A COMPACTNESS THEOREM FOR $SO(3)$ ANTI-SELF-DUAL EQUATION WITH TRANSLATION SYMMETRY

GUANGBO XU

Abstract
Motivated by the Atiyah–Floer conjecture, we consider $SO(3)$ anti-self-dual instantons on the product of the real line and a three-manifold with cylindrical end. We prove a Gromov–Uhlenbeck type compactness theorem, namely, any sequence of such instantons with uniform energy bound has a subsequence converging to a type of singular objects which may have both instanton and holomorphic curve components. This result is the first step towards constructing a natural bounding cochain proposed by Fukaya for the $SO(3)$ Atiyah–Floer conjecture.

CONTENTS
1. Introduction .......................................................... 1
2. Preliminaries .......................................................... 7
3. The rescaled equation and interior compactness .................... 19
4. The isoperimetric inequality ........................................ 23
5. Compactness modulo energy blowup .................................. 35
6. Stable scaled instantons .............................................. 40
7. Proof of the compactness theorem .................................... 43
References ........................................................................ 49

1. INTRODUCTION

1.1. The Atiyah–Floer conjecture. In 1980s Floer [Flo88a][Flo88b][Flo89] introduced several important invariants of different types of geometric objects. These invariants are now generally called the Floer (co)homology. The Atiyah–Floer conjecture [Ati88] asserts that two such invariants, the instanton Floer homology of a three-dimensional manifold and the Lagrangian intersection Floer homology associated to a splitting of the three-manifold, are isomorphic. This conjecture has become a central problem in the field of symplectic geometry, gauge theory, and low-dimensional topology. The principle underlying the Atiyah–Floer conjecture has also motivated a number of important constructions in low-dimensional topology, such as the Heegaard–Floer homology [OS04]. There are two principal versions of the Atiyah–Floer conjecture, the $SU(2)$ case and the $SO(3)$ case, corresponding respectively to the two choices of the gauge group. Despite many progresses made in recent years, the general cases of both versions are still open. On the level of Euler characteristics, the Atiyah–Floer conjecture was proved by Taubes [Tau90].

Let us briefly review Atiyah’s intuitive argument [Ati88] leading to the identification of the two Floer homologies. Let $M$ be a closed oriented three-manifold. Let $\Sigma \subset M$ be an embedded surface separating $M$ into two pieces $M^-$ and $M^+$ which share the common boundary $\Sigma$ (see Figure 1.1). Consider a $G$-bundle $P \to M$ where $G$ is either $SU(2)$ or $SO(3)$. The moduli space of flat connections on $P|_{\Sigma}$, denoted by $R_{\Sigma}$, is naturally a
(singular) symplectic manifold. Inside $R_\Sigma$ there are two Lagrangian submanifolds $L_{M^-}$ and $L_{M^+}$ associated to the splitting, i.e., the set of gauge equivalence classes of flat connections on $\Sigma$ which can be extended to a flat connection on $P|_{M^+}$ resp. $P|_{M^-}$. The generators of the instanton Floer chain complex, which are gauge equivalence classes of flat connections on $M$, correspond naturally to intersections of the two Lagrangian submanifolds. These intersection points are generators of the Lagrangian Floer chain complex. On the other hand, the differential map of the instanton Floer homology $I(M,P)$ is defined by counting solutions to the anti-self-dual equation (the ASD equation) on the product $\mathbb{R} \times M$ of the real line $\mathbb{R}$ and $M$. This equation depends on the metric on $M$ while the resulting homology is independent of the metric. If one “stretches the neck,” namely one chooses a family of metrics $g_T$ on $M$ such that a fixed neighborhood of $\Sigma$ is isometric to $[-T,T] \times \Sigma$, then one expects that solutions to the ASD equation converge as $T \to \infty$ to holomorphic maps $u : [-1,1] \times \mathbb{R} \to R_\Sigma$ with boundary condition $u(\{\pm 1\} \times \mathbb{R}) \subset L_{M^\pm}$—the counting of these holomorphic maps defines the differential map of the Lagrangian intersection Floer chain complex with resulting homology group $HF(L_{M^-},L_{M^+})$. The correspondence between instantons and holomorphic strips shows that the two Floer chain complexes, and hence the homology groups, should be isomorphic.

This paper is motivated by the $SO(3)$ case of the Atiyah–Floer conjecture. We first remark on a crucial difference between the $SO(3)$ case and the $SU(2)$ case. When $G = SU(2)$, the moduli space of flat connections over a surface has singularities corresponding to reducible connections. This fact makes the Lagrangian Floer homology $HF(L_{M^-},L_{M^+})$ difficult to define (see [MW12] for an equivariant construction of this Floer homology). When $G = SO(3)$, the moduli space $R_\Sigma$ of flat connections on a nontrivial $SO(3)$-bundle over a surface is smooth. Consequently the Lagrangian Floer homology can be defined in the traditional way and the corresponding Atiyah–Floer conjecture has been proved in certain cases. For example, Dostoglou–Salamon [DS94] proved the $SO(3)$ case for mapping cylinders.

Another motivation of this paper comes from the original neck-stretching argument sketched above. As people become more interested in the alternative approach using Lagrangian boundary conditions for the instanton equation (see [Sal95] [Fuka, Fuk98] [Weh05a, Weh05b, Weh05c] [SW08] [Fukb] [DF18]), the neck-stretching argument is more or less abandoned. Furthermore, the neck-stretching argument can provide a direct comparison between the moduli spaces, potentially leading to Atiyah–Floer type correspondences for more refined invariants. We would like to see how far Atiyah’s original idea can go beyond the situation of [DS94]. Meanwhile, the analytic problems involved in the neck-stretching limit have their own interests and deserve to be explored.
1.2. **Bounding cochain on the symplectic side.** A more direct motivation of this paper is a conjecture of Fukaya [Fuk18, Conjecture 6.7]. This conjecture is related to an additional complexity in the study of the Atiyah–Floer conjecture, that is, the regularity of the Lagrangian submanifolds. From now on we assume $G = SO(3)$, which grants a smooth moduli space $R_\Sigma$ of flat connections on the nontrivial $SO(3)$-bundle over the surface $\Sigma$. If both Lagrangians $L^{+}_M$ and $L^{-}_M$ are embedded and intersect transversely, then one can define $HF(L^{+}_M, L^{-}_M)$ in a standard way as in [Flo88b] [Oh93a, Oh93b], as the two Lagrangians are both monotone. In general, however, the natural maps $L^{\pm}_M \to R_\Sigma$ are not embeddings. After a generic perturbation, one can only achieve immersions $L^{\pm}_M \hookrightarrow R_\Sigma$ which satisfy the monotonicity condition in a weak sense. In this situation we need to apply the more complicated construction of immersed Lagrangian Floer homology developed by Akaho–Joyce [AJ10] with the appearance of *bounding cochains*. A general Lagrangian immersion may not have bounding cochains; even it has, the Floer homology involving the immersed Lagrangian depends on the choice of a bounding cochain.

The necessity of considering bounding cochains in the Atiyah–Floer conjecture can also be seen via a closer look at the neck-stretching process. As one stretches the neck, the energy of ASD instantons can be distributed in three parts, $R \times M^{-}, R \times M^{+}$, and $R \times [-T, T] \times \Sigma$ (corresponding to the left, the right, and the middle parts of the second picture of Figure 1.1). The energy stored on the left and on the right can form ASD instantons on $R \times M^{-}_X$ and $R \times M^{+}_X$, where $M^{\pm}_X$ is the completion of $M^{\pm}$ by adding the cylindrical end $\Sigma \times [0, +\infty)$. On the other hand, the energy stored in the middle part produces holomorphic strips in $R_\Sigma$ as predicted by the Atiyah–Floer conjecture. Hence a general limiting object could be a complicated configuration having components corresponding to either instantons over $R \times M^{-}_X$ or holomorphic strips (see Figure 1.2); we ignore bubbling of instantons over $R^4$, instantons over $C \times \Sigma$, and holomorphic spheres, as they happen in high codimensions. While in the case when $L^{\pm}_M$ are immersed, instantons over $R \times M^{\pm}_X$ may give nontrivial contributions which happen in codimension zero.

![Figure 1.2](image)  

**Figure 1.2.** A typical limiting configuration in the adiabatic limit: $u$ is a holomorphic strip in $R_\Sigma$ with boundary in $L^{-}_M$ and $L^{+}_M$, $A_1$ and $A_2$ are ASD instantons over $R \times M^{-}_X$, $A_3, A_4, A_5$ are ASD instantons over $R \times M^{+}_X$. The limits of the instantons $A_i$ are double points of the Lagrangian immersions $L^{\pm}_M$.

The geometric picture discussed above suggests that one has to modify the Lagrangian Floer chain complex to match with the instanton Floer chain complex. The differential map of the modified chain complex counts not only usual holomorphic strips, but also...
strips with additional boundary constraints associated with the instantons over $R \times M^\pm$. This kind of boundary constraints can be regarded as bounding cochains. As introduced in [FOOO09], for any cochain model chosen for the Lagrangian Floer theory (such as Morse cochains), a bounding cochain is a cochain of odd degree which can cancel all contributions of disk bubbling; these disk bubbles may obstructed $d^2 = 0$ in the original Floer chain complex. In a symplectic manifold, if $L_-, L_+$ are Lagrangian submanifolds, then a choice of a pair of bounding cochains $b_-, b_+$ leads to a deformed complex $CF((L_-, b_-), (L_+, b_+))$. The differential map of the deformed complex counts holomorphic strips with boundary “insertions,” i.e., points on the two boundary components satisfying geometric constraints prescribed by the cochains $b_-$ and $b_+$ (see Figure 1.3).

![Figure 1.3. A holomorphic strip with boundary insertions lying in constraints given by the bounding cochains.](image)

The following conjecture of Fukaya summarizes the above discussion and directly motivates the work of the current paper.

**Conjecture 1.1.** [Fuk18, Conjecture 6.7] Let $M$ be a three-manifold with a cylindrical end isometric to $\Sigma \times [0, +\infty)$ and $P \to M$ be an $SO(3)$-bundle whose restriction to every connected component of $\Sigma$ is nontrivial. Let $L_M$ be the moduli space of flat connections over $M$. Suppose the natural map $L_M \to R_\Sigma$ is an immersion with transverse double points. Then “counting” instantons on $R \times M$ defines a bounding cochain

$$b_M \in CF(L_M).$$

of the $A_\infty$-algebra associated to the immersed Lagrangian $L_M \hookrightarrow R_\Sigma$.

Indeed, for the situation of immersed Lagrangians considered in [AJ10], the cochain model has summands corresponding to the double points of the immersion. Hence a priori a bounding cochain is in general a linear combination of ordinary cochains on $L_M$ and double points of the immersion. As observed by Fukaya [Fuk18], due a weak version of monotonicity, the bounding cochain $b_M$ in the above conjecture is expected to be a linear combination of only double points.

A refined version of the $SO(3)$ Atiyah–Floer conjecture can be stated as follows.

**Conjecture 1.2** (The $SO(3)$ Atiyah–Floer conjecture). For a closed three-manifold $M$ with a suitable $SO(3)$-bundle $P \to M$ and a suitable splitting $M = M^- \cup M^+$, there is a natural isomorphism of abelian groups

$$I(M, P) \cong HF((L_M^-, b_M^-), (L_M^+, b_M^+)).$$

(1.1)
Remark 1.3. In [Fukb, Fuk18] a different strategy of proving the existence of a bounding
cochain was sketched. Instead of considering instantons over $R \times M$ where $M$ has a
cylindrical end, consider the ASD equation over $R \times M_0$ where $M_0$ is the corresponding
manifold with boundary, imposing the boundary condition given by a testing Lagrangian
$L \subset R\Sigma$. The advantage of this approach is that it can avoid certain difficult analysis
associated to the ASD equation on $R \times M$, while a simpler moduli space is enough to
produce a chain map. However, this approach lacks a direct comparison between the
moduli space of instantons and moduli space of holomorphic strips as indicated by the
straightforward neck-stretching argument. Such a comparison between moduli spaces
can be useful in establishing relations between more refined invariants. The approach of
using Lagrangian boundary conditions have also been adopted to solve the Atiyah–Floer
conjecture, see [Sal95] [Fuka, Fuk98, Fukb] [Weh05c, Weh05a, Weh05b] [SW08] [DF18].

1.3. Main results of this paper. The purpose of this paper is to take the first step
towards the resolution of Conjecture 1.1, namely, to compactify the moduli space of ASD
instantons over $R \times M$ where $M$ is a three-manifold with cylindrical end. More precisely,
given a sequence of ASD instantons $A_i$ over $R \times M$ with uniformly bounded energy,
we study the possible limiting configurations as $i \to \infty$. There are several phenomena
preventing $A_i$ from converging to an ASD instanton. (i) As in the usual situation of the
ASD equation, energy may concentrate in small scales and bubble off instantons on $R^4$.
(ii) Since $R \times M$ is noncompact, the energy may concentrate at different regions of the
same scale which move apart from each other. This is similar to the situation in Morse
theory, where a sequence of gradient lines can converge to a broken gradient line. There
can also be instantons over $C \times \Sigma$ appearing as energy may escape in the noncompact
direction of $M$. (iii) A nontrivial amount of energy may escape from any finite region
of $R \times M$ and spread over larger and larger domains; after rescaling such energy form
either holomorphic spheres or holomorphic disks in $R\Sigma$. In general a combination of
these phenomena can happen in the limit. The hierarchy of different speeds of energy
concentration or spreading is captured by the combinatorial type of the limiting object
described by a tree. See Figure 1.4 for a typical configuration of the limiting object, which
we will call stable scaled instantons.

One can see that the limiting configurations are very similar to objects appearing in
the adiabatic limit of the symplectic vortex equation (see [GS05] and [Zil05, Zil14] for
the closed case and [WX17] [WX] for the case with Lagrangian boundary condition).
Combinatorially these objects are also similar to certain objects appearing in the compact-
ification of pseudoholomorphic quilts studied by [WW15] and [BW18]. To describe such
limiting objects, we need to define certain singular configurations which have components
corresponding to energy concentrations in different scales (see [WX17, Section 4] and
Section 6 of the current paper). Having this picture in mind, in this paper we define the
notion of stable scaled instantons (see Definition 6.2) as the expected limiting objects,
and a Gromov–Uhlenbeck type convergence (see Definition 6.3). Then we can state our
main theorem as follows.

Theorem 1.4. Let $M$ be a three-manifold with cylindrical end and $P \to M$ be an $SO(3)$-
bundle. Suppose $(M, P)$ satisfies assumptions of Conjecture 1.1. Then given a sequence
of anti-self-dual instantons on $R \times P \to R \times M$ with uniformly bounded energy, there is
a subsequence which converges modulo gauge transformation and translation to a stable
scaled instanton (in the sense of Definition 6.3).

Besides proving the above compactness theorem, using the same method and more
simplified argument, we can also prove a counterpart for instantons over $C \times \Sigma$. 
Figure 1.4. A typical stable scaled instanton with eight components. The dark regions are holomorphic curve components while the corners are where the curves meet the double points of immersed Lagrangian. The white regions are instanton components: the two on the left are instantons over $\mathbb{R} \times M$ and the one on the right (the teardrop) is an instanton over $\mathbb{C} \times \Sigma$.

Theorem 1.5. Let $\Sigma$ be a compact Riemann surface (not necessarily connected) and $Q \to \Sigma$ be an $SO(3)$-bundle which is nontrivial over every connected component of $\Sigma$. Then given a sequence of anti-self-dual instantons on $\mathbb{C} \times Q \to \mathbb{C} \times \Sigma$ with uniformly bounded energy, there is a subsequence which converges modulo gauge transformation and translation to a stable scaled instanton.

We remark that certain compactness problems in gauge theory with respect to adiabatic limit or neck-stretching which are of similar nature have been considered by other people, for example Chen [Che98], Nishinou [Nis10], and Duncan [Dun13, Dun]. Comparing to these previous works, the main contribution of this paper is the treatment of the compactness problem near the “boundary” (the compact part of the three-manifold with cylindrical ends). The argument is based on the isoperimetric inequality (Theorem 4.4), the annulus lemma (Proposition 4.5 and extensions), and the boundary diameter estimate (Lemma 4.10). The method of using boundary diameter estimate to establish boundary compactness would also be useful in other situations. For example, for the compactness problem about the strip-shrinking limit of pseudoholomorphic quilts, this method potentially leads to a simplified argument as opposed to the method of [BW18] which appeals to hard elliptic estimates over varying domains. Another contribution of the current paper is to define the correct notion of singular configurations (stable scaled instantons) that may appear in the limit and provide detailed argument of constructing the limiting bubble tree. Last but not the least, a modification of our construction will lead to a proof of a compactness theorem about the neck-stretching limit for the $SO(3)$ instanton equation. The details will be completed in future works.

This paper is organized as follows. In Section 2 we recall basic notions and facts about the anti-self-dual equation, holomorphic curves, state the main assumption of the three-manifold, and recall a few technical results. In Section 3 we recall a basic compactness theorem for the rescaled ASD equation over the product of two surfaces in the adiabatic limit and prove a refinement of an interior estimate of Dostoglou–Salamon.
In Section 4 we (re)prove an isoperimetric inequality for a closed three-manifold and the annulus lemma, and establish a boundary diameter estimate. In Section 5 we prove the compactness modulo energy blowup theorem for the ASD equation over the noncompact four-manifold $\mathbb{R} \times M$. In Section 6 we state the main theorem in technical terms. In Section 7 we finish the proof of the main theorem (Theorem 1.4 and Theorem 1.5).

Acknowledgments. The author would like to thank Kenji Fukaya for suggesting this problem, for many stimulating discussions, and for his warm encouragement and generous support. The author would like to thank Simons Center for Geometry and Physics for creating a wonderful environment for mathematical research. The author would like to thank Donghao Wang and David Duncan for helpful discussions.

This work is partially supported by the Simons Collaboration Grant on Homological Mirror Symmetry.

2. Preliminaries

In this paper we study gauge theory for a smooth $SO(3)$-bundles $P \to U$ where $U$ is some smooth manifold of dimension at most four. As in the usual treatment of $SO(3)$-gauge theory, we modify the definition of gauge transformations as follows. The conjugation of $SO(3)$ can be extended to an $SO(3)$-action on $SU(2)$. A gauge transformation on $P$ is regarded as $SU(2)$-valued, i.e., a map $g : P \to SU(2)$ satisfying

$$g(ph) = h^{-1}g(p)h, \forall h \in SO(3), p \in P.$$ 

Since $\mathfrak{so}(3) \cong \mathfrak{su}(2)$, such $SU(2)$-valued gauge transformations acts on $SO(3)$-connections in the usual way. Let $A(P)$ be the space of smooth connections on $P$ and $\mathcal{G}(P)$ the space of $SU(2)$-valued smooth gauge transformations. The gauge equivalence class of a connection $A \in A(P)$ is usually denoted by $[A]$.

In this paper, when there is no extra explanation, the sequential convergence of smooth objects are always regarded as convergence in the $C^0_{\text{loc}}$-topology.

2.1. Chern–Simons functional and the anti-self-dual equation.

2.1.1. The Chern–Simons functional. The instanton Floer cochain complex can be formally viewed as the Morse cochain complex for the Chern–Simons functional. Let $M$ be a smooth oriented four-manifold. Let $P \to M$ be a smooth $SO(3)$-bundle. By Chern–Weil theory, the Pontryagin class can be represented by the differential form

$$p_1(A) = -\frac{1}{8\pi^2} \text{tr}(F_A \wedge F_A) \in \Omega^4(M)$$

for any smooth connection $A$ on $P$. When $M$ is closed, the integral of this differential form is an integer and is a topological invariant. On the other hand, suppose $M$ has a nonempty boundary $\partial M \cong M$ where $M$ inherits a natural orientation. Denote $P = P|_M$. Then for any smooth connection $A \in A(P)$, for any smooth extension $\tilde{A}$ of $A$ to the interior, the Chern–Weil integral

$$\frac{1}{2} \int_M \text{tr}(F_A \wedge F_A) \in \mathbb{R}$$

only depends on the gauge equivalence class of $A$. This is called the Chern–Simons functional, denoted by

$$CS_P : A(P)/\mathcal{G}(P) \to \mathbb{R}.$$
Often we will omit the dependence on the bundle $P$ from the notation. When $M$ has several connected components $M_1, \ldots, M_k$ and $[A_1], \ldots, [A_k]$ are the restrictions of $[A] \in \mathcal{A}(P)/\mathcal{G}(P)$ to these components, we denote this action by

$$CS([A_1], \ldots, [A_k]).$$

In particular, for a three-manifold $N$ with an $SO_p^3$-bundle $P \to N$, for a reference connection $A_0$ and any $A \in \mathcal{A}(P)$, denote the relative Chern–Simons action of $A$ to be

$$CS_{rel}^{[A_0]}([A]) = CS_P([A_0], [A])$$

where $P = [0, 1] \times P$ is the product bundle over $[0, 1] \times N$. When $A_0$ is flat, the action is equal to

$$\int_M \text{tr} \left[ \frac{1}{2} d_{A_0}(A - A_0) \wedge (A - A_0) + \frac{1}{3} (A - A_0) \wedge (A - A_0) \wedge (A - A_0) \right].$$

(2.1)

This coincides with the classical expression of the Chern–Simons functional.

