union of smooth, oriented circles (see Figure 21). To see some initial point of contact with Witten's theory we note that a compact Riemann surface Σ with boundary $\partial\Sigma$ consisting n circles is homotopically equivalent to a compact Riemann surface $\Sigma_{p_1,\ldots,p_n}$ with n punctures.

Now, if $\sqcup S$ and $\sqcup S'$ are two objects in the category $\mathcal{C}$, then a morphism from $\sqcup S$ to $\sqcup S'$ is a Riemann surface Σ with boundary $\partial\Sigma$ together with an orientation preserving diffeomorphism of $(-\sqcup S\,) \sqcup (\,\sqcup S'\,)$ onto $\partial\Sigma$. Σ is called an *oriented, conformal cobordism* from $\sqcup S$ to $\sqcup S'$. The image of $\sqcup S$ is called the *incoming boundary* of Σ and denoted $\partial_{\,\mathrm{in}}\,\Sigma$, while the image of $\sqcup S'$ is the *outgoing boundary* of Σ and denoted $\partial_{\,\mathrm{out}}\,\Sigma$.

Intuitively, one can think of the cobordism as describing the “evolution” of the incoming strings into the outgoing strings. Two such cobordisms are identified if they are conformally equivalent. If the boundary of Σ consists of n circles, p of which are incoming and q of which are outgoing, then the evolution of the string states is given by an operator from $\mathcal{H}^{\otimes p}$ to $\mathcal{H}^{\otimes q}$. A compact Riemann surface Σ without boundary is regarded as a conformal cobordism from $\emptyset$ to $\emptyset$ and is therefore mapped by the CFT to a linear map from $\mathbb{C}$ to $\mathbb{C}$, that is, to a complex number, called the *partition function* corresponding to Σ. Composing conformal cobordisms Σ and Σ' requires some care due to the presence of the complex structures. We have seen that gluing two smooth cobordisms is easy in the topological category, but it requires some work to show that the resulting manifold admits a natural smooth structure (Theorem 1.4 of [Miln2]). One can show that, in our present context, the smooth composition $\Sigma' \circ \Sigma$ admits the structure of a Riemann surface with boundary that is compatible with the structures of Σ and Σ' and for which the embedding of the glued circles in $\Sigma' \circ \Sigma$ is smooth. This defines the composition of the two conformal cobordisms. The CFT functor carries this composition onto the composition of the corresponding morphisms (bounded linear maps) in the Hilbert space category.

As was the case for Atiyah’s axioms, this categorical definition of CFT provides only a context within which to view the subject. In any physically motivated, concrete model of CFT the objects of central interest are fields ϕ, each component of which is an operator on the states, and their so-called *vacuum expectation values*

$$\langle \Psi_0,\, \phi_{\alpha_1}(z_1)\cdots\phi_{\alpha_n}(z_n)\,\Psi_0\rangle. \tag{42}$$

Here Ψ_0 is the *vacuum state* of the theory and each $\phi_{\alpha_j}(z_j)$ represents a component of a field thought of as a function on an open subset of a Riemann surface Σ with local conformal coordinate z_j and taking values in the, generally unbounded and self-adjoint, operators.

Remark 1.59. Describing the fields in this way is common practice (see, for example, Section 9.1 of [Schott]), but is also both mathematically and physically unsound. We will see later that, in order to proceed rigorously, quantum fields cannot be thought of as operator-valued *functions*, but rather must be regarded as operator-valued *distributions* (see [Wight] for the history of how this realization came about). For the moment we will not worry about this.

The vacuum expectation value (42) is also referred to as a *correlation function* or an *n-point function*. We should point out that there are alternative axioms for CFT that focus on these correlation functions, specifying properties that they should have and then proving that these properties determine the underlying Hilbert space and the fields (see Chapter 9 of [Schott]). It turns out that in CFT all of these correlation functions can be expressed in terms of the correlation functions for what are called *primary fields* so one can restrict one’s attention to these. Conformal invariance determines the 2-point functions and the 3-point functions for primary fields up to a numerical constant, but for $n \geq 4$ this is not the case. Nevertheless, Belavin, Polyakov, and Zamolodchikov [BPZ] have isolated a family of functions, called *conformal blocks*, that are completely determined by conformal invariance and in terms of which any n-point function can be constructed (see page 353 of [BPZ]). The collection of linearly independent conformal blocks on $\Sigma_{g,n}$ spans a complex vector space $V(\Sigma_{g,n})$ and, if these are finite-dimensional, the conformal field theory is said to be *rational*. We will be interested only in rational conformal field theories, the best known of which are the so-called *Wess-Zumino-Witten models* to which we referred in Remark 1.55.

In Segal's scheme a crucial role is played by what are called "modular functors" (see Section 5 of [Segal2] and Section 3 of [Segal4]). Arriving at a precise definition of these objects requires some considerable technical work (see [BK]) so, for the moment, we must be satisfied with a modest indication of what they are intended to accomplish. A *label set* is some finite set $\mathcal{L}$ furnished with an involution $a \mapsto \hat{a}$ and containing an element 1 with the property that $\hat{1} = 1$. One might think, for example, of a finite set of representations of some group with the involution operation taking a representation to its dual and 1 corresponding to the trivial representation. Next consider a compact Riemann surface Σ with boundary $\partial\Sigma$ consisting of n oriented, closed curves $\partial_i\Sigma$, $i = 1, \ldots n$, each diffeomorphic to the circle $\mathbb{S}^1$. To each $\partial_i\Sigma$ we assign a label $a_i \in \mathcal{L}$ and fix a parametrization of $\partial_i\Sigma$, that is, a diffeomorphism $\psi_i : \partial_i\Sigma \to \mathbb{S}^1$ of $\partial_i\Sigma$ onto $\mathbb{S}^1$. If it is necessary to keep this data at the forefront we may write $\Sigma_{a,\psi}$, where $a = (a_1, \ldots, a_n)$ and $\psi = (\psi_1, \ldots, \psi_n)$, but when all of the decorations (g, n, a, ψ) are understood we will usually just write Σ. Segal defines a *modular functor* based on the label set $\mathcal{L}$ as a rule V that assigns to each $\Sigma_{a,\psi}$ a finite-dimensional complex vector space $V(\Sigma_{a,\psi})$. The assignment $\Sigma_{a,\psi} \mapsto V(\Sigma_{a,\psi})$ is assumed to satisfy a number of properties that one would expect of a symmetric, monoidal functor. It must, for example, take disjoint unions to tensor products and must send $-\Sigma_{a,\psi}$ to the dual of $V(\Sigma_{a,\psi})$. There are also certain holomorphicity requirements. For example, (3.4) of [Segal4] requires that, if $\{\Sigma^\tau\}_{\tau\in M}$ is a holomorphic family of surfaces parametrized by the points τ in some complex manifold M, then the corresponding vector spaces $\{V(\Sigma^\tau)\}_{\tau\in M}$ form a holomorphic vector bundle over M. The gluing procedure is required to respect the given parametrizations, is to be applied only to boundary components with the same label and it too is required to be holomorphic (see page 464 of [Segal2]).

Theorem 3.8 of [Segal4] asserts that, for any modular functor V, if $\{\Sigma^\tau\}_{\tau\in M}$ is a holomorphic family of surfaces, then the holomorphic vector bundle $\{V(\Sigma^\tau)\}_{\tau\in M}$ admits a projectively flat connection (see Remark 1.51). He then credits Witten with Corollary 3.9, according to which the projectivization of any $V(\Sigma)$ is naturally associated to the smooth surface Σ without any choice of complex structure on it. We will sketch the construction of a particularly important class of modular functors shortly.

Finally, we would like to say a few words about the following rather provocative statement of Segal in [Segal3].

In fact it is not much of an exaggeration to say that the mathematics of two-dimensional conformal field theory is almost the same thing as the representation theory of loop groups.

To understand what is behind this recall that in our discussion of $Z_k(\Sigma)$ when there are no links present the appropriate space of fields was $\mathcal{A}(\Sigma)$, that is, the space of connections on the trivial SU(2)-bundle over Σ and that the gauge group $\mathcal{G}(\Sigma)$ was identified with the group $C^\infty(\Sigma; \mathrm{SU}(2))$ under pointwise multiplication. In conformal field theory Σ is replaced by $\mathbb{S}^1$ so that the appropriate gauge group is $C^\infty(\mathbb{S}^1; \mathrm{SU}(2))$, that is, the loop group $L(\mathrm{SU}(2))$ of SU(2). The representations of the gauge group provide the objects on which the gauge group acts in various ways and consequently the appropriate notion of gauge invariant quantities with which physical theories deal. We understand already that it is really only the projective representations that are of interest and will find somewhat later that, in conformal field theory, there is a further restriction to the so-called *positive energy representations*. The restriction comes about due to the natural action of $\mathrm{Diff}(\mathbb{S}^1)$ on $L(\mathrm{SU}(2))$ that rotates the loops. These positive energy representations are all projective and have very attractive properties quite like the representations of compact groups. They are, for example, completely reducible and unitary (see Theorem 9.3.1 of [PreSeg]). In the following example, taken from [Segal4], we will say a bit more about positive energy representations and see how they give rise to an important class of modular functors. For some of the details that we omit in our quick sketch one can turn to Section 7 of [Segal4].

