Knots, three-manifolds and instantons

P. B. Kronheimer and T. S. Mrowka

Low-dimensional topology is the study of manifolds and cell complexes in dimensions four and below. Input from geometry and analysis has been central to progress in this field over the past four decades, and this article will focus on one aspect of these developments in particular, namely the use of Yang-Mills theory, or gauge theory. These techniques were pioneered by Simon Donaldson [**7, 9**] in his work on 4-manifolds, but the past ten years have seen new applications of gauge theory, and new interactions with more recent threads in the subject, particularly in 3-dimensional topology.

This is a field where many mathematical techniques have found applications, and sometimes a theorem has two or more independent proofs, drawing on more than one of these techniques. We will focus primarily on some questions and results where gauge theory plays a special role.

1. Representations of fundamental groups

1.1. Knot groups and their representations. Knots have long fascinated mathematicians. In topology, they provide blueprints for the construction of manifolds of dimension three and four. For this exposition, a *knot* is a smoothly embedded circle in 3-space, and a *link* is a disjoint union of knots. The simplest examples, the trefoil knot and the Hopf link, are shown in Figure 1, alongside the trivial round circle, the "unknot".

Knot theory is a subject with many aspects, but one place to start is with the *knot group*, defined as the fundamental group of the complement of a knot $K \subset \mathbb{R}^3$. We will write it as $\pi(K)$. For the unknot, $\pi(K)$ is easily identified as $\mathbb{Z}$. One of the basic tools of 3-dimensional topology is Dehn's Lemma, proved by Papakyriakopoulos in 1957, which provides a converse:

THEOREM 1.1 (Papakyriakopoulos, [**38**]). *If the knot group $\pi(K)$ is $\mathbb{Z}$, then K is the unknot.*

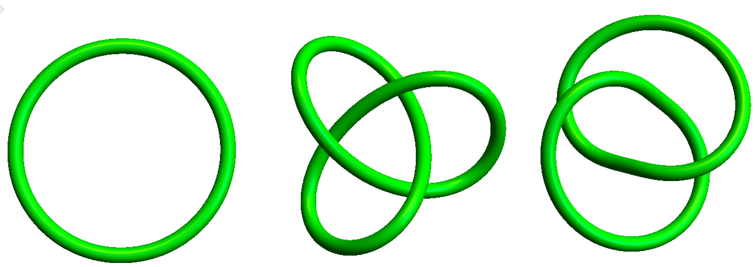


FIGURE 1. The unknot, trefoil and Hopf link.

The work of the first author was supported by the National Science Foundation through NSF grants DMS-1405652 and DMS-1707924. The work of the second author was supported by NSF grant DMS-1406348 and by a grant from the Simons Foundation, grant number 503559 TSM..

It is a consequence of Alexander duality that the *abelianization* of $\pi(K)$ is $\mathbb{Z}$ for any knot. (This is the first homology of the complement.) So we may restate the result above as saying that the unknot is characterized by having abelian fundamental group. In particular, if we are able to find a homomorphism $\pi(K) \to G$ with non-abelian image in any target group G, then K must be genuinely knotted. We begin our account of more modern results with the following theorem.

Theorem 1.2 (Kronheimer-Mrowka, [**29**]). *If K is a non-trivial knot, then there exists a homomorphism,*

$$\rho : \pi(K) \to SO(3),$$

with non-abelian image, from the knot group to the 3-dimensional rotation group.

There are two aspects to why this result is interesting. First, representations of the knot group in particular types of target groups are a central part of the subject: the case that G is dihedral leads to the "Fox colorings" [**16**], and the more general case of a two-step solvable group is captured by the Alexander polynomial and related invariants. But there are non-trivial knots with no Fox colorings and trivial Alexander polynomial. It is known that $\pi(K)$ is always a residually finite group, so there are always non-trivial homomorphisms to finite groups; but it is perhaps surprising that the very smallest simple Lie group is a target for all non-trivial knots.

Secondly, the theorem is of interest for the techniques that are involved in its proof, some of which we will describe later. A rich collection of tools from gauge theory are needed, and these are coupled with more classical tools from 3-dimensional topology, namely the theory of incompressible surfaces and decomposition theory, organized in Gabai's theory of sutured manifolds [**17**].

1.2. Orbifolds from knots. The knot group $\pi(K)$ has a distinguished conjugacy class, namely the class of the *meridional elements.* A meridional element m is one represented by a small loop running once around a circle linking K. If we take a planar diagram of a knot (a generic projection of K into $\mathbb{R}^2$), and take our basepoint for the fundamental group to lie above the plane, then there is a distinguished meridional element m_e for each *arc* e of the diagram (a path running from one undercrossing to the next). The elements m_e generate the knot group and satisfy a relation at each crossing, the *Wirtinger relations* [**16**]. (See Figure 2.)

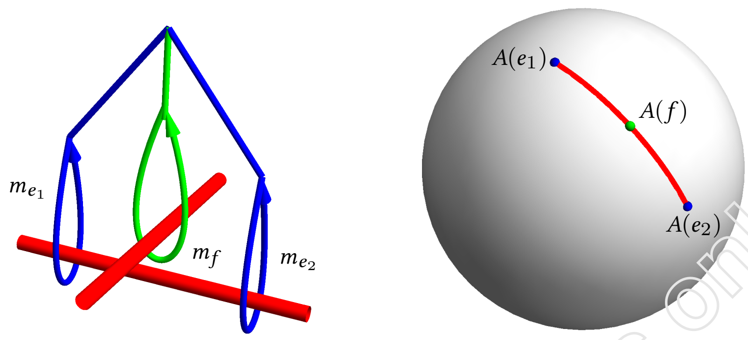
Theorem 1.2 can be refined to say that ρ can be chosen so that $\rho(m)$ has order 2 in $SO(3)$, for one (and hence all) meridional elements. This refinement can be helpfully reinterpreted in terms of the fundamental group of an *orbifold.* Recall that an orbifold is a space locally modeled on the quotient of a manifold by a finite group, and that its *singular* set is the locus of points which have non-trivial stabilizer in the local models. Given a knot or link K in a 3-manifold Y, one can equip Y with the structure of an orbifold whose non-trivial stabilizers are all $\mathbb{Z}/2$ and whose singular set is K. Let us write $\mathrm{Orb}(Y, K)$ for this orbifold. The *orbifold fundamental group* in this situation can be described as the fundamental group of the complement of the singular set with relations

$$m^2 = 1$$

imposed, for all meridional elements. Thus the refinement we seek can be stated:

Theorem 1.3 (Kronheimer-Mrowka, [**29**]). *If K is a non-trivial knot in S^3, and $O = \mathrm{Orb}(S^3, K)$ is the corresponding orbifold with $\mathbb{Z}/2$ stabilizers, then there exists a homomorphism from the orbifold fundamental group,*

$$\rho : \pi_1(O) \to SO(3),$$



(A) (B)

FIGURE 2. (A) The *Wirtinger relation*, $m_f m_{e_2} = m_{e_1} m_f$ holds in the fundamental group of the complement. (B) The corresponding points on the sphere lie on a geodesic arc. The reflection about the green axis interchanges the two blue points.

with non-abelian image.

Using the Wirtinger presentation described above, this result can be given a concrete interpretation. An element of order 2 in $SO(3)$ is a 180° rotation about an axis A in $\mathbb{R}^3$, an these are therefore parametrized by the points A of $\mathbb{RP}^2$. So if we are given a diagram of K, then a representation $\rho : \pi(K) \to SO(3)$ which sends meridians to elements of order 2 can be described by giving a point $A(e)$ in $\mathbb{RP}^2$ for each arc e, satisfying a collection of constraints coming from the Wirtinger relations at the crossings. So the concrete version of Theorem 1.3 is the following.

THEOREM 1.4. *Given any diagram of a non-trivial knot K, we can find a non-trivial assignment $e \mapsto A(e)$,*

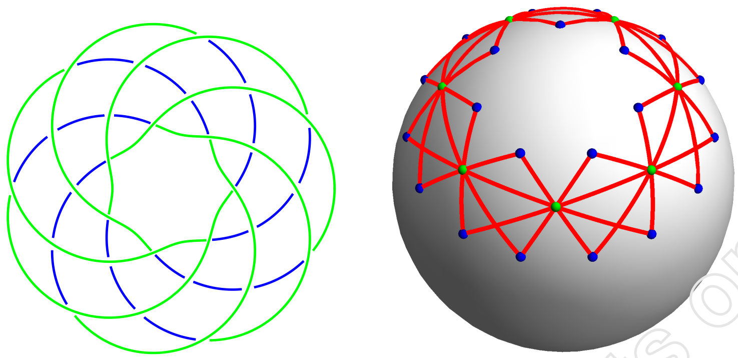
$$\{\text{arcs of the diagram}\} \to \mathbb{RP}^2,$$

so that the following condition holds: whenever e_1, e_2, f are arcs meeting at a crossing, with f being the overcrossing arc (see Figure 2), the point $A(e_2)$ is the reflection of $A(e_1)$ in the point $A(f)$.

The last condition in the theorem means that $A(e_1)$, $A(f)$, $A(e_2)$ are equally spaced along a geodesic, see Figure 2. The case that all the $A(e)$ are equal is the trivial case, and corresponds to the abelian representation. Dihedral representations arise when the points $A(e)$ lie at the vertices of a regular polygon on $\mathbb{RP}^1$. A configuration corresponding to a non-dihedral representation of the $(5, 7)$-torus knot is illustrated in Figure 3.