The Chern–Simons functional can be viewed as a multivalued function on the space of connections on the three-manifold. For each $A \in \mathcal{A}(P)$ and a deformation $\alpha \in \Omega^1(\text{ad}P)$, the directional derivative of the Chern–Simons functional in the direction of $\alpha$ is

$$D_A CS(\alpha) = \int_M \text{tr}(F_A \wedge \alpha).$$

Hence critical points of the Chern–Simons functional are flat connections.

2.1.2. The anti-self-dual equation. Suppose $M$ is equipped with a Riemannian metric. A connection $A \in \mathcal{A}(P)$ is called an anti-self-dual connection (ASD connection for short) if

$$F_A + \# F_A = 0 \in \Omega^2(M, \text{ad}P).$$

Here $F_A$ is the curvature of $A$ and $\#$ is the Hodge start operator on differential forms on $M$. When $M$ is noncompact, we also impose the finite energy condition. Then define the Yang–Mills functional, also called the energy, of a connection $A$ by

$$E(A) := \frac{1}{2} \int_M |F_A|^2.$$ 

Here the norm of the curvature is induced from the metric on $M$ and the Killing metric on the Lie algebra. We always assume that ASD connections have finite energy.

A particular case is when $M = \mathbb{R}^4$ which is equipped with the standard Euclidean metric. We call an ASD connection over $\mathbb{R}^4$ an $\mathbb{R}^4$-instanton.

The ASD equation can be viewed as the gradient flow equation of the Chern–Simons functional. A direct consequence of this perspective is the following energy identity.

**Lemma 2.1.** Let $M$ be a compact oriented four-manifold with boundary and $P \to M$ be an $SO(3)$-bundle. Then for any Riemannian metric on $M$ and any ASD connection on $P$ with respect to this metric, one has

$$E(A) = CS([A|_{\partial M}]).$$

\footnote{When the domain is $\mathbb{R}^4$, or the product of the complex plane and a closed surface, or the product of the real line and a three-manifold with cylindrical end, we call an ASD connection an instanton.}
2.1.3. Convergence and compactness. We discuss the topology of the space of connections and recall the celebrated Uhlenbeck compactness theorem. Let $M$ be a manifold and $P \to M$ be a principal bundle. The convergence of smooth $SO(3)$-connections $A_i \in \mathcal{A}(P)$ towards a limit $A_\infty \in \mathcal{A}(P)$ (in the $C^\infty_{\text{loc}}$ topology) means for any precompact open subset $K \subset M$, $A_i|_K$ converges uniformly with all derivatives to $A_\infty|_K$. We define the more general notion of convergence in the Uhlenbeck sense. Let $M$ be a Riemannian four-manifold. Let $M_i \subset M$ be an exhausting sequence of open subsets, meaning that every compact subset of $M$ is contained in $M_i$ for sufficiently large $i$. Let $P_i \to M_i$ be $SO(3)$-bundles and $A_i \in \mathcal{A}(P_i)$ be a sequence of ASD connections on $P_i$. Let $P_\infty \to M$ be an $SO(3)$-bundle and $A_\infty \in \mathcal{A}(P_\infty)$ be an ASD connection on $P$. Let $m_\infty$ be a positive measure on $M$ supported at finitely many points.

Definition 2.2. We say that $A_i$ converges to $(A_\infty, m_\infty)$ in the Uhlenbeck sense if

(a) the sequence of functions $|F_{A_i}|^2$ converge as measures to $|F_{A_\infty}|^2 + 2m_\infty$, and
(b) there are bundle isomorphisms $\rho_i : P_\infty \to P_i$ over $M_i \setminus \text{Supp}m_\infty$ such that $\rho_i^* A_i$ converges to $A_\infty$.

This notion of convergence is independent of the choices of representatives in their gauge equivalence classes. Therefore if $A_i$ and $(A_\infty, m_\infty)$ satisfy conditions of Definition 2.2, we will say that $[A_i]$ converges to $([A_\infty], m_\infty)$ in the Uhlenbeck sense. Further, the measure $m_\infty$ in the limit, called the bubbling measure, is nonzero at $x \in M$ if and only if a nontrivial $R^4$-instanton bubbles off in the limit. We know that the masses of $m_\infty$ are in $4\pi^2 Z_+$. The convergence implies the following energy identity: for any compact subset $K \subset M$ containing the support of $m_\infty$, there holds

$$\lim_{i \to \infty} E(A_i; K) = E(A_\infty; K) + \int_K m_\infty.$$ 

We summarize the celebrated Uhlenbeck compactness theorem as follows.

Theorem 2.3. (cf. [DK90, Section 4.4]) Let $M, M_i, P_i, A_i$ be as above. Suppose the energy of $A_i$ is uniformly bounded from above, i.e,

$$\limsup_{i \to \infty} E(A_i) < +\infty.$$ 

Then there exist a subsequence (still indexed by $i$), a positive measure $m_\infty$ on $M$ with finite support, an $SO(3)$-bundle $P_\infty \to M$, and an ASD connection $A_\infty \in \mathcal{A}(P_\infty)$, such that $A_i$ converges to $(A_\infty, m_\infty)$ in the Uhlenbeck sense. In particular, there is a positive constant $h > 0$ with the following property: if $E(A_i) < h$, then a subsequence of $A_i$ converges to a limiting ASD connection $A_\infty$ without bubbling.

2.2. Product of two surfaces. The Atiyah–Floer conjecture can be regarded as an effect of the reduction from 4D gauge theory to 2D sigma model observed by physicists [BJSV95]. Consider the special case that $M = S \times \Sigma$ where $S$ and $\Sigma$ are oriented surfaces, equipped with a product metric. We assume for simplicity that $S$ is an open subset of either the complex plane $C$ or the upper half plane $H$, in which cases $S$ is equipped with the standard holomorphic coordinate $z = s + it$ and the standard flat metric. Let $Q \to \Sigma$ be an $SO(3)$-bundle and $P = S \times Q \to M$ be the pullback of $Q$ via the projection $M \to \Sigma$. Then we can write a connection $A \in \mathcal{A}(P)$ as

$$A = dS + \phi ds + \psi dt + B,$$

where $dS$ is the exterior differential in $S$, $B : S \to \mathcal{A}(Q)$ is a smooth map, and

$$\phi, \psi \in \Omega^0(S \times \Sigma, \text{ad}Q).$$
Now we look at the ASD equation with respect to the product metric. Introduce
\[
A_s = \partial_s B - d_B \phi, \quad A_t = \partial_t B - d_B \psi.
\] (2.2)
and
\[
\kappa_A = \partial_s \psi - \partial_t \phi + [\phi, \psi], \quad \mu_A = F_B.
\] (2.3)
Then with respect to the product metric, the ASD equation can be written in the local form
\[
A_s + * A_t = 0, \quad \kappa_A + * \mu_A = 0.
\] (2.4)
Here \(*\) is the Hodge star on \(\Sigma\).

Remark 2.4. The equation (2.4) can be viewed as an infinite dimensional version of the symplectic vortex equation introduced by Cieliebak–Gaio–Salamon [CGS00] and Mundet [Mun99, Mun03]. Indeed this perspective is one of the motivation of [CGS00] to propose the symplectic vortex equation.

2.3. Holomorphic curves. We recall a few basic facts about pseudoholomorphic maps from Riemann surfaces to almost complex manifolds. In this subsection, \((X, J)\) always denotes a compact almost complex manifold. We fix a Riemannian metric \(h\) on \(X\). Let \(S\) be a smooth Riemann surface with possibly nonempty boundary. A \(J\)-holomorphic map from \(S\) to \(X\) is a continuous map \(u: S \rightarrow X\) which is smooth in the interior and satisfies the Cauchy–Riemann equation
\[
\bar{\partial}_J u := \frac{1}{2} \left( \frac{\partial u}{\partial s} + J \frac{\partial u}{\partial t} \right) d\bar{z} = 0.
\]
Here \(z = s + it\) is a local holomorphic coordinate on \(S\). In this paper \(S\) is always an open subset of either the complex plane \(C\) or the upper half plane \(H\). The energy of \(u\) is
\[
E(u; S) := \frac{1}{2} \| du \|_{L^2(S)} = \frac{1}{2} \int_S |du|^2_h dsdt.
\]
When \(S\) is understood from the context, we abbreviate \(E(u; S)\) by \(E(u)\).

2.3.1. Compactness. We would like to define the notion of convergence of \(J\)-holomorphic maps over a bordered surface without any appropriate boundary condition. Let \(S \subset H\) be an open subset and \(S_i \subset S\) be an exhausting sequence of open subsets.

Definition 2.5. Let \(u_i : S_i \rightarrow X\) be a sequence of \(J\)-holomorphic maps. Let \(u_\infty : S \rightarrow X\) be another \(J\)-holomorphic map. Then we say that \(u_i\) converges to \(u_\infty\) if \(u_i\) converges to \(u_\infty\) in \(C^0_{\text{loc}}(S) \cap C^\infty_{\text{loc}}(S \cap \text{Int} H)\).

Now we recall a compactness result about holomorphic maps on bordered surfaces without imposing a boundary condition and give a proof.

Proposition 2.6. Let \(u_i : S_i \rightarrow X\) be a sequence of \(J\)-holomorphic maps with
\[
\limsup_{i \to \infty} E(u_i; S_i) < +\infty.
\]
(a) Assume there is no energy concentration in the interior, namely, for all \(z \in S \cap \text{Int} H\), one has
\[
\lim_{r \to 0} \limsup_{i \to \infty} E(u_i; B_r(z) \cap S_i) = 0.
\] (2.5)
Then there exist a subsequence (still indexed by \(i\)) and a holomorphic map \(u_\infty : S \cap \text{Int} H \rightarrow X\) such that \(u_i\) converges to \(u_\infty\) in \(C^\infty_{\text{loc}}(S \cap \text{Int} H)\).
(b) In addition, suppose for each \( z \in \partial S \) there holds
\[
\lim_{i \to \infty} \limsup_{r \to 0} \text{diam}(u_i(B^+_r(z) \cap S_i)) = 0 \tag{2.6}
\]
(here \( B^+_r(z) \) is the half disk \( B_r(z) \cap H \)). Then the limit \( u_\infty \) extends continuously to \( S \cap \partial H \) and \( u_i \) converges to \( u_\infty \) in the sense of Definition 2.5.

**Proof.** Part (a) is the classical Gromov compactness result (see for example [MS04] [IS00]). For part (b), we first show that \( u_\infty \) has limits at all boundary points. Choose \( z \in \partial S \). By (2.6), for any \( \epsilon > 0 \), there exists an \( r > 0 \) such that
\[
\lim_{i \to \infty} \text{diam}(u_i(B^+_r(z))) < \epsilon.
\]
Then for sufficiently large \( i \), for all \( z', z'' \in B^+_r(z) \cap S \cap \text{Int} H \), there holds
\[
d(u_i(z'), u_i(z'')) < \epsilon.
\]
Since \( u_i(z') \to u_\infty(z') \) and \( u_i(z'') \to u_\infty(z'') \) as \( i \to \infty \). It implies that
\[
z', z'' \in B^+_r(z) \implies d(u_\infty(z'), u_\infty(z'')) < \epsilon.
\]
Hence \( u_\infty \) has limits at all boundary points. It is a similar argument to show that the boundary limits define a continuous extension of \( u_\infty \) and \( u_i \) converges to \( u_\infty \) in \( C^0_{\text{loc}}(S) \).
We leave the details to the reader. \( \square \)

2.3.2. **Immersed Lagrangian boundary condition.** We recall basic notions of pseudoholomorphic curves with an immersed Lagrangian boundary condition. We fix a compact symplectic manifold \((X, \omega)\). A **Lagrangian immersion** is a smooth immersion \( \iota : L \hookrightarrow X \) such that \( \iota^* \omega = 0 \) and such that \( \dim X = 2 \dim L \). We assume \( L \) is compact. We assume that \( L \) only has **transverse double points.** Namely, the map
\[
(\iota, \iota) : L \times L \to X \times X
\]
is transverse to the diagonal \( \Delta_X \subset X \times X \) away from the diagonal \( \Delta_L \subset L \times L \). The compactness of \( L \) implies that there are finitely many **double points** of \( \iota(L) \), each of which has exactly two preimages \((p, q), (q, p) \in L \times L \setminus \Delta_L \). Elements of the set
\[
R_L := \{(p, q) \in L \times L \mid \iota(p) = \iota(q), \; p \neq q\}.
\]
are called **ordered double points.** The map \( (p, q) \mapsto (q, p) \) which preserves the set \( \Delta_L \cup R_L \) is called the **transpose.**

Now we define the notion of holomorphic curves with boundary lying in the immersed Lagrangian. One can see that this notion coincides with that in [AJ10] after ordering the set of marked points \( W \). Fix an \( \omega \)-tamed almost complex structure \( J \) on \((X, \omega)\), namely
\[
\xi \mapsto \omega(\xi, J\xi) > 0, \; \forall \xi \in TX, \; \xi \neq 0.
\]
The symplectic form \( \omega \) and the almost complex structure \( J \) determines a metric
\[
g(\xi, \xi') := \frac{1}{2} (\omega(\xi, J\xi') + \omega(\xi', J\xi)).
\]
We also assume that \( L \) is **totally real** with respect to \( J \), namely, for every \( x \in L \),
\[
T_{\iota(x)}X = d\iota(T_xL) \oplus Jd\iota(T_xL).
\]

**Definition 2.7.** (cf. [AJ10, Definition 4.2]) Let \( S \) be a Riemann surface with possibly nonempty boundary. A **marked \( J \)-holomorphic map** from \( S \) to \( X \) with boundary in \( \iota(L) \) is a quadruple
\[
\mathbf{u} = (u, W, \gamma)
\]
where $u : S \to X$ is a $J$-holomorphic map with $u(\partial S) \subset \iota(L)$, $W \subset S$ is a finite subset, and $\gamma : \partial S \setminus W \to L$ is a smooth map, satisfying the following **boundary condition:**

$$u|_{\partial S \setminus W} = \iota \circ \gamma.$$ 

Given such a marked $J$-holomorphic map, for each $w \in W \cap \partial S$, there is a local holomorphic coordinate chart $U_w \cong \phi_w(U_w) \subset H$. The boundary condition implies that the limit

$$\text{ev}_w(u) := (\lim_{s \to 0^-} \gamma(\phi_w^{-1}(s)), \lim_{s \to 0^+} \gamma(\phi_w^{-1}(s))) \in \Delta_L \cup R_L \subset L \times L$$

exists. We call this limit the **evaluation** of $u$ at $w$. If $\text{ev}_w(u) \in R_L$, we call $w$ a **switching point** of $u$. On the other hand, the evaluation of $u$ at an interior marking $w \in W \cap \text{Int} S$ is the value $\text{ev}_w(u) = u(w) \in X$.

We also include a nontrivial Dirac measure in the datum of a holomorphic curve. If $m : W \to R_1$ is a function, regarded as a positive measure on $S$ whose support is contained in $W$, then we call the tuple

$$\tilde{u} = (u, W, \gamma, m)$$

a **marked holomorphic curve with mass**. When $\partial S = \emptyset$, we simplify the notation as $\tilde{u} = (u, W, m)$.

2.4. **Flat connections on three-manifolds.** Now we introduce the basic assumptions on the three-manifolds. Let $M_0$ be a connected, oriented three-manifold with a nonempty and not necessarily connected boundary $\partial M_0 \cong \Sigma$. Let $M$ be the completion, i.e.,

$$M := M_0 \cup ([0, +\infty) \times \Sigma),$$

where the two parts are glued along the common boundary. Then we always identify $M_0$ with a closed subset of $M$. Let $P \to M$ be an $SO(3)$-bundle. Let $P_0 \to M_0$ be the restriction of $P$ to $M_0 \subset M$ and

$$Q \to \Sigma$$

be the restriction of $P_0$ to the boundary. Let $L_M$ be the moduli space of gauge equivalence classes of flat connections on $P_0$, i.e.,

$$L_M := \left\{ A \in \mathcal{A}(P_0) \mid F_A = 0 \right\}/\mathcal{G}(P_0).$$

Let $R_\Sigma$ be the moduli space of gauge equivalence classes of flat connections on $Q$, i.e.,

$$R_\Sigma := \left\{ B \in \mathcal{A}(Q) \mid F_B = 0 \right\}/\mathcal{G}(Q).$$

Both $L_M$ and $R_\Sigma$ have natural topology. There is a natural continuous map

$$\iota : L_M \to R_\Sigma$$

induced by boundary restriction.

2.4.1. **Transversality assumption.** Now we consider moduli spaces of flat connections on $M$ and $\Sigma$. We impose certain extra conditions to guarantee that these moduli spaces are smooth. For any flat connection $A \in \mathcal{A}(P_0)$, the covariant derivative $d_A$ makes $adP_0$ a flat bundle with a twisted de Rham complex

$$0 \longrightarrow \Omega^0(M_0, adP_0) \xrightarrow{d_A} \Omega^1(M_0, adP_0) \xrightarrow{d_A} \Omega^2(M_0, adP_0) \xrightarrow{d_A} \Omega^3(M_0, adP_0) \longrightarrow 0.$$ 

Similarly, when $B$ is a flat connection on $Q$, there is a complex

$$0 \longrightarrow \Omega^0(\Sigma, adQ) \xrightarrow{d_B} \Omega^1(\Sigma, adQ) \xrightarrow{d_B} \Omega^2(\Sigma, adQ) \longrightarrow 0.$$
When $A|\Sigma = B$, one can form the relative complex with
\[
\Omega^k_{\text{rel}}(\text{ad}P_0) = \Omega^k(M_0, \text{ad}P_0) \oplus \Omega^{k-1}(\Sigma, \text{ad}Q)
\]
and differential $d_{A,B}$, which is defined as
\[
d_{A,B} \begin{bmatrix} \alpha \\ h \end{bmatrix} = \begin{bmatrix} d_A & 0 \\ -r^* & d_B \end{bmatrix} \begin{bmatrix} \alpha \\ h \end{bmatrix}.
\]
Here $r^* : \Omega^k(M_0, \text{ad}P_0) \to \Omega^k(\Sigma, \text{ad}Q)$ is the pullback. Then there is a long exact sequence (in real coefficients)
\[
\cdots \longrightarrow H^0(d_A) \longrightarrow H^0(d_B) \longrightarrow H^1(d_{A,B}) \longrightarrow H^1(d_A) \longrightarrow H^1(d_B) \longrightarrow \cdots
\]

We assume the following conditions throughout this paper.

**Hypothesis 2.8.** The three-manifold with boundary $M_0$ and the $SO(3)$-bundle $P_0 \to M_0$ satisfy the following conditions.

(a) For any flat connection $B$ on $Q$, $H^0(d_B) \oplus H^2(d_B)$ vanishes. This implies that $R_\Sigma$ is a smooth manifold with
\[
\dim R_\Sigma = -3\chi(\Sigma).
\]

(b) For any flat connection $A$ on $P_0$, the operator $d_A : \Omega^1(M_0, \text{ad}P_0) \to \Omega^2(M_0, \text{ad}P_0)$ is surjective. This implies that the moduli space $L_M$ of flat connections on $P$ is smooth and
\[
\dim L_M = -\frac{3}{2}\chi(\Sigma).
\]

(c) For any flat connection $A$ on $P_0$ whose boundary restriction is $B$, the map $H^1(d_A) \to H^1(d_B)$ is injective. This implies that the natural map $L_M \to R_\Sigma$ is an immersion.