Example 1.4. Let G be a compact, simple Lie group and $G^{\mathbb{C}}$ its complexification. The example of most interest to us is $G = \mathrm{SU}(2)$ so that $G^{\mathbb{C}} = \mathrm{SL}(2, \mathbb{C})$. We consider the loop group $LG^{\mathbb{C}}$ of $G^{\mathbb{C}}$ and a positive energy representation $\rho : LG^{\mathbb{C}} \to \mathrm{Aut}\,(E)$ of it. Denote by U_γ the action of $\gamma \in LG^{\mathbb{C}}$ on E and by γ_α the loop γ rotated by $\alpha \in \mathbb{R}$. Thus, $U_\gamma(\xi) = \rho(\gamma)(\xi)$ and $\gamma_\alpha(\theta) = \gamma(\theta - \alpha)$. The positive energy condition implies that there is defined on E a "Hamiltonian" operator $H : E \to E$ that satisfies

$$\frac{d}{d\alpha}\, U_{\gamma_\alpha} = i\,[H, U_{\gamma_\alpha}]$$

and has positive integer eigenvalues with each

$$E_n = \{\, \xi \in E : H\xi = n\xi \,\}, \quad n = 0, 1, 2, \ldots,$$

finite-dimensional. Intuitively, one thinks of rotation on a 1-dimensional "space" consisting of a loop as playing the role of time translation. Since positive energy representations are all projective one can think of ρ as an honest representation of a central extension $\tilde{L}G^{\mathbb{C}}$ of $LG^{\mathbb{C}}$ by the nonzero complex numbers $\mathbb{C}^\times$. An element u of the central subgroup $\mathbb{C}^\times$ of $\tilde{L}G^{\mathbb{C}}$ acts on E by multiplication by u^k for some positive integer k called the *level* of the representation. One can show that the representation ρ on $\tilde{L}G^{\mathbb{C}}$ can be reconstructed from the level k and its "lowest energy part" E_0, which is a representation space for $G^{\mathbb{C}}$ and is irreducible if ρ is irreducible (see page 35 of [Segal4]). Positive energy representations of $LG^{\mathbb{C}}$ are therefore determined by a positive integer and a representation of $G^{\mathbb{C}}$. There are, moreover, only finitely many irreducible positive energy representations of $LG^{\mathbb{C}}$ of a given level k. Positive energy is a sort of finiteness condition.

With this we can now describe Segal's basic example of a modular functor. Fix a level k and let the label set $\mathcal{L}$ consist of the finitely many equivalence classes of irreducible representations of $G^{\mathbb{C}}$ on E_0 whose highest weight λ satisfies $\|\lambda\|^2 \leq 2k$. Denote by $\{E^a\}_{a \in \mathcal{L}}$ the corresponding positive energy representations of $LG^{\mathbb{C}}$. Now suppose that Σ is a Riemann surface with m boundary circles $\partial_i\Sigma$, $i = 1, \ldots, m$ and select a parametrization of each boundary component. Let $G^{\mathbb{C}}_\Sigma$ denote the group of holomorphic maps from Σ to $G^{\mathbb{C}}$. Being holomorphic, these maps are determined by their boundary values so $G^{\mathbb{C}}_\Sigma$ can be regarded as a subgroup of $LG^{\mathbb{C}} \times \overset{m}{\cdots} \times LG^{\mathbb{C}}$. If we label the i^{th} boundary circle $\partial_i\Sigma$ with one of these representation E^{a_i}, then $G^{\mathbb{C}}_\Sigma$ acts projectively on

$$E^a = E^{a_1} \otimes \cdots \otimes E^{a_m}.$$

Finally, we let $V(\Sigma)$ denote the subspace of E^a that is fixed by $G^{\mathbb{C}}_\Sigma$. Segal proves that the assignment

$$\Sigma \mapsto V(\Sigma)$$

is a modular functor (see pages 36-37 of [Segal4] for the crucial gluing axiom which depends on a version of the Peter-Weyl Theorem for loop groups).

This will have to suffice for the time being. We have noted above that all of this information on Segal's Axioms requires rather substantial work to make entirely rigorous and we must leave any more detailed discussion to another time. We will mention a few more features of 2-dimensional, rational CFT as we proceed to follow the remainder of Witten's argument in [Witt2] and will then take up the subject again in a subsequent chapter.

1.5.4. *The Jones-Witten Invariants.* At this point we have reached the stage at which simply trying to translate Witten's ideas into rigorous mathematics encounters a seemingly impenetrable wall. Perhaps one day the relevant path integrals will yield to rigorous treatment, but this would seem to be in the distant future. This leaves us with two options. One might develop a rigorous mathematical structure which evades the path integrals and yet reproduces the results to which Witten has been led by his formal arguments. This has, in fact, been done in a variety of ways (see, for example, [KM], [Kohno], [PraSos], [RT] and [Tur2]). This important work is, however, technically quite involved and rather far removed from the physically motivated ideas in [Witt2]. On the other hand, one could adopt the point of view that it is not so much the results themselves, but rather the fact that they arose out of intuitions born in physics that demands the attention of mathematicians. One could, in other words, forbear the natural inclination to dismiss arguments relying on properties of an object that has no rigorous mathematical definition. Mathematics is, after all, also born out of intuitions that have no rigorous justification; the rigor is an afterthought. For the remainder of our discussion we will adopt this more casual attitude and simply listen to what Witten is trying to tell us by providing a brief outline of some of the remarkable ideas in [Witt2]. We recall the context. M is a compact, oriented, smooth 3-manifold, $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \to M$ is the trivial SU(2)-bundle over M, $\mathcal{A}$ is the affine space of all connections on $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \to M$, identified with globally defined gauge potentials $A \in \Omega^1(M, \mathfrak{su}(2))$, $\mathcal{G}$ is the group of gauge transformations of $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \to M$, identified with smooth maps $g : M \to \mathrm{SU}(2)$ from M into SU(2), and k is a positive integer called the level.

We will begin with the partition function

$$Z_k(M) = \int_{\mathcal{A}/\mathcal{G}} e^{\,i\,S_{CS}(A,k)}\,\mathcal{D}A,$$

where

$$S_{CS}(A,k) = \frac{k}{4\pi}\int_M \mathrm{tr}\,(\,A \wedge dA + \frac{2}{3}\,A \wedge A \wedge A\,).$$

Taking $\int_{\mathcal{A}/\mathcal{G}} e^{\,i\,S_{CS}(A,k)}\,\mathcal{D}A$ seriously as an integral one is struck by its resemblance to the finite-dimensional oscillatory integrals that one encounters in analysis as well as in classical and quantum physics. These integrals generally cannot be evaluated exactly and one must have recourse to what is called the *Stationary Phase Approximation* which we record now as motivation for what is coming (a proof is available in Appendix C of [Nab4]).

Theorem 1.16. *(Stationary Phase Approximation) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Morse function, that is, a smooth function with finitely many nondegenerate critical points $p_1, \ldots, p_N$. Let $g : \mathbb{R}^n \to \mathbb{R}$ be a smooth function with compact support and let $T > 0$ be a real number. Then*

$$\int_{\mathbb{R}^n} e^{\,i\,T f(x)} g(x)\, d^n x = \sum_{j=1}^{N} \left(\frac{2\pi}{T}\right)^{n/2} e^{\pi\, i\, sgn\,(H_f(p_j))/4}\, \frac{e^{\,i\,T f(p_j)}}{\sqrt{|\det\,(H_f(p_j))|}}\, g(p_j) \;+\; O\,(T^{-n/2-1}), \tag{43}$$

where $H_f(x) = \left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right)_{i,j=1,\ldots,n}$ *is the Hessian of* f *and* $sgn\,(H_f(p_j))$ *is the signature of the matrix* $H_f(p_j)$*.*

The sum on the right-hand side of (43) is called the *stationary phase approximation* to the integral. The rationale behind the theorem goes something like this. Near a point x_0 at which $df_{x_0} \neq 0$ the rate at which $e^{iTf(x)}$ is oscillating varies with both x and T and, as $T \to \infty$, these oscillations become more and more rapid near x_0. One might expect that these very rapid oscillations near x_0 with varying frequencies will (approximately) cancel (real and imaginary parts are negative as often as they are positive) so that the dominant contributions to the integral as $T \to \infty$ will come from neighborhoods of points where $df_{x_0} = 0$ and f is approximately constant. Near such a point the Morse Lemma (Theorem B.0.2 (3) of [Nab4]) asserts that f is quadratic in some coordinates and one might hope to compute the contribution to the integral at this point by evaluating a Gaussian integral (Gaussian integrals are reviewed in Appendix B of [Nab4]). This is how one arrives at the first term on the right-hand side of (43).

Needless to say, Theorem 1.16 has nothing whatever to say about the Feynman "integral"

$$Z_k(M) = \int_{\mathcal{A}/\mathcal{G}} e^{\,i\,k/4\pi \int_M \mathrm{tr}\,(\,A\wedge dA + \frac{2}{3}\,A\wedge A\wedge A\,)}\,\mathcal{D}A.$$

Nevertheless, in an effort to understand how topological information might be contained in $Z_k(M)$, Witten appealed to the common practice in QFT of formally adapting the ideas behind the stationary phase approximation to the infinite-dimensional context in which we now find ourselves. Recall that we have already pointed out that the level k in $Z_k(M)$ plays the role of a reciprocal coupling constant (analogous to $1/\hbar$) so that its $k \to \infty$ asymptotic behavior, called the *weak coupling limit*, corresponds roughly to what one might identify with a classical limit ($\hbar \to 0$). Also recall that we have seen that the stationary points of the Chern-Simons action $S_{CS}(A,k)$ are precisely the flat connections. One therefore fixes a flat connection A_f and computes the contribution to the integral from the gauge equivalence class $[A_f]$. In order for this to make sense, of course, the integral must be gauge invariant. We know that this is the case for the integrand, but one must now assume that the same is true of the non-existent "measure" $\mathcal{D}A$.

Remark 1.60. Witten assumes, for simplicity, that there are only finitely many such $[A_f]$. This is a topological assumption on M since the gauge equivalence classes of flat connections on $\mathrm{SU}(2) \hookrightarrow M \times \mathrm{SU}(2) \to M$ are in one-to-one correspondence with the conjugacy classes of homomorphisms from the fundamental group $\pi_1(M)$ of M into SU(2). Intuitively, this correspondence comes about by assigning to any loop in M the holonomy of the flat connection around it (see Section 5.1 of [Mich] for the details).

Not having a Morse Lemma to appeal to Witten expands $S_{CS}(A,k)$ near the flat connection A_f in a power series. Specifically, he writes nearby connections as $A = A_f + a$, expands $S_{CS}(A_f + a, k)$ and retains only the quadratic terms in a. This reduces the integral defining $Z_k(M)$ to a formal, infinite-dimensional analogue of a Gaussian integral. The evaluation of the Gaussian integral proceeds along lines familiar in QFT and requires that one select a representative from each gauge equivalence class in $\mathcal{A}/\mathcal{G}$ near $[A_f]$. One must, in other words, fix a local gauge for the $\mathcal{G}$-bundle $\mathcal{A} \to \mathcal{A}/\mathcal{G}$. For this Witten had to choose a metric on the underlying 3-manifold M. Since the end result of Witten's calculation will turn out to be a known 3-manifold invariant it is, in particular, independent of this choice. Having fixed a gauge Witten's calculations led him to the stationary point contribution, denoted $\mu(A_f)$. These calculations do not appear in [Witt2], but a few of them can be found in Section 5, Chapter XII, of [Nash]. Witten writes the result as

$$\mu(A_f) = e^{iS_{CS}(A_f,k)} \cdot \frac{\det \Delta}{\sqrt{\det L_-}} \tag{44}$$

so that the stationary phase approximation to $Z_k(M)$ is taken to be the sum of these over the gauge equivalence classes of flat connections on $SU(2) \hookrightarrow M \times SU(2) \to M$. The phase factor in (44) is just the value of the integrand in $Z_k(M)$ at the flat connection A_f, whereas Δ and L_- are both differential operators. Specifically, Δ is a Laplacian and L_- is a twisted Dirac operator. Their determinants are defined by zeta function regularization which we will not discuss here (see [Robles] for an introduction to the subject).