1.3. Three-manifolds and $SO(3)$. Having considered a knot or link in $\mathbb{R}^3$ and an associated orbifold, we consider next a closed 3-manifold Y and its fundamental group $\pi_1(Y)$. From the solution of the Poincaré conjecture [**33**], we know that $\pi_1(Y)$ is non-trivial if Y is not the 3-sphere. Motivated by the discussion of knot groups in the previous section, one might ask:

QUESTION 1.5. *Let Y be a closed 3-manifold with non-trivial fundamental group. Does there exist a non-trivial homomorphism $\rho : \pi_1(Y) \to SO(3)$?*



(A) (5,7)-torus knot diagram (B) Axes of one representation

FIGURE 3. The green arcs in the diagram on the left contain all the over-crossings. The blue arcs consist only of under-crossings. In the right-hand picture, the green vertices are axes $A(e)$ corresponding to green arcs e in the knot diagram. The blue vertices correspond to blue arcs.

It is not known whether the answer is yes in general. Stated this way, the interesting case for this question is when Y is a *homology 3-sphere*, i.e. a 3-manifold with the same (trivial) homology as S^3. (If Y has non-trivial homology, then $\pi_1(Y)$ has a cyclic group as a quotient, and there will always be a representation in $SO(3)$ with cyclic image.) For homology 3-spheres, an affirmative answer to the question is known when Y has non-zero Rohlin invariant [**1**], when Y is obtained by Dehn surgery on a knot in S^3 [**27, 29**], or when Y carries a taut foliation [**28**]. See also [**3**].

There is an interesting variant of this question, for 3-manifolds with non-trivial homology. A representation $\rho : \pi_1(Y) \to SO(3)$ defines a flat vector bundle on Y with fiber $\mathbb{R}^3$, and such a vector bundle has a second Stiefel-Whiney class w_2. Thus the representations ρ can be grouped by this class,

$$w_2(\rho) \in H^2(Y; \mathbb{Z}/2).$$

which is the obstruction to lifting ρ to the double cover $SU(2) \to SO(3)$. The following result completely describes the classes which arise as $w_2(\rho)$.

THEOREM 1.6. [**29**] *Let Y be a closed, oriented 3-manifold and let $\omega \in H^2(Y; \mathbb{Z}/2)$ be given. Suppose that for every embedded 2-sphere S in Y, the pairing $\omega \cdot [S]$ is zero mod 2. Then there exists a homomorphism $\rho : \pi_1(Y) \to SO(3)$ with $w_2(\rho) = \omega$.*

REMARK 1.7. The restriction on ω is also necessary as well as sufficient, because a flat vector bundle on a 2-sphere is trivial and must therefore have trivial Stiefel-Whitney class on the sphere.

REMARK 1.8. The condition on $\omega \cdot [S]$ is automatically satisfied if Y is *irreducible*, i.e. if every 2-sphere in Y bounds a ball. To prove the theorem it is enough to consider only irreducible 3-manifolds.

REMARK 1.9. If ρ has cyclic image, then $w_2(\rho)$ has a lift to a *torsion* class in the integer homology, $H^2(Y)$. The case that there is no such lift is the case that $w_2(\rho)$ has non-trivial image in $\mathrm{Hom}(H_2(Y), \mathbb{Z}/2)$, and in this case ρ must be non-cyclic. The case that ω has non-trivial image in $\mathrm{Hom}(H_2(Y), \mathbb{Z}/2)$ is the difficult case for the theorem.

An interesting special case of this theorem is the case that Y is the mapping torus of a diffeomorphism, $h : \Sigma_g \to \Sigma_g$, of a surface of genus g. The conjugacy classes of representations $\pi_1(\Sigma_g) \to SO(3)$ with non-zero w_2 are parametrized by an orbifold $M(\Sigma_g)$ of dimension $6g - 6$, and the diffeomorphism h gives rise to a map $h^* : M(\Sigma_g) \to M(\Sigma_g)$ by pull-back. It is then a consequence of the above theorem that the diffeomorphism h^* has fixed points in $M(\Sigma_g)$. In this form, the result was proved independently and with different methods by Ivan Smith [**42**].

1.4. Spatial graphs. A spatial graph is a graph (tamely) embedded as a topological space in $\mathbb{R}^3$. We will be interested here in finite, trivalent graphs (also called cubic graphs). Thus we are generalizing classical knots and links by allowing vertices of valence 3. We allow that the set of vertices may be empty, so knots and links are included as a special case, regarded as vertexless graphs. There is a significant literature on spatial graphs: see for example [**4**].

As with knots and links, we write $\pi(K)$ for the fundamental group of the complement of a spatial graph $K \subset \mathbb{R}^3$. For each edge e of K, there is corresponding distinguished conjugacy class of meridional curves m_e, obtained from the small circles linking e. Following the same lines as before, we wish to study representations of $\pi(K)$ in $SO(3)$, with the constraint that the meridional elements map to elements of order 2. Representations of this sort are parametrized by a topological space, the *representation variety*,

$$\mathcal{R}(K) = \{\, \rho : \pi(K) \to SO(3) \mid \rho(m_e) \text{ has order 2 for all edges } e \} \tag{1}$$

As with knots and links, this representation variety for a spatial graph can be interpreted as a space of representations for the fundamental group of an orbifold. Given a trivalent graph K in a 3-manifold Y, we may construct a 3-dimensional orbifold $\mathrm{Orb}(Y, K)$ whose underlying topological space is Y, whose singular set is K, and whose local stabilizer groups are $\mathbb{Z}/2$ at the interior points of edges of K. At vertices of K where three edges meet, the local model for the orbifold is the quotient of the 3-ball by the Klein four-group, V_4. In this way, $\mathcal{R}(K)$ becomes the space of homomorphisms from the orbifold fundamental group,

$$\rho : \pi_1(\mathrm{Orb}(S^3, K)) \to SO(3),$$

with the additional property that ρ is injective on each of the non-trivial local stabilizer groups.

The Klein 4-group V_4 is contained in $SO(3)$ as the subgroup of diagonal matrices, and representations $\rho : \pi(K) \to SO(3)$ with image in V_4 play a special and already subtle role. Since V_4 is abelian, a representation into the Klein four-group factors through the abelianization of $\pi(K)$, namely the homology $H_1(S^3 \setminus K))$, so we are considering homomorphisms

$$\tau : H_1(S^3 \setminus K) \to V_4$$

which map meridional elements to elements of order 2. If we write the elements of V_4 as $\{1, A, B, C\}$, then τ assigns one of three "colors" $\{A, B, C\}$ to each edge e, and this coloring must satisfy the constraint that, at a vertex, the colors of the three incident edges are all different (because the sum of the corresponding elements of H_1 is zero). Such a 3-coloring of the edges of a trivalent graph is a called a *Tait coloring*. So we have:

PROPOSITION 1.10. *For a trivalent spatial graph K, the representations $\rho \in \mathcal{R}(K)$ whose image is contained in the Klein 4-group V_4 are in one-to-one correspondence with Tait colorings of K.*

Notice in particular that this set depends only on the abstract graph K, independent of the embedding. This is a reflection of the fact that the homology group $H_1(S^3 \setminus K)$ is isomorphic to $H^1(K)$ by Lefschetz duality.

The question of whether a cubic graph admits a Tait coloring is difficult. Indeed Tait [**43**] observed that the four color theorem [**2, 40**] could be reframed as a question about the existence of Tait colorings as follows. A planar map determines a graph, giving the borders of the countries. A map is called proper if no country has a border with itself, in which case the graph of the borders is *bridgeless*; that is, no edge (or "bridge") can be removed making the graph disconnected. It is also elementary to see that it suffices to verify the four color theorem for planar maps whose border graph is trivalent. Tait's observation is that the four-colorability of the regions of the map is equivalent to the existence of a Tait coloring of the border graph. So the four color theorem is equivalent to the statement that every bridgeless trivalent graph admits a Tait coloring.

While the methods of gauge theory at the time of writing have not given a proof the four color theorem, one can prove some suggestive results. To state the main result we observe that for spatial graphs there is a natural extension of being bridgeless. A *spatial bridge* is an edge of a spatial graph K for which the meridional loop is contractible in the complement of the graph. Equivalently, it is an edge e for which we can find a sphere S for which $K \cap S$ is a single point of e, with transverse intersection. Note that the existence of such a spatial bridge implies that $\mathcal{R}(K)$ is empty. The converse is the following non-trivial theorem.

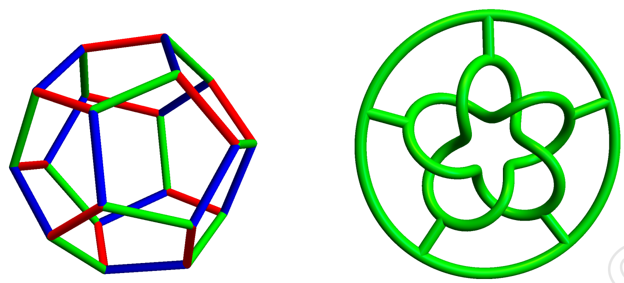
THEOREM 1.11. [**26**] *For any trivalent graph $K \subset \mathbb{R}^3$ without a spatial bridge, the representation variety $\mathcal{R}(K)$ is non-empty.*

We conclude this section with some remarks to put this result in context. There is an action of $SO(3)$ on $\mathcal{R}(K)$ by conjugacy. Representation with image V_4 are characterized by the fact their stabilizer is exactly V_4 under this action. For a graph with at least one vertex, the possible groups that can arise as stabilizers are V_4, $\mathbb{Z}/2$ and the trivial group. One can show (see [**26**]) that, for planar graphs, if there is a representation with non-trivial stabilizer then there is also one with V_4 stabilizer, and hence a Tait coloring of K. Theorem 1.11 is agnostic regarding the possible stabilizers of the representation that it guarantees. There are planar graphs with only V_4 representations, as well as many with both V_4 representations and irreducible representations (the simplest being the 1-skeleton of the dodecahedron.)