(d) The immersion $\iota : L_M \hookrightarrow R_\Sigma$ has transverse double points.

**Lemma 2.9.** Item (a) of Hypothesis 2.8 holds if and only if $Q$ is nontrivial over each connected component of $\Sigma$. In this case $\Sigma$ necessarily has an even number of connected components.

**Proof.** If $Q$ is trivial over some component $\Sigma_i \subset \Sigma$, then there exist reducible flat connections, i.e., there are flat connections $B \in A_{\text{flat}}(Q)$ with $H^0(d_B) \neq 0$. On the other hand, one has the Poincaré duality $H^2(d_B) \cong H^0(d_B)$. Hence (a) is equivalent to the nontriviality of $Q$ over each $\Sigma_i$.

To show that $\Sigma$ has an even number of connected components, consider the exact sequence in $\mathbb{Z}_2$ coefficients
\[
\cdots \longrightarrow H^2(M_0) \longrightarrow H^2(\partial M_0) \longrightarrow H^3(M_0, \partial M_0) \longrightarrow \cdots
\]
It follows that the second Stiefel–Whitney class $w_2(Q)$, which is the image of $w_2(P) \in H^2(M_0)$, is sent to zero in $H^3(M_0, \partial M_0) \cong \mathbb{Z}_2$. On the other hand, since $Q$ is nontrivial over each component $\Sigma_i$, $w_2(Q)$ restricts to the generator of $H^2(\Sigma_i; \mathbb{Z}_2)$, while each generator is sent to the generator of $H^3(M_0, \partial M_0) \cong \mathbb{Z}_2$. Hence $\Sigma$ has an even number of connected components. \qed

**Remark 2.10.** In general Item (b), (c), and (d) of Hypothesis 2.8 do not hold. However, one can perturb the Chern–Simons functional by the so-called holonomic perturbation supported away from the boundary, so that critical points of the Chern–Simons functional (i.e., certain perturbed flat connection) are non-degenerate in the Bott sense. Let $L_M$ still denote the moduli space of critical points of the perturbed Chern–Simons functional.
Pairs of flat connections on two three-manifolds with boundary induce certain flat connections on a closed three-manifold. Let $N^-, N^+$ be connected oriented three-manifolds with boundary such that
\[ \partial N^- \cong \Sigma \cong \partial (N^+)^{op}. \]
Here $(N^+)^{op}$ is a copy of $N^+$ with the reversed orientation. Let $P_{N^-} \to N^-$, $P_{N^+} \to N^+$, and $Q \to \Sigma$ be $SO(3)$-bundles such that
\[ P_{N^-}|_{\partial N^-} \cong Q \cong P_{N^+}|_{\partial N^+}. \]
Suppose $(P_{N^-}, N^-)$ and $(P_{N^+}, N^+)$ both satisfy Hypothesis 2.8. Then one obtains Lagrangian immersions
\[ t_- : L_{N^-} \cong R\Sigma, \quad t_+ : L_{N^+} \cong R\Sigma. \]
We assume in addition that

**Hypothesis 2.11.** The two three-manifolds with boundary $N^-$, $N^+$ with diffeomorphic boundary and the $SO(3)$-bundles $P_{N^\pm} \to N^\pm$ with isomorphic boundary restrictions satisfy the following condition.

- The immersions $t_- : L_{N^-} \cong R\Sigma$ and $t_+ : L_{N^+} \cong R\Sigma$ intersect cleanly.

Define a closed three-manifold and together with an $SO(3)$-bundle as follows. Let $N$ be the closed three-manifold defined by
\[ N := N^- \cup N_{\text{neck}} \cup N^+ := N^- \cup [\partial N^- \times \Sigma] \cup N^+. \] (2.8)
Here we identify the common boundaries of the two components via $\partial N^- \cong \{-1\} \times \Sigma$, $\partial (N^+)^{op} \cong \{1\} \times \Sigma$. The bundles $P_{N^\pm}$ can be glued similarly to give an $SO(3)$-bundle
\[ P_N := P_{N^-}|_{N^-} \cup P_{N^-}|_{N_{\text{neck}}} \cup P_{N^+}|_{N^+} := P_{N^-} \cup [\partial N^- \times Q] \cup P_{N^+}. \] (2.9)
Let $L_N$ be the moduli space of gauge equivalence classes of flat connections on $P_N$. Then $L_N$ is a compact manifold (with possibly varying dimensions) with a diffeomorphism
\[ L_N \cong (t_- \times t_+)^{-1}(\Delta R\Sigma) \subset L_{N^-} \times L_{N^+}. \]

**Remark 2.12.** There are two special situations when Hypothesis 2.11 is satisfied. The first special case is when $N^- \cong (N^+)^{op} \cong M_0$ and $P_{N^-} \cong P_{N^+} \cong M_0$. In this case $N$ is diffeomorphic to the doubling of $M_0$, denoted by $M_{\text{double}}$ and one has
\[ L_N \cong \Delta_{LM} \cup R_{LM} \subset L_M \times L_M. \]
The second special case is when $N^- \cong (N^+)^{op} \cong [0, \pi] \times \Sigma$ whose boundary is two copies of $\Sigma$ and $P_{N^-} \cong P_{N^+} \cong [0, \pi] \times Q$. In this case $N$ is diffeomorphic to $S^1 \times \Sigma$ and $L_N$ is diffeomorphic to $R\Sigma$.

It is convenient to allow certain piecewise smooth connections. Define
\[ \mathcal{A}^{p,s}(P_N) := \{(A_-, A_0, A_+) \in \mathcal{A}(P_N|_{N^-} \cup P_{N^-}|_{N_{\text{neck}}} \cup P_{N^+}|_{N^+}) \mid (A_-, A_+)|_{\partial N^- \cup \partial N^+} = A_0|_{\partial N_{\text{neck}}}) \]
whose elements are called **piecewise smooth connections**. Define the space of piecewise smooth gauge transformations $\mathcal{G}^{p,s}(P_N)$ in a similar way. Then one has
\[ L_N \cong \{ A \in \mathcal{A}^{p,s}(P_N) \mid F_A = 0 \}/\mathcal{G}^{p,s}(P_N). \]
2.4.2. Almost flat connections on $M_0$. Now we turn to the analytical part of the gauge theory. For all three-manifolds with boundary, say $M_0$ for example, we fix a Riemannian metric on $M_0$ such that a neighborhood of the boundary is isometric to $[0, \varepsilon) \times \partial M_0$. The metric on $M_0$ induces a metric on manifold with cylindrical ends $M$ defined by (2.7) which is of the product type on the cylindrical end. When discussing a pair of three-manifolds $N^-$ and $N^+$ sharing the same boundary, we assume the boundary restrictions of the metrics are isometric. The pair of metrics induce a metric on the closed manifold $N$ defined by (2.9) which is of product type over the neck.

The differentiation of gauge fields depends on the choice of a covariant derivative. From now on we fix a smooth flat connection $A_{\text{ref}}$ on $P \to M$ as a reference connection. By applying gauge transformations, we can arrange that the restriction of $A_{\text{ref}}$ to the cylindrical end $[0, +\infty) \times \Sigma$ is equal to $d + B_{\text{ref}}$ where $B_{\text{ref}}$ is a flat connection on $Q \to \Sigma$. Sobolev norms of sections of ad$P$ or ad$Q$, without further explanation, will be taken with respect to these reference connections.

The following lemma shows that near an almost flat connection on the three-manifold with boundary there is always a flat connection. It essentially follows from the transversality assumption Hypothesis 2.8 and the implicit function theorem.

**Lemma 2.13.** Let $p \geq 2$. There exist $\varepsilon = \varepsilon_p > 0$ and $C = C_p > 0$ satisfying the following properties. Let $A$ be a smooth connection on $P_0 \to M_0$ with

$$\|F_A\|_{L^p(M_0)} \leq \varepsilon.$$  

Then there exists a flat connection $A^*$ on $P_0$ of regularity $W^{1,p}$ satisfying

$$\|A - A^*\|_{W^{1,p}(M_0)} \leq C\|F_A\|_{L^p(M_0)}.$$  

**Proof.** Consider the space

$$W^{1,p}_c(M_0, \Lambda^1 \otimes \text{ad}P_0) = \{ \alpha \in W^{1,p}(M_0, \text{ad}P_0) \mid * \alpha|_{\partial M_0} = 0 \}.$$  

Consider the linear operator

$$D_A : W^{1,p}_c(M_0, \Lambda^1 \otimes \text{ad}P_0) \to L^p(M_0, \Lambda^1 \otimes \text{ad}P_0) \oplus L^p(M_0, \Lambda^1 \otimes \text{ad}P_0)$$

defined by $D_A \alpha = (*d_A \alpha, d_A \alpha \wedge \alpha)$. This is Fredholm with index $-\frac{3}{2} \chi(\Sigma)$. We claim that there exist $\varepsilon = \varepsilon_p > 0$ and $C = C_p > 0$ such that when $\|F_A\|_{L^p(M_0)} \leq \varepsilon$, $D_A$ is surjective and there is a right inverse $Q_A$ with $\|Q_A\| \leq C_p$. Suppose this is not the case, then there exist a sequence of connections $A_i$ with $\|F_{A_i}\|_{L^p} \to 0$ but $D_{A_i}$ is not surjective. Then the weak Uhlenbeck compactness theorem (see [Web03, Theorem A]) in three dimensions implies that a subsequence converges modulo gauge to a flat connection $A_\infty$ on $M_0$, and the convergence is weakly in $W^{1,p}$. This implies that $D_{A_i}$ converges to $D_{A_\infty}$ in operator norm. However, by Hypothesis 2.8, $D_{A_\infty}$ is surjective as its kernel is the tangent space of $L^p$ at $[A_\infty]$. This contradiction means as long as $\|F_A\|_{L^p}$ is sufficiently small, $D_A$ is surjective. Same argument further guarantees the existence of a right inverse with bounded norm as $L^p$ is compact. Then one can apply the implicit function theorem (see for example [MS04, Proposition A.3.4]) to find a nearby flat connection $A^*$ satisfying our requirement.

**Remark 2.14.** Throughout this paper, we adopt the convention that $C$ and $\varepsilon$ represent constants which are allowed to vary from line to line.
2.5. The representation variety. We recall basic facts about the moduli space of flat connections over a surface, which we often call by the name representation variety. Let $\Sigma$ be a closed surface and $Q \to \Sigma$ be the nontrivial $SO(3)$-bundle which is nontrivial over each connected component of $\Sigma$. The space $\mathcal{A}(Q)$ of smooth connections on $Q$ is an affine space modelled on $\Omega^1(\text{ad}Q)$. There is a symplectic form defined by

$$\omega_{\mathcal{A}(Q)}(\alpha, \beta) = \int_{\Sigma} \text{tr}(\alpha \wedge \beta), \; \alpha, \beta \in \Omega^1(\text{ad}Q). \quad (2.10)$$

The conformal class of the Riemannian metric on $\Sigma$ defines a compatible almost complex structure, i.e.,

$$J_{\mathcal{A}(Q)} \alpha = * \alpha$$

where the Hodge star operator on 1-forms on $\Sigma$ only depends on the complex structure. $\omega_{\mathcal{A}(Q)}$ and $J_{\mathcal{A}(Q)}$ make $\mathcal{A}(Q)$ an (infinite dimensional) Kähler manifold.

The space of gauge transformations $\mathcal{G}(Q)$ acts on $\mathcal{A}(Q)$ through gauge transformations. The action is Hamiltonian, with a moment map

$$\mu(B) = -F_B \in \Omega^2(\text{ad}Q) \cong \left( \text{Lie}\mathcal{G}(Q) \right)^*$$

The representation variety associated to $Q$ can be identified with the symplectic quotient $R_{\Sigma} \cong \mu^{-1}(0)/\mathcal{G}(Q)$.

The representation variety $R_{\Sigma}$ inherits a Kähler structure from the symplectic form $\omega_{\mathcal{A}(Q)}$ and the complex structure $J_{\mathcal{A}(Q)}$ (see [Gol84]). The associated Kähler metric on $R_{\Sigma}$ is called the $L^2$-metric. It is standard knowledge that the immersion $\iota : L_M \hookrightarrow R_{\Sigma}$ is Lagrangian and totally real with respect to the Kähler structure of $R_{\Sigma}$.

2.5.1. Projection onto the representation variety. We review the complexification of gauge transformations on connections on $Q$ (see also discussions in [Fuk98] [Dun13]). Take an open subset $U \subset \Sigma$ over which $Q$ can be trivialized. Then we can write $B \in \mathcal{A}(Q)$ as $B = d + \beta$ where $\beta \in \Omega^1(U, \mathfrak{so}(3)) = \Omega^1(U, \mathfrak{su}(2))$. Then $B$ can be viewed as a connection on a rank two complex vector bundle $E \to U$ equipped with a Hermitian metric. A purely imaginary gauge transformation on $Q$, written as

$$g = e^{ih}, \; h : U \to \mathfrak{su}(2)$$

can be viewed as a change of metric on $E$. Then the Chern connection associated to this new metric reads

$$g^* B = d + e^{-ih} \nabla_B e^{ih} - e^{-ih} \nabla_B e^{ih}.$$  

Here $\nabla_B + \bar{\nabla}_B$ is the covariant derivative on $E$ associated to $B$. This definition extends globally to all sections $h$ of $\text{ad}Q$, and induces another $SO(3)$-connection on $Q$, denoted by $g^* B$. The infinitesimal version of this action is

$$h \mapsto - * d_B h.$$  

Following the terminology of Duncan [Dun13, Dun], we define a nonlinear map which assign to each almost flat connection on the surface to a flat connection via a unique imaginary gauge transformation. For $p > 1$ and $\epsilon > 0$, define

$$\mathcal{A}^{1,p}_{\epsilon}(Q) = \left\{ B \in \mathcal{A}^{1,p}(Q) \mid \|F_B\|_{L^p(\Sigma)} < \epsilon \right\}$$

and

$$\mathcal{A}^{1,p}_{\text{flat}}(Q) = \left\{ B \in \mathcal{A}^{1,p}(Q) \mid F_B = 0 \right\}.$$
Lemma 2.15. Let \( p > 1 \). There exist constants \( \epsilon = \epsilon_p > 0 \) and \( C = C_p > 0 \) satisfying the following conditions. For each \( B \in \mathcal{A}^{1,p}_p(Q) \), there exists a unique complex gauge transformation of the form \( g = e^{ih}B \) where \( h_B \in W^{2,p}(\Sigma, \text{ad}(Q)) \) such that \( g^*B \in \mathcal{A}^{1,p}_p(Q) \). Moreover, \( h_B \in W^{2,p}(\Sigma, \text{ad}(Q)) \) is a \( C^1 \) function of \( B \in \mathcal{A}^{1,p}_p(Q) \) and there holds the estimate

\[
\|h_B\|_{W^{2,p}(\Sigma)} \leq C\|F_B\|_{L^p(\Sigma)}.
\] (2.11)

Proof. We use the implicit function theorem. Consider the map

\[
W^{2,p}(\Sigma, \text{ad}(Q)) \ni h \mapsto F_ghB \in L^p(\Sigma, \text{ad}(Q))
\]

where \( g = e^{ih} \). Its linearization at \( h = 0 \) is the linear operator

\[
\Delta_B := d_g^*d_B : W^{2,p}(\Sigma, \text{ad}(Q)) \to L^p(\Sigma, \text{ad}(Q)).
\]

This is a Fredholm operator of index zero. If \( B \) is flat, then Hypothesis 2.8 implies that \( \Delta_B \) is invertible. Since \( \Delta_B \) depends on \( B \in \mathcal{A}^{1,p}_p(Q) \) smoothly, and since \( R_\Sigma \) is compact, there is a constant \( C = C_p > 0 \) such that

\[
\|\Delta_B h\|_{L^p(\Sigma)} \geq 2C\|h\|_{W^{2,p}(\Sigma)}, \quad \forall B \in \mathcal{A}^{1,p}_p(Q).
\]

Then when \( \epsilon \) is sufficiently small, by Uhlenbeck’s weak compactness, any \( B \in \mathcal{A}^{1,p}_p(Q) \) is sufficiently close to a flat \( W^{1,p} \)-connection in the \( W^{1,p} \)-norm. It follows that

\[
\|\Delta_B h\|_{L^p(\Sigma)} \geq C\|h\|_{W^{2,p}(\Sigma)}, \quad \forall B \in \mathcal{A}^{1,p}_p(Q).
\]

Then by applying the implicit function theorem (see for example [MS04, Proposition A.3.4]), one obtains the unique \( h_B \) satisfying (2.11). The \( C^1 \)-dependence of \( h_B \) on \( B \) is also a consequence of the implicit function theorem. \( \square \)

Definition 2.16. Let \( p > 1 \) and \( \epsilon_p \) be the one in Lemma 2.15. Define the Narasimhan–Seshadri map

\[
NS_p : \mathcal{A}^{1,p}_p(Q) \to \mathcal{A}^{1,p}_p(Q)
\]

by

\[
NS_p(B) = (e^{ih}B)^*B.
\]

We know that each flat connection in \( \mathcal{A}^{1,p}_p(Q) \) is gauge equivalent via a gauge transformation of class \( W^{2,p} \) to a smooth flat connection. Then there is a homeomorphism

\[
\mathcal{A}^{1,p}_p(Q)/\mathbb{G}^{2,p}(Q) \cong R_\Sigma.
\]

The composition of \( NS_p \) with the projection \( \mathcal{A}^{1,p}_p(Q) \to R_\Sigma \) is denoted by

\[
\overline{NS}_p : \mathcal{A}^{1,p}_p(Q) \to R_\Sigma.
\]

By abusing names we still call \( \overline{NS}_p \) the Narasimhan–Seshadri map. It is well-known that the derivative \( D\overline{NS}_p \) is complex linear and annihilates infinitesimal complex gauge transformations. Therefore an ASD instanton over the product \( S \times Q \to S \times \Sigma \) with fibrewise small curvature projects down to a holomorphic curve in the representation variety via the Narasimhan–Seshadri map. This fact is stated in precise terms as follows.

Proposition 2.17. Let \( S \subset C \) be an open subset and \( A = d_S + \phi ds + \psi dt + B \) be a smooth connection on \( S \times Q \) satisfying the first equation of (2.4), i.e.

\[
A_s + *A_t = 0.
\]

Suppose for sufficiently small \( \epsilon \) one has \( B(z) \in \mathcal{A}^{1,2}_p(Q) \) for all \( z \in S \), then the map \( z \mapsto NS_2(B(z)) \) defines a holomorphic map \( u : S \to R_\Sigma \).
2.5.2. Energy identity for pseudoholomorphic curves in the representation variety. We show that the energy of holomorphic curves in the representation variety can be expressed in terms of the Chern–Simons functional. For simplicity we only discuss a special case but the general formula can be obtained using the same argument. Let \( P_{N^\pm} \to N^\pm \) be a pair of \( SO(3) \)-bundles over three-manifolds with boundary which satisfy Hypothesis 2.8 and Hypothesis 2.11. Let the domain of the holomorphic curve be the strip

\[
S_{[a,b]} := [a, b] \times [-1, 1]
\]

whose coordinates are \((s, t)\). Let \( \partial^\pm S_{[a,b]} \subset \partial S_{[a,b]} \) be the \( \pm 1 \) side of the boundary. Let \( u : S_{[a,b]} \to R_S \) be a holomorphic map and \( \gamma_\pm : \partial^\pm S_{[a,b]} \to L_{N^\pm} \) be continuous maps satisfying

\[
u|_{\partial^\pm S_{[a,b]}} = t_\pm \circ \gamma_\pm.
\]

For each \( s \in [a, b] \), one can define a piecewise smooth connection \( A_s \in \mathcal{A}^{p,s}(P_N) \) as follows. First, one can find a smooth path of smooth flat connections \( B_s(t) \in \mathcal{A}_{\text{flat}}(Q) \) parametrized by \( t \in [-1, 1] \) such that

\[
[B_s(t)] = u(s, t).
\]

Then there is a unique map \( \psi_s : [-1, 1] \to \Omega^0(\text{ad}Q) \) such that

\[
\frac{\partial B_s}{\partial t} - d_{B_s(t)} \psi_s = 0.
\]

Define

\[
A_{s,0} = dt + \psi_s(t) dt + B_s(t) \in \mathcal{A}(P_N|_{N_{\text{neck}}}).
\]

On the other hand, the boundary values \( \gamma_\pm \) allows one to find a pair \( A_{s,\pm} \in \mathcal{A}_{\text{flat}}(P_N|_{N^\pm}) \) such that

\[
[A_{s,\pm}] = \gamma(\pm s) \in \mathcal{A}(P_{N^\pm})/\mathcal{G}(P_{N^\pm}), A_{s,\pm}|_{\partial N^\pm} = B_s(\pm 1).
\]

Then \( A_{s,-}, A_{s,0}, A_{s,+} \) define a piecewise smooth connection \( A_s \in \mathcal{A}^{p,s}(P_N) \). It is easy to show that its gauge equivalence class is well-defined.