Remark 1.61. Witten points out (pages 361-362 of [Witt2]) that this argument assumes that these determinants are nonzero and that this requires additional topological assumptions on M which are not satisfied even for $M = \mathbb{S}^3$. Without such assumptions the finite sum of the $\mu(A_f)$ must be replaced by a path integral over the space of gauge equivalence classes of flat connections. One should keep in mind that our current discussion concerns only the large k asymptotic behavior of $Z_k(M)$ and is intended only to *suggest* that $Z_k(M)$ contains topological information.

Witten argues that each of the contributions $\mu(A_f)$ is a differential topological invariant. The ratio of determinants in (44) had, in fact, appeared in earlier work of Albert Schwarz [Schw1]. Schwarz considered the same problem we are discussing now, but in the Abelian case, that is, when SU(2) is replaced by U(1). In this context Schwarz proved that the ratio $\det \Delta / \sqrt{\det L_-}$ is, up to a phase factor, an invariant known as the *Ray-Singer analytic torsion*. Witten simply notes that "this aspect of [Schw1] generalizes" (see page 358 of [Witt2]). The phase factor $e^{iS_{CS}(A_f,k)}$, however, is considerably more subtle and to understand the role it plays here one must look carefully at the phase of the determinant $\det L_-$. We will not attempt to sketch Witten's arguments (see pages 358-361 of [Witt2]), but will only mention that they require yet one more ingredient. It is known that any compact, orientable 3-manifold M is parallelizable, that is, has trivial tangent bundle TM (see Problem 12-B of [MilSta]). A particular choice of trivialization for TM is called a *framing* of M. The framings of M fall into homotopy classes determined by an integer related to the number of relative "twists". At one point in his argument (page 360 of [Witt2]) Witten introduces what is called the *gravitational Chern-Simons term* and this is unambiguously defined only with some choice of framing. Consequently, the argument regarding the phase factor strictly applies only to compact, oriented, smooth, *framed* 3-manifolds, although Witten computes an explicit formula for the effect on $Z_k(M)$ of a change of framing and concludes that this is "more or less as good as a topological invariant of oriented three manifolds without a choice of framing".

The conclusion one draws from Witten's analysis to this point is that at least the weak coupling limit of quantum SU(2)-Chern-Simons theory on M contains known topological information about M. This inspires the hope that *new* topological information might be contained in the full quantum theory. To further reinforce this point Witten considers, as another "test case", the weak coupling limit of the Abelian theory (SU(2) replaced by U(1)) when the 3-manifold M is $\mathbb{S}^3$ and there is an oriented link $L = K_1 \sqcup \cdots \sqcup K_m$ in

$\mathbb{S}^3$. Choosing a representation ρ_a for each K_a, $a = 1, \ldots, m$, one obtains a Wilson line observable $W_{K_a,\rho_a}(A)$, $A \in \mathcal{A}$, given by (18), and a corresponding expectation value

$$Z_k(\mathbb{S}^3, L, \rho_1, \ldots, \rho_m) = \int_{\mathcal{A}/\mathcal{G}} e^{i\,S_{CS}(A,k)} \prod_{a=1}^{m} W_{K_a,\rho_a}(A)\, \mathcal{D}A.$$

Since U(1) is 1-dimensional and Abelian, the cubic term $A \wedge A \wedge A$ in $S_{CS}(A,k)$ vanishes and every representation is 1-dimensional and given by $\rho_a(\theta) = e^{i\,n_a\theta}$ for some integer n_a. The integrand is therefore the exponential of some quadratic in A and can, by completing the square, be treated formally as a Gaussian integral. Witten, referring to a result of Polyakov where this integral was also considered, writes the result of this calculation for the stationary phase approximation as

$$\exp\Big(\frac{i}{2k}\sum_{a,b=1}^{m} n_a\, n_b \int_{K_a}\int_{K_b} \varepsilon_{ijk}\, \frac{x^k - y^k}{\|\, x - y\,\|^3}\, dy^i\, dx^j\Big), \tag{45}$$

where ε_{ijk} is the Levi-Civita symbol (1 if ijk is an even permutation of 123, -1 if it is an odd permutation of 123, and 0 otherwise) and $x = (x^1, x^2, x^3)$ and $y = (y^1, y^2, y^3)$ are local Euclidean coordinates on the knots K_a and K_b, respectively. Now notice that, if $a \neq b$, then

$$lk\,(K_a, K_b) = \frac{1}{4\pi}\int_{K_a}\int_{K_b} \varepsilon_{ijk}\, \frac{x^k - y^k}{\|\, x - y\,\|^3}\, dy^i\, dx^j$$

is the Gauss integral formula for the linking number of K_a and K_b. However, the sum in (45) includes terms for which $a = b$ and these are clearly undefined due to the presence of the term $\|\, x - y\,\|^3$. The stationary phase approximation must be *regularized*, that is, redefined, and Witten does this by interpreting the integrals with $a = b$ as *self-linking numbers*. In order to do this one must assume that each of the knots K_a comes equipped with a *framing*, that is, a trivialization of its normal bundle in $\mathbb{S}^3$ or, equivalently, a non-vanishing, normal vector field along K_a (see Figure 27). The framing determines another knot K'_a disjoint from K_a by displacing K_a along the vector field and we define the *self-linking number* $lk\,(K_a, K_a)$ by

$$lk\,(K_a, K_a) = lk\,(K_a, K'_a).$$

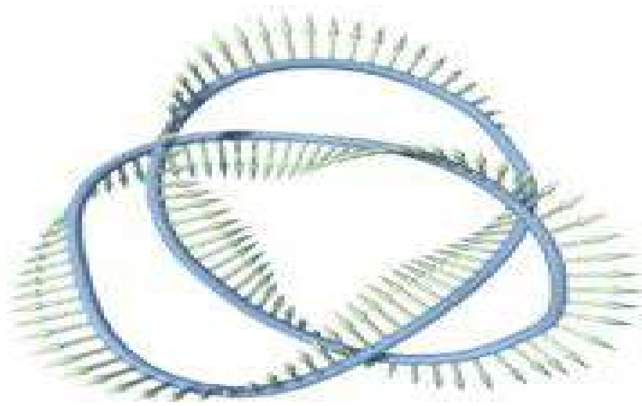


FIGURE 27. Framing of a Knot

Remark 1.62. This definition is clearly not unique since a different normal vector field may determine a knot with more or fewer twists depending on how many times it wraps around K_a and this would give a different $lk\,(K_a, K'_a)$. One must therefore regard the choice of framing as part of the input. We will see in a moment that, as in the case of $Z_k(M)$, this is not as serious a problem as it might appear. Moreover, for links in $\mathbb{S}^3$ there is a *standard framing* for which the self-linking number of each component is zero and we will always assume this choice of framing for $\mathbb{S}^3$.

With this we can regularize (45) by replacing the ill-defined integrals by self-intersection numbers and obtain what we will take to be the *definition* of the stationary phase approximation of $Z_k(\mathbb{S}^3, L, \rho_1, \ldots, \rho_m)$.

$$Z_k(\mathbb{S}^3, L, \rho_1, \ldots, \rho_m) = \exp\left(\frac{2\pi i}{k} \sum_{a,b=1}^{m} n_a\, n_b\, lk\,(K_a, K_b) \right) \tag{46}$$

Consequently, one sees that the expectation value of the Wilson line observable $\prod_{a=1}^{m} W_{K_a,\rho_a}(A)$ corresponding to the link L contains the classical topological information expressed by the linking numbers of the knots in L. Finally, Witten notes that selecting different framings for the knots in L has a simple effect on $Z_k(\mathbb{S}^3, L, \rho_1, \ldots, \rho_m)$ so that a topological invariant of framed, oriented links is "more or less as good as a topological invariant" of links without a choice of framing. Specifically, if the framing of the knot K_a is replaced by another framing that differs from the first by t twists, the the effect on the expectation value is, in the Abelian case we are considering,

$$Z_k(\mathbb{S}^3, L, \rho_1, \ldots, \rho_m) \mapsto e^{2\pi i t\, n_a^2/k} \cdot Z_k(\mathbb{S}^3, L, \rho_1, \ldots, \rho_m).$$

Witten remarks that there is a non-Abelian analogue of this result, considered in Section 5.1 of [Witt2], in which n_a^2/k is replaced by "the conformal weight of a certain primary field in 1+1 dimensional current algebra".

Our discussion up to this point has been essentially motivational and based on Sections 1 and 2 of [Witt2]. We have seen that weak coupling limits of quantum Chern-Simons theory give one reason to hope that there is important topological information contained in the full quantum theory. Section 3 of Witten's paper deals with the canonical quantization of Chern-Simons on 3-manifolds of the form $\Sigma\times\mathbb{R}$, first when Σ is a Riemann surface and then when Σ is a so-called *punctured (or marked) Riemann surface*. The first case is relevant to the problem of producing the 3-manifold invariants that we described in Sections 1.5.2 and 1.5.3. When the 3-manifold contains an oriented link and one hopes to produce link invariants the Riemann surfaces with which one cuts the 3-manifold are "punctured" by knots entering and leaving it (see the discussion at the end of Section 1.5.2). We saw in Section 1.5.3 that when Σ is an unmarked Riemann surface the canonical quantization of $\Sigma \times \mathbb{R}$ yields, for each level k, a Hilbert space $Z_k(\Sigma)$. According to Witten these Hilbert spaces are precisely those assigned to Σ by Segal's modular functors. Indeed, in Sections 3.3 and 3.4 of [Witt2], Witten briefly describes the canonical quantization in the case of a Riemann surface Σ with marked points and concludes that, once again, the resulting Hilbert space is precisely the space of conformal blocks of a WZW-model on Σ.