Bridgeless trivalent graphs with no Tait colorings are called snarks. The simplest one is the Petersen graph, shown in a spatial embedding in Figure 4.

Trivially, the four color theorem says that there are no planar snarks. For any spatial embedding of a snark, Theorem 1.11 guarantees the existence of a representation $\pi(K) \to SO(3)$.

Theorem 1.11 also says that even a graph with a bridge, when embedded in a spatially bridgeless manner, will have a nontrivial representation. An example is shown Figure 5. On the left, the "handcuffs" are shown embedded in $\mathbb{R}^3$ with a spatial bridge, and the representation variety $\mathcal{R}(K_1)$ is empty. On the right, the same abstract graph is shown with a more interesting embedding. The representation variety $\mathcal{R}(K_2)$ in this case consists



(A) (B)

FIGURE 4. The dodecahedral graph admits a Tait coloring, while the Petersen graph, the simplest snark, has none.

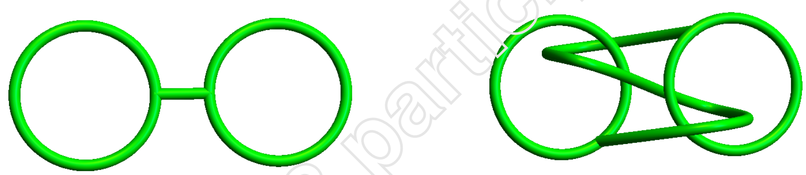


FIGURE 5. The standard handcuffs and the tangled handcuffs.

of the $SO(3)$ orbit of a single representation ρ whose image in $SO(3)$ is the symmetry group of a cube.

2. Background on instantons and four-manifolds

The theorems discussed above are proved by means of more general results based on non-vanishing theorems for Floer's instanton homology for 3-manifolds, first introduced in [**15**]. Before introducing the instanton homology groups, we discuss their natural historical precursor, the invariants of smooth 4-manifolds developed by Simon Donaldson in the 1980's [**9**].

2.1. Instanton moduli spaces. The story begins with the work of Donaldson and his use of gauge theory in 4-dimensional topology. On an oriented Riemannian manifold X of dimension $2n$, the Hodge $*$-operator maps n-forms to n-forms,

$$* : \Omega^n(X) \to \Omega^n(X)$$

and satisfies $*^2 = (-1)^n$. When n is even, this gives rise to a decomposition into the ± 1 eigenspaces, the *self-dual* and *anti-self-dual* n-forms,

$$\Omega^n(X) = \Omega^n_+(X) \oplus \Omega^n_-(X).$$

The case of dimension 4 and $n = 2$ plays a special role, because if $E \to X$ is a vector bundle and A is a connection in E, then the *curvature* of the connection is a 2-form with values in the endomorphisms of E:

$$F_A \in \Omega^2\big(X; \mathrm{End}(E)\big).$$

Only in dimension 4, therefore, we can decompose the curvature into its self-dual and anti-self-dual parts, $F_A^+ + F_A^-$ and we can consider the *anti-self-dual Yang-Mills equations*,

$$F_A^+ = 0.$$

The solutions are the *anti-self-dual* connections A, sometimes called *instantons* on X.

We shall first consider the case that the structure group G for the bundle is $SU(N)$, so that E is a rank-N bundle with a hermitian metric and trivialized determinant. Our connections A will be $SU(N)$ connections: they will respect the trivialization. The isomorphism classes of pairs (E, A) consisting of an $SU(N)$ bundle E with an anti-self-dual connection A are parametrized by a *moduli space* $M_N(X)$ (which depends also on the Riemannian metric). When X is closed and connected, the bundles E themselves are classified by a single integer k, the second Chern number, or instanton number,

$$k = c_2(E)[X].$$

We therefore have a decomposition

$$M_N(X) = \bigcup_k M_{N,k}(X).$$

Each $M_{N,k}(X)$ is finite-dimensional, and for generic choice of Riemannian metric it will be a smooth manifold, except at *reducible* solutions: i.e. those where A preserves some orthogonal decomposition of E. An index calculation yields a formula for the dimension of $M_{N,k}$,

$$\dim M_{N,k} = 4Nk - (N^2 - 1)(\chi + \sigma)/2, \tag{2}$$

in which χ and σ are the signature an Euler number of X. The quantity $(\chi + \sigma)/2$ is an integer, which can also be written as

$$b^2_+ - b^1 + 1$$

where b^i is the rank of $H^i(X)$ and b^2_+ is the dimension of a maximal positive-definite subspace for the quadratic form on $H^2(X; \mathbb{R})$ defined by the cup-square.

The space $M_{N,k}(X)$ will usually be non-compact, because there may be sequences of solutions (E_n, A_n) in which the point-wise norm of the curvature, $|F_{A_n}|$, diverges near finitely many points in X, a "bubbling" phenomenon analyzed by Uhlenbeck in [**47**].

Having associated to each closed Riemannian 4-manifold an infinite sequence of new spaces, one is led to ask whether the moduli spaces $M_{N,k}(X)$ are non-empty. Do the anti-self-dual Yang-Mills equations have solutions? This question was answered in the affirmative by Taubes [**44, 45**], who constructed solutions on general 4-manifolds X using a grafting technique to transfer standard solutions from flat $\mathbb{R}^4$. Taubes' results tell us in particular that $M_{2,k}(X)$ is non-empty for all $k \geq k_0$, where the value of k_0 depends only on the topology of X. The resulting solutions have curvature concentrated near points in X, the same situation that is allowed in Uhlenbeck's work.

2.2. Donaldson's polynomial invariants. Although they may be non-compact, Donaldson showed that the moduli spaces $M_{N,k}(X)$ have sufficient compactness properties as a consequence of Uhlenbeck's theorems that they may (under mild conditions) be regarded as possessing a *fundamental class* $[M_{N,k}(X)]$ in the homology of the ambient space in which they sit, namely the space $\mathcal{B}_{N,k}(X)$ which parametrizes *all* isomorphism classes of $SU(N)$ connections with instanton number k.

To elaborate on this, the space $\mathcal{B}_{N,k}(X)$ (or more relevantly, the open subspace $\mathcal{B}^*_{N,k}(X)$ of irreducible connections) has a well-understood topology, and the "fundamental class" $[M_{N,k}(X)]$ gives rise to a rich collection of invariants, the *Donaldson invariants* of X. For these to be defined, it is important that the moduli space be contained in the irreducible part, $\mathcal{B}^*_{N,k}(X)$, and this will be true for $N = 2$ and for generic choice of Riemannian metric, as long as $b^2_+(X) \geq 1$ and $k > 0$. (We discuss this point again in the next subsection.) Furthermore, if $b^2_+(X) \geq 2$, then the fundamental class of the moduli space in $\mathcal{B}^*_{2,k}(X)$ is independent of the choice of metric, and so can be regarded as an *invariant* of the underlying smooth 4-manifold X.

These invariants, defined originally using the $N = 2$ moduli spaces, are usually referred to as Donaldson's *polynomial invariants*. In the $N = 2$ case, the rational cohomology of $\mathcal{B}^*_{2,k}$ contains a polynomial algebra [**9**]. More specifically, let us introduce the symmetric algebra

$$\mathbb{A}(X) = \mathrm{Sym}(H_{even}(X;\mathbb{Q}))$$

graded so that $H_r(X;\mathbb{Q})$ lies in $\mathbb{A}_{4-r}(X)$. Then there is an injection, for each k,

$$\mu : \mathbb{A}_d(X) \to H^d(\mathcal{B}_{2,k}(X);\mathbb{Q}).$$

The polynomial invariants defined by the moduli spaces $M_{2,k}(X)$ are linear maps

$$q_{X,k} : \mathbb{A}_{d(k)}(X) \to \mathbb{Q}$$

where $d(k)$ is the dimension of the moduli space (2). If we accept that $M_{2,k}(X)$ carries a fundamental class in homology, then we can regard the definition as:

$$q_{X,k}(z) = \langle \mu(z), [M_{2,k}(X)] \rangle.$$

The definition can be generalized in various way, in particular by considering $N > 2$. We can omit k from the notation by taking the sum,

$$q_X = \bigoplus_k q_{X,k} : \mathbb{A}(X) \to \mathbb{Q}. \tag{3}$$

There is an understanding that q_X is zero on $\mathbb{A}_d(X)$ for integers d not of the form $d(k)$.

The invariants q_X of smooth 4-manifolds, together with some closely-related invariants [**8**], were the first tools which were able to show that the the diffeomorphism type of a simply-connected compact 4-manifold is not determined by its cohomology ring alone.

2.3. A generalization, $U(N)$ bundles. In the discussion above, the bundle E had structure group $SU(N)$. We now consider $U(N)$ bundles with non-trivial determinant instead. A $U(N)$ connection A can be described locally as the sum of an $SU(N)$ connection and a $U(1)$ connection. More invariantly, and globally, A is determined by

- a $PU(N)$ connection A^o; and
- a connection $\mathrm{tr}(A)$ in the determinant line bundle, the top exterior power $\Lambda^N E$.

Whether there are anti-self-dual connections on the line bundle $\Lambda^N E$ is determined by Hodge theory on X, and usually they will not exist if the line bundle is non-trivial. The appropriate set-up is to ask only that A^o is anti-self-dual. More specifically, we *fix* a line bundle $W \to X$ and the data we seek is:

- a $U(N)$ bundle $E \to X$ with $c_1(E) = c_1(W)$;
- a chosen isomorphism $\iota : \Lambda^N E \to L$;
- an anti-self-dual $PU(N)$ connection A' in the associated $PU(N)$ bundle of E.