To the strip one can associate a four-manifold with boundary as follows. Define

\[
N_{[a,b]} := (S_{[a,b]} \times \Sigma) \cup (\partial^+ S_{[a,b]} \times N^+) \cup (\partial^- S_{[a,b]} \times N^-)
\]

where we glue along the common boundaries \( \partial^\pm S_{[a,b]} \times \Sigma \). One defines an \( SO(3) \)-bundle \( P \to N_{[a,b]} \) as

\[
P = (S_{[a,b]} \times Q) \cup (\partial^+ S_{[a,b]} \times P_{N^+}) \cup (\partial^- S_{[a,b]} \times P_{N^-}).
\]

The restriction of \( P \) to each boundary component of \( N_{[a,b]} \) is isomorphic to \( P_N \to N \). Then one can define the Chern–Simons action

\[
CS : \left( \mathcal{A}^{p,s}(P_N)/\mathcal{G}^{p,s}(P_N) \right)^2 \to \mathbb{R}
\]

defined by

\[
CS([A_a], [A_b]) = \frac{1}{2} \int_{N_{[a,b]}} \text{tr}(F_A \wedge F_A) \tag{2.13}
\]

where \( A \) is any piecewise smooth connection on \( N_{[a,b]} \) extending the boundary values \([A_a]\) and \([A_b]\). One has the following formula for holomorphic maps defined over the strip.

**Proposition 2.18.** The energy of \( u \) is equal to \( CS([A_a], [A_b]) \).
Proof. For completeness we give a proof. We define a piecewise smooth connection $A$ on $N_{[a,b]}$ as follows. Since $u$ is smooth, one can find a family of smooth connections $B(z) \in \mathcal{A}(Q)$ parametrized smoothly by $z \in S_{[a,b]}$ such that

$$[B(z)] = u(z).$$

Since $R_\Sigma$ is a free symplectic quotient, there are unique $\phi(z), \psi(z) \in \Omega^0(\text{ad}Q)$ such that

$$\partial_s B(z) - d_{B(z)} \phi(z) \in \ker d_{B(z)} \cap \ker^*_{B(z)}, \quad \partial_t B(z) - d_{B(z)} \psi(z) \in \ker d_{B(z)} \cap \ker^*_{B(z)}.$$

Then define

$$A_1 = ds + \phi(z) ds + \psi(z) dt + B(z)$$

which is a smooth connection on $P$ restricted to $S_{[a,b]} \times \Sigma$. Since the map $z \mapsto [B(z)]$ is holomorphic, one has

$$\partial_s B(z) - d_{B(z)} \phi(z) + * (\partial_t B(z) - d_{B(z)} \psi(z)) = 0.$$ 

Moreover, since the linear map $D\overline{N}_2 : \ker d_{B(z)} \cap \ker^*_{B(z)} \to T_{[B(z)]} R_\Sigma$ is an isometry, we have

$$|\partial_s u|^2 = |\partial_s B - d_B \phi(z)|^2 = \int_{\Sigma} \text{tr} \left( (\partial_s B - d_B \phi) \wedge (\partial_t B - d_B \psi) \right).$$

(2.14)

On the other hand, for each boundary point $(s, \pm 1) \in \partial \pm S_{[a,b]}$, using the map $\gamma_{\pm}$ one can find a family of flat connections $A_{s,\pm} \in \mathcal{A}_{\text{flat}}(P_N)$ parametrized smoothly by $s$ such that

$$[A_{s,\pm}] = \gamma_{\pm}(s) \forall (s, \pm 1) \in \partial \pm S_{[a,b]}.$$ 

Then define

$$A_\pm = ds + A_{s,\pm} \in \mathcal{A}(\partial \pm S_{[a,b]} \times \partial N_{\pm}).$$

This is a smooth connection on $P$ restricted to $\partial \pm S_{[a,b]} \times \partial N$. Then $A_1$ and $A_\pm$ together define a piecewise smooth connection $A$ on $P$.

We calculate the Chern–Simons action using $A$. Indeed, since $A_{s,\pm}$ is flat, one has

$$\text{tr}(F_A \wedge F_A)|_{\partial \pm S_{[a,b]} \times \partial N_{\pm}} = 0.$$ 

On the other hand, over $S_{[a,b]} \times \Sigma$ one has

$$F_A = F_{B(z)} + ds \wedge (\partial_s B(z) - d_B \phi) + dt \wedge (\partial_t B(z) - d_B \psi) + (\partial_s \psi - \partial_t \phi + [\phi, \psi]) \wedge dsdt.$$ 

Since $F_{B(z)} \equiv 0$, one has

$$\text{tr}(F_A \wedge F_A) = \text{tr} \left( (\partial_s B - d_B \phi) \wedge (\partial_t B(z) - d_B \psi) \right) \wedge dsdt.$$ 

By (2.14), we see

$$\text{CS}([A_u], [A_b]) = \frac{1}{2} \int_{N_{[a,b]}} \text{tr}(F_A \wedge F_A) = \int_{S_{[a,b]}} |\partial_s u|^2 dsdt = E(u).$$

\[ \square \]

3. The rescaled equation and interior compactness

In this section we consider the ASD equation over the product of two surfaces. In the adiabatic limit of the rescaled version, we recall an interior estimate by Dostoglou–Salamon [DS94, DS07] which leads to the compactness modulo bubbling results (Theorem 3.3). We also give a refined version of the interior estimate near the boundary (Corollary 3.6) which will be useful in the next section.
3.1. An interior estimate for the rescaled equation. First we recall the notion of the rescaled ASD equations. Let \( \rho \) be a positive number. Let \( S \subset C \) be an open subset. Let

\[
\varphi_\rho : C \to C
\]

be the multiplication \( z \mapsto \rho z \). Recall that a connection \( A \) on \( \varphi_\rho(S) \times Q \) can be written in components as

\[
A = d\varphi_\rho(s) + \phi ds + \psi dt + B.
\]

Denote

\[
A' = \varphi_\rho^* A = dS + \phi' ds + \psi' dt + B'.
\]

Recall the notations introduced in (2.2) and (2.3). Then the ASD equation on \( A' \)

\[
A'_s + \ast A'_t = 0, \quad \ast \kappa A' + \rho^2 \mu A' = 0. \tag{3.1}
\]

We call this equation the \( \rho \)-\( \text{ASD} \) equation. Define the rescaled energy density function

\[
e_\rho(z) := \|A'_s\|_{L^2((z) \times \Sigma)}^2 + \rho^2 \|\mu A'\|_{L^2((z) \times \Sigma)}^2 \tag{3.2}
\]

and the rescaled energy

\[
E_\rho(A'; S) := \int_S e_\rho(z) ds dt.
\]

The basic relation between the rescaled energy density and the original energy density is

\[
e_\rho(z) = \rho^2 \|F A\|_{L^2((\rho z) \times \Sigma)}^2, \quad \forall z \in S.
\]

Therefore we have

\[
E(A; \varphi_\rho(S)) = \|F A\|_{L^2(\varphi_\rho(S) \times \Sigma)}^2 = E_\rho(A'; S).
\]

When the set \( S \) is understood from the context, we abbreviate \( E_\rho(A'; S) \) by \( E_\rho(A') \). We also introduce another density in terms of fibrewise \( L^\infty \)-norm:

\[
e_\rho^{\infty}(z) := \|A'_s\|_{L^\infty((z) \times \Sigma)}^2 + \rho^2 \|\mu A'\|_{L^\infty((z) \times \Sigma)}^2.
\]

In [DS94, DS07] Dostoglou–Salamon obtain several important estimates about the rescaled equation. These estimates are essential in proving the convergence of rescaled instantons towards holomorphic curves. Here we recall one of their estimates as follows.

**Theorem 3.1.** [DS07, Corollary 1.1] Let \( S \subset C \) be an open set, \( K \subset S \) be a compact subset, and \( c_0 > 0 \). Then there exist constants \( C > 0 \) and \( T > 0 \) such that the following holds. Let \( A' = (B', \phi', \psi') \) be a solution to (3.1) over \( S \times \Sigma \) with \( \rho \geq T \) satisfying

\[
 \sup_{z \in S} e_\rho(z) = \sup_{z \in S} \left( \|A'_s\|_{L^2(\Sigma)}^2 + \rho^2 \|\mu A'\|_{L^2(\Sigma)}^2 \right) \leq c_0. \tag{3.3}
\]

Then there holds

\[
\|A'_s\|_{L^\infty(K \times \Sigma)} + \rho \|\mu A'\|_{L^\infty(K \times \Sigma)} \leq C \sqrt{E_\rho(A'; S)}.
\]

**Remark 3.2.** Dostoglou–Salamon’s estimates are valid for the case that the \( \rho \)-\( \text{ASD} \) equation has a type of holonomic perturbation term. Such a perturbation induces a family of Hamiltonian function on \( R_\Sigma \). See [DS94, Section 7] for more details.
3.2. Adiabatic limit. Now we consider the adiabatic limit of the rescaled ASD equation towards holomorphic maps in the representation variety. Let $S \subset C$ be an open subset and $S_i \subset S$ be an exhausting sequence of open subsets. The conclusions of the following theorem is well-known and is essentially a consequence of Dostoglou–Salamon’s estimates in [DS94, Section 7] and [DS07].

**Theorem 3.3.** (cf. [DS94, Theorem 9.1] [Dun, Lemma 3.5] [Nis10, Theorem 1.2]) Let $\rho_i \to +\infty$ be a sequence of positive numbers diverging to infinity. Let $S \subset C$ be an open subset and $S_i \subset S$ be an exhausting sequence of open subsets. Let $A_i^\rho = dS + \phi_i^\rho ds + \psi_i^\rho dt + B_i^\rho(z)$ be a sequence of solutions to the $\rho_i$-ASD equation over $S_i \times \Sigma$. Let $e_{\rho_i} : S_i \to [0, +\infty)$ the rescaled energy density function of $A_i^\rho$. Suppose

$$\lim_{i \to \infty} \sup E_{\rho_i}(A_i^\rho) < \infty.$$  

Then there exists a subsequence (still indexed by $i$) and a holomorphic map with mass 

$$(u_\infty, W_\infty, m_\infty) : S \to R_\Sigma$$

satisfying the following condition.

(a) For every point $z \in S$ there holds

$$\lim_{r \to 0} \lim_{i \to \infty} E_{\rho_i}(A_i^\rho; B_r(z) \times \Sigma) = m_\infty(z).$$

(b) For any precompact open subset $K \subset S \setminus W_\infty$, for $i$ sufficiently large, $B_i^\rho(z)$ is contained in the domain of the Narasimhan–Seshadri map $\overline{\text{NS}}_2$ for all $z \in K$, hence $A_i^\rho$ induces a holomorphic map

$$u_i^\rho : K \to R_\Sigma, \quad u_i^\rho(z) := \overline{\text{NS}}_2(B_i^\rho(z)).$$

Moreover, the sequence of maps $u_i^\rho$ converges to $u_\infty$ in $C^\infty(K)$ and the energy density $e_{\rho_i}$ converges to the energy density of $u_\infty$ in $C^0(K)$.

(c) For any compact subset $K \subset S$ containing $W_\infty$, there holds

$$\lim_{i \to \infty} E_{\rho_i}(A_i^\rho; K \times \Sigma) = E(u_\infty; K) + \int_K m_\infty.$$

In particular, $W_\infty = \emptyset$ if and only if for all compact $K \subset S$ there holds

$$\lim_{i \to \infty} \sup \| e_{\rho_i} \|_{L^\infty(K)} < \infty.$$

**Definition 3.4.** Let $S \subset C$ be an open subset and $S_i \subset S$ be an exhaustive sequence of open subsets. Let $\rho_i \to +\infty$ be a sequence of positive numbers. Let $A_i$ be a sequence of solutions to the ASD equation over $\varphi_{\rho_i}(S_i) \times \Sigma$. Let $\tilde{u}_\infty = (u_\infty, W_\infty, m_\infty)$ be a holomorphic map from $S$ to $R_\Sigma$ with mass. We say that $A_i$ converges to $\tilde{u}_\infty$ along $\{\rho_i\}$ if for the corresponding sequence of solutions $A_i^\rho$ of the $\rho_i$-ASD equation over $S_i \times \Sigma$, conditions (a), (b), and (c) of Theorem 3.3 are satisfied.

3.3. Estimates near the boundary. In this section we extend certain estimates of Dostoglou–Salamon near the boundary of the domain. Such an extension will be useful in the next section. Let $B_R = B_R(0) \subset C$ be the radius $R$ open disk centered at the origin.

**Lemma 3.5.** There exist $\epsilon > 0$, $T > 0$, and $C > 0$ satisfying the following conditions. Let $A = A_\infty = dC + \phi ds + \psi dt + B(z)$ be a solution to the $\rho$-ASD equation over $B_1 \times \Sigma$ satisfying $\rho \geq T$, $E_\rho(A) \geq \epsilon$ and

$$\sup_{B_1} e_\rho \leq \rho^2.$$
Then for all \( r \in (0, 1) \) there holds
\[
\sup_{B_r} \varepsilon_{\rho}^\infty \leq \frac{C}{(1-r)^2}.
\] (3.5)

**Proof.** Suppose this is not the case. Then there exist a sequence \( \rho_i \to \infty \), a sequence of solutions \( A_i \) to the \( \rho_i \)-ASD equation over \( B_1 \times \Sigma \), and a sequence of points \( z_i \in B_1 \) such that \( E_{\rho_i}(A_i) \to 0 \), \( \sup \varepsilon_{\rho_i}^\infty \leq \rho_i^2 \), and
\[
\lim_{i \to \infty} (1-r_i)^2 \varepsilon_{\rho_i}^\infty(z_i) = \infty, \quad \text{where } r_i = |z_i|.
\] (3.6)

Then it follows that
\[
\rho_i^2 \geq \varepsilon_{\rho_i}^\infty(z_i) \geq \frac{1}{(1-r_i)^2}.
\]

Here \( a_i \gg b_i \) (equivalently \( b_i \ll a_i \)) means \( a_i/b_i \to +\infty \). Denote \( \tilde{\rho}_i := \rho_i(1-r_i) \) which diverges to infinity. Then define the rescaling
\[
\varphi_i : B_1 \to B_{1-r_i}(z_i), \quad w \mapsto z_i + (1-r_i)w.
\]

The sequence \( A_i \) is pulled back to a sequence of solutions \( \tilde{A}_i \) to the \( \tilde{\rho}_i \)-ASD equation over \( B_1 \times \Sigma \) whose energy converges to zero. On the other hand, (3.6) implies that
\[
\varepsilon_{\tilde{\rho}_i}^\infty(0) \to +\infty.
\] (3.7)

By Theorem 3.3, this sequence converges to the constant holomorphic map, implying that \( \varepsilon_{\tilde{\rho}_i} \) converges to zero uniformly over compact subsets of \( B_1 \). This contradicts the interior estimate of Theorem 3.1 applied to \( \tilde{A}_i \) and (3.7). \( \square \)

There is also a more refined case of the inequality (3.1) of Theorem 3.1 which will be useful in the next section. Denote the open half disk in the upper half plane by
\[
B_r^+ := B_r^+(0) := \{ z \in \mathbb{C} \mid \mathrm{Im} z \geq 0, \ |z| < r \}.
\]

**Corollary 3.6.** There exist \( \epsilon > 0 \), \( C > 0 \), \( T > 0 \) satisfying the following conditions. Suppose \( A = dA + \phi ds + \psi dt + B(z) \) is a solution to the \( \rho \)-ASD equation over \( \text{Int} B_r^+ \times \Sigma \) for \( \rho \geq T \) satisfying
\[
E_{\rho}(A) \leq \epsilon, \quad \sup_{z \in \text{Int} B_r^+} \| F_B(z) \|_{L^2(\Sigma)} \leq \epsilon, \quad \sup_{B_r^+} \varepsilon_{\rho}^\infty \leq \rho^2.
\]

Then for all \( z = s + it \in \text{Int} B_r^+ \) there holds
\[
\| A_s \|_{L^2((z) \times \Sigma)} + \rho \| \mu A \|_{L^2((z) \times \Sigma)} \leq \frac{C}{T} \sqrt{E_{\rho}(A; B_t(z) \times \Sigma)}.
\] (3.8)

**Proof.** The proof is essentially from the proof of [DS94, Theorem 7.1]. Define two functions \( h_0, h_1 : \text{Int} B_r^+ \to [0, +\infty) \) by
\[
h_0(z) = \frac{1}{2} \left( \| A_s \|_{L^2((z) \times \Sigma)}^2 + \rho^2 \| F_B(z) \|_{L^2(\Sigma)}^2 \right) = \frac{1}{2} \varepsilon_{\rho}(z),
\]
\[
h_1(z) = \frac{1}{2} \left( \rho^2 \| d_B(z) A_s \|_{L^2(\Sigma)}^2 + \| d_B(z) \ast A_s \|_{L^2(\Sigma)}^2 + \| \nabla_s A_s \|_{L^2(\Sigma)}^2 + \| \nabla_t A_s \|_{L^2(\Sigma)}^2 \right.
\]
\[
\left. + \| d_B(z) \kappa A \|_{L^2(\Sigma)}^2 + \rho^{-2} \| \nabla_s \kappa A \|_{L^2(\Sigma)}^2 + \rho^{-2} \| \nabla_t \kappa A \|_{L^2(\Sigma)}^2 \right).
\]

Then by the explicit calculation in [DS94, p. 616], one obtains
\[
\Delta h_0 = 2h_1 + 5 \langle A_s, [A_s \wedge \kappa A] \rangle \geq \| d_B(z) \kappa A \|_{L^2(\Sigma)}^2 + 5 \langle A_s, [A_s \wedge \kappa A] \rangle.
\]
When $\epsilon$ is small enough, the condition $\|F_{B(z)}\|_{L^2(\Sigma)} \leq \epsilon$ implies that $B(z)$ is sufficiently close to a flat connection on $Q \to \Sigma$ in the $L^2$-metric. Since there is no reducible flat connections on $Q$ and $R^2_\Sigma$ is compact, there is a constant $a > 0$ such that

$$\|d_{B(z)} \kappa_A\|_{L^2(\Sigma)} \geq \frac{1}{a} \|\kappa A\|_{L^2(\Sigma)}.$$ 

On the other hand, by Lemma 3.5, when $\rho$ is sufficiently large and the total energy of $A$ is sufficiently small, for some $C > 0$ there holds

$$\|A_s\|_{L^2(\Sigma)} \leq C \rho^{-1}, \forall z = s + it \in \text{Int} B^+_1.$$

Hence

$$\langle A_s, [A_s \wedge \kappa_A] \rangle \leq \frac{Ca}{t}\|d_{B(z)} \kappa_A\|_{L^2(\Sigma)} \|A_s\|_{L^2(\Sigma)}.$$

Therefore, for a suitably modified value of $C$ there holds

$$\Delta h_0 \geq -\frac{C}{2t^2} \|A_s\|_{L^2(\Sigma)}^2 \geq -\frac{C}{t^2} h_0.$$ 

Then by Lemma 3.7 below, one has

$$h_0(z) \leq \frac{C}{t^2} \int_{B_t(z)} h_0 = \frac{C}{t^2} E_\rho(A; B_t(z) \times \Sigma).$$

By the definition of $h_0$, (3.8) follows. \qed

The following mean value estimate was used in the above proof.