This is 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.

-Witten [Witt2]*, page 370*

We should point out that Witten does not offer a detailed argument to support this claim in [Witt2] (see [LR] for a discussion of the case $M = \mathbb{S}^3$). Our discussion in Section 1.5.3 of the case in which Σ is an unmarked Riemann surface was based on the rigorous treatment in [ADPW] which does *not* cover the case of a marked surface. In the case of marked surfaces an analogous treatment has not been carried out and we, along with Witten, will simply appeal to results from 2-dimensional conformal field theory and point out once again that a rigorous discussion of the Jones polynomial in this CFT context is available in [Kohno]..

Remark 1.63. Witten briefly describes the problem in CFT that gives rise to these Hilbert spaces on pages 371-372 of [Witt2], but defers "a fuller treatment for another occasion". Roughly, the procedure involves to an appeal to what is called the *Borel-Weil Theorem* which we will briefly describe (for the details see Section 7.4 of [Sepan]). One begins with a compact, connected Lie group G. Select a maximal Abelian subgroup T of G, that is, a maximal torus in G, and consider the homogeneous space G/T. This is an example of what is called a *flag manifold* (see [Alek]). It can be shown that, if $G^{\mathbb{C}}$ is the complexification of G, then $G^{\mathbb{C}}$ has a subgroup B, called a *Borel subgroup*, such that $G^{\mathbb{C}}/B$ is diffeomorphic to G/T (Theorem 7.50 of [Sepan]). Since $G^{\mathbb{C}}/B$ is a complex manifold (Section 7.4.2 of [Sepan]), G/T also admits the structure of a complex manifold. Topological considerations and the *Kodaira Vanishing Theorem* (see [Math]) show that any complex line bundle over G/T admits a unique holomorphic structure and so one can consider its space of holomorphic sections. The essential content of the Borel-Weil Theorem (Theorem 7.58 of [Sepan]) can be described as follows. An irreducible representation ρ of G on a complex vector space V canonically determines a complex line bundle L_λ over G/T, namely, the line bundle associated to the principal bundle $T \hookrightarrow G \to G/T$ by the highest weight $\lambda : T \to \mathrm{U}(1)$ of the restricted action of ρ on T. The Borel-Weil Theorem asserts that the space $\Gamma_{\mathrm{Hol}}(L_\lambda)$ of holomorphic sections of L_λ is a realization of the representation space V. Consequently, every irreducible representation of G can be thought of as acting on a space of holomorphic sections of some complex line bundle. Witten's interest in the Borel-Weil construction is that the bundle L_λ gives rise to a symplectic form ω_λ on G/T that is invariant under the left G-action on G/T (see the Proposition in Section IV.3 of [Alek]) . Thus, $(G/T, \omega_\lambda)$ can be thought of a classical mechanical system whose quantization gives rise to $\Gamma_{\mathrm{Hol}}(L_\lambda)$. In this way the representation ρ can "be seen as a quantum object" (page 372 of [Witt2]). Somewhat later we will encounter an extension of the Borel-Weil Theorem to the loop group LG which plays an important role in conformal field theory (see Chapter 11 of [PreSeg] for a thorough discussion of the loop group version of the Borel-Weil Theorem).

At last we are in a position to sketch a path from Chern-Simons to knot invariants. Before getting started, however, we should reiterate that our goal is *not* a rigorous reformulation of results to which Witten was led by other means; for this one should consult Section 2.2 of [Kohno]. Rather, we would like to provide a brief, and necessarily rather crude, guide to some of the underlying ideas in a spirit that does not clash too seriously with [Witt2].

Witten's results are extremely general and produce invariants for knots and links in any compact, oriented 3-manifold M for any choice of compact, simple Lie group G. Our discussion will be much more limited in scope. We will only attempt to uncover the underlying ideas by considering the very special case that leads to the classical Jones polynomial (Section 1.2). For this we will assume henceforth that

$$M = \mathbb{S}^3$$

and

$$G = \mathrm{SU}(2).$$

Remark 1.64. We should point out, however, that more general 3-manifolds M are dealt with by employing classical surgery techniques. Specifically, one appeals to a theorem, proved independently by Lickorish [Lick] and Wallace [Wall], on a procedure known as *Dehn surgery* on $\mathbb{S}^3$. Roughly, this procedure can be described as follows. Let L be a link in $\mathbb{S}^3$ (this one should be regarded as an auxiliary tool and not as the link for which we are hoping to define an invariant). Delete from $\mathbb{S}^3$ an open tubular neighborhood of L. The result is a compact 3-manifold with boundary components that are all 2-dimensional tori $\mathbb{S}^1 \times \mathbb{S}^1$. Now, for each boundary component, glue in a solid torus $\mathbb{S}^1 \times \mathbb{D}^2$ by some diffeomorphism of $\mathbb{S}^1 \times \mathbb{S}^1$ onto itself. The result is a compact, connected, orientable 3-manifold M. The theorem of Lickorish and Wallace is that *every* such 3-manifold can be obtained in this way.

Theorem 1.17. *(Lickorish-Wallace) Every compact, connected, orientable 3-manifold M can be obtained by Dehn surgery on a link in* $\mathbb{S}^3$.

Witten then studies the behavior of his invariants under Dehn surgery on $\mathbb{S}^3$ in order to define them on M (see Section 4 of [Witt2]). It is worth noting that, prior to the appearance of [Witt2], many attempts had been made to construct a reasonable theory of knots and links in 3-manifolds other than $\mathbb{S}^3$ or $\mathbb{R}^3$, all of which essentially came to nought.

Now we suppose that $L = K_1 \sqcup \cdots \sqcup K_m$ is an oriented, framed link in $\mathbb{S}^3$ (see Remark 1.62 regarding the framing). To implement Witten's procedure we must choose a representation ρ_i of SU(2) for each knot K_i, $i = 1, \ldots m$. We will take each ρ_i to be the defining (spin $\frac{1}{2}$) representation ρ of SU(2) on $\mathbb{C}^2$, that is, just matrix multiplication of the column vectors in $\mathbb{C}^2$ by the 2×2 matrices in SU(2).

$$\rho_1 = \cdots = \rho_m = \rho$$

Fix a level k. The object of interest to Witten is then the heuristic Feynman path integral

$$Z_k(\mathbb{S}^3, L, \vec{\rho}\,) = \int_{\mathcal{A}/\mathcal{G}} e^{\,i\,S_{CS}(A,k)} \prod_{i=1}^{m} W_{K_i,\rho}(A)\, \mathcal{D}A, \tag{47}$$

where $\mathcal{A}/\mathcal{G}$ is the moduli space of all gauge equivalence classes of connections on the trivial SU(2)-bundle over $\mathbb{S}^3$ and we have written $\vec{\rho}$ for $\rho, \overset{m}{\cdots}, \rho$. This is the unnormalized expectation value of the Wilson line observables corresponding to the knots $K_1, \ldots, K_m$. The normalized expectation values are given by

$$\frac{Z_k(\mathbb{S}^3, L, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)}, \tag{48}$$

where $Z_k(\mathbb{S}^3)$ is the partition function for $\mathbb{S}^3$ in the case in which there are no links present. Witten proposes to "evaluate" the integrals. The evaluation, however, does not amount to the sort of formal manipulations or asymptotic expansions that we have seen earlier, but rather to the application of general properties that these integrals are assumed to possess and that have been axiomatized by Segal and Atiyah. In particular, a central role is played by results from 2-dimensional, rational conformal field theory. For example, Witten asserts (page 375 of [Witt2]) that the theory of affine Lie algebras, which is at the heart of conformal field theory, determines $Z_k(\mathbb{S}^3)$ to be

$$Z_k(\mathbb{S}^3) = \sqrt{\frac{2}{k+2}}\, \sin\left(\frac{\pi}{k+2}\right).$$

The objective is to show that the path integrals give rise to functions $V_L(t)$ that satisfy the skein relations (7) for the Jones polynomial as well as the appropriate normalization condition (8) and therefore must, in fact, be the Jones polynomials. The first step is to show that the normalized expectation values (48) "behave multiplicatively" in the following sense. The link L might very well be the union of a number of disjoint links no two of which are linked to each other (two unlinked knots, for example); for want of a better term we will refers to these as *sublinks*. We decompose $\mathbb{S}^3$ into a connected sum $\mathbb{S}^3 = M_1 \# M_2$ of two 3-manifolds in such a way that none of the knots in L pass through the 2-sphere $\mathbb{S}^2$ along which M_1 and M_2 are glued. One way to do this is to use the compactness of the sublinks in L to choose an equatorial $\mathbb{S}^2$ on $\mathbb{S}^3$ that does not intersect L. Then the closed hemispheres $\check{M}_1$ and $\check{M}_2$ determined by $\mathbb{S}^2$ can be regarded 3-spheres M_1 and M_2 with open 3-discs removed which, when glued together along the boundary copies of $\mathbb{S}^2$ by an orientation reversing diffeomorphism, give $\mathbb{S}^3$.

Remark 1.65. The 3-manifold $\mathbb{S}^3$ can be shown to be *irreducible* in the sense that any embedded $\mathbb{S}^2$ in $\mathbb{S}^3$ bounds a 3-dimensional ball. It follows from this that $\mathbb{S}^3$ is a *prime* 3-manifold meaning that, for any connected sum decomposition $\mathbb{S}^3 = M_1 \# M_2$ of it, at least one of M_1 or M_2 must be a copy of $\mathbb{S}^3$. But spheres act as identity elements for the connected sum so it follows that the other summand must be a copy of $\mathbb{S}^3$ as well. We have simply chosen a particularly convenient way of arriving at such a decomposition.