The resulting moduli space of solutions (E, ι, A') depends on the Riemannian manifold X and the choice of the class $w = c_1(W)$ in $H^2(X; \mathbb{Z})$. The topology of E is detemined by w and the the *instanton number*, defined now as an appropriately normalized Pontryagin number of the associated $PU(N)$ bundle. With a standard normalization, the instanton number k is not an integer but satisfies a congruence

$$k = -\left(\frac{N-1}{2N}\right) w \cdot w \pmod{\mathbb{Z}}.$$

We write $M_{N,k}(X)^w$ for this moduli space of anti-self-dual connections with instanton number k. It leads to polynomial invariants,

$$q^w_{X,k} : \mathbb{A}_{d(k)}(X) \to \mathbb{Q},$$

generalizing the $q_{X,k}$ (defined using the $N = 2$ moduli spaces) and we can combine these again as

$$q^w_X = \bigoplus_k q^w_{X,k} : \mathbb{A}(X) \to \mathbb{Q}.$$

This extra generality is introduced not so much for its own sake, but because it serves to avoid the difficulty that was mentioned in our discussion of the polynomial invariants above. The difficulty is the possible presence of *reducible* connections A in E. We will say that E or w is *admissible* if there is an integer homology class σ in $H_2(X)$ such that

$$(w \cdot \sigma) \text{ is prime to } N. \tag{4}$$

The relevance of admissibility is in the following result.

Proposition 2.1. *If E is admissible and $b^2_+(X) > 0$, then the moduli space of anti-self-dual connections contains no reducible solutions, for a generic metric on X. The same is true for a generic path of metrics if $b^2_+(X) > 1$.*

Remark 2.2. A prototype which captures part of this is the more elementary statement, that if E is a $U(N)$ bundle on a closed, oriented 2-manifold and the degree of E is prime to N, then the associated $PU(N)$ bundle E' admits no reducible flat connections. In this way, reducible solutions can be avoided, and the invariants q^w_X can be generalized to higher-rank bundles with admissible w.

Even for $N = 2$, introducing w is helpful, and necessary too for instanton homology.

Remark 2.3. In the case $N = 2$, the group $PU(2)$ is $SO(3)$ and elements of the moduli space $M_{2,k}(X)^w$ give rise to anti-self-dual $SO(3)$ connections with second Stiefel-Whitney class w mod 2. There is a distinction between the two setups however. The automorphisms of the pair (E, ι) are the bundle automorphisms of E that have determinant 1 on each fiber. If $H^1(X; \mathbb{Z}/2)$ is non-zero, then not every automorphism of the associated $PU(2)$ bundle E' lifts to a determinant-1 automorphism of E. The moduli space of anti-self-dual $SO(3)$ or $PU(2)$ connections is the quotient of $M_{2,k}(X)^w$ by an action of the finite group $H^1(X; \mathbb{Z}/2)$.

2.4. Non-vanishing for the polynomial invariants. While Taubes' results inform us that $M_{N,k}(X)$ is non-empty for large enough k, one can now ask a different question whose answer reflects the non-triviality of the moduli space in a different way: one can ask whether the Donaldson invariants of X are non-zero; or equivalently, is the fundamental class $[M_{N,k}(X)]$ non-zero? Donaldson proved the following non-vanishing theorem for the polynomial invariants q_X arising from the $SU(2)$ instanton moduli spaces.

Theorem 2.4 (Donaldson, [**9**]). *If X is the smooth 4-manifold underlying a simply-connected complex projective algebraic surface, then the polynomial invariants $q_{X,k}$ are non-zero for all sufficiently large k. In particular, $q_{X,k}(h^{d/2})$ is non-zero, where h is the hyperplane class and $d = d(k)$ is the dimension of the moduli space.*

This result moves us from the simple non-emptiness of a moduli space to non-triviality in homology. Donaldson's proof uses the fact that, for the Kähler metrics adapted to the complex-algebraic structure, the moduli spaces of instantons can be identified with moduli spaces of stable holomorphic bundles, which are quasi-projective varieties. The non-vanishing eventually derives from the positivity of intersections in complex geometry.

The theorem generalizes to $U(2)$ bundles and the corresponding invariants $q^{w}_{X,k}$ for non-zero w. The authors believe that, using later results and constructions from [**19**], [**36**] and [**25**], the restriction to $N = 2$ can be dropped, and that the hypothesis that X is simply-connected is also unnecessary.

2.5. Non-vanishing for symplectic 4-manifolds. The non-vanishing theorem for algebraic surfaces was proved when Donaldson's invariants were first introduced in [**9**]. A class of 4-manifolds that is in many ways closely related are the *symplectic* 4-manifolds, i.e. those which carry a closed 2-form ω for which $\omega \wedge \omega$ is a volume form. Although the original proof of Theorem 2.4 does not extend to the symplectic case, the more general theorem does hold:

Theorem 2.5. *The non-vanishing statement of Theorem 2.4 continues to hold for the larger class of symplectic 4-manifolds, with the role of the hyperplane class h now played by the de Rham class $[\omega]$ of the symplectic form.*

Part of the history of this result is as follows. A non-vanishing theorem was proved by Taubes [**46**] for the Seiberg-Witten invariants of symplectic 4-manifolds, and the above theorem should then follow from Witten's conjecture [**49**] relating the Donaldson invariants to the Seiberg-Witten invariants. A weakened version of Witten's conjecture has been proved by Feehan and Leness [**14**], building on ideas of Pydstrigach and Tyurin, and this work can be used to deduce Theorem 2.5 for a large class of symplectic 4-manifolds. The general version of Theorem 2.5 was later proved more cleanly, and without use of the Seiberg-Witten invariants: an argument was outlined in [**29**] and a variant is given in [**41**]. These later proofs make use of another theorem of Donaldson, on the existence of Lefschetz pencils for symplectic 4-manifolds [**10**].

3. Instanton homology for 3-manifolds

3.1. Formalities, and non-vanishing. The instanton homology groups of an oriented 3-manifold Y arise naturally when one seeks to understand the Donaldson invariants of a 4-manifold X which is decomposed as a union of two manifolds with common boundary Y:

$$\begin{aligned} X &= X_+ \cup_Y X_- \\ \partial X_+ &= Y \\ \partial X_- &= -Y. \end{aligned} \tag{5}$$

(We use $-Y$ to denote Y equipped with the opposite orientation.) There are several variants of Floer's construction depending on, among other choices, the gauge group and the coefficient ring, and some variants are applicable only to certain 3-manifolds (such as homology spheres) or only allow certain bundles.

Floer's first construction worked only for homology 3-spheres and structure group $SU(2)$. To each oriented connected homology 3-sphere Y it gave a finitely-generated abelian group $I(Y)$. The simplest property of this invariant is that the instanton homologies of Y and $-Y$ are related as the homology and cohomology of a complex, so that in particular there is a perfect pairing

$$I(Y) \otimes I(-Y) \to \mathbb{Z}. \tag{6}$$

If we work with rational coefficients, as we often will, then these are dual vector spaces.

When a connected, oriented 4-manifold X is decomposed as in (5) where Y is a homology sphere and $b^2_+(X_\pm) > 0$ then the Donaldson invariant q_X (3) can be expressed in terms of the *relative invariants* of the two pieces $X_\pm$. These relative invariants take the form of linear maps

$$q_{X_+} : \mathbb{A}(X_+) \to I(Y)$$
$$q_{X_-} : \mathbb{A}(X_-) \to I(-Y)$$

and the Donaldson invariant of the closed manifold X is expressed using the pairing (6) as

$$q_X(z) = \langle q_{X_+}(z_+), q_{X_-}(z_-) \rangle, \tag{7}$$

where $z = i_+(z_+)i_i(z_-)$ and $i_\pm : \mathbb{A}(X_\pm) \to \mathbb{A}(X)$ arise from the inclusion maps. This pairing formula has a straightforward corollary:

PROPOSITION 3.1. *If $q_X \neq 0$ for some 4-manifold X, and a homology 3-sphere Y can be placed into X in such a way that X is decomposed into two pieces, each with $b^2_+ > 0$, then it must be that $I(Y) \otimes \mathbb{Q}$ is non-zero.*

REMARK 3.2. Unlike ordinary homology, the instanton homology groups are not $\mathbb{Z}$-graded. The version $I(Y)$ discussed here has a cyclic grading by $\mathbb{Z}/8$. Some versions we will encounter later have not grading at all, as they arise as the homology $\ker(d)/\mathrm{im}(d)$ for a differential d on an ungraded abelian group rather than a chain complex.