**Lemma 3.7.** [DS94, Lemma 7.3] Let $u : \overline{B}_r \to [0, +\infty)$ be a $C^2$-function satisfying

$$\Delta u \geq -au, \text{ where } a \geq 0.$$

Then there holds

$$u(0) \leq \frac{2}{\pi} \left( a + \frac{4}{t^2} \right) \int_{B_r} u.$$ 

4. **The isoperimetric inequality**

It is not too far away from results of the previous section to the compactification of the moduli space of instantons over the product of the complex plane and the compact surface, namely Theorem 1.5. For the other type of noncompact four-manifold, namely the product $R \times M$ where $M$ is the three-manifold with cylindrical end, there is another difficulty. As approaching to the infinity of the $R$-direction, the ASD equation over the compact part $R \times M_0$ is almost like a Lagrangian boundary condition for the part ASD equation over $H \times \Sigma$. In the theory of holomorphic curves, the Lagrangian boundary condition allows one to extend interior elliptic estimates to the boundary which leads to compactness near the boundary. However, in our situation, the failure of being an actual Lagrangian boundary condition prevents one to have elliptic estimates near the boundary, at least not in a straightforward way. Indeed, one should view the “seam” $R \times \Sigma$ between $R \times M_0$ and $H \times \Sigma$ as giving a Lagrangian seam condition given by an infinite dimensional Lagrangian correspondence (see [WW15]). One possible approach of our compactness problem would be based on certain hard estimates as did in [BW18] for finite dimensional holomorphic quilts (see another approach in [Dun, Section 4]).

In this paper, instead, we take a less analytic approach. The main idea is to view the ASD equation over $R \times M$ as a gradient line of the Chern–Simons functional on the closed three-manifold $M^{\text{double}}$ with respect to a time-dependent metric. The asymptotic behavior as well as the compactness problem over the part $R \times M_0$ can both be treated via an isoperimetric inequality. Roughly speaking, for a closed Riemannian three-manifold
N, if the Chern–Simons functional for connections on an $SO(3)$-bundle $P \to N$ is Morse or Morse–Bott, then for an almost flat connection $A \in \mathcal{A}(P)$, one can define a “local Chern–Simons action,” which is the Chern–Simons action of $A$ relative to a certain nearby flat connection. The isoperimetric inequality says that the local action can be controlled by $\|F_A\|_{L^2(L)}^2$. This is analogous to the isoperimetric inequality in symplectic geometry (see [MS04, Section 4.4] [Poz94, Chapter 3]).

In this section we prove the isoperimetric inequality and certain monotonicity properties of solutions to the ASD equation (which we call the annulus lemma). We will also derive a diameter estimate which will be needed for the compactness theorem as well as the asymptotic behavior and energy quantization property of instantons over $R \times M$.

4.1. The isoperimetric inequality. In this subsection we derive the isoperimetric inequality. Let $N^-, N^+$ be three-manifold with boundary with diffeomorphic boundary and $P_{N^\pm} \to N^\pm$ be $SO(3)$-bundles with isomorphic boundary restrictions, all of which satisfy Hypothesis 2.8 and Hypothesis 2.11. Let $N$ be the closed three-manifold obtained by gluing $N^-, (N^+)\op$ and a neck region $N_{\text{neck}} = [-1, 1] \times \Sigma$ and $P_N \to N$ be the glued $SO(3)$-bundle. For any piecewise smooth connection $A \in \mathcal{A}^{p,s}(P_N)$, if its restriction to the neck region is

$$A|_{[-1, 1] \times \Sigma} = d_t + \eta(t)dt + B(t)$$

where $d_t$ is the exterior differentiation in the coordinate $t \in [-1, 1]$, we define the $L^2$-length of $A$ to be

$$l_2(A) := \int_{-1}^1 \|B'(t) - d_B(t)\eta(t)\|_{L^2(\Sigma)} dt.$$ 

**Lemma 4.1.** Given $p \geq 2$, there exist $\epsilon > 0$ and $C > 0$ satisfying the following condition. Let $A$ be a piecewise smooth connection on $P_N$ whose restriction to $N_{\text{neck}} = [-1, 1] \times \Sigma$ is

$$A|_{[-1, 1] \times \Sigma} = d_t + \eta(t)dt + B(t).$$

Suppose

$$\|F_A\|_{L^p(N^- \cup N^+)} + l_2(A) \leq \epsilon. \tag{4.1}$$

Then there exists a piecewise smooth flat connection $A^* \in \mathcal{A}^{p,s}_{\text{flat}}(P_N)$ such that

$$\|A - A^*\|_{W^{1,p}(N^- \cup N^+)} \leq C \left( \|F_A\|_{L^p(N)} + l_2(A) \right). \tag{4.2}$$

Moreover, $A^*$ is gauge equivalent to another connection $A_0 \in \mathcal{A}^{p,s}_{\text{flat}}(P_N)$ satisfying

(a) The restriction of $A_0$ to $N_{\text{neck}}$ is $d_t + \eta(t)dt + B_0(t)$.

(b) There holds

$$\|B(t) - B_0(t)\|_{L^2(\Sigma)} \leq C \left( \|F_A\|_{L^p(N^- \cup N^+)} + l_2(A) \right) \tag{4.3}$$

and

$$\|A - A_0\|_{L^3(N^- \cup N^+)} \leq C \left( \|F_A\|_{L^p(N^- \cup N^+)} + l_2(A) \right). \tag{4.4}$$

**Proof.** First, notice that the statement of this lemma is gauge invariant. Hence we may assume that $\eta(t) = 0$ and look for a flat connection whose neck restriction has no $dt$ component. Then for any pair $a, b \in [-1, 1]$ there holds

$$\|B(b) - B(a)\|_{L^2(\Sigma)} = \int_a^b \|B'(t)\|_{L^2(\Sigma)} dt \leq l_2(A).$$
Let $C > 0$ be a constant which is independent of $A$ but may vary in the context. Let $\epsilon$ be smaller than the $\epsilon_p$ of Lemma 2.13. Then by Lemma 2.13 there exist $W^{1,p}$-connections $A_\pm$ on $N^\pm$ such that

$$\|A - A_\pm\|_{W^{1,p}(N^\pm)} \leq C\|F_A\|_{L^p(N^\pm)}.$$ 

We can replace $A_\pm$ by nearby smooth connections such that this bound is still valid for a slightly larger $C$. Let $B_\pm$ be the boundary restrictions of $A_\pm$ over $\partial N^\pm \cong \Sigma$. Then

$$\|B_\pm - B(\pm 1)\|_{L^2(\Sigma)} \leq C\|B_\pm - B(\pm 1)\|_{L^p(\Sigma)} \leq C\|A - A_\pm\|_{W^{1,p}(N^\pm)} \leq C\|F_A\|_{L^p(N^\pm)}.$$ 

On the other hand, one has

$$\|B_+ - B_-\|_{L^2(\Sigma)} \leq \|B_+ - B(1)\|_{L^2(\Sigma)} + \|B(1) - B(-1)\|_{L^2(\Sigma)} \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right). \quad (4.5)$$

Let $d_{R_\Sigma}$ be the distance function on $R_\Sigma$ induced from the $L^2$-metric. The estimate (4.5) implies that for some $C > 0$,

$$d_{R_\Sigma}\left(([B_-], [B_+]), \Delta_{R_\Sigma}\right) \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right).$$

The immersions $\iota_{\pm}: L_{N^\pm} \to R_\Sigma$ pulls back a metric which induces a distance function $d_{L_{N^\pm}}$ on $L_{N^\pm}$. Let $d_{L_{N^\pm} \times L_{N^+}}$ be the product metric on $L_{N^\pm} \times L_{N^+}$. Since $\iota_-$ and $\iota_+$ intersect cleanly, there holds

$$d_{L_{N^\pm} \times L_{N^+}}\left(([A_-], [A_+]), (\iota_-, \iota_+)^{-1}(\Delta_{R_\Sigma})\right) \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right).$$

Moreover, the topology on $L_{N^\pm}$ induced by $d_{L_{N^\pm}}$ is the same as the topology induced from the Banach space topology on $A^{1,p}(P_{N^\pm})$. Hence one can find a pair of smooth flat connections $A^*_+ $ on $N^\pm$ such that

$$\|A_\pm - A^*_+\|_{W^{1,p}(N^\pm)} \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right) \quad (4.6)$$

and such that

$$[A^*_+ |_{\partial N^-}] = [A^*_+ |_{\partial N^+}] \in R_\Sigma.$$ 

Denote $B^*_\pm := A^*_\pm |_{\partial N^\pm}$. Then by (4.5) and (4.6) one also has

$$\|B^*_+ - B^*_\mp\|_{L^2(\Sigma)} \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right). \quad (4.7)$$

Since $B^*_-$ and $B^*_+$ are gauge equivalent and the group of gauge transformations $G(Q)$ is connected (since we only take $SU(2)$-valued gauge transformations), there is a gauge transformation $g$ relating $B^*_-$ and $B^*_+$ which can be homotoped to the identity. Such a homotopy can be used to extend $A^*_+$ to a piecewise smooth flat connection $A^* \in A^{0,\text{flat}}(P_N)$ which satisfies (4.2).

Now we would like to modify $A^*$ via a suitable gauge transformation so that (4.3) and (4.4) holds. Indeed, the gauge transformation $g$ can be written as $g = e^h$ and (4.7) implies that

$$\|h\|_{W^{1,2}(\Sigma)} \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right). \quad (4.8)$$

Recall that for any smooth manifold $Y$ with boundary $\partial Y$, the boundary restriction defines a trace operator

$$H_s(Y) \to H^{s-\frac{1}{2}}(\partial Y)$$

which admits a bounded right inverse. Therefore, by the bound (4.8) there exists a smooth extension of $h$ to $N^+$, denoted by $h_+$ such that

$$\|h_+\|_{H^{s+\frac{1}{2}}(N^+)} \leq C\left(\|F_A\|_{L^p(N^{-\cup N^+})} + l_2(A)\right).$$
Then we can define a piecewise smooth flat connection

\[ A_0 = \begin{cases} 
A^*, & N^-, \\
B^*, & [-1, 1] \times \Sigma, \\
(e^{\lambda})^* A^+, & N^+
\end{cases} \]

Then by the Sobolev embedding \( W^{1,2}(N^-) \to L^3(N^-) \) and the Sobolev embedding (for fractional Sobolev norms, see [NPV, Theorem 6.5]) \( H^{1/2}(N^+) = W^{1/2,2}(N^+) \to L^3(N^+) \), there holds

\[ \|A - A_0\|_{L^3(N^- \cup N^+)} \leq C \left( \|F_A\|_{L^p(N^- \cup N^+)} + l_2(A) \right). \]

The condition (4.3) is also easy to verify.

The above lemma allows one to define a local Chern–Simons action.

**Definition 4.2.** Let \( A \) be a piecewise smooth connection on \( P_N \) satisfying (4.1). Then there exists a nearby flat connection \( A_0 \) on \( P_N \) satisfying properties of Lemma 4.1. Then we define

\[ F_{\text{loc}}(A) = CS([A^*], [A]) \]

where \( CS([A^*], [A]) \) is the Chern–Simons action of \([A]\) relative to \([A^*]\).

It is easy to verify that the value of the local Chern–Simons functional is independent of the choice of the reference flat connection as long as the reference connection belongs to the same connected component \( L_N \). We denote

\[ F_{\text{loc}}(A) = F_Z(A) \]

where \( Z \subset L_N \) is the connected component of the nearby flat connection \([A^*]\).

Lemma 4.1 allows us to derive the following isoperimetric inequality. The proof of this inequality (for the SU(2) case) has appeared in [Flo88a] for the case that \( L_N \) is a transverse intersection (see also [Don02, Proof of Theorem 4.2]) and in the proof of [Fuk96, Lemma 7.13] for the case that \( L_N \) is a clean intersection.

**Theorem 4.3 (Isoperimetric Inequality).** Given \( p \geq 2 \), there exist \( \epsilon > 0 \) and \( C > 0 \) such for all piecewise smooth connection \( A \) on \( P_N \) satisfying (4.1), there holds

\[ |F_{\text{loc}}(A)| \leq C \left( \|F_A\|_{L^p(N^- \cup N^+)} + l_2(A)^2 \right). \]

**Proof.** By Lemma 4.1, there is a nearby flat connection \( A_0 \) and the local action is defined by Definition 4.2. Then by Lemma 4.1, (4.4), and formula (2.1), one has

\[
\int_{N^\pm} \text{tr} \left[ \frac{1}{2} d_{A_0} (A - A_0) \wedge (A - A_0) + \frac{1}{3} (A - A_0) \wedge (A - A_0) \wedge (A - A_0) \right] \\
\leq C \|F_A\|_{L^2(N^\pm)} \|A - A_0\|_{L^3(N^\pm)} + C \|A - A_0\|_{L^3(N^\pm)}^3 \\
\leq C \left( \|F_A\|_{L^p(N^- \cup N^+)}^2 + l_2(A)^2 \right).
\]
Using the fact that $A - A_0|_{N_{neck}} = B(t) - B_0$ and (4.3), one has
\[
\int_{N_{neck}} \text{tr} \left[ \frac{1}{2} d_{A_0} (A - A_0) \wedge (A - A_0) + \frac{1}{3} (A - A_0) \wedge (A - A_0) \wedge (A - A_0) \right] \\
= \int_{[-1,1] \times \Sigma} \text{tr} \left[ \frac{1}{2} d_{A_0} (A - A_0) \wedge (A - A_0) \right]
\leq C \int_{[-1,1] \times \Sigma} |B'(t)| |B(t) - B_0| dt \Sigma \leq C \int_{-1}^{1} \|B'(t)\|_{L^2(\Sigma)} \|B(t) - B_0\|_{L^2(\Sigma)} dt
\leq C \left( \|F_A\|_{L^p(N^{-\cup}N^+)} + l_2(A) \right) \int_{-1}^{1} \|B'(t)\|_{L^2(\Sigma)} dt
\leq C \left( \|F_A\|^2_{L^p(N^{-\cup}N^+)} + l_2(A)^2 \right).
\]
Then (4.9) follows. \hfill \square

In practice it is more convenient to use the following version of the isoperimetric inequality. Notice that the $L^2$-length $l_2(A)$ can be bounded by $\|F_A\|_{L^2(N_{neck})}$. So the $p = 2$ case of Lemma 4.1 and Theorem 4.3 imply the following more standard form of the isoperimetric inequality. Indeed, by using the fact that the moduli space $L_N$ is a clean intersection, the following version follows from an estimate of the Hessian of the Chern–Simons functional (see [Fuk96]).

**Theorem 4.4 (Isoperimetric Inequality).** There exist $\epsilon > 0$ and $c_p > 0$ satisfying the following condition. Let $A$ be a piecewise smooth connection on $P_N \to N$ satisfying
\[\|F_A\|^2_{L^2(N)} \leq \epsilon.\]
Then the local Chern–Simons action $F_{loc}(A)$ is well-defined and there holds
\[|F_{loc}(A)| \leq c_p \|F_A\|^2_{L^2(N)}\]

### 4.2. The annulus lemma.

Now we turn to the decay property of the energy for ASD instantons. For $0 < r < R < \infty$ define the open annulus and the half annulus by
\[\text{Ann}(r, R) := \{ z \in C \mid r < |z| < R \}, \quad \text{Ann}^+(r, R) = \text{Ann}(r, R) \cap H.\]
The half annulus has two boundary components
\[\partial^\pm \text{Ann}^+(r, R) := \{ s \in \partial H \cong R \mid r < \pm s < R \}.\]
Define a noncompact four-manifold $N_{r,R}$ as
\[N_{r,R} = (\text{Ann}^+(r, R) \times \Sigma) \cup (\partial^+ \text{Ann}^+(r, R) \times N^+) \cup (\partial^- \text{Ann}^+(r, R) \times N^-).\]
Equip $N_{r,R}$ with the product metric. The bundle $P_N \to N$ also extends to a bundle $P \to N_{r,R}$.

It is convenient to use the polar coordinates over the neck region. We identify the neck region with
\[N_{neck} \cong [-1, 1] \times \Sigma \cong [0, \pi] \times \Sigma\]
with the coordinate $t \in [-1, 1]$ corresponding to $\theta := \frac{\pi}{2}(t + 1) \in [0, \pi]$. Suppose $A$ is a smooth connection on $N_{r,R}$. One can identify $A$ with a piecewise smooth connection on the product $(\log r, \log R) \times N$ as follows. Over $(\log r, \log R) \times N^\pm$, define $A'$ as the pullback of $A$ via the diffeomorphism
\[(\log r, \log R) \times N^\pm \cong \partial^\pm \text{Ann}^+(r, R) \times N^\pm, \quad (\tau, x) \mapsto (e^\tau, x).\]
Over \((\log r, \log R) \times [0, \pi] \times \Sigma\), define \(A'\) as the pullback of \(A\) via the diffeomorphism 
\((\log r, \log R) \times [0, \pi] \times \Sigma \to \text{Ann}^+ (r, R) \times \Sigma, \ (\tau, \theta, w) \to (e^{\tau + i\theta}, w)\).
Then \(A'\) is a piecewise smooth connection, giving a family of piecewise smooth connections \(A_{e^r}\) on the bundle \(P_N \to N\) parametrized by \(\rho = e^\tau \in (r, R)\). Then we can write 
\[A_{\rho}|_{[0,\pi] \times \Sigma} = d\theta + \eta_\rho(\theta) d\theta + B_\rho(\theta).\]
It is easy to see that if \(A\) is a solution to the ASD equation over \(N_{r,R}\), then 
\[E(A; N_{r,R}) = \int_{\log r}^{\log R} \left( e^\tau \|F_{A_{e^\tau}}\|^2_{L^2(N^- \cup N^+)} + \|\frac{\partial B_{e^\tau}}{\partial \theta} - dB_{e^\tau} \eta_{e^\tau}\|^2_{L^2(N_{\text{neck}})} + e^{2\tau} \|F_{B_{e^\tau}}\|^2_{L^2(N_{\text{neck}})} \right) d\tau.\]
Then one has 
\[E(A; N_{r,R}) \geq \int_{\log r}^{\log R} \|F_{A_{e^\tau}}\|^2_{L^2(N)} d\tau, \ \forall R > r \geq 1. \quad (4.10)\]
Moreover, denote the slice of \(N_{r,R}\) at radius \(\rho \in (r, R)\) by 
\[N_{\rho} := N^- \cup ([0, \rho \pi] \times \Sigma) \times N^+ \subset N_{r,R},\]
which is diffeomorphic to \(N\) with a different neck length. One can see that 
\[\|F_{A_{\rho}}\|^2_{L^2(N)} \leq \rho \|F_{A_{\rho}}\|^2_{L^2(N_{\rho})}, \ \forall \rho \geq 1. \quad (4.11)\]
Now we state and prove the annulus lemma. Define the exponential factor 
\[\delta_P := c_P^{-1}\]
where \(c_P\) is the isoperimetric constant of Theorem 4.4. The constant depends on the bundle \(P_N \to N\) and the Riemannian metric on \(N\). But by abuse of notation we do not distinguish them since we only consider finitely many examples of such a bundle \(P_N \to N\).

**Proposition 4.5 (Annulus Lemma).** There exist \(\epsilon > 0, C > 0\) satisfying the following conditions. Let \(A\) be a solution to the ASD equation over \(N_{r,R}\) with \(r \geq 1\) and \(\log R - \log r \geq 4\). Assume 
\[E(A) < \epsilon.\]
Then for \(1 \leq s \leq \frac{1}{2}(\log R - \log r)\) there holds 
\[E(A; N_{e^{r,s},e^{-s}R}) \leq Ce^{-\delta_P s} E(A; N_{r,R}). \quad (4.12)\]

**Proof.** One can identify \(A\) with a piecewise smooth connection over \((r, R) \times N\) as explained above. In particular, one obtains a family of connections \(A_{\rho}\) on \(P_N \to N\) parametrized by \(\rho \in (r, R)\). Define 
\[I_\epsilon(r, R) := \{\rho \in (r, R) \mid \|F_{A_{\rho}}\|^2_{L^2(N)} < \epsilon\}.\]
By (4.10) one has 
\[\epsilon > E(A) \geq \int_{\log r}^{\log R} \|F_{A_{e^\tau}}\|^2_{L^2(N)} d\tau.\]
Therefore in every subinterval \((a, a + 1) \subset (\log r, \log R)\) of length one there exists a point \(\tau \in (a, a + 1)\) with \(e^\tau \in I_\epsilon(r, R)\). Then by Lemma 4.1 and Definition 4.2, the local action of \(A_{e^\tau}\), denoted temporarily by \(F(A_{e^\tau})\), can be defined for \(e^\tau \in I_\epsilon(r, R)\).