With this decomposition L splits into a disjoint union $L = L_1 \sqcup L_2$, where each L_i is a link and we can regard $L_1 \subseteq M_1 \cong \mathbb{S}^3$ and $L_2 \subseteq M_2 \cong \mathbb{S}^3$. The "multiplicative" property of normalized expectation values that we referred to above is

$$\frac{Z_k(\mathbb{S}^3, L, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)} = \frac{Z_k(M_1, L_1, \vec{\rho}_1)}{Z_k(\mathbb{S}^3)}\, \frac{Z_k(M_2, L_2, \vec{\rho}_2)}{Z_k(\mathbb{S}^3)}\,, \tag{49}$$

where $\vec{\rho}_i$ has one ρ for each knot in M_i. Equivalently,

$$Z_k(\mathbb{S}^3)\, Z_k(\mathbb{S}^3, L, \vec{\rho}\,) = Z_k(M_1, L_1, \vec{\rho}_1)\, Z_k(M_2, L_2, \vec{\rho}_2). \tag{50}$$

Witten's argument proceeds in the following way. Consider first the 3-manifold M_1 with boundary before the manifolds M_1 and M_2 are glued together to obtain $\mathbb{S}^3$. M_1 contains the link L_1 which does not intersect the boundary $\mathbb{S}^2$ so $\mathbb{S}^2$ has no marked points. Canonical quantization of Chern-Simons on M_1 yields a Hilbert space associated with $\mathbb{S}^2$ that we will denote simply $\mathcal{H}_{\mathbb{S}^2}$. This coincides with the Hilbert space associated with the unmarked $\mathbb{S}^2$ by Segal's modular functor and we have already mentioned (Remark 1.53) that the complex dimension of $\mathcal{H}_{\mathbb{S}^2}$ is 1. Similar remarks apply to M_2, but the orientation of the boundary is opposite that of M_1 so the associated Hilbert space is identified with the *dual* of $\mathcal{H}_{\mathbb{S}^2}$; this is Segal's analogue of Atiyah's axiom **(A2)**. Next we employ the CFT analogue of Atiyah's functoriality axiom **(A4)** which, in this context, is most conveniently expressed in the form we described in Remark 1.27. Thus, the Chern-Simons path integral over M_1 determines a vector ξ_1 in $\mathcal{H}_{\mathbb{S}^2}$, the integral over M_2 determines a vector ξ_2 in the dual of $\mathcal{H}_{\mathbb{S}^2}$ and

$$Z_k(\mathbb{S}^3, L, \vec{\rho}\,) = \langle\, \xi_1, \xi_2 \,\rangle.$$

Applying this same argument to $\mathbb{S}^3$ in the case in which $\mathbb{S}^3$ contains no link gives

$$Z_k(\mathbb{S}^3) = \langle\, \eta_1, \eta_2 \,\rangle,$$

where η_1 is in $\mathcal{H}_{\mathbb{S}^2}$ and η_2 is in its dual. Now, since $\mathcal{H}_{\mathbb{S}^2}$ is 1-dimensional, η_1 is a multiple of ξ_1 and η_2 is a multiple of ξ_2 so we can write

$$\langle\, \eta_1, \eta_2 \,\rangle \langle\, \xi_1, \xi_2 \,\rangle = \langle\, \xi_1, \eta_1 \,\rangle \langle\, \eta_2, \xi_2 \,\rangle = Z_k(M_1, L_1, \vec{\rho}_1)\, Z_k(M_2, L_2, \vec{\rho}_2)$$

which gives (50) and therefore (49).

Proceeding by induction we can decompose $\mathbb{S}^3$ into a connected sum

$$\mathbb{S}^3 = M_1 \# \cdots \# M_p$$

in such a way that each M_i, $i = 1, \ldots, p$, is a copy of $\mathbb{S}^3$ and contains precisely one of the sublinks of L. We then find that

$$\frac{Z_k(\mathbb{S}^3, L, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)} = \frac{Z_k(M_1, L_1, \vec{\rho}_1)}{Z_k(\mathbb{S}^3)} \cdots \frac{Z_k(M_p, L_p, \vec{\rho}_p)}{Z_k(\mathbb{S}^3)}\,. \tag{51}$$

Since $M_i \cong \mathbb{S}^3$ for each $i = 1, \ldots, p$ we will write this as

$$\frac{Z_k(\mathbb{S}^3, L, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)} = \frac{Z_k(\mathbb{S}^3, L_1, \vec{\rho}_1)}{Z_k(\mathbb{S}^3)} \cdots \frac{Z_k(\mathbb{S}^3, L_p, \vec{\rho}_p)}{Z_k(\mathbb{S}^3)}\,. \tag{52}$$

Apply this to the special case in which $L = C_1 \sqcup \cdots \sqcup C_p$ is the p-component unlink, that is, consists of p *unknotted, unlinked* circles, each with associated representation ρ and write the result as

$$\frac{Z_k(\mathbb{S}^3, C_1 \sqcup \cdots \sqcup C_p, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)} = \frac{Z_k(\mathbb{S}^3, C_1, \rho)}{Z_k(\mathbb{S}^3)} \cdots \frac{Z_k(\mathbb{S}^3, C_p, \rho)}{Z_k(\mathbb{S}^3)}\,. \tag{53}$$

Next we turn to the skein relations (7) for the Jones polynomial. Skein relations, of course, deal with "crossings" and these do not exist in $\mathbb{S}^3$, but rather in a link diagram for L so we begin by fixing some regular projection of L and the corresponding link diagram D. Every crossing in D is the projection of precisely two points on L. We fix such a pair of points and choose a copy S of the 2-sphere $\mathbb{S}^2$ that encloses only two segments on the knots in L through these points. Figure 28 (a), which is taken from [Witt2], illustrates this.

Remark 1.66. The Figure shows only one knot, but one should imagine additional links and sublinks some of which may lie inside S and some outside.

S is therefore a 2-sphere with four punctures, two leaving S and two entering. Since $\mathbb{S}^3$ is irreducible (see Remark 1.65), S bounds a 3-dimensional disc in $\mathbb{S}^3$. Cutting along S therefore decomposes $\mathbb{S}^3$ into two 3-manifolds with boundary S, denoted M_L and M_R in Figure 28 (b), at least one of which is a disc (the hatched region indicates the complicated part of the link outside this disc). Since S inherits opposite orientations from M_L and M_R, conformal field theory assigns to ∂M_L a Hilbert space $\mathcal{H}_S$ and to ∂M_R the dual of $\mathcal{H}_S$.

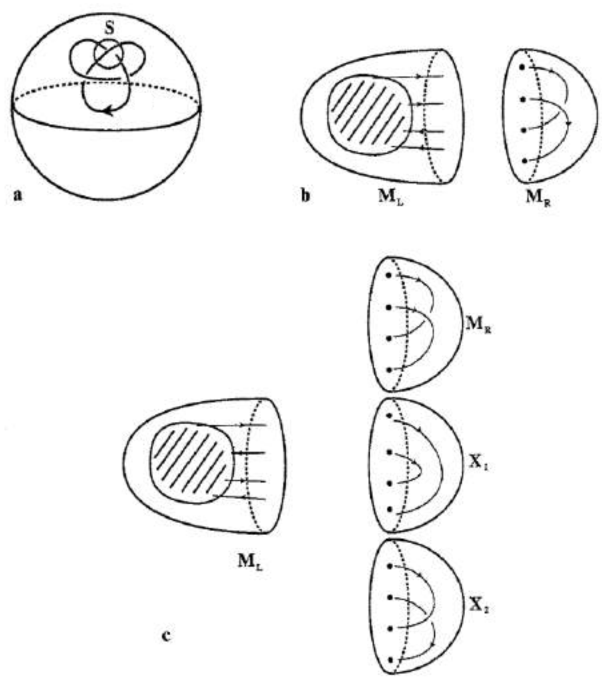


FIGURE 28. 4-Punctured $\mathbb{S}^2$

Now the argument proceeds in much the same way as that for multiplicativity given above. The Chern-Simons path integrals for M_L and M_R determine vectors ξ_1 in $\mathcal{H}_S$ and ξ_2 in the dual of $\mathcal{H}_S$ with

$$\langle\, \xi_1, \xi_2 \,\rangle = Z_k(\mathbb{S}^4, L, \vec{\rho}\,).$$

We mentioned earlier that techniques from conformal field theory compute the dimensions of the Hilbert spaces associated to marked Riemann surfaces; Witten lists a few examples at the end of Section 3 of [Witt2]. We will not attempt to describe these results, but will simply quote the one we need. Specifically, when $G = \mathrm{SU}(2)$, $\Sigma = \mathbb{S}^2$, ρ is the defining representation of SU(2) on $\mathbb{C}^2$, and S is a 2-sphere with four marked points, one has

$$\dim \mathcal{H}_S = 2$$

for $k \geq 2$. For $k = 1$ the dimension is 1 and we will assume henceforth that the level k is taken to be greater than 1. To quote Witten (page3 377-378 of [Witt2]),

"A two dimensional vector space has the marvelous property that any three vectors obey a relation of linear dependence."

Witten's procedure for arriving at two vectors ξ_1' and ξ_1'' in $\mathcal{H}_S$ to add to ξ_1 in order to apply this strategy is illustrated in Figure 28 (c). Rather than gluing M_L and M_R together along S to obtain the original picture Figure 28 (a), one replaces M_R with the 3-manifolds X_1 and X_2. These are topologically the same as M_R, but the configuration of arcs inside is altered to produce, with the same projection, link diagrams identical to D except that the crossings are of the three types indicated in Figure 18. The path integrals for X_1 and X_2 then give vectors ξ_1' and ξ_1'' in $\mathcal{H}_S$ so, for some complex constants α, β, and γ, we have

$$\alpha\,\xi_1 + \beta\,\xi_1' + \gamma\,\xi_1'' = 0$$

and therefore

$$\alpha\,\langle\,\xi_1, \xi_2\,\rangle + \beta\,\langle\,\xi_1', \xi_2\,\rangle + \gamma\,\langle\,\xi_1'', \xi_2\,\rangle = 0.$$

Gluing each of M_R, X_1 and X_2 to M_L along S gives three copies of $\mathbb{S}^3$ containing links that we will now denote L_+, L_0 and L_-, where L_+ is L (the labeling of these links is chosen simply to be consistent with the projection suggested by Figure 28 (c)). Consequently, we have

$$\alpha\, Z_k(\mathbb{S}^3, L_+, \vec{\rho}\,) + \beta\, Z_k(\mathbb{S}^3, L_0, \vec{\rho}\,) + \gamma\, Z_k(\mathbb{S}^3, L_-, \vec{\rho}\,) = 0 \tag{54}$$

which is beginning to take on the appearance of a skein relation.