3.2. Sketch of the construction. The basic idea for instanton Floer homology [**15**] can be motivated by thinking of solutions to the anti-self-duality equations on a closed 4-manifold X, decomposed as above, but with a Riemannian metric containing long cylinder $[-L, L] \times Y$. By means of a gauge transformation on this cylinder we can assume that a connection A in $E \to [-L, L] \times Y$ is pulled back from path of connections $B(t)$, for $t \in [-L, L]$, in $E \to Y$. The anti-self-duality equation for A becomes the equation

$$\frac{\partial B}{\partial t} + *F_B. \tag{8}$$

In particular, translationally invariant solutions to the ASD equation (i.e. solutions with $\frac{\partial B}{\partial t} = 0$) are flat connections B on Y, so that $F_B = 0$. A key observarion in [**15**] is that the above equation for a path $B(t)$ is formally the downward gradient flow for a functional – the *Chern-Simons functional* – on a space of connections on the 3-manifold. To see this, consider for simplicity a trivial bundle on a 3-manifold Y. We write a connection B as sum of the connection Γ coming from a trivialization and a 1-form with values in the Lie algebra $\mathfrak{su}(N)$:

$$B = \Gamma + b, \text{ where } b \in \Omega^1(Y) \otimes \mathfrak{su}(N).$$

In this form, the Chern-Simons function is given by

$$CS(B) = -\frac{1}{2}\int_Y \mathrm{tr}(b \wedge db + \frac{1}{3} b \wedge b \wedge b).$$

The first variation of CS is given by

$$\begin{aligned}\frac{d}{dt}CS(B+t\beta)|_{t=0} &= -\int_Y \operatorname{tr}(\beta \wedge (db + \frac{1}{2} b \wedge b)) \\ &= -\int_Y \operatorname{tr}(\beta \wedge F_B),\end{aligned}$$

so the stationary points are flat connections, $F_B = 0$. If Y is given a Riemannian metric, then the standard inner product on $su(N)$-valued one forms can be written

$$\langle \alpha, \beta \rangle = -\int_Y \operatorname{tr}(\alpha \wedge *\beta).$$

With respect to this inner product, the gradient of CS at the connection B is $*F_B$, which verifies that the equation (8) is indeed the downward gradient flow.

Floer's construction applies the ideas of Morse theory to the Chern-Simons functional. In the case of a finite-dimensional compact manifold B carrying a Morse function f, the ordinary homology of B can be computed as the *Morse homology* of f. This is the homology of a complex whose generators correspond to critical points of f and whose boundary map records intersection numbers between ascending a descending submanifolds. After a generic perturbation, these intersection numbers can be interpreted as counting gradient flow lines between critical points of adjacent index.

When this framework from finite dimensions is applied to the Chern-Simons functional on the space of gauge-equivalence classes of connections on Y, the relation with ordinary homology is lost. The "instanton homology" which results is something new. It is the homology of a complex whose generators are critical points of the (perturbed) Chern-Simons functional and whose boundary map counts gradient flow lines. In line with the discussion above, the critical points are flat $SU(N)$ connections on Y (corresponding to representations of $\pi_1(Y)$ in $SU(N)$) and the boundary map counts solutions of the anti-self-duality equations on the cylinder.

Floer's original construction ignores the trivial connection, and the assumption that Y is a homology sphere is used in a crucial manner to prove that $I(Y)$ is independent of the choice of metric on Y and the perturbation. This restriction on Y also means that the non-trivial representations of $\pi_1(Y)$ in $SU(2)$ are all irreducible. It follows then from the construction of $I(Y)$ as a Morse homology that if there are no irreducible representations, i.e. no critical points for CS used in the construction of $I(Y)$, then $I(Y)$ is trivial. So Proposition 3.1 has the following straightforward corollary:

Corollary 3.3. *If $q_X \neq 0$ for some 4-manifold X, and a homology 3-sphere Y can be placed into X in such a way that X is decomposed into two pieces, each with $b^2_+ > 0$, then there is a non-trivial homomorphism $\rho : \pi_1(Y) \to SU(2)$.*

In particular (from the case that Y is S^3), if a 4-manifold admits a connected sum decomposition $X = X_1 \# X_2$ where $b^2_+(X_i) > 0$, then $q_X = 0$, which is an earlier vanishing theorem due to Donaldson.

As a tool to prove existence of non-trivial homomorphisms, the Corollary is useful, but it is not a completely general tool. For example, it is not known which homology 3-spheres can be embedded in complex algebraic surfaces or in symplectic 4-manifolds. (See Theorem 2.4 and Theorem 2.5.)

In a positive direction, it is shown [**11, 13, 28**] that 3-manifolds carrying a *taut foliation* can always be embedded in a symplectic 4-manifold. The argument combines deep and difficult work of Eliashberg-Thurston [**12**] and Giroux [**20**]. One can deduce:

Corollary 3.4. *If the homology 3-sphere Y admits a taut foliation, then there is a non-trivial homomorphism $\rho : \pi_1(Y) \to SU(2)$.*

As indicated earlier (Question 1.5), the general case of a 3-manifold with non-zero fundamental group remains open.

3.3. Using $U(N)$ bundles. The instanton homology we have just described is the first version which Floer defined. Rather than work with homology spheres where the unique reducible connection can be excluded, Floer observed that there is an alternative setup for 3-manifolds with $b_1(Y) \neq 0$ where reducible flat connections can be avoided entirely. One works with 3-manifolds that carry an $SO(3)$ bundle whose the second Stiefel-Whitney class $\omega \in H^2(Y; \mathbb{Z}/2)$ has non-zero evaluation on some integral homology class. Almost the same, as we did for Donaldson's invariants in section 3.3, we may fix a line bundle $W \to Y$ and work with triples (E, ι, A'), where E a $U(2)$ bundle, ι is an isomorphism $\Lambda^2 E \to W$, and A' is a connection in the associated bundle $PU(2)$. We define *admissibility* for W – or equivalently for its first Chern class $w = c_1(W)$ – just as in the 4-dimensional situation, equation (4). In the admissible case there are no reducible flat connections A'. (See Remark 2.2.)

We arrive at instanton homology groups $I^w(Y)$, labeled by admissible classes w. A pairing formula similar to equation 7 holds in this context. Suppose again that X is decomposed along Y as in (5), and suppose now that v is a class in $H^2(X; \mathbb{Z})$ whose restriction, w, to Y is also admissible. Then we have relative invariants,

$$q^{v_+}_{X_+} : \mathbb{A}(X_+) \to I^w(Y)$$
$$q^{v_-}_{X_-} : \mathbb{A}(X_-) \to I^w(-Y)$$

and a pairing formula,

$$q^v_X(z) = \langle q^{v_+}_{X_+}, q^{v_-}_{X_-} \rangle, \tag{9}$$

where $z = i_+(z_+)\iota_-(z_-)$ as before.

Along the same lines as Proposition 3.1, we now have:

Proposition 3.5. *Let Y be given, and let w be an admissible class on Y. Suppose that Y can be embedded as a separating hypersurface in X in such a way that the class w extends to a class $v \in H^2(X; \mathbb{Z})$, and suppose that the Donaldson invariant q^v_X is non-zero. Then the instanton homology group $I^w(Y) \otimes \mathbb{Q}$ is non-zero.*

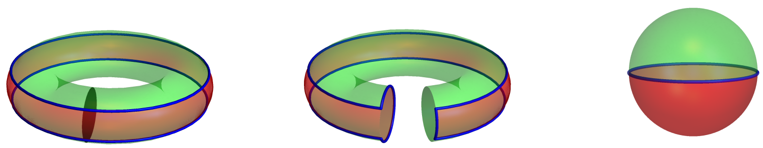
As in the case of a homology 3-sphere (see Corollary 3.4), one can deduce that if Y admits a taut foliation, then $I^w(Y)$ is non-zero for any admissible w, and there exists a representation $\rho : \pi_1(Y) \to SO(3)$ with $w_2(\rho) = w$ mod 2. Unlike the case of homology 3-spheres however, irreducible 3-manifolds with non-zero Betti number *all* carry taut foliations, by a deep existence result due to Gabai [**17**]. So we have:

Corollary 3.6. *If Y is irreducible with $b_1 \neq 0$ and w is any admissible class, then $I^w(Y) \otimes \mathbb{Q}$ is non-zero.*

In this way we arrive at an existence result for representations that is sufficiently general to deduce the necessary and sufficient condition, Theorem 1.6.

4. Sutured manifolds

In [**29**], the authors found a much more efficient proof of Corollary 3.6 and Theorem 1.6. The original arguments outlined above used the existence of a taut foliation,



(A) (B) (C)

FIGURE 6. An example of sutured manifold decomposition: a sutured solid torus is decomposed along an embedded disk. The new sutured manifold (center) is isomorphic to a standard sutured ball (right). Red and green indicate R_+ and R_-, while blue curves are sutures.

which had been proved by Gabai [**17**] using his theory of sutured manifolds. The later argument in [**29**] uses Gabai's sutured manifold theory more directly. This strategy is inspired in part by the construction of sutured Heegaard Floer homology by Juhasz [**22**] and its precursors in the work of Ni and Ghiggini [**35, 18**]. On the gauge theory side, the non-vanishing theorem for Donaldson invariants and the difficult proof of the relation between Donaldson and Seiberg-Witten invariants is replaced by work of Munoz [**34**] in computing of the instanton homology groups of $S^1 \times \Sigma_g$. We will explain how some of this works, beginning with Gabai's work.

4.1. Sutured manifold decompositions. An important idea in 3-manifold topology going back to Haken and Waldhausen [**21, 48**] and further developed by Gabai is that of surface decomposition: cutting a 3-manifold along a surface may result in a simpler 3-manifold. In order to organize a sequence of surface decompositions of manifolds with boundary, Gabai defined a notion of *sutured manifold.* This is an oriented 3-manifold with boundary, Y, together with a decomposition of its boundary into two parts,

$$\partial Y = R_+ \cup R_-,$$

intersecting along a union of simple closed curves $\gamma \subset \partial Y$. These simple closed curves are the *sutures.* The simplest example is a 3-ball, with its boundary divided into upper and lower hemispheres, meeting at the equator (Figure 6c). We can always orient γ by first orienting R_+ as the boundary of Y and then orienting γ as the boundary of R_+.