We would like to show that these local actions, which *a priori* are not defined for all \(\tau\), extend to a smooth function in \(\tau\). In fact there is a map 
\[I_\epsilon(r, R) \to \pi_0(L_N) =: \{Z_1, \ldots, Z_m\}\]
such that
\[ F(A_\rho) = F_{Z_\rho}(A_\rho), \quad \forall \rho \in I_\epsilon(r, R) \]
where \( Z_\rho \) is the image of \( \rho \) and \( F_{Z_\rho}(A_\rho) \) is the Chern–Simons action of \( A_\rho \) relative to the connected component \( Z_\rho \). We claim that
\[ F_{Z_\rho'}(A_\rho) = F_{Z_\rho}(A_\rho), \quad \forall \rho, \rho' \in I_\epsilon(r, R). \tag{4.13} \]
Indeed, by Lemma 2.1 we know that (suppose \( \rho < \rho' \))
\[ |F_{Z_\rho'}(A_\rho) - F_{Z_\rho}(A_\rho)| = E(A; N_{\rho, \rho'}) < \epsilon. \]
Moreover, by the isoperimetric inequality (Theorem 4.4) we know that
\[ |F_{Z_\rho'}(A_\rho) - F_{Z_\rho}(A_\rho)| \leq c_{\rho} \left( \|F_{A_\rho}\|_{L^2(N)}^2 + \|F_{A_\rho'}\|_{L^2(N)}^2 \right) \leq C_\epsilon. \]
Therefore, when \( \epsilon \) is small enough, the difference between \( F_{Z_\rho}(A_\rho) \) and \( F_{Z_\rho'}(A_\rho) \) is smaller than the minimal difference between critical values of the Chern–Simons functional on \( N \). Hence (4.13) is true. Then we can fix a connected component \( Z = Z_{\rho_0} \) for some \( \rho_0 \in I_\epsilon(r, R) \) and define
\[ F(\rho) := F_Z(A_\rho), \quad \forall \rho \in (r, R). \]
This is a smooth non-increasing function and \( F(\rho) \) agrees with the local action of \( A_\rho \) when \( \rho \in I_\epsilon(r, R) \). Then for \( s \in [1, \frac{1}{2}(\log R - \log r)] \), define
\[ J(s) := F(re^s) - F(Re^{-s}) = \int_{re^s}^{Re^{-s}} \|F_A\|_{L^2(N_\rho)}^2 d\rho. \]
We would like to derive a certain differential inequality of \( J(s) \). By (4.11) and the fact that \( Re^{-s} \geq re^s \geq 1 \) one obtains
\[ J'(s) = -re^s\|F_A\|^2_{L^2(N_{re^s})} - Re^{-s}\|F_A\|^2_{L^2(N_{Re^{-s}})} \leq -\|F_{A_{re^s}}\|^2_{L^2(N)} - \|F_{A_{Re^{-s}}}\|^2_{L^2(N)}. \]
If both \( re^s \) and \( Re^{-s} \) are in \( I_\epsilon(r, R) \), then by the isoperimetric inequality, one has
\[ J'(s) \leq -\|F_{A_{re^s}}\|^2_{L^2(N)} - \|F_{A_{Re^{-s}}}\|^2_{L^2(N)} \leq -\delta P(F(re^s) - F(Re^{-s})) = -\delta P J(s). \]
If \( re^s \) and/or \( Re^{-s} \) are not in \( I_\epsilon(r, R) \), i.e., when
\[ \|F_{A_{re^s}}\|^2_{L^2(N)} > \epsilon \quad \text{and/or} \quad \|F_{A_{Re^{-s}}}\|^2_{L^2(N)} > \epsilon, \]
there exists \( s' \) and/or \( s'' \) \([s - 1, s] \subset [0, \frac{1}{2}(\log R - \log r)]\) such that
\[ \|F_{A_{re^{s'}}}\|^2_{L^2(N)} \leq \epsilon \quad \text{and/or} \quad \|F_{A_{Re^{-s''}}}\|^2_{L^2(N)} \leq \epsilon. \]
Then by the isoperimetric inequality and the monotonicity of \( F \), one obtains
\[ J'(s) \leq -\|F_{A_{re^s}}\|^2_{L^2(N)} - \|F_{A_{Re^{-s}}}\|^2_{L^2(N)} \leq -\|F_{A_{re^{s'}}}\|^2_{L^2(N)} - \|F_{A_{Re^{-s''}}}\|^2_{L^2(N)} \leq -\delta P(F(re^{s'}) - F(Re^{-s''})) \leq -\delta P(F(re^s) - F(Re^{-s})) = -\delta P J(s). \]
It follows that \( J(s) \) decays exponentially as \( s \) increases and hence (4.12) is proved. \( \square \)

The above annulus lemma contains two special cases corresponding to the two special cases of Remark 2.12. In the first case when \( N^- \cong (N^+)^\text{op} \cong M_0 \), the four-manifold \( N_{r,R} \) is an open subset of \( R \times M \). In this case we denote
\[ M_{r,R} := N_{r,R}, \quad \text{where} \quad N \cong M^\text{double}. \]
In the second case when \( N^- \cong (N^+)^\text{op} \cong [0, \pi] \times \Sigma \) (whose boundary is two copies of \( \Sigma \) instead of one), one has a diffeomorphism \( N \cong S^1 \times \Sigma \). Although the four-manifold \( N_{r,R} \) is not isometric to \( Ann(r, R) \times \Sigma \) there is a constant \( a > 0 \) independent of large \( r \)
and there holds

\[ E(A; \text{Ann}(e^r, e^{-r} R) \times \Sigma) \leq Ce^{-\delta R^s} E(A; \text{Ann}(r, R) \times \Sigma). \]

In both of the two special cases, we would like to extend the annulus lemma to allow \( r = 0 \). For \( R > 0 \), define

\[ M_R := (B_R^+ \times \Sigma) \cup ((-R, R) \times M_0) \]

where we glue the common boundary \((-R, R) \times \Sigma\) in the obvious manner. The four-manifold can be formally viewed as the annulus \( M_{0,R} \) considered above.

**Proposition 4.7.** There exist \( \epsilon > 0, C > 0 \) satisfying the following conditions. Let \( A \) be a solution to the ASD equation over \( M_R \) with \( \log R \geq 4 \) and \( E(A) < \epsilon \). Then for \( s \geq 0 \) there holds

\[ E(A; M_{R^{s-}}) \leq Ce^{-\delta R^s} E(A; M_R). \]  

\[ (4.14) \]

*Proof.* Similar to the proof of Proposition 4.5, for all \( \rho \in [1,R] \), one can define a relative Chern–Simons action \( F(A_\rho) \) which agrees with the local action whenever \( \|F_{A_\rho}\|_{L^2(N)} \leq \epsilon \).

We claim that when \( Re^{-s} \geq 1 \) there holds

\[ E(A; M_{R^{s-}}) = -F(Re^{-s}). \]  

\[ (4.15) \]

Notice that the connection \( A_0 := A|_{\{0\} \times M_0} \) defines a not necessarily flat connection \( A_0^{\text{double}} \) over \( M^{\text{double}} \) by doubling \( A_0 \) and by Lemma 2.1 one has

\[ E(A; M_{R^{s-}}) = CS([A_0^{\text{double}}], [A_{R^{s-}}]). \]

On the other hand, by Uhlenbeck compactness, if \( \epsilon \) is sufficiently small, the restriction of \( A|_{M_L} \) is sufficiently close to a flat connection. Moreover, for every \( \rho \in [1,2] \), \( \|F_{A_\rho}\|_{L^2(N)} \) is sufficiently small. Hence there exists a nearby flat connection \( A_\rho^* \in \mathcal{A}_{\text{flat}}^*(P_N) \) such that

\[ F(\rho) = F_{\text{loc}}(A_\rho) = CS([A_\rho^*], [A_\rho]). \]

The fact that \( A|_{M_L} \) is close to a flat connection implies that \( \|A_\rho|_{N^-} - A_\rho|_{N^+}\|_{W^{1,2}(M_0)} \) is very small. Therefore, the gauge equivalence class of the nearby flat connection \( A_\rho^* \) must be in \( \Delta_{LM} \subset L_M \times L_M \), but not a double point in \( R_{LM} \). Hence \( A_\rho^* \) is gauge equivalent to the double of a flat connection on \( M_0 \). Therefore

\[ CS([A_0^{\text{double}}], [A_\rho^*]) = 0 \]

as it is the sum of an integral of the same differential form over two copies of the same three-manifold with boundary having the opposite orientations. Therefore

\[ E(A; M_{R^{s-}}) = CS([A_0^{\text{double}}], [A_{R^{s-}}]) = CS([A_\rho^*], [A_{R^{s-}}]) = -F(Re^{-s}). \]

Abbreviate \( J(s) = -F(Re^{-s}) \). By (4.15), (4.11), and the condition \( Re^{-s} \geq 1 \) one has

\[ J'(s) = -R e^{-s} \|F_A\|_{L^2(N_{R^{s-}})}^2 \leq -\|F_{A_{R^{s-}}}\|_{L^2(N)}^2. \]

Assume \( s \geq 1 \). If \( \|F_{A_{R^{s-}}}\|_{L^2(N)}^2 \leq \epsilon \), then by the isoperimetric inequality one has

\[ J'(s) \leq -\delta P J(s). \]
If $\|F_{A_{Re^{-s}}}\|^2_{L^2(N)} > \epsilon$, then one can find $s' \in (s-1, s) \subset [0, \log R]$ such that $\|F_{A_{Re^{-s'}}}\|^2_{L^2(N)} < \epsilon$. Then

$$J'(s) \leq -\|F_{A_{Re^{-s}}}\|^2_{L^2(N)} \leq -\|F_{A_{Re^{-s'}}}\|^2_{L^2(N)} \leq -\delta \rho J(s') \leq -\delta \rho J(s).$$

This shows that $J(s)$ decays exponentially for $s \in [1, \log R]$. So for some $C > 0$,

$$E(A; M_{Re^{-s}}) \leq Ce^{-\delta \rho s} E(A; M_R), \ \forall s \in [0, \log R]. \quad (4.16)$$

In particular,

$$E(A; M_2) \leq CR^{-\delta \rho} E(A; M_R).$$

Then by the standard interior estimate for the ASD equation, one has

$$\|F_A\|_{L^\infty(M)} \leq C \sqrt{E(A; M_2)} \leq CR^{-\frac{\delta \rho}{2}} \sqrt{E(A; M_R)}.$$ 

On the other hand, for $r \leq 1$, one has $\text{Volume}(M_r) \leq Cr$. Hence for $s \geq \log R$, $r = Re^{-s} \leq 1$ and one has

$$E(A; M_r) \leq CrR^{-\delta \rho} E(A; M_R) \leq Ce^{-s}R^{1-\delta \rho} E(A; M_R) \leq Ce^{-\delta \rho s} E(A; M_R). \quad (4.17)$$

Combining (4.16) and (4.17) we obtain (4.14). \hfill \Box

Similarly one has the following monotonicity property of ASD equation over the product of a disk with the compact surface, although it can be proved by using the mean value estimate instead of the isoperimetric inequality.

**Proposition 4.8.** There exist $\epsilon > 0$, $C > 0$ satisfying the following conditions. Let $A$ be a solution to the ASD equation over $B_R \times \Sigma$ with $\log R \geq 4$ and $E(A) < \epsilon$. Then for $s \geq 0$ there holds

$$E(A; B_{Re^{-s}} \times \Sigma) \leq Ce^{-\delta \rho s} E(A; B_R \times \Sigma).$$

**4.3. Diameter bound.** To ensure the convergence towards holomorphic curves on the boundary, we need a further diameter control. We first define the notion of diameter. Let $S \subset H$ be an open subset and let $A$ be a solution to the ASD equation over $S \times \Sigma$ identified with a triple $(B, \phi, \psi)$. Suppose for each $z \in S$ the connection $B(z) \in A(Q)$ is contained in the domain of the Narasimhan–Seshadri map $NS_2$. Then $A$ projects to a continuous map $u : S \to R\Sigma$. We define

$$\text{diam}(A; S \times \Sigma) := \text{diam}(u(S)) := \sup_{p, q \in S} d(u(p), u(q)).$$

Clearly this notion is gauge invariant. Moreover, denote

$$M_S := (S \times \Sigma) \cup (\partial S \times M_0)$$

where the two parts are glued along the common boundary $\partial S \times \Sigma$. If $A$ is a solution to $M_S$ whose restriction to $S \times \Sigma$ is $d_S + \phi ds + \psi dt + B(z)$ such that all $B(z)$ is contained in the domain of the Narasimhan–Seshadri map $NS_2$, then we define

$$\text{diam}(A; M_S) := \text{diam}(A; S \times \Sigma).$$

By using Theorem 3.1, one has the following interior diameter estimate.

**Lemma 4.9** (Interior diameter bound). There exist $C > 0$, $T > 0$, and $\epsilon > 0$ such that for any solution to the ASD equation over $B_{2R} \times \Sigma$ with $R \geq T$ and

$$E(A) \leq \epsilon, \quad \sup_{B_{2R}} \|F_B(z)\|_{L^2(\Sigma)} \leq \epsilon$$

there holds

$$\text{diam}(A; B_R \times \Sigma) \leq C \sqrt{E(A; B_{2R} \times \Sigma)}.$$
Next, we prove the following diameter estimate near the boundary.

**Lemma 4.10 (Boundary diameter bound).** There exist \( C > 0, \, T > 0, \) and \( \epsilon > 0 \) such that, for any solution to the ASD equation \( A \) over \( M_{3R} = M_{B_{3R}} \) with \( E(A) \leq \epsilon \) and \( R \geq T \) satisfying

\[
\|F_A\|_{L^\infty(M_{3R})} \leq 1, \quad \sup_{B_{3R}^+} \|F_B(z)\|_{L^2(\Sigma)} \leq \epsilon
\]

(4.18)

there holds

\[
\text{diam}(A; M_R) \leq C \sqrt{E(A; M_{3R})}.
\]

**Proof.** Let \( R \) and \( A \) satisfy the assumption of this lemma with \( \epsilon \) and \( T \) undetermined. Suppose the restriction of \( A \) to \( B_{3R}^+ \times \Sigma \) by \( dC + \phi ds + \psi dt + B(z) \). Denote by \( u : B_{3R}^+ \rightarrow R_\Sigma \) the holomorphic map defined by \( z \mapsto \overline{NS}_2(B(z)) \). Define the segment

\[
Z_R := \{ z = s + it \mid -R \leq s \leq R, \, t = R \}.
\]

Since \( Z_R \) can be covered by a fixed number of radius \( R \) disks contained in \( B_{3R}^+ \), by Lemma 4.9, for \( T \) large and \( \epsilon \) small, one has

\[
\text{diam}(A; Z_R \times \Sigma) \leq C \sqrt{E(A; M_{3R})}.
\]

(4.19)

Hence it remains to show that

\[
\sup_{-R \leq s \leq R} \sup_{0 \leq t_1, t_2 \leq R} \text{dist}(u(s + it_1), u(s + it_2)) \leq C \sqrt{E(A; M_{3R})}.
\]

(4.20)

To estimate the distance between \( u(s + it_1) \) and \( u(s + it_2) \) for \( 0 \leq t_1, t_2 \leq R \), we consider the rescaled equation and use the estimate of Corollary 3.6. The restriction of \( A \) to \( B_{3R}^+ \times \Sigma \) can be pulled back to a solution \( A' = d_C + \phi' ds + \psi' dt + B'(z) \) to the R-ASD equation over \( B_3^+ \times \Sigma \) satisfying \( E_R(A') \leq \epsilon \) and

\[
\sup_{B_3^+} \|F_{A'}(z)\|_{L^2(\Sigma)} \leq \epsilon.
\]

Then by Corollary 3.6, when \( T \) is sufficiently large and \( \epsilon \) is sufficiently small one has

\[
\|A_s'\|_{L^2((s+it) \times \Sigma)} \leq \frac{C}{t} \sqrt{E_R(A'; B_t(z) \times \Sigma)}, \quad \forall |s| \leq 1, \, 0 < t \leq 1.
\]

Moreover, by Proposition 4.7 for the case \( N = M_{\text{double}} \), for some constant \( C > 0 \) independent of \( A \) and \( R \) one has

\[
E_R(A'; B_t(z) \times \Sigma) \leq E_R(A'; B_{2t}(s) \times \Sigma) = E(A; M_{B_{2t}(s)}) \leq C t^{\delta_p} \sqrt{E(A; M_{3R})}, \quad \forall t \in (0, 1].
\]

Therefore one has

\[
\|A_s'\|_{L^2((s+it) \times \Sigma)} \leq C t^{-1+\frac{\delta_p}{2}} \sqrt{E(A; M_{2R})}, \quad \forall t \in (0, 1].
\]

Now for \( z_1 = s + it_1 \) and \( z_2 = s + it_2 \) with \( |s| \leq 1 \) and \( 0 < t_1 \leq t_2 \leq 1 \) we would like to estimate the distance between \( u(z_1) \) and \( u(z_2) \). Using a suitable gauge transformation we can assume \( \psi'(s + it) = 0 \) for all \( t \in [0, 1] \). Then

\[
\|B'(z_1) - B'(z_2)\|_{L^2(\Sigma)} = \int_0^1 \|\partial_s B'(s + it)\|_{L^2(\Sigma)} d\tau = \int_0^1 \|A'(z)\|_{L^2(\Sigma)} d\tau
\]

\[
= \int_0^1 \|A_s'\|_{L^2(\Sigma)} d\tau \leq C \sqrt{E(A; M_{3R})} \int_0^1 \tau^{-1+\frac{\delta_p}{2}} d\tau = C \sqrt{E(A; M_{3R})}.
\]
Then by Lemma 2.15, for $|s| \leq 1$ and $t_1, t_2 \in (0, 1]$, one has
\[
\text{dist}(u(Rs + iT_1), u(Rs + iT_2)) \leq \|NS_2(B(z_1)) - NS_2(B(z_2))\|_{L^2(\Sigma)}
\]
\[
\leq \|NS_2(B(z_1)) - B(z_1)\|_{L^2(\Sigma)} + \|B(z_1) - B(z_2)\|_{L^2(\Sigma)} + \|B(z_2) - NS_2(B(z_2))\|_{L^2(\Sigma)}
\]
\[
\leq C\left(\|F_{B(z_1)}\|_{L^2(\Sigma)} + \|F_{B(z_2)}\|_{L^2(\Sigma)} + \sqrt{E(A; M_{3R})}\right).
\]
It is standard to use the energy to bound the norm $\|F_{B(z)}\|_{L^2(\Sigma)}$. Hence one has
\[
\sup_{-R \leq s \leq R} \sup_{0 \leq t_1, t_2 \leq R} \text{dist}(u(s + iT_1), u(s + iT_2)) \leq C\sqrt{E(A; M_{3R})}.
\]
This proves (4.20). Then together with (4.19) the lemma is proved. \qed

4.4. Consequences of the annulus lemma and the diameter estimate. We first prove the asymptotic behavior of solutions to the ASD equation.

**Theorem 4.11** (Asymptotic Behavior). Let $A$ be a solution to the ASD equation over $M = R \times M$. Then there exists a point $(a_-, a_+) \in (t \times i)^{-1}(\Delta_{R_0}) \cong L^2_{\text{double}}$ whose image in $R_{\Sigma}$ is denoted by $b_{\infty}$ such that the following conditions hold.