To proceed further one must explicitly determine values for the coefficients α, β and γ. This can be done, but it requires quite nontrivial results in conformal field theory due to Moore and Seiberg [MoSei] and quoted by Witten on page 381 of [Witt2]. We will say only this much. Witten shows that the three configurations shown in M_R, X_1 and X_2 of Figure 28 (c) can be interchanged by certain diffeomorphisms of S that change the positions of the punctures (see Figure 11 (a) and (b) of [Witt2]). Functoriality implies that these diffeomorphisms give rise to isomorphisms of $\mathcal{H}_S$, but, since $\mathcal{H}_S$ is 2-dimensional, these isomorphisms are simply matrices acting on $\mathcal{H}_S$. It is these matrices that Moore and Seiberg study in [MoSei]. Their analysis is subtle and involves aspects of conformal field theory that we do not have at our disposal so we will, as Witten did, simply record the result.

$$\begin{aligned}
\alpha &= -\exp\left(\frac{2\pi i}{k+2}\right)\\
\beta &= -\exp\left(\frac{-\pi i}{k+2}\right) + \exp\left(\frac{\pi i}{k+2}\right)\\
\gamma &= \quad \exp\left(\frac{-2\pi i}{k+2}\right)
\end{aligned}$$

Now, if we define

$$q = \exp\Big(\frac{2\pi i}{k+2}\Big),$$

then these become $\alpha = -q, \beta = q^{1/2} - q^{-1/2}$, and $\gamma = q^{-1}$. Substituting these into (54) gives

$$-q\, Z_k(\mathbb{S}^3, L_+, \vec{\rho}\,) + (q^{1/2} - q^{-1/2})\, Z_k(\mathbb{S}^3, L_0, \vec{\rho}\,) + q^{-1}\, Z_k(\mathbb{S}^3, L_-, \vec{\rho}\,) = 0. \tag{55}$$

This is equation (4.22) of [Witt2] with $N = 2$ (Witten works more generally with $G = \mathrm{SU}(N)$).

Recall that the Jones skein relation (7) is

$$t^{-1} V_{L_+}(t) + (t^{-1/2} - t^{1/2}) V_{L_0}(t) - t V_{L_-}(t) = 0.$$

If we make the substitution $t^{1/2} = -q^{-1/2} = -\exp\big(\frac{-\pi i}{k+2}\big)$, then $t = q^{-1}$, $t^{-1} = q$ and $t^{-1/2} - t^{1/2} = -q^{1/2} + q^{-1/2}$ so this becomes

$$q V_{L_+}(q^{-1}) + (-q^{1/2} + q^{-1/2}) V_{L_0}(q_{-1}) - q^{-1} V_{L_-}(q^{-1}) = 0,$$

or

$$-q V_{L_+}(q^{-1}) + (q^{1/2} - q^{-1/2}) V_{L_0}(q^{-1}) + q^{-1} V_{L_-}(q^{-1}) = 0$$

which has precisely the same form as (55). The unnormalized expectation values $Z_k(\mathbb{S}^3, L, \vec{\rho})$ satisfy the same skein relation as the value of the Jones polynomial $V_L(t)$ at $t = q^{-1} = \exp\big(\frac{-\pi i}{k+2}\big)$. Of course, the same is true of the normalized expectation values.

$$-q\left(\frac{Z_k(\mathbb{S}^3, L_+, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)}\right) + (q^{1/2} - q^{-1/2})\left(\frac{Z_k(\mathbb{S}^3, L_0, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)}\right) + q^{-1}\left(\frac{Z_k(\mathbb{S}^3, L_-, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)}\right) = 0. \tag{56}$$

Now we would like to apply (56) to the case in which $L = C$ is a single unknotted circle. In this case, $L_+ = L_- = C$ and L_0 is the 2-component unlink, that is, the disjoint union of two unlinked, unknotted circles (see Figure 19). Using (53) for the L_0 term we write (56) as

$$-q\left(\frac{Z_k(\mathbb{S}^3, C, \rho\,)}{Z_k(\mathbb{S}^3)}\right) + (q^{1/2} - q^{-1/2})\left(\frac{Z_k(\mathbb{S}^3, C, \rho\,)}{Z_k(\mathbb{S}^3)}\right)^2 + q^{-1}\left(\frac{Z_k(\mathbb{S}^3, C, \rho\,)}{Z_k(\mathbb{S}^3)}\right) = 0 \tag{57}$$

Solving for the normalized expectation value gives

$$\frac{Z_k(\mathbb{S}^3, C, \rho\,)}{Z_k(\mathbb{S}^3)} = \frac{q - q^{-1}}{q^{1/2} - q^{-1/2}}.$$

Now we would like to regard q, that is k, as an independent variable rather than a special value of t and write this as a function of t as follows.

$$\frac{Z_k(\mathbb{S}^3, C, \rho\,)}{Z_k(\mathbb{S}^3)} = \frac{t^{-1} - t}{t^{1/2} - t^{-1/2}} = -\,(t^{1/2} + t^{-1/2})$$

More generally, for any link $C_1 \sqcup \cdots \sqcup C_p$ that is a disjoint union of p unknotted, unlinked circles, (53) gives

$$\frac{Z_k(\mathbb{S}^3, C_1 \sqcup \cdots \sqcup C_p, \vec{\rho}\,)}{Z_k(\mathbb{S}^3)} = [-\,(t^{1/2} + t^{-1/2})]^p.$$

Motivated by these we define, for any link L in $\mathbb{S}^3$,

$$V'_L(t) = [-(t^{1/2} + t^{-1/2})]^{-1}\left(\frac{Z_k(\mathbb{S}^3, L, \vec{\rho})}{Z_k(\mathbb{S}^3)}\right). \tag{58}$$

Since the functions $V'_L(t)$ satisfy the skein relation (7) and agree with the Jones polynomial $V_L(t)$ when L consists of a single unknotted circle (and a disjoint union of p unknotted, unlinked circles) and since the Jones polynomial is characterized by this skein relation and normalization, we must have $V'_L(t) = V_L(t)$. We have, therefore, at long last uncovered Witten's quantum field theory version of the Jones polynomial.

We conclude with a few reminders of what we have *not* done here. We have considered only a very special case of what Witten has actually done in [Witt2], that is, SU(2) Chern-Simons theory on the 3-sphere $\mathbb{S}^3$ and its relation to the Jones polynomial invariant. There are, for example, two variable generalizations of the Jones polynomial that arise in Witten's analysis when SU(2) is replaced by SU(N) and the representation is the defining representation of SU(N) on $\mathbb{C}^N$. The two variables are N and the level k. Similarly, the so-called *Kauffman polynomial invariant* arises when the group is SO(N) and ρ is its N-dimensional representation. Furthermore, we have already pointed out (Remark 1.64) that Witten's ideas also extend, by surgery techniques, to knots and links in more general compact, oriented 3-manifolds. Finally, we should emphasize once again that the expectation values in (58) are given in [Witt2] by heuristic and mathematically ill-defined path integrals. Nevertheless, formal manipulations of these objects led Witten to various concrete results that can be, and have been, verified by independent and quite rigorous means. This is the fundamental mystery that still resides far below the surface of what is currently understood: What is it in the structure of quantum field theory that accounts for its "unreasonable effectiveness in mathematics"?

References

[Abik] Abikoff, W., The Uniformization Theorem, Amer. Math. Monthly, **88**(1981), 574-592.

[AM] Abraham, R. and J.E. Marsden, *Foundations of Mechanics*, Second Edition, Addison-Wesley, Redwood City, CA, 1987.

[AB] Aharonov, Y. and D. Bohm, Significance of Electromagnetic Potentials in the Quantum Theory, Phys. Rev. **115**(1959), 485-491.

[AHS] Albevario, S., A. Hahn and A.N. Sengupta, Rigorous Feynman Path Integrals, with Applications to Quantum Theory, Gauge Fields and Topological Invariants, In *Stochastic Analysis and Mathematical Physics (samp/anestoc 2002)*, World Scientific, Singapore, 2004.

[AHKM] Albevario, S., R.H. Hoegh-Krohn, and S. Mazzucchi, *Mathematical Theory of Feynman Path Integrals: An Introduction*, Second Edition, Springer, New York, NY, 2008.

[Alek] Alekseevsky, D.V., Flag Manifolds. Available at http://elib.mi.sanu.ac.rs/files/journals/zr/14/n014p003.pdf.

[Alex1] Alexander, J.W., A Lemma on a System of Knotted Curves, Proc Nat. Acad. Sci. USA **9**(1923), 93-95.

[Alex2] Alexander, J.W., Topological Invariants of Knots and Links, Trans. Amer. Math. Soc. **30**(1928), 275-306.

[AlBr] Alexander, J.W. and G.B. Briggs, On Types of Knotted Curves, Ann. Math., Second Series **28**(1926), 562-586.

[Atiy1] Atiyah, M.F., Circular Symmetry and Stationary Phase Approximation, Proceedings of Conference in Honor of L. Schwartz, Astérisque **131**(1985), 43-59.

[Atiy2] Atiyah, M.F., New Invariants of 3- and 4-Dimensional Manifolds, Proc. Symp. Pure Math. **48**(1988), 285-299.

[Atiy3] Atiyah, M.F., Topological Quantum Field Theory, Publications Mathématiques de l'I.H.É.S. **68**(1988), 175-186.

[Atiy4] Atiyah, M.F., *The Geometry and Physics of Knots*, Cambridge University Press, Cambridge, UK, 1990.

[Atiy5] Atiyah, M.F., The Jones-Witten Invariants of Knots, Séminaire N. Bourbaki, 1989-1990, exp. no.715, pages 7-16.

[AtBott] Atiyah, M.F. and R.Bott, The Yang-Mills Equations over Riemann Surfaces, Phil. Trans. R. Soc. Lond. A **308**(1983), 523-615.

[ADPW] Axelrod, S., S. Della Pietra, and E. Witten, Geometric Quantization of Chern-Simons Gauge Theory, J. Diff. Geo. **33**(1991), 787-902.

[BM] Baez, J. and J.P. Muniain, *Gauge Fields, Knots and Gravity*, World Scientific, Singapore, 2006.

[BK] Bakalov, B. and A. Kirillov, Jr., *Lectures on Tensor Categories and Modular Functors*, Amer. Math. Soc., Providence, RI, 2001.

[Bas] Baseilhac, S., Chern-Simons Theory in Dimension Three. Available at http://www.math.univ-montp2.fr/ baseilhac/CS.pdf.