A *decomposing surface* S for a sutured manifold (Y, γ) is an oriented embedded surface $S \subset Y$ with $\partial S \subset \partial Y$. It is required that S and ∂Y meet transversally, so that ∂S a union of simply closed curves in ∂Y; and each of these is required to either meet the sutures γ transversally, or to coincide with a component of γ as an oriented 1-manifold. If a component of ∂S is disjoint from γ, then it is required that this circle does not bound a disk in $R_\pm$ nor a disk in S. Given such a decomposing surface, one obtains a new sutured manifold Y' by cutting Y open along S and smoothing the corners. The new decomposition of $\partial Y'$ as $R'_+ \cup R'_-$ is defined by setting

$$R'_+ = R_+ \cup S_+ \qquad R'_- = R_- \cup S_-$$

where S_+ is the copy of S in $\partial Y'$ picked out by the oriented normal to $S \subset Y$. The process of forming (Y', γ') from (Y, γ) in this way is called a *sutured manifold decomposition.* (See Figure 6.)

The following is a slightly special case of one of Gabai's central results about sutured manifolds.

Theorem 4.1 ([**17**]). *Let Y be a closed irreducible 3-manifold, regarded as a sutured manifold without boundary. Suppose that the Betti number $b_1(Y)$ is non-zero. Then we can find a sequence of sutured manifolds, starting with Y, each obtained from the previous one by sutured manifold decomposition, and ending with a disjoint union of 3-balls with one equatorial suture each:*

$$(Y,\varnothing) = (Y^0,\gamma^0) \rightsquigarrow (Y^1,\gamma^1) \rightsquigarrow \cdots \rightsquigarrow (Y^k,\gamma^k) = \coprod_1^m (B^3, \text{equator}). \tag{10}$$

Furthermore, for the first decomposition in the sequence, the decomposing surface $S \subset Y$ can be chosen to be any connected genus-minimizing surface: a surface that achieves the minimum genus among all oriented surfaces in the same homology class. Conversely, the genus-minimizing property for the first cut is a necessary condition for the existence of such a decomposition ending with standard 3-balls.

This result provides a broad framework for proving existence results for structures on an irreducible 3-manifold Y, by starting with existence (of whatever structure) on the trivial (Y^k,γ^k), and working back up to Y. In Gabai's work this framework is used to prove the existence of taut foliations, and in our context it can be used to prove that the instanton homology $I^w(Y)$ is non-zero (Corollary 3.6). What needs to be done is:

(1) extend the definition of instanton homology $I^w(Y)$ to the case of sutured manifolds;
(2) show that the rank of the instanton homology of sutured manifolds is monotone decreasing in any sequence of decompositions such as (10);
(3) show that the instanton homology has non-zero rank for the disjoint union of sutured balls.

Once one has item (1) of the above three, the remaining pieces fall into place quite easily. The non-trivial ingredient needed for the definition in (1) is what we turn to next.

4.2. Munoz' computation of $I^w(S^1 \times \Sigma)$ and its consequences. In [**34**] Munoz gave a description of the instanton homology $I^w(S^1\times\Sigma)$, where Σ is a surface of genus $g \ge 1$ and w is the 2-dimensional cohomology class Poincaré dual to $S^1 \times \{\text{point}\}$.

In general, ordinary homology classes h in a 3-manifold Y give rise to operators $\hat{h}$ on $I^w(Y)$. This can be seen as arising naturally from the formalism of Donaldson's invariants, as follows. We working with coefficients in a field and consider (roughly speaking) the relative Donaldson invariants of the 4-manifold $Z = [-1,1]\times Y$ as defining a linear map

$$\mathbb{A}(Z) \to I^w(\partial Z)\otimes\mathbb{Q},$$

or equivalently

$$\mathbb{A}(Y) \to I^w(-Y)\otimes I^w(Y)\otimes\mathbb{Q}.$$

The algebra $\mathbb{A}(Y)$ contains $H_*(Y)$ as a linear subspace, and $I^w(-Y)$ is dual to $I^w(Y)$, so we obtain the operators we seek:

$$\hat{h} : I^w(Y)\otimes\mathbb{Q} \to I^w(Y)\otimes\mathbb{Q}, \qquad h\in H_*(Y).$$

If we have two homology classes h_1 and h_2 then the corresponding operators commute in the graded sense.

Munoz' work can be used to determine the spectrum, and indeed joint spectrum, of the operators coming from $H_*(S^1 \times \Sigma)$. Let $s \in H_2(S^1 \times \Sigma)$ be the homology class of $(\text{point}) \times \Sigma$ and $y \in H_0(S^1 \times \Sigma)$ be the homology class of a point.

Theorem 4.2. *Then the simultaneous eigenvalues of the action of $\hat{s}$ and $\hat{y}$ on $I^w(S^1 \times \Sigma_g) \otimes \mathbb{C}$ are the pairs of complex numbers $\big(i^m(2k),\ (-1)^m 2\big)$ for all the integers k in the range $0 \le k \le g-1$ and all $m = 0, 1, 2, 3$. Here i denotes $\sqrt{-1}$.*

Furthermore the generalized eigenspace corresponding to $(2g-2, 2)$ is one-dimensional hence simple.

As a corollary of this one can deduce an result for a general 3-manifold Y. If we have a 2-dimensional homology class s represented by a connected surface S in Y with genus $g > 1$, and a point y thought of as a 0-dimensional homology class, then the following holds.

Theorem 4.3. *For any admissible class $w \in H^2(Y)$ with $w \cdot s = 1$, the simultaneous eigenvalues of the action of $\hat{s}$ and $\hat{y}$ on $I^w(Y) \otimes \mathbb{C}$ are **contained** in the pairs that arise in the case of the product manifold $S^1 \times S$. That is, they are pairs of complex numbers*

$$\big(i^m(2k),\ (-1)^m 2\big)$$

where k is in the range $0 \le k \le g-1$.

We return now to a *sutured* manifold (Y, γ), which we shall suppose satisfies the condition that Juhasz [**22**] calls *balanced*, namely we require that $\chi(R_+) = \chi(R_-)$, that no component of Y is a closed 3-manifold, and that every component of ∂Y contains a suture. The first of these conditions holds automatically for the sutured manifolds (Y^i, γ^i) in Theorem 4.1, and one can arrange that the other two mild conditions hold from Y^2 onwards. For such balanced sutured manifold one can form (not uniquely) a *closure* $\bar{Y}$ as follows.

Choose an oriented connected surface T whose boundary admits an orientation-reversing diffeomorphism $\partial T \to \gamma$. Extend this diffeomorphism to a diffeomorphism ϕ of $[-1, 1] \times \partial T$ with a tubular neighborhood (in ∂Y) of γ. Then form the new 3-manifold with boundary

$$\tilde{Y} = Y \cup_\phi [-1, 1] \times \partial T.$$

Note that our assumptions imply that the boundary of $\tilde{Y}$ has two connected components $\bar{R}_\pm$ formed from $R_\pm$ and $\pm 1 \times T$. The balanced assumption implies that $\bar{R}_+$ $\bar{R}_-$ have the same Euler characteristic, and since these are connected these surfaces are diffeomorphic. Choosing a diffeomorphism ψ we construct an closed, connected 3-manifold

$$\bar{Y} = \tilde{Y}|_\psi$$

where the two boundary components are glued together using ψ. The original sutured manifold Y can be obtained from the closed manifold $\bar{Y}$ by a sequence of two sutured manifold decompositions, decomposing first along $\bar{R}$ and then along the annuli $[-1, 1] \times \partial T$.

Formed in this way, the closure $\bar{Y}$ contains a distinguished non-separating connected surface $\bar{R}$ carrying a homology class $r \in H_2(\bar{Y})$. Let w be an admissible class with $w \cdot r = 1$, and consider the application of Theorem 4.3 to the operators

$$\hat{r}, \hat{y} : I^w(\bar{Y}) \otimes \mathbb{C} \to I^w(\bar{Y}) \otimes \mathbb{C}.$$

According to the theorem, the integers that arise in the spectrum of $\hat{r}$ are bounded above by $2g - 2$. We make the following defintion:

Definition 4.4. The *sutured instanton homology* of the sutured manifold (Y, γ), written $SHI(Y, \gamma)$ is defined to be the simultaneous eigenspace for the pair $(2g-2, 2)$ for the operators $(\hat{r}, \hat{y})$ on $I^w(\bar{Y}) \otimes \mathbb{Q}$ for any closure $\bar{Y}$.

In showing that this definition is good (i.e. is independent of the choice of closure) an important role is played by last clause of Theorem 4.2. Note in particular that it tells us that the dimension of $SHI(Y, \gamma)$ is 1 in the case that the sutured manifold is $(B^3, \text{equator})$ or a union of such, for in this case we can take the closure to be $S^1 \times \bar{R}$.

Returning to the three-step plan (1)–(3) from the end of section 4.1, we see that what remains for a proof of the non-vanishing theorem, Corollary 3.6, is item (2) there. That is, one must show that if (Y, γ) is decomposed along S to obtain (Y', γ'), then $SHI(Y', \gamma')$ has rank no larger than the rank of $SHI(Y, \gamma)$. The idea of the proof here is to construct a particular closure $\bar{Y}$ for Y so that S becomes a closed surface $\bar{S}$, and consider the operators $\hat{s}$ that it gives rise to on $SHI(Y, \gamma)$. One then seeks to identify $SHI(Y', \gamma')$ with an eigenspace of the operator $\hat{s}$, thus exhibiting it as a subspace of $SHI(Y, \gamma)$.