(a) In the quotient topology of the configuration spaces $\mathcal{A}(P_0)/\mathcal{G}(P_0)$ one has
\[
\lim_{s \to \pm \infty} [A]_{\{s\} \times M_0} = a_\pm \in L_M \subset \mathcal{A}(P_0)/\mathcal{G}(P_0).
\]

(b) In the quotient topology of the configuration space $\mathcal{A}(Q)/\mathcal{G}(Q)$ one has
\[
\lim_{z \to \infty} [A]_{\{z\} \times \Sigma} = b_{\infty} \in R_{\Sigma} \subset \mathcal{A}(Q)/\mathcal{G}(Q).
\]

**Proof.** Denote the restriction of $A$ to $\text{Ann}^+(r, \infty) \times \Sigma$ by $d_H + \phi ds + \psi dt + B$. We first prove the convergence of the gauge equivalence class $[B(z)]$. By the strong Uhlenbeck compactness for Yang–Mills connections, we know that for each sequence $z_i \to \infty$, there is a subsequence (still indexed by $i$) for which $[B_{z_i}]$ converges in $\mathcal{A}(Q)/\mathcal{G}(Q)$ to a limit in $R_{\Sigma}$. We would like to show that the subsequential limit is unique. By the annulus lemma (Proposition 4.5), there exist $C > 0$ and $\delta_P > 0$ such that for $r \geq 1$ there holds
\[
E(A; M \setminus M_r) \leq C r^{-\delta_P}.
\]
Moreover, we claim that
\[
\lim_{r \to \infty} \|F_A\|_{L^\infty(M \setminus M_r)} = 0.
\]
Indeed, if this is not the case, then a nontrivial ASD instanton over $C \times \Sigma$, a nontrivial ASD instanton over $R \times M$, or a nontrivial $R^3$-instanton bubbles off at infinity, contradicting (4.21). Then by the diameter estimates (Lemma 4.9 and Lemma 4.10), there exist a constant $C > 0$ and sufficiently large $r_0$ such that for all $r \geq \log r_0$, there holds
\[
\text{diam}(A; M_{e^r-1, e^r+1}) \leq C \sqrt{E(A; M_{e^r-3, e^r+3})} \leq C e^{-\frac{\delta_P r}{2}}
\]
(this is because the half annulus $\text{Ann}^+(e^{r-2}, e^{r+2})$ can be covered by a fixed number of half disks and a fixed number of disks with radii being comparable to $e^r$ which are contained in the half annulus $\text{Ann}^+(e^{r-3}, e^{r+3})$). Hence
\[
\lim_{r \to \infty} \text{diam}(A; M \setminus M_r) = 0.
\]
Therefore the subsequential limit of $[B(z)]$ in $R_{\Sigma}$ is unique. Denote the limit by $b_{\infty}$.

Now we prove the convergence of $[A]_{\{s\} \times M_0}$ as $s \to \pm \infty$. Abbreviate the restriction $A_{\{s\} \times M_0}$ by $A_s$. First, the finiteness of energy implies that
\[
\lim_{s \to \pm \infty} \|F_A\|_{L^2([s-1, s+1] \times M_0)} = 0.
\]
Hence by Uhlenbeck compactness, for any sequence \( s_i \to \pm \infty \) there is a subsequence for which the sequence \([A_{s_i}]\) converges to a limit in \( L_M \subset A(P_0)/\mathcal{G}(P_0) \). Then any subsequential limit must be in \( \iota^{-1}(b_x) \) where \( \iota : L_M \hookrightarrow R_\Sigma \) is the Lagrangian immersion. We would like to show that as \( s \to +\infty \) or \( s \to -\infty \) the subsequential limit of \([A_s]\) is unique. Indeed, if there are two sequences \( s'_i, s''_i \to +\infty \) such that \([A_{s_i}]\) and \([A_{s'_i}]\) converges to two different preimages of \( \iota^{-1}(b_x) \), denoted by \( a_+, a'_+ \), then since the configuration space \( A(P_0)/\mathcal{G}(P_0) \) is Hausdorff (see Lemma 4.12 below), one can choose two disjoint neighborhoods \( U, U' \) of \( a_+ \) and \( a'_+ \) and a sequence \( s''_i \to +\infty \) with \([A_{s''_i}] \notin U \cup U'' \). Then a subsequence of \([A_{s''_i}]\) converges to a limiting flat connection different from \( a_+ \) and \( a'_+ \). However since \( b_x \) has at most two preimages, this cannot happen. Hence the subsequential limit of \([A_s]\) is unique and hence \([A_s]\) converges to a limit \( a_{\pm} \) as \( s \to \pm \infty \).

The following lemma is used in the previous proof.

**Lemma 4.12.** Let \( M_0 \) be a three-manifold with boundary and \( P_0 \to M_0 \) be an \( SO(3) \)-bundle. Then the configuration space \( A(P_0)/\mathcal{G}(P_0) \) is Hausdorff with respect to the quotient topology induced from the \( C^0 \)-topology of \( A(P_0) \).

**Proof.** For \( A_1, A_2 \in A(P_0) \), define

\[
\text{dist}_{C^0}(A_1, A_2) := \inf_{g \in \mathcal{G}(P_0)} \| g^* A_1 - A_2 \|_{C^0(M_0)}.
\]  

(4.22)

This clearly descends to a symmetric function on the quotient \( A(P_0)/\mathcal{G}(P_0) \) satisfying the triangle inequality. We claim that this is a metric, namely, if \( A_1 \) and \( A_2 \) are not gauge equivalent, then there is some \( \delta > 0 \) such that \( \text{dist}_{C^0}(A_1, A_2) > \delta \). Suppose this is not the case, then there exists a sequence of smooth gauge transformations \( g_i \) such that \( \| g_i^* A_1 - A_2 \|_{C^0(M_0)} \to 0 \). Since \( g_i \) takes value in a compact group \( SU(2) \), this implies that a subsequence of \( g_i \) converges to a continuous gauge transformation \( g \) on \( P_0 \). Moreover, in the weak sense, \( g_i^* A_1 = A_2 \). Since both \( A_1 \) and \( A_2 \) are smooth, \( g \) is also smooth and hence \( A_1 \) is gauge equivalent to \( A_2 \), which contradicts our assumption. Hence \( \text{dist}_{C^0} \) is a metric on the configuration space and the lemma is proved. \( \square \)

Theorem 4.11 is the analogue of the following results about the asymptotic behavior of instantons over \( C \times \Sigma \).

**Theorem 4.13.** [DS94, Proof of Theorem 9.1] Let \( A \) be an ASD instanton over \( (C \setminus B_R) \times \Sigma \) for some \( R > 0 \) which can be written as \( d_C + \phi ds + \psi dt + B(z) \). Then there is a gauge equivalence class of flat connections \( b_x \in R_\Sigma \subset A(Q)/\mathcal{G}(Q) \) such that in the quotient topology one has

\[
\lim_{|z| \to \infty} [B(z)] = b_x.
\]

**Definition 4.14.** In the situation of Theorem 4.11 resp. Theorem 4.13, the point \( (a_-, a_+) \in L_N \) resp. \( b_x \in R_\Sigma \) is called the **evaluation at infinity** of the solution \( A \) to the ASD instanton over \( M \) resp. \( (C \setminus B_R) \times \Sigma \).

Finally we have the energy quantization property for instantons over \( M = R \times M \).

**Theorem 4.15.** There exists \( h > 0 \) which only depends on the bundle \( P \to M \) and which satisfies the following property. For any ASD instanton \( A \) over \( M \) there holds

\[
E(A) \in \{0\} \cup [h, +\infty).
\]

**Proof.** We claim that \( h \) being the \( \epsilon \) of Proposition 4.7 satisfies the condition of this theorem. Indeed, suppose there is an ASD instanton over \( M \) with \( E(A) < h \). For all
$r > 0$ and $R = e^s r$ with $s$ sufficiently large, the restriction of $A$ to $M_R$ satisfies the hypothesis of Proposition 4.7. Then
\[
E(A; M_r) = E(A; M_{e^{-s} R}) \leq Ce^{-\delta p s}E(A; M_R) \leq C e^{-\delta p s}.
\]
As $s$ can be arbitrarily large, it follows that $E(A; M_r) = 0$. Therefore $A$ is a trivial solution. $\square$

5. Compactness modulo energy blowup

In this section we prove a compactness theorem modulo bubbling. It is the analogue of Theorem 3.3 for the domain being $M = R \times M$. We first set up the problem. Recall that for any open subset $S \subset H$, $M_S$ denotes
\[
M_S = (S \times \Sigma) \cup (\partial S \times M_0)
\]
where the two parts are glued over the common boundary $\partial S \times \Sigma$. Let $S_i \subset H$ be an exhaustive sequence of open subsets. They may or may not intersect the boundary of $H$. Let $\rho_i \to +\infty$ be a sequence of positive numbers diverging to infinity. Recall that $\phi_\rho : H \to H$ is the map corresponding to multiplying by $\rho$. In the proof of the main theorem of this paper we will only use the case that $S_i = H$.

**Theorem 5.1.** Suppose $A_i$ is a sequence of ASD instantons on $M_{\phi_\rho_i(S_i)}$ with
\[
\limsup_{i \to \infty} E(A_i; M_{\phi_\rho_i(S_i)}) < +\infty.
\]
Then there exist a subsequence (still indexed by $i$) and a holomorphic map
\[
\tilde{u}_\infty = (u_\infty, W_\infty, \gamma_\infty, m_\infty),
\]
from $H$ to $R_\Sigma$ with mass (see Definition 2.7) satisfying the following conditions.

(a) For each $z \in H$, one has the convergence
\[
\lim_{r \to 0} \lim_{i \to \infty} E(A_i; M_{\phi_\rho_i(B_r(z))}) = m_\infty(z). \tag{5.1}
\]

(b) Over $\phi_\rho_i(S_i) \times \Sigma$ we write $A_i = d_H + \phi_i ds + \psi_i dt + B_i$. Then for any precompact open subset $K \subset H \setminus W_\infty$, for $i$ sufficiently large, $B_i(z)$ is in the domain of the Narasimhan–Seshadri map $NS_2$ (see Definition 2.16) for all $z \in \phi_\rho_i(K)$, hence projects down to a sequence of holomorphic maps $u_i : \phi_\rho_i(K) \to R_\Sigma$. Further $u_i \circ \phi_\rho_i$ converges to $u_\infty|_K$ in the sense of Definition 2.5.

(c) There is no energy lost in the following sense:
\[
\lim_{R \to \infty} \lim_{i \to \infty} E(A_i; M_{\rho_i R}) = E(u_\infty) + \int_H m_\infty. \tag{5.2}
\]

These two theorems motivates the following notion of convergence.

**Definition 5.2.** Let $A_i$ be a sequence of ASD instantons over $M$. Let $\tilde{u}_\infty$ be a holomorphic map from $S$ to $R_\Sigma$ with mass in the sense of Definition 2.7. Suppose $\{\rho_i\}$ be a sequence of positive numbers diverging to $\infty$. $A_i$ is said to converge to $u_\infty$ along with $\{\rho_i\}$ if conditions (a), (b), and (c) of Theorem 5.1 hold.
5.1. Energy blowup threshold. First we define the notion of energy blowup.

**Definition 5.3 (Energy blowup).** Let $\rho_i, A_i$ satisfy the assumptions of Theorem 5.1. For each $w \in H$, we say that energy blows up at $w$ if
\[
\lim_{r \to 0} \limsup_{i \to \infty} E(A_i; M_{\varphi_{\rho_i}(B^+_r(w))}) > 0.
\]

**Lemma 5.4 (Energy blowup threshold).** There exists $\hbar > 0$ satisfying the following property. Let $\rho_i$ and $A_i$ be as in Theorem 5.1. Then for all $w \in H$ there holds
\[
\lim_{r \to 0} \limsup_{i \to \infty} E(A_i; M_{\varphi_{\rho_i}(B^+_r(w))}) \in \{0\} \cup [h, +\infty).
\]  

(5.3)

**Proof.** When energy concentrates at $w \in \text{Int} H$, (5.3) follows from the bubbling analysis in [DS94, Proof of Theorem 9.1] (or via Proposition 4.8 as argued below). We consider the case when $w \in \partial H$. We claim that $\hbar$ being the $\epsilon$ of Proposition 4.7 satisfies the property. Indeed, suppose on the contrary
\[
\lim_{r \to \infty} \limsup_{i \to \infty} E(A_i; M_{\varphi_{\rho_i}(B^+_r(w))}) = a < \epsilon.
\]
Then there exists $r_0 > 0$ such that
\[
\limsup_{i \to \infty} E(A_i; M_{\varphi_{\rho_i}(B^+_r(w))}) < \epsilon.
\]
Then by Proposition 4.7 for $r < r_0$ and sufficiently large $i$, one has
\[
E(A_i; M_{\varphi_{\rho_i}(B^+_r(w))}) \leq C \left( \frac{r}{r_0} \right)^{\delta_p}.
\]
This contradicts the assumption that energy concentrates at $w$. \qed

One can use the threshold to select a subsequence for which energy concentration happens at only finitely many points. Hence one can assume for a subsequence (still indexed by $i$) energy concentrates at points of a finite subset $W_\infty \subset H$ and a bubbling measure $m_\infty : H \to [0, +\infty)$ supported over $W_\infty$ satisfying (5.1).

5.2. Constructing the limiting holomorphic curve. Next we construct the limiting holomorphic curve. For any precompact open subset $K \subset H \setminus W_\infty$

(which may intersect the boundary of $H$), we claim that
\[
\lim_{i \to \infty} \sup_{z \in \nu_{\rho_i}(K)} \|F_{B_i(z)}\|_{L^2(\Sigma)} = 0.
\]
Indeed, if this is not true, then some subsequence will bubble off an $R^4$-instanton, an instanton over $C \times \Sigma$, or an instanton over $M_i$, contradicting that no energy concentrates at points away from $W_\infty$. Then for $i$ sufficiently large, $B_i(z)$ is in the domain of the Narasimhan–Seshadri map $\overline{NS}_2 : A^{1,2}_k(Q) \to R_{\Sigma}$. By Proposition 2.17, the map
\[
u_i : \varphi_{\rho_i}(K) \to R_{\Sigma}, \quad \nu_i(z) := \overline{NS}_2(B_i(z))
\]
is holomorphic. Denote the reparametrized holomorphic maps by
\[
u'_i := \nu_i \circ \varphi_{\rho_i} : K \to R_{\Sigma}.
\]
One can choose a subsequence (still indexed by $i$) and an exhausting sequence of precompact open subsets $K_i \subset H \setminus W_\infty$ such that $\nu'_i$ is defined over $K_i$. 
Lemma 5.5. There is a subsequence of $u'_i$ which converges to a holomorphic map $u_\infty : H \setminus W_\infty \to R_\Sigma$ in the sense of Definition 2.5. Moreover, the energy density function $e_{\rho_i}$ of $A'_i$ converges to the energy density of $u_\infty$ in $C^0_{\text{loc}}(\text{Int} H \setminus W_\infty)$.

Proof. By Theorem 3.3, a subsequence of $u'_i : K_i \to R_\Sigma$ (still indexed by $i$) converges to a holomorphic map $u : \text{Int} H \setminus W_\infty \to R_\Sigma$ in $C^\infty_{\text{loc}}(\text{Int} H \setminus W_\infty)$ and the rescaled energy density function converges in $C^0_{\text{loc}}(\text{Int} H \setminus W_\infty)$ to the energy density of $u_\infty$. To prove the boundary convergence, we need to verify that there is no diameter blowup along the boundary. Indeed, by the fact that no energy blowup happens in $H \setminus W_\infty$ and Lemma 4.10, there holds

$$\lim_{r \to 0} \lim_{i \to \infty} \sup \text{diam}(u'_i(B^+(z))) = 0, \ \forall z \in \partial H \setminus W_\infty.$$ 

Then by Proposition 2.6 the map $u_\infty$ extends continuously to the boundary and $u'_i$ converges to $u_\infty$ in the sense of Definition 2.5.

We would like to use Gromov’s removal of singularity theorem to extend the limiting holomorphic map over the bubbling points. We first show that the limit $u_\infty$ satisfies the Lagrangian boundary condition.

Lemma 5.6. The holomorphic map $u_\infty$ maps the boundary of $H \setminus W_\infty$ into $\iota(L_M)$.

Proof. For each $s \in \partial H \setminus W_\infty$, denote $A_{i;s} = A_{i|s \times M_0}$. Since there is no energy blow up at $s$, one has

$$\lim_{i \to \infty} \|F_{A_{i;s}}\|_{L^2(M_0)} = 0.$$ 

Then there is a subsequence of $A_{i;s}$ (still indexed by $i$) which converges modulo gauge transformation to a flat connection $A_{\iota(L_M)}$ on $M_0$. Since the map induced from boundary restriction

$$A(P_0)/G(P_0) \to A(Q)/G(Q)$$

is continuous, one has

$$\lim_{i \to \infty} [B_{i;s}] = \lim_{s \to \iota(L_M)} [A_{i;s}|\Sigma] = \left(\lim_{i \to \infty} [A_{i;s}]\right)|_{\iota(L_M)}.$$ 

On the other hand, since $\|F_{B_{i,s}}\|_{L^2(\Sigma)} \to 0$, the point $u_i(s) \in R_\Sigma$ has a flat connection representative which is arbitrarily $L^2$-close to $B_{i,s}$. Since $u_i(s)$ converges to $u_\infty(s)$, $u_\infty(s)$ agrees with the above limit which is in $\iota(L_M)$.

Lemma 5.7. There is a subsequence (still indexed by $i$) such that for all $s \in \partial H \setminus W_\infty$, the sequence $[A_{i;s}]$ converges in $A(P_0)/G(P_0)$ to a flat connection on $P_0 \to M_0$.

Proof. For each connected component $I_\alpha \subset \partial H \setminus W_\infty$, pick one point $s_\alpha \in I_\alpha$. Then for the finitely many chosen points $z_\alpha$, we can find a subsequence of $A_i$, still indexed by $i$, such that $[A_{i;s_\alpha}]$ converges to $[A_{s_\alpha}]$ for all $\alpha$. Then consider the sets

$$I^*_\alpha = \left\{ s \in I_\alpha \mid [A_{i;s}] \text{ converges in } A(P_0)/G(P_0) \right\}.$$ 

Then $s_\alpha \in I^*_\alpha$. Notice that, by the proof of Lemma 5.6, for each $s \in I^*_\alpha$ one has

$$u_\infty(s) = \iota(\lim_{i \to \infty} [A_{i;s}]). \quad (5.4)$$

Moreover, if $u_\infty(s)$ is not a double point of $\iota(L_M)$, then $s \in I^*_\alpha$. The lemma follows if we can show that $I^*_\alpha$ is both open and closed.

We prove $I^*_\alpha$ is open. Suppose this is not the case. Then there exist $s^* \in I^*_\alpha$ and a sequence $s_k \notin I^*_\alpha$ with $s_k \to s^*$. Since $u_\infty$ is continuous, $u_\infty(s_k)$ converges to $u_\infty(s^*)$. Since
$u_\infty(s_k)$ is a double point for all $k$ and the set of double points is discrete, $u_\infty(s^*) = u_\infty(s_k)$ for sufficiently large $k$. Assume $i^{-1}(u_\infty(s^*)) = \{a', a''\} \subset L_M$ and assume that $[A_{i;s^*}]$ converges to $a'$. Then for each $k$, there is a subsequence of $[A_{i;s_k}]$ which converges to $a''$. By Lemma 4.12 which says the space $\mathcal{A}(P_0)/\mathcal{G}(P_0)$ is Hausdorff, we can choose disjoint open neighborhoods $\mathcal{U}', \mathcal{U}'' \subset \mathcal{A}(P_0)/\mathcal{G}(P_0)$. Then for each $k$, we can choose $i_k$ inductively such that 1) $i_k+1 > i_k$; 2) $[A_{i_k;s^*}] \in \mathcal{U}'$; 3) $[A_{i_k;s_k}] \in \mathcal{U}''$. Then there exists a point $w_k$ between $s^*$ and $s_k$ such that $[A_{i_k;w_k}] \notin \mathcal{U}' \cup \mathcal{U}''$. Hence a subsequence (still indexed by $k$) of $[A_{i_k;w_k}]$ converges to a point in $L_M$ which is neither $a'$ or $a''$. Denote $[B_{i_k;w_k}] = [A_{i_k;w_k}]|_{\partial M_0}$. Then this convergence implies

$$\lim_{k \to \infty} u_{i_k}'(w_k) = \lim_{k \to \infty} [B_{i_k;w_k}] \notin i(\mathcal{U}' \cup \mathcal{U}'' \cap L_M)$$

(5.5)

which is different from $u_\infty(s^*)$. This contradicts the convergence $u_i'$ towards $u_\infty$ and the continuity of $u_\infty$. This proves the closedness of $I^*_\alpha$.