[Bax] Baxter, R.J., *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, UK, 1989.

[BL] Beauville, A. and Y. Laszlo, Conformal Blocks and Generalized Theta Functions, Commun. Math. Phys. **164**(1994), 385-419.

[BPZ] Belavin, A.A., A.M. Polyakov, and A.B. Zamolodchikov, Infinite Conformal Symmetry in Two-Dimensional Quantum Field Theory, Nucl. Phys. B **241**(1984), 333-380.

[Berry] Berry, M.V., Quantal Phase Factors Accompanying Adiabatic Changes, Proc. Roy. Soc. Lond. A **392**(1984), 45-57.

[BB] Birman, J.S. and T.E. Brendle, Braids: A Survey, Available at https://arxiv.org/abs/math/0409205.

[Blau] Blau, M., *Symplectic Geometry and Geometric Quantization*, Available at http://www.blau.itp.unibe.ch/Lecturenotes.html

[Borde] Borde, A., Topology Change in Classical General Relativity, Available at https://arxiv.org/abs/gr-qc/9406053v1

[Bott] Bott, R., Homogeneous Vector Bundles, Ann. of Math. **66**(1957), 51-101.

[Brown] Brown, L.M. (Editor), *Feynman's Thesis: A New Approach to Quantum Theory*, World Scientific, Singapore, 2005.

[Bryant] Bryant, R.L., An Introduction to Lie Groups and Symplectic Geometry. Pages 1-181 of [FU2].

[BZ] Burde, G. and H. Zieschang, *Knots*, Second Edition, de Gruyter, Berlin, Germany, 2003.

[CMV] Cherednikov, I.O., T. Mertens, and F.F. Van der Veken, *Wilson Lines in Quantum Field Theory*, de Gruyter, Berlin, Germany, 2014.

[CS] Chern, S-S and J. Simons, Characteristic Forms and Geometric Invariants, Ann. Math. **99**(1974), 48-69.

[Chev] Chevalley, C., *Theory of Lie Groups*, Princeton University Press, Princeton, NJ, 1946.

[CJ] Chruściński, D. and A. Jamiołkowski, *Geometric Phases in Classical and Quantum Physics*, Springer, New York, NY, 2004.

[Collins] Collins, D.J., A Simple Presentation of a Group with Unsolvable Word Problem, Illinois J. Math. **30**(1986), 230-234.

[Con] Conway, J.H., An Enumeration of Knots and Links, *Computational Problems in Abstract Algebra*, Pergamon Press, New York, NY, 1970, 329-358.

[CF] Crowell, R.H. and R.H. Fox, *Introduction to Knot Theory*, Springer, New York, NY, 1963.

[DS] Dewitt, B.S. and R. Stora (Editors), *Relativity, Groups and Topology II*, Elsevier, The Netherlands, 1984.

[DFMS] Di Francesco, P., P. Mathieu, and D. Sénéchal, *Conformal Field Theory*, Springer, New York, NY, 1996.

[Dijk] Dijkgraaf, R., *A Geometrical Approach to Two-Dimensional Conformal Field Theory*. Available at https://dspace.library.uu.nl/handle/1874/210872

[Dug] Dugundji, J., *Topology*, Allyn and Bacon, Boston, MA, 1970.

[Fiore] Fiore, T.M., *Pseudo Limits, Biadjoints, and Pseudo Algebras: Categorical Foundations of Conformal Field Theory*, Amer. Math. Soc., Providence, RI, 2006.

[Fisher] Fisher, G.M., On the Group of all Homeomorphisms of a Manifold, Trans. Amer. Math. Soc. **97**(1960), 193-212.

[Folland] Folland, G.B., *Quantum Field Theory: A Tourist Guide for Mathematicians*, American Mathematical Society, Providence, RI, 2008.

[Fox] Fox, R.H., A Quick Trip Through Knot Theory, in *Topology of 3-Manifolds and Related Topics*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[Freed] Freed, D.S., Classical Chern-Simons Theory, Part 1, Available at https://arxiv.org/abs/hep-th/9206021.

[FU1] Freed, D.S. and K.K. Uhlenbeck, *Instantons and Four-Manifolds*, Second Edition, Springer, New York, NY, 1991.

[FU2] Freed, D.S. and K.K. Uhlenbeck, Editors, *Geometry and Quantum Field Theory*, American Mathematical Society, Providence, RI, 1995.

[Fuj1] Fujiwara, D., A Construction of the Fundamental Solution for the Schrödinger Equations, Proc. Jap. Acad. **55**, Ser A(1979), 10-14.

[Fuj2] Fujiwara, D., On the Nature of Convergence for Some Feynman Path Integrals I, Proc. Jap. Acad. **55**, Ser A(1979), 195-200 .

[Fuj3] Fujiwara, D., On the Nature of Convergence for Some Feynman Path Integrals II, Proc. Jap. Acad. **55**, Ser A(1979), 273-277.

[Gluck] Gluck, H., Unknotting $\mathbb{S}^1$ in $\mathbb{S}^4$, Bull. Amer. Math. Soc. **69**(1963), 91-94.

[GG] Golubitsky, M. and V. Guillemin, *Stable Mappings and their Singularities*, Third Printing, Springer, New York, NY, 1986.

[G-M] Gonzales-Meneses, J., Basic Results on Braid Groups, Available at https://arxiv.org/abs/1010.0321.

[GL] Gordon, C. McA. and J. Luecke, Knots are Determined by their Complements, Proc. Amer. Math. Soc. **35**(1972), 287-288.

[Gotay] Gotay, M., A Class of Nonpolarizable Symplectic Manifolds, Monatsh. Math. **103**(1987), 27-30.

[Green] Greenberg, M., *Lectures on Algebraic Topology*, W.A. Benjamin, Inc., New York, NY, 1967.

[GP] Guillemin, V. and A. Pollack, *Differential Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[GS] Guillemin, V. and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, England, 1984.

[Hahn] Hahn, A., Torus Knots and the Chern-Simons Path Integral: A Rigorous Treatment. Available at https://arxiv.org/pdf/1508.03804.pdf.

[Haken] Haken, W., Some Results on Surfaces in 3-Manifolds, in *Studies in Modern Topology*, Mathematical Association of America, Washington, D.C., 1968.

[Hall] Hall, B.C., *Quantum Theory for Mathematicians*, Springer, New York, NY, 2013.

[Ham] Hamilton, R.S., The Inverse Function Theorem of Nash and Moser, Bull. Amer. Math. Soc. **7**(1982), 65-222.

[Hatch] Hatcher, A., *Algebraic Topology*, Available at https://www.math.cornell.edu/ hatcher/AT/AT.pdf.

[Herm] Herman, M., Simplicité du groupe des difféomorphismes de classe C^∞ isotopes a l'identité du tore de dimension n, C.R. Acad. Sci. Paris, Sér A, **273**(1971), 232-234.

[Hirsch] Hirsch, M.W., *Differential Topology*, Springer, New York, NY, 1976.

[Hitchin] Hitchin, N., Flat Connections and Geometric Quantization, Commun. Math. Phys. **131**(1990), 347-380.

[HOMFLY] Freyd, P., Yetter, D., Hoste, J., Lickorish, W.B.R., Millett, K., and Ocneanu, A., A New Polynomial Invariant for Knots and Links, Bull. Amer. Math. Soc. **12**(1985), 239-246.

[Hu] Hu, S., *Lecture Notes on Chern-Simons-Witten Theory*, World Scientific, Singapore, 2001.

[Huy] Huybrechts, D., *Complex Geometry: An Introduction*, Springer, New York, NY, 2005.

[IT] Imayoshi, Y. and M. Taniguchi, *An Introduction to Teichmïller Spaces*, Springer, New York, NY, 1992.

[Jimbo] Jimbo, M., Introduction to the Yang-Baxter Equation, Inter. J. Math. Phys. A **4**(1989), 3759-3777.

[Jones1] Jones, V.F.R., Index for Subfactors, Invent. Math. **72**(1983), 1-25.

[Jones2] Jones, V.F.R., Braid Groups, Hecke Algebras, and Type Π_1 Factors, Japan-U.S. Conf. Proc. (1983).

[Jones3] Jones, V.F.R., A Polynomial Invariant for Knots Via Von Neumann Algebras, Bull. Amer. Math. Soc. **12**(1985), 103-111.

[Jones4] Jones, V.F.R., Hecke Algebra Representations of Braid Groups and Link Polynomials, Ann. Mathe **126**(1987), 335-388.

[Jones5] Jones, V.F.R., On Knot Invariants Related to Some Statistical Mechanical Models, Pacific J. Math. **137**(1989), 311-334.

[Jones6] Jones, V.F.R., The Jones Polynomial for Dummies, Available at https://math.berkeley.edu/ vfr/jonesakl.pdf

[Jost] Jost, J., *Compact Riemann Surfaces*, Third Edition, Springer, New York, NY, 2006.

[Kawa] Kawauchi, A., *A Survey of Knot Theory*, Birkhäuser, Basel, 1996.

[KM] Kirby, R. and P. Melvin, The 3-Manifold Invariants of Witten and Reshetikhin-Turaev for $\mathfrak{sl}(2,\mathbb{C})$, Invent. Math. **105**(1991), 473-545.

[Kiri] Kirillov, A.A., Geometric Quantization, in *Dynamical Systems IV, Encyclopedia of Mathematical Sciences* **4**, pages 139-176, Springer, New York, NY, 2001.

[Kob] Kobayashi, S., *Differential Geometry of Complex Vector Bundles*, Iwanami Shoten Publishers and Princeton University Press, Princeton, NJ, 1987. Available at http://mathsoc.jp/publication/PublMSJ/PDF/Vol15.pdf

[KN1] Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry, Volume I*, Interscience Publishers, New York, NY, 1963.

[KN2] Kobayashi, S. and K. Nomizu, *Foundations of Differential Geometry, Volume II*, Interscience Publishers, New York, NY, 1969.

[Kohno] Kohno, T., *Conformal Field Theory and Topology*, American Mathematical Society, Providence, RI, 2002.

[Kost] Kostant, B., Lie Algebra Cohomology and the Generalized Borel-Weil Theorem, Ann. of Math. **74**(1961), 329-387.