Note that in nearly all cases, this line of proof gives a considerable strengthening of the non-vanishing theorem, Corollary 3.6. As a very simple example:

Corollary 4.5. *Let Y be an irreducible 3-manifold containing a non-separating connected surface S of genus at least 2 which is genus-minimizing in its homology class s. Then $I^w(Y)$ has rank at least 4, for every admissible w with $w \cdot s = 1$.*

Proof. Consider the operators $(\hat{s}, \hat{y})$ again. The proof of non-vanishing shows that the simultaneous eigenspace for the pair $(2g-2, 2)$ for $\hat{s}$ is non-zero. For formal reasons, the eigenspaces of the pair $(i^r(2g-2), (-1)^r 2)$ are all of the same dimension, and if $2g-2$ is non-zero then these four pairs are distinct. □

4.3. Sutured manifolds and knots. As Juhasz observed in [**22**], one can use the sutured manifold formalism to define an instanton homology for knots. For simplicity, let us consider a classical knot K in S^3, and let Y be the "knot complement": the manifold with torus boundary obtained by removing from S^3 an open tubular neighborhood of K. Let γ be the union of two disjoint meridional curves on ∂Y, with opposite orientations. In this way we associate a sutured manifold (Y, γ) to $K \subset S^3$, and the sutured instanton homology of (Y, γ) is an invariant, which we can call the knot instanton homology $KHI(K)$.

To understand this knot invariant better, one can describe an explicit closure $\bar{Y}$ in this case, using an annulus for the auxiliary surface T. To describe this closure, consider the 3-torus T^3 as $S^1 \times T^2$, and let $\lambda \subset T^3$ be the circle $S^1 \times \{\text{point}\}$. Remove an open neighborhood of λ and glue the resulting torus boundary to the boundary of knot complement Y. The gluing should be chosen so that the longitudinal curves in ∂Y are glued to the meridional curves of λ and vice versa. In this closed 3-manifold $\bar{Y}$, the distinguished surface $\bar{R} \subset \bar{Y}$ is a standard torus disjoint from Y. Note if we are given a Seifert surface for K (an oriented surface with boundary a longitude of K), then we obtain a closed surface $S \subset \bar{Y}$ as the union of the Seifert surface and the punctured torus $T^2 \setminus \{\text{point}\}$.

Following [**22**], one can show that the invariant KHI can be used to detect the unknot:

Proposition 4.6. *For a classical knot K, the dimension of $KHI(K)$ is greater than or equal to 1, with equality if and only if K is the unknot.*

Proof. Unwrapping the definitions we see that $KHI(K)$ is the simultaneous eigenspace for the eigenvalues $(0, 2)$ of the operators $(\hat{r}, \hat{y})$. Since $\hat{r}$ has genus 1, this is simply the eigenspace of $\hat{y}$ for the eigenvalue 2. In the case of the unknot, $\bar{Y}$ is a 3-torus, the instanton homology $I^w(\bar{Y})$ has rank 2, and the $+2$ and -2 eigenspaces of $\hat{y}$ are both 1-dimensional. For a non-trivial knot, the dimensions are larger, by the argument of Corollary 4.5, for there is a genus-minimizing surface $S \subset Y$ of genus at least 2, obtained from a Seifert surface for K as described above. □

Since instanton homology is defined ultimately in terms of flat connections, one can use the above proposition to deduce that, if K is a non-trivial knot, then the $SU(2)$ or $SO(3)$ representations of the knot group $\pi(K)$ is strictly larger than the case of the unknot. In this way, one can derive Theorem 1.2 from the first section.

Remark 4.7. In the Heegaard Floer homology setting, Juhasz's construction recovers the simplest version of the Heegaard knot homology of Ozsváth-Szabó and Rasmussen [**37, 39**]. On the basis of the few existing calculations in the instanton case, one can conjecture that the rank of $SHI(K)$ is equal to the rank of the Heegaard knot homology group.

Remark 4.8. The generalized eigenspaces of the operator defined by S give a direct sum decomposition of $KHI(K)$. There is also a $\mathbb{Z}/2$ grading, so it makes sense to compute the Euler characteristic of the summands. In this way one recovers the coefficients of the Alexander polynomial, just as one does in the case of the Heegaard knot homology.

Remark 4.9. There is work of Daemi and Xie [**6**] on generalizing the sutured instanton homology *SHI* by using the gauge groups $SU(N)$. The essential step is in establishing an appropriate replacement for the results of Munoz.

5. Instanton homology for knots, links and spatial graphs

We have explained above that Theorem 1.6, which asserts the existence of a non-abelian representation of the fundamental group of a 3-manifold in $SO(3)$, can be seen as a corollary of a non-vanishing theorem for the instanton Floer homology group $I^w(Y)$ (Corollary 3.6). The non-vanishing theorem can be proved using sutured manifold decompositions, as outlined in section 4; and from these results about sutured manifolds, we deduced results about representations of $\pi(K)$ for knots and links such as Proposition 4.6 via the knot homology group $KHI(K)$.

There is a more direct approach to defining an instanton homology group for knots, which we will outline next. Although the invariant defined by this approach turns out to be isomorphic to $KHI(K)$, the alternative approach plays dividends in its extra flexibility.

5.1. Instanton homology for orbifolds. Our strategy is to define an instanton Floer homology group for a certain class or orbifolds, and the class we have in mind are closed, oriented 3-dimensional orbifolds O whose singular set is a knot or link and whose non-trivial local stabilizers are all $\mathbb{Z}/2$. Thus our class includes the orbifolds that appear in the statement Theorem 1.3 in the introduction. The material here is drawn from [**30**].

There is no particular difficulty in studying bundles, connections and the anti-self-duality equations on orbifolds. To avoid reducible connections we need to work again with $U(2)$ bundles with fixed determinant, as in subsection 3.3. At points of the singular set, where the stabilizer is $\mathbb{Z}/2$, the local model will be the quotient of a smooth $U(2)$ bundle over the ball, and we ask that the action on the fiber of the bundle be by the element

$$\tau = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The first Chern class of such an orbifold bundle E is a class dual to a relative 1-cycle w (thought of geometrically as a 1-manifold in the smooth part of O with possible endpoints on the singular part).

As long as w is admissible one can define an instanton homology group $I^w(O)$ much as before. The critical points of the Chern-Simons functional are flat $PU(2)$ connections on the complement of the singular set whose monodromy around the meridional links has order 2 and whose Stiefel-Whitney class is w mod 2.

Recall that, for a classical link K in S^3, we write $\mathrm{Orb}(S^3, K)$ for the orbifold whose singular set is K. A first Chern class w on $\mathrm{Orb}(S^3, K)$ will be admissible if it is dual to an arc joining two components of K, because such a w evaluates to 1 on an oriented surface separating the components. To achieve admissibility in general, we adopt the following device. Given classical knot or link K, we form a new link by taking the union of K with a small meridional loop L, linking K at chosen point $x \in K$. We take w to be the admissible class dual to an arc joining the new loop L to K. We may then define:

$$I^\natural(K) = I^w(K \cup L). \tag{11}$$

If K has more than one component, then the choice of the point x may be material, but we still omit x from the notation.

To understand the definition a little, observe that when K is an unknot, the union $K \cup L$ is a Hopf link. It is not hard to verify in this case that there is exactly one critical point of the Chern-Simons functional on the corresponding orbifold with the correct determinant w. This flat connection corresponds to the Klein 4-group representation of the fundamental group of the complement, $\pi(K \cup L) = \mathbb{Z} \oplus \mathbb{Z}$. This unique critical point is the generator for $I^\natural(K)$. Having just one generator, $I^\natural(K)$ is infinite cyclic.

Although the definition is different, the same techniques of cut-and-paste topology that are used in the construction of sutured instanton Floer homology can be used to show that the two approaches yield the same result, for knots:

Proposition 5.1. *For a knot K, the homology groups $KHI(K)$ and $I^\natural(K)$ are isomorphic.*

As a trivial corollary, the orbifold version also detects knottedness:

Corollary 5.2. *For a knot K in S^3, the rank of $I^\natural(K)$ is at least 1, with equality if and only K is the unknot.*

5.2. Khovanov homology. An advantage of the orbifold approach to the definition of $I^\natural(K)$ is that it allows a straightforward approach to functoriality. A *cobordism* between classical links K_0 and K_1 is an embedded surface Σ in $[0, 1] \times S^3$, meeting the boundary transversely in K_0 and K_1 at the two ends. Without any requirement of orientability, such a cobordism gives rise to a homomorphism $I^\natural(K_0) \to I^\natural(K_1)$.

This functoriality is the starting point in making an unexpected connection between this instanton homology and a knot homology group from a quite different stable, namely the *Khovanov homology* groups introduced in [**23**]. The Khovanov homology $Kh(K)$ for a classical knot or link is a "categorification" of the Jones polynomial. It has a definition which is entirely algebraic, and eventually elementary, but $Kh(K)$ and its generalizations have turned out to have deep connections with geometry, in several directions. In our particular context, we have the following result:

Theorem 5.3. *([**30**]) For a classical knot or link K, there is a spectral sequence whose E_2 page is the (reduced variant of) the Khovanov homology of K and which abuts to the orbifold instanton homology, $I^\natural(K)$.*

Like $I^\natural(K)$, the reduced Khovanov homology has rank 1 if K is the unknot. From Corollary 5.2 we and the existence of the spectral sequence, we therefore obtain:

Corollary 5.4. *For a knot K in S^3, the rank of the reduced Khovanov homology is at least* 1*, with equality if and only K is the unknot.*

It is an open question whether the Jones polynomial itself is an unknot-detector. Although many geometric techniques can be used to characterize the unknot algorithmically (starting with Haken's work in [**21**]), the above corollary stands somewhat apart, because of the origins of Khovanov homology in quantum algebra and representation theory.

An interesting avenue to pursue is to replace $U(2)$ in the orbifold setup with $U(N)$ and to explore the relationship to generalizations such as Khovanov-Rozansky homology [**24**]. See [**5, 50**].