[LR] Labastida, J.M.F. and A.V. Ramallo, Chern-Simons Theory and Conformal Blocks, CERN-TH. 5432/89. Available at http://www-fp.usc.es/theory/labastida/papers/p20.pdf

[Lev] Levine, J., Unknotting Spheres in Codimension Two, Topology **4**(1965), 9-16.

[Lick] Lickorish, W.B.R., A Representation of Orientable, Combinatorial 3-Manifolds, Ann. Math. **76**(1962), 531-540.

[Mack] Mackey, G.W., *The Theory of Unitary Group Representations*, University of Chicago Press, Chicago, IL, 1976.

[MT] Madsen, I. and J. Tornehave, *From Calculus to Cohomology: De Rham Cohomology and Characteristic Classes*, Cambridge University Press, Cambridge, England, 1997.

[Mark] Markov, A.A., Uber die freie Aquivalenz geschlossener Zopfe, Mat. Sb. **1**(1935), 73-78.

[Math] Mathew, A., Notes on the Kodaira Vanishing Theorem. Available at http://math.uchicago.edu/ amathew/kodaira.pdf.

[MS] McDuff, D. and D. Salamon, *Introduction to Symplectic Topology*, Oxford University Press, Oxford, England, 1998.

[Merm] Mermin, N.D., Could Feynman Have Said This?, Physics Today, **57**, Issue 5, (2004), 10-12.

[Mich] Michiels, D., *Moduli Spaces of Flat Connections*, Master's Thesis, KU Leuven, 2013. Available at https://faculty.math.illinois.edu/ michiel2/docs/thesis.pdf

[MilSta] Milnor, J.W. and J.D. Stasheff, *Characteristic Classes*, Princeton University Press, Princeton, NJ, 1974.

[Miln1] Milnor, J.W., *Topology from the Differentiable Viewpoint*, Princeton University Press, Princeton, NJ, 1997.

[Miln2] Milnor, J.W., *Lectures on the h-Cobordism Theorem*, Princeton University Press, Princeton, NJ, 2015.

[MoSei] Moore, G. and N. Seiberg, Polynomial Equations for Rational Conformal Field Theories, Phys. Lett., **212B**(1988), 451-460.

[Morita] Morita, S., *Geometry of Characteristic Classes*, American Mathematical Society, Providence, RI, 2001.

[Mour] Mourão, J.M., Aspects of the Connections between Path Integrals, Quantum Field Theory, Topology and Geometry, $https://www.math.tecnico.ulisboa.pt/~jmourao/PI.pdf$

[Nab1] Naber, G., *Topological Methods in Euclidean Spaces*, Dover Publications, Mineola, NY, 2000.

[Nab2] Naber, G., *Topology, Geometry and Gauge Fields: Foundations*, Second Edition, Springer, New York, NY, 2011.

[Nab3] Naber, G., *Topology, Geometry and Gauge Fields: Interactions*, Second Edition, Springer, New York, NY, 2012.

[Nab4] Naber, G., *Foundations of Quantum Mechanics: An Introduction to the Physical Background and Mathematical Structure*, Available at http://gregnaber.com/wp-content/uploads/Simple-Harmonic-Oscillator.pdf.

[NS] Narasimhan, M.S. and C.S. Seshadri, Stable and Unitary Bundles on a Compact Riemann Surface, Ann. Math. **82**(1965), 540-567.

[Nash] Nash, C., *Differential Topology and Quantum Field Theory*, Academic Press (Elsevier), San Diego, CA, 2003.

[NN] Newlander, A. and L. Nirenberg, Complex Analytic Coordinates in Almost Complex Manifolds, Ann. Math. **65**(1957), 391-404.

[Nir] Nirenberg, L., A Complex Frobenius Theorem, Conference on Analytic Functions, Institute for Advanced Study, 1958. Available at https://archive.org/details/complexfrobenius00nire.

[PraSos] Prasolov, V.V. and A.B. Sossinsky, *Knots, Links, Braids and 3-Manifolds*, American Mathematical Society, Providence, RI, 1997.

[PreSeg] Pressley, A. and G. Segal, *Loop Groups*, Oxford University Press, 1986.

[Quinn] Quinn, F., Lectures on Axiomatic Topological Quantum Field Theory, in *Geometry and Quantum Field Theory*, pages 325-453, American Mathematical Society, Providence, RI, 1995.

[RSW] Ramadas, T.R., I.M. Singer and J. Weitsman, Some Comments on Chern-Simons Gauge Theory, Commun. Math. Phys. **126**(1989), 409-420.

[Reid] Reidemeister, K., Elementare Begründung der Knotentheorie, Abh. Math. Sem. Univ. Hamburg **5**(1927), 24-32.

[RT] Reshetikhin, N.Yu. and V.G. Turaev, Invariants of 3-Manifolds via Link Polynomials and Quantum Groups, Invent. Math. **103**(1991), 547-597.

[RN] Ricca, R.L. and B. Nipoti, Gauss' Linking Number Revisited, J. Knot Theory and its Ramifications **20**(2011), 1325-1343.

[Robles] Robles, N.M., Zeta Function Regularization. Available at http://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/theoretical-physics/msc/dissertations/2009/Nicolas-Robles-Thesis.pdf

[Rolf] Rolfsen, D., *Knots and Links*, AMS Chelsea Publishing, American Mathematical Society, Providence, RI.

[Scher] Scherich, N., The Alexander Polynomial, Available at https://ir.library.oregonstate.edu/xmlui/handle/1957/40791.

[Schm] Schmid, R., The Geometry of BRS Transformations, Illinois J. Math. **34**(1990), 87-97.

[Schott] Schottenloher, M., *A Mathematical Introduction to Conformal Field Theory*, Second Edition, Springer, New York, NY, 2008.

[Schul] Schulman, L.S., *Techniques and Applications of Path Integration*, Dover Publications, Mineola, NY, 2005.

[Schw1] Schwarz, A., The Partition Function of Degenerate Quadratic Functional and Ray-Singer Invariants, Lett. Math. Phys. 2(1978), 247-252.

[Schw2] Schwarz, A., Topological Quantum Field Theories, Available at https://arxiv.org/abs/hep-th/0011260

[Schw3] Schwarz, A., *Quantum Field Theory and Topology*, Springer, New York, NY, 1993.

[Segal1] Segal, G., Geometric Aspects of Quantum Field Theory, Proceedings of the International Congress of Mathematicians, Kyoto, Japan, 1388-1396, The Mathematical Society of Japan, 1991.

[Segal2] Segal, G., The Definition of Conformal Field Theory, in *Topology, Geometry and Quantum Field Theory*, 421-577, Lond. Math. Soc. Lecture Notes Ser. **308**, Cambridge University Press, Cambridge, England, 2004.

[Segal3] Segal, G., Loop Groups. Available at http://people.mpim-bonn.mpg.de/zagier/files/doi/10.1007/BFb0084581/chapter08.pdf

[Segal4] Segal, G., Two-Dimensional Conformal Field Theory and Modular Functors, IX International Congress on Mathematical Physics, Swansea, 1988, 22-37.

[Seng] Sengupta, A.N., Chern-Simons Models: Rigorous Results. In *Encyclopedia of Mathematical Physics*, Volume 1, 496-499. Editors: Françoise, J-P, G.L. Naber, and Tsou Sheung Tsun, Elsevier, Amsterdam, The Netherlands, 2006.

[Sepan] Sepanski, M.R., *Compact Lie Groups*, Springer, New York, NY, 2007.

[Simon] Simon, B., Holonomy, the Quantum Adiabatic Theorem, and Berry's Phase, Phy. Rev. Lett. **51**(1983), 2167-2170.

[Spivak] Spivak, M., *A Comprehensive Introduction to Differential Geometry*, Volume I, First Edition, Publish or Perish, Houston, TX, 1970.

[TL] Temperley, H.N.V. and E.H. Lieb, Relations Between the ‘Percolation’ and ‘Colouring’ Problem and Other Graph-Theoretical Problems Associated with Regular Planar Lattices: Some Exact Results for the ‘Percolation’ Problem, Proc. Roy. Soc. (London) A **322**(1971), 251-280.

[Tur1] Turaev, V.G., The Yang-Baxter Equation and Invariants of Links, Invent. Math. **92**(1988), 527-553.

[Tur2] Turaev, V.G., *Quantum Invariants of Knots and 3-Manifolds*, Studies in Mathematics, Walter de Gruyter, Berlin, 1994.

[TW] Tuynman, G.M. and W.A.J.J. Wiegerinck, Central Extensions and Physics, J. Geo. Phys. **4**(1987), 207-258.

[Vais] Vaisman, I., Basics of Lagrangian Foliations, Publicacions Matemàtiques, **33**(1989), 559-575.

[Walk] Walker, K., On Witten’s 3-Manifold Invariants. Available at http://www.maths.ed.ac.uk/ v1ranick/papers/walker3.pdf

[Wall] Wallace, A.H., Modifications and Cobounding Manifolds, Can. J. Math. **12**(1960), 503-528.

[Wight] Wightman, A.S, How it was Learned that Quantized Fields are Operator-Valued Distributions, Fortchr. Phys. **44** (2), 1996, 143-178.

[Witt1] Witten, E., 2+1 Dimensional Gravity as an Exactly Soluble System, Nucl. Phys. B **311**(1988), 46-78.

[Witt2] Witten, E., Quantum Field Theory and the Jones Polynomial, Comm. Math. Phys. **121**(1989), 351-399.

[Witt3] Witten, E., New Results in Chern-Simons Theories, In *Geometry of Low-Dimensional Manifolds 2*, Edited by S.K. Donaldson and C.B. Thomas, Cambridge University Press, Cambridge, England, 1990.

[Wood] Woodhouse, N.M.J., *Geometric Quantization*, Second Edition, Oxford University Press, Oxford, England, 1991.

[Wu1] Wu, F.Y., Knot Theory and Statistical Mechanics, Rev. Mod. Phys. **64**(1992), 1099-1131.

[Wu2] Wu, F.Y., The Yang-Baxter Equation in Knot Theory, Inter. J. Mod. Phys. B **7**, No. 20 and 21, 3737-3750.

[Zee] Zee, A., *Quantum Field Theory in a Nutshell*, Second Edition, Princeton University Press, Princeton, NJ, 2010.

[Zeeman] Zeeman, E.C., Unknotting Spheres, Ann. Math. **72**(1960), 350-361.

INDEX