5.3. Instanton homology for spatial graphs. We return to the material of section 1.4, to consider a trivalent spatial graph $K \subset S^3$. As we did for knots and links, we can apply instanton Floer homology to the associated orbifold $O = \mathrm{Orb}(S^3, K)$. Allowing trivalent vertices in K leads to new issues, related in particular to the possibility of the Uhlenbeck bubbling phenomenon occurring at orbifold points corresponding to vertices of K. In order to have a well-defined instanton homology, it turns out to be necessary to use a ring of coefficients of characteristic 2.

Following this line, the authors defined in [**26**] an invariant of trivalent spatial graphs K which takes the form of a $\mathbb{Z}/2$ vector space $J^\sharp(K)$. This variant of instanton homology arises from a Chern-Simons functional whose set of critical points can be identified with the space of $SO(3)$ representations $\mathcal{R}(K)$ considered at (1). (In particular, this is essentially an $SO(3)$ gauge theory, not the type of $U(2)$ gauge theory used in the definition of $I^w(Y)$ before.)

Once again, by reducing the question to one about the instanton homology of a sutured manifold (essentially the complement of K) one can prove a non-vanishing theorem for graphs that are spatially bridgeless in the sense of section 1.4:

Theorem 5.5. *If $K \subset \mathbb{R}^3$ is a spatially bridgeless trivalent graph, then the instanton homology group $J^\sharp(K)$ is non-zero.*

An immediate corollary is that the space of representations $\mathcal{R}(K)$ is non-empty, which is the statement of Theorem 1.11 in the introduction.

As mentioned in section 1.4, the space of $SO(3)$ representations $\mathcal{R}(K)$ contains the set of representations, $\pi(K) \to V_4$, into the Klein 4-group, which are in one-to-one correspondence with Tait colorings of K. It is difficult to compute $J^\sharp(K)$, but an examination of the simplest examples prompts this question.

Question 5.6. *For a spatial trivalent graph K that is* planar *(that is, embedded in a plane $\mathbb{R}^2$ in $\mathbb{R}^3$), is it the case that the dimension of $J^\sharp(K)$ is equal to the number of Tait colorings?*

It is known that, if the answer is no, then a minimal counterexample can have no bigons, triangles or squares [**26**]. Various equivalent forms of the question are given in [**31**] and [**32**]. It is also known [**32**] that the number of Tait colorings is a lower bound for the dimension of $J^\sharp(K)$. Because of the connection between Tait coloring the edges and four-coloring the regions of a planar trivalent graph (see section 1.4 again), an affirmative answer to the question would provide a new proof that every planar map can be four-colored.

References

[1] S. Akbulut and J. D. McCarthy. *Casson's invariant for oriented homology 3-spheres*, volume 36 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 1990. An exposition.

[2] K. Appel and W. Haken. *Every planar map is four colorable*, volume 98 of *Contemporary Mathematics*. American Mathematical Society, Providence, RI, 1989. With the collaboration of J. Koch.

[3] J. A. Baldwin and S. Sivek. Stein fillings and SU(2) representations. preprint, 2016.

[4] J. H. Conway and C. M. Gordon. Knots and links in spatial graphs. *J. Graph Theory*, 7(4):445–453, 1983.

[5] A. Daemi. *Symmetric Spaces and Knot Invariants from Gauge Theory*. PhD thesis, Harvard University, 2014.

[6] A. Daemi and Y. Xie. Sutured manifolds and polynomial invariants from higher rank bundles. preprint, 2017.

[7] S. K. Donaldson. An application of gauge theory to four-dimensional topology. *J. Differential Geom.*, 18(2):279–315, 1983.

[8] S. K. Donaldson. Irrationality and the h-cobordism conjecture. *J. Differential Geom.*, 26(1):141–168, 1987.

[9] S. K. Donaldson. Polynomial invariants for smooth four-manifolds. *Topology*, 29(3):257–315, 1990.

[10] S. K. Donaldson. Lefschetz pencils on symplectic manifolds. *J. Differential Geom.*, 53(2):205–236, 1999.

[11] Y. Eliashberg. A few remarks about symplectic filling. *Geom. Topol.*, 8:277–293, 2004.

[12] Y. M. Eliashberg and W. P. Thurston. *Confoliations*, volume 13 of *University Lecture Series*. American Mathematical Society, Providence, RI, 1998.

[13] J. B. Etnyre. On symplectic fillings. *Algebr. Geom. Topol.*, 4:73–80, 2004.

[14] P. M. N. Feehan and T. G. Leness. Witten's conjecture for many four-manifolds of simple type. *J. Eur. Math. Soc. (JEMS)*, 17(4):899–923, 2015.

[15] A. Floer. An instanton-invariant for 3-manifolds. *Comm. Math. Phys.*, 118(2):215–240, 1988.

[16] R. H. Fox. A quick trip through knot theory. In *Topology of 3-manifolds and related topics (Proc. The Univ. of Georgia Institute, 1961)*, pages 120–167. Prentice-Hall, Englewood Cliffs, N.J., 1962.

[17] D. Gabai. Foliations and the topology of 3-manifolds. *J. Differential Geom.*, 18(3):445–503, 1983.

[18] P. Ghiggini. Knot Floer homology detects genus-one fibred knots. *Amer. J. Math.*, 130(5):1151–1169, 2008.

[19] D. Gieseker and J. Li. Moduli of high rank vector bundles over surfaces. *J. Amer. Math. Soc.*, 9(1):107–151, 1996.

[20] E. Giroux. Géométrie de contact: de la dimension trois vers les dimensions supérieures. In *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*, pages 405–414, Beijing, 2002. Higher Ed. Press.

[21] W. Haken. Theorie der Normalflächen. *Acta Math.*, 105:245–375, 1961.

[22] A. Juhász. Floer homology and surface decompositions. *Geom. Topol.*, 12(1):299–350, 2008.

[23] M. Khovanov. A categorification of the Jones polynomial. *Duke Math. J.*, 101(3):359–426, 2000.

[24] M. Khovanov and L. Rozansky. Matrix factorizations and link homology. *Fund. Math.*, 199(1):1–91, 2008.

[25] P. B. Kronheimer. Four-manifold invariants from higher-rank bundles. *J. Differential Geom.*, 70(1):59–112, 2005.

[26] P. B. Kronheimer and T. S. Mrowka. Tait colorings, and an instanton homology for webs and foams. To appear in J. Eur. Math. Soc.

[27] P. B. Kronheimer and T. S. Mrowka. Dehn surgery, the fundamental group and SU(2). *Math. Res. Lett.*, 11(5-6):741–754, 2004.

[28] P. B. Kronheimer and T. S. Mrowka. Witten's conjecture and property P. *Geom. Topol.*, 8:295–310, 2004.

[29] P. B. Kronheimer and T. S. Mrowka. Knots, sutures, and excision. *J. Differential Geom.*, 84(2):301–364, 2010.

[30] P. B. Kronheimer and T. S. Mrowka. Khovanov homology is an unknot-detector. *Publ. Math. IHES*, 113:97–208, 2012.

[31] P. B. Kronheimer and T. S. Mrowka. Exact triangles for $SO(3)$ instanton homology of webs. *J. Topol.*, 9(3):774–796, 2016.

[32] P. B. Kronheimer and T. S. Mrowka. A deformation of instanton homology for webs. preprint, 2017.

[33] J. Lott. The work of Grigory Perelman. In *International Congress of Mathematicians. Vol. I*, pages 66–76. Eur. Math. Soc., Zürich, 2007.

[34] V. Muñoz. Ring structure of the Floer cohomology of $\Sigma \times \mathbf{S}^1$. *Topology*, 38(3):517–528, 1999.

[35] Y. Ni. Knot Floer homology detects fibred knots. *Invent. Math.*, 170(3):577–608, 2007.

[36] K. G. O'Grady. Moduli of vector bundles on projective surfaces: some basic results. *Invent. Math.*, 123(1):141–207, 1996.

[37] P. Ozsváth and Z. Szabó. Holomorphic disks and knot invariants. *Adv. Math.*, 186(1):58–116, 2004.

[38] C. D. Papakyriakopoulos. On Dehn's lemma and the asphericity of knots. *Ann. of Math. (2)*, 66:1–26, 1957.

[39] J. Rasmussen. *Floer homology and knot complements*. PhD thesis, Harvard University, 2003.

[40] N. Robertson, D. Sanders, P. Seymour, and R. Thomas. The four-colour theorem. *J. Combin. Theory Ser. B*, 70(1):2–44, 1997.
[41] S. Sivek. Donaldson invariants of symplectic manifolds. *Int. Math. Res. Not. IMRN*, 6:1688–1716, 2015.
[42] I. Smith. Floer cohomology and pencils of quadrics. *Invent. Math.*, 189(1):149–250, 2012.
[43] P. G. Tait. Listing's Topologie. *Philosophical Magazine (5th Series)*, 17:30–46, 1884.
[44] C. H. Taubes. Self-dual Yang-Mills connections on non-self-dual 4-manifolds. *J. Differential Geom.*, 17(1):139–170, 1982.
[45] C. H. Taubes. Self-dual connections on 4-manifolds with indefinite intersection matrix. *J. Differential Geom.*, 19(2):517–560, 1984.
[46] C. H. Taubes. The Seiberg-Witten invariants and symplectic forms. *Math. Res. Lett.*, 1(6):809–822, 1994.
[47] K. K. Uhlenbeck. Connections with L^p bounds on curvature. *Comm. Math. Phys.*, 83(1):31–42, 1982.
[48] F. Waldhausen. On irreducible 3-manifolds which are sufficiently large. *Ann. of Math. (2)*, 87:56–88, 1968.
[49] E. Witten. Monopoles and four-manifolds. *Math. Res. Lett.*, 1(6):769–796, 1994.
[50] Y. Xie. *On the Framed Singular Instanton Floer Homology from Higher Rank Bundles*. PhD thesis, Harvard University, 2016.