We prove $I^*_\alpha$ is closed. Suppose this is not the case. Then there exists $s^* \notin I^*_\alpha$ and a sequence $s_k \in I^*_\alpha$ with $s_k \to s^*$. Then $u_\infty(s^*)$ is a double point of $i(L_M)$. Let $\mathcal{U}', \mathcal{U}'' \subset \mathcal{A}(P_0)/\mathcal{G}(P_0)$ be disjoint open neighborhoods of the two preimages of $u_\infty(s^*)$. Then there must be a subsequence of $s_k$ (still indexed by $k$), and one of $\mathcal{U}'$ and $\mathcal{U}''$, say $\mathcal{U}'$, such that for all $k$,

$$\lim_{i \to \infty} [A_{i;s_k}] \in \mathcal{U}'.$$ 

By (5.4) and the non-convergence of $[A_{i;s^*}]$, there must also be a subsequence of $i$ (still indexed by $i$), such that

$$\lim_{i \to \infty} [A_{i;s^*}] \in \mathcal{U}''.$$ 

Then for each $k$, one can find $i_k$ such that

$$[A_{i_k;w_k}] \in \mathcal{U}', \quad [A_{i_k;s^*}] \in \mathcal{U}''.$$ 

Choose a sequence $i_k$ such that $i_{k+1} > i_k$. Then since $\mathcal{U}'$ and $\mathcal{U}''$ are disjoint, one can find for each $k$ a point $w_k$ between $z^*$ and $z_k$ with

$$[A_{i_k;w_k}] \notin \mathcal{U}' \cup \mathcal{U}''.$$ 

Denote $[B_{i_k;w_k}] = [A_{i_k;w_k}]|_{\partial M_0}$. Then (5.5) still holds and one derives the same contradiction. This proves the closedness of $I^*_\alpha$.

It follows from Lemma 5.7 that the boundary map can be defined.

**Corollary 5.8.** There exist a map $\gamma_\infty : \partial H \setminus W_\infty \to L_M$ and a subsequence (still indexed by $i$) such that

$$\lim_{i \to \infty} [A_{i;s}] = \gamma_\infty(s), \quad \forall s \in \partial H \setminus W_\infty$$

(5.6)

and

$$u_\infty|_{\partial H \setminus W_\infty} = i \circ \gamma_\infty.$$ 

**Lemma 5.9.** $\gamma_\infty$ is continuous.

**Proof.** One only needs to show the continuity at points whose image is a double point of $i(L_M)$. The argument is similar as in the proof of Lemma 5.7, using the Hausdorffness of the configuration space $\mathcal{A}(P_0)/\mathcal{G}(P_0)$. The details are left to the reader.

**Corollary 5.10.** The holomorphic map $u_\infty : H \setminus W_\infty \to R_\Sigma$ extends to a holomorphic map with immersed Lagrangian boundary condition with mass, denoted by

$$\tilde{u}_\infty = (u_\infty, \gamma_\infty, W_\infty, m_\infty).$$
Proof. For each point \( w \in W_\infty \cap \text{Int} H \), by Gromov theorem of removal of singularity, \( u_\infty \) extends smoothly over \( w \). On the other hand, given \( w \in W_\infty \cap \partial H \) and consider a punctured neighborhood of \( w \) that is biholomorphic to a strip \((0, +\infty) \times [0, \pi]\) with coordinates \((s, t)\). Then by the standard elliptic estimate for holomorphic maps with Lagrangian boundary condition, one can prove that the length of the paths \( u_\infty(s, \cdot) \) converges to zero. Hence the pair of points \((\gamma_\infty(s, 0), \gamma_\infty(s, \pi))\) is either close to the diagonal of \( L_M \) or is close to an ordered double point. In either case one can define a local action for the path \( u_\infty(s, \cdot) \) and this path satisfies the isoperimetric inequality similar to [MS04, Theorem 4.4.1 (ii)]. The isoperimetric inequality implies that as \( s \to +\infty \), the paths \( u_\infty(s, \cdot) \) converge to a constant path. Hence \( u_\infty \) extends continuously to \( w \). □

5.3. Conservation of energy. The last step of proving Theorem 5.1 is to verify the conservation of energy.

Lemma 5.11. There holds
\[
\lim_{R \to \infty} \lim_{i \to \infty} E(A_i; M_{\rho_i R}) = E(u_\infty) + \sum_{w \in W_\infty} m_\infty(w).
\]

Proof. To simplify the notations, assume that \( W_\infty \) contains only one point \( \{0\} \in \partial H \). The proof of the general case can be derived easily from the proof of the special case, which we are going to present.

Recall the closed three-manifold \( N \) and the \( SO(3) \)-bundle \( P_N \to N = M^{\text{double}} \). For any positive number \( r \), the restriction of \( A_i \) to \( \partial M_{\rho_i r} \) can be identified with a piecewise smooth connection \( A_{i,r} \) on \( P_N \to N \). Then by the energy identity (Lemma 2.1) and the definition of energy concentration, one has
\[
\lim_{R \to \infty} \lim_{i \to \infty} E(A_i; M_{\rho_i R}) = m_\infty(0) + \lim_{r \to 0, R \to \infty} \lim_{i \to \infty} CS([A_i,r], [A_i,R])
\]
where the last term is the Chern–Simons functional associated to the four-manifold with boundary \( M_{r,R} \) (and the \( SO(3) \)-bundle over it). On the other hand, for each pair \((r, R)\), the holomorphic map \( u_\infty \) defines piecewise smooth connections \( A_r, A_R \in \mathcal{A}^{p,s}(P_N) \) whose gauge equivalence classes are well-defined. Then by Proposition 2.18, one has
\[
E(u_\infty) = \lim_{r \to 0, R \to \infty} CS([A_r], [A_R]).
\]
Then by the property of Chern–Simons functional, to prove (5.2), it suffices to show that
\[
\lim_{r \to 0} \lim_{i \to \infty} CS([A_r], [A_{i,r}]) = 0
\]
and
\[
\lim_{R \to \infty} \lim_{i \to \infty} CS([A_R], [A_{i,R}]) = 0.
\]
We will only prove (5.7) as follows. The proof of (5.8) is similar.

First, the evaluation of \( u_\infty \) at the origin defines a gauge equivalence class of flat connections \([A_0] \in \mathcal{A}^{p,s}(P_N)/\mathcal{G}^{p,s}(P_N)\). From the finiteness of the energy of \( u_\infty \) one can see that
\[
\lim_{r \to 0} CS([A_0], [A_r]) = 0.
\]
Hence it suffices to prove that
\[
\lim_{r \to 0} \lim_{i \to \infty} CS([A_0], [A_{i,r}]) = 0.
\]
Second, we would like to identify the relative Chern–Simons action \( CS([A_0], [A_{i,r}]) \) with the local action defined by Definition 4.2.
Claim. For any \( \varepsilon > 0 \), there exists \( r_\varepsilon > 0 \) such that for all \( r' \in (0, r_\varepsilon) \), one has

\[
\sup_{r' \leq r \leq r_\varepsilon} \limsup_{i \to \infty} \left( \|F_{A_{i,r'}}\|_{L^4(N^- \cup N^+)} + l_2(A_{i,r'}) \right) \leq \varepsilon.
\]

Proof of the claim. The smallness of \( \|F_{A_{i,r'}}\|_{L^4(N^- \cup N^+)} \) follows from the fact that no energy blows up away from the origin and the \( \varepsilon \)-regularity of the Yang–Mills equation. The smallness of \( l_2(A_{i,r'}) \) follows from the annulus lemma (Proposition 4.5) and the diameter estimate of Lemma 4.10.

End of the proof of the claim.

Then for each \( r \in (0, r_\varepsilon] \), for sufficiently large \( i \), the \( p = 4 \) case of Lemma 4.1 implies that there exists a nearby flat connection \( A^*_{i,r} \) on \( P_N \) satisfying

\[
\|A_{i,r} - A^*_{i,r}\|_{W^1,4(N^- \cup N^+)} \leq C\varepsilon.
\]

Then by the Sobolev embedding \( W^{1,4} \to C^0 \) in dimension three, for any given \( \delta > 0 \), for \( \varepsilon \) sufficiently small, \( r \) sufficiently small, and \( i \) sufficiently big, one has

\[
\text{dist}_{C^0}(\{A_{i,r}|_{N^\pm}\}, \{A^*_{i,r}|_{N^\pm}\}) < \delta
\]

(\text{here } \text{dist}_{C^0} \text{ is defined by (4.22)). On the other hand, by the property of the pseudoholomorphic curve } u_{\infty} \text{ and its boundary map } \gamma_{\infty}, \text{ for } r \text{ sufficiently small, one has}

\[
\text{dist}_{C^0}(\gamma_{\infty}(\pm r), [A_0|_{N^\pm}]) < \delta.
\]

Further, by the convergence \( [A_{i,r}|_{N^-}] \to \gamma_{\infty}(\pm r) \), for sufficiently large \( i \), one has

\[
\text{dist}_{C^0}(\{A_{i,r}|_{N^\pm}\}, \gamma_{\infty}(\pm r)) < \delta.
\]

It follows that

\[
\text{dist}_{C^0}(\{A^*_{i,r}|_{N^\pm}\}, [A_0|_{N^\pm}]) < 3\delta.
\]

Therefore, if \( \delta \) is small enough, \( [A^*_{i,r}] \) and \( [A_0] \) are in the same connected component of \( L_N \). Hence

\[
\text{CS}([A_0], [A_{i,r}]) = \text{CS}([A^*_{i,r}], [A_{i,r}]) = F_{\text{loc}}(A_{i,r}).
\]

Then by the first version of the isoperimetric inequality (Theorem 4.3) one has

\[
\lim_{r \to 0} \limsup_{i \to \infty} |F_{\text{loc}}(A_{i,r})| \leq C \limsup_{r \to 0} \limsup_{i \to \infty} \left( \|F_{A_{i,r}}\|_{L^4(N)}^2 + l_2(A)^2 \right) = 0.
\]

Hence (5.9) and therefore (5.7) follow. \( \square \)

6. Stable scaled instantons

In this section we define the Gromov–Uhlenbeck convergence for ASD instantons. In fact we define a notion called stable scaled instantons, which is a combination of stable maps in symplectic geometry and “instantons with Dirac measures” in gauge theory. The combinatorial model of trees is also used in the study of the vortex equation (see for example [WX17] [WX]) and holomorphic quilts (see [BW18]).

In this and the next sections, we will emphasize on the discussion of instantons over \( R \times M \). The situation for instantons over \( C \times \Sigma \) can be dealt with similarly and most of the details will be left to the reader.
6.1. Trees. Let us fix a few notations. A tree, usually denoted by \( \Gamma \), consists of a set of vertices \( V_\Gamma \) and a set of edges \( E_\Gamma \). One can associate to each tree a 1-complex whose 0-cells are vertices and whose 1-cells are edges. A rooted tree is a tree \( \Gamma \) together with a distinguished vertex \( v_\infty \in V_\Gamma \) called the root. A rooted subtree of a rooted tree \((\Gamma, v_\infty)\) is a subtree which contains \( v_\infty \). A ribbon tree is a tree \( \Gamma \) together with an isotopy class of embeddings of \( \Gamma \) into the complex plane. A based tree is a tree \( \Gamma \) with a rooted subtree \( \hat{\Gamma} \) with \( \Gamma \) equipped with the structure of a ribbon tree. The subtree \( \hat{\Gamma} \) is called the base of \( \hat{\Gamma} \). A based tree can be used to model stable holomorphic disks such that vertices in the base correspond to disk components and vertices not in the base correspond to sphere components.

We only consider rooted trees and often skip the term “rooted.” Notice that the root \( v_\infty \in V_\Gamma \) induces a natural partial order \( \preceq \) among all vertices: \( v \preceq v' \) if \( v \) is closer to the root \( v_\infty \). We write \( v' > v \) if \( v \preceq v' \) and \( v, v' \) are adjacent; in this case we denote the corresponding edge by \( e = e_{v' > v} \in E_\Gamma \).

**Definition 6.1.** Consider the set \( \{1, \infty\} \) ordered as \( 1 \leq \infty \). A scaling on a based tree \( \Gamma \) is a map \( s : V_\Gamma^1 \rightarrow \{1, \infty\} \) satisfying the following condition.

- Denote \( V_\Gamma^1 = s^{-1}(1) \) and \( V_\Gamma^\infty = s^{-1}(\infty) \). Then \( V_\Gamma^\infty \) forms a (possibly empty) rooted subtree of \( \hat{\Gamma} \) and vertices in \( V_\Gamma^1 \) are all disconnected from each other.

A based tree \( \Gamma \) with a scale \( s \) is called a scaled tree.

Each vertex of a scaled tree is supposed to support an instanton or a holomorphic map over a certain domain, which we specify as follows. Consider a scaled tree \( \Gamma = (\Gamma, s) \) with possibly empty base \( \Gamma_0 \). For each \( v \in V_\Gamma \), define \( M_v = R \times M \), \( S_v = H \), and \( \overline{S_v} = H \cup \{\infty\} \cong D^2 \); for each \( v \in V_\Gamma \setminus V_\Gamma^1 \), define \( M_v = C \times \Sigma \), \( S_v = C \), and \( \overline{S_v} = C \cup \{\infty\} \cong S^2 \). For each \( v \), there is an \( SO(3) \)-bundle \( P_v \rightarrow M_v \) specified previously.

**Definition 6.2.** A stable scaled instanton modelled on a scaled tree \((\Gamma, s)\) consists of a collection

\[
\mathcal{C} = \left\{ (v, a_v, m_v) \mid v \in V_\Gamma^1, \{ w_e \mid e \in E_\Gamma \}, \{ u_v = (w_v, W_v, \gamma_v) \mid v \in V_\Gamma^\infty \} \right\}
\]

where the symbols denote the following objects.

- For each \( v \in V_\Gamma^1 \), \( a_v \) is a gauge equivalence class of ASD instantons on \( P_v \rightarrow M_v \) and \( m_v \) is a positive measure on \( M_v \) with finite support.
- For each edge \( e = e_{v' > v} \in E_\Gamma \), \( w_e \) is a point of \( S_v \).
- For each \( v \in V_\Gamma^\infty \), \( W_v \) is the subset of \( S_v \) defined by

\[
W_v := \{ w_e \mid e = e_{v' > v} \in E_\Gamma \}
\]  

where \( u_v \) is a holomorphic map from \( S_v \) to \( R_{\Sigma} \) with boundary in \( \iota(L_M) \) (see Definition 2.7).

These objects also satisfy the following conditions.

(a) If \( e = e_{v' > v} \in E_\Gamma \), namely both \( v' \) and \( v \) are in the base, then \( w_e \in \partial S_v \).
(b) For every \( v \in V_\Gamma^\infty \), the collection of points defining \( W_v \) in (6.1) are distinct.
(c) The measure \( m_v \) takes value in \( 4\pi^2 \mathbb{Z} \).
(d) **Matching Condition** For each interior edge \( e = e_{v' > v} \in E_\Gamma \setminus E_\Gamma^1 \) the evaluation at infinity of \( a_v \) or \( u_v \), which is a point of \( R_\Sigma \), is equal to the evaluation of \( u_v \) at \( w_e \). For each boundary edge \( e = e_{v > v} \in E_\Gamma^1 \) the evaluation at infinity of \( a_v \) or \( u_v \), which is a point in \( L_N \cong \Delta_{L_M} \cup R_{L_M} \) is equal to the transpose of \( e_v w_e(u_v) \) (see Definition 2.7 and Definition 4.14 for the definitions of different evaluations.)
(e) **Stability Condition** If \( a_v \) has zero energy, then \( m_v \neq 0 \); if \( v \in V^\infty_1 \) and \( u \) is a constant map, then the number of boundary nodal points plus twice of the interior nodal points attached to \( S_v \) is at least two; if \( v \in V^\infty_1 \setminus V^\infty_1 \) and \( u_v \) is a constant map, then the number of nodal points on \( S_v \) is at least two.

A typical configuration of a stable scaled instanton has been shown in Figure 1.4.

One can see that on the combinatorial level a stable scaled instanton is very similar to a stable affine vortex over \( C \) (when \( \Gamma \) has an empty base) or over \( H \) (when \( \Gamma \) has a nonempty base). These two constructions appeared in [Zil14] and [WX17] respectively. We can define a notion of equivalence among stable scaled instantons by incorporating the translation symmetry of ASD instantons and the conformal invariance of holomorphic maps. The details are left to the reader. It is then easy to check that the automorphism group of a stable scaled instanton is finite.

### 6.2. Sequential convergence

Now we define the notion of convergence of a sequence of ASD instantons over \( M = R \times \tilde{M} \) towards a stable scaled instanton. Let

\[
C = \left\{ (a_v, m_v) \mid v \in V^1_1 \right\} \cup \left\{ w_e \mid e \in E_\Gamma \right\} \cup \left\{ u_v \mid v \in V^\infty_1 \right\}
\]

be a stable scaled instanton modelled on a based scaled tree \((\Gamma, s)\). For each \( v \in V^\infty_1 \), we have the set of nodes \( W_v \subset S_v \) defined by (6.1). Then we can define a measure \( m_v \) on \( S_v \) supported in \( W_v \) as follows. For each \( w_e \in W_\delta \) with \( e = e_{v'v} \), the mass \( m_v \) at \( w_e \) is the sum of the energy of all components in \( C \) labelled by \( v'' \in V_1 \) with \( v'' \geq v' \). Denote

\[
\tilde{u}_v = (u_v, m_v)
\]

which is a holomorphic curve with mass (see discussion after Definition 2.7).

**Definition 6.3** (Convergence towards stable scaled instantons). Let \( a_i = [A_i] \) be a sequence of gauge equivalence classes of ASD instantons on \( P \to M \). Let \((\Gamma, s)\) be a based scaled tree. Let

\[
C = \left\{ (a_v, m_v) \mid v \in V^1_1 \right\} \cup \left\{ w_e \mid e \in E_\Gamma \right\} \cup \left\{ u_v \mid v \in V^\infty_1 \right\}
\]

be a stable scaled instanton modelled on \((\Gamma, s)\). We say that \( a_i \) converges (modulo gauge transformation) to \( C \) if the following conditions are satisfied.

(a) For each \( v \in V^1_1 \), there exist a sequence of real translations \( \phi_{i,v}(z) = z + Z_{i,v} \) such that \( \phi_{i,v}^* a_i \) converges in the Uhlenbeck sense to \((a_v, m_v)\) over the manifold \( M_v = R \times \tilde{M} \) (see Definition 2.2).

(b) For each \( v \in V^1_1 \setminus V^1_1 \), there exist a sequence of complex translations \( \phi_{i,v}(z) = z + Z_{i,v} \) satisfying the following properties.

(i) \( \text{Im} Z_{i,v} \to +\infty \);

(ii) Viewing \( \phi_{i,v} \) as a diffeomorphism from \( C \times \Sigma \) to itself, for all \( R > 0 \), \( \phi_{i,v}^* (a_i |_{B_R \times \Sigma}) \) converges to \((a_v |_{B_R \times \Sigma}, m_v |_{B_R \times \Sigma})\) in the Uhlenbeck sense.

(c) For each \( v \in V^\infty_1 \), there exists a sequence of real affine transformations \( \phi_{i,v}(z) = \rho_{i,v}(z + Z_{i,v}) \) satisfying the following properties.

(i) \( \rho_{i,v} \in R, Z_{i,v} \in R, \rho_{i,v} \to +\infty \);

(ii) Denoting the translation \( z \mapsto z + Z_{i,v} \) by \( \psi_{i,v} \) and viewing it as a diffeomorphism from \( \tilde{M} \) to itself, \( \psi_{i,v}^* a_i \) converges to \( \tilde{u}_v \) along with \( \{\rho_{i,v}\} \) in the sense of Definition 5.2.

(d) For each \( v \in V^\infty_1 \setminus V^\infty_1 \), there exist a sequence of complex affine transformations \( \phi_{i,v}(z) = \rho_{i,v}(z + Z_{i,v}) \) satisfying the following properties.

(i) \( \rho_{i,v} \in R, Z_{i,v} \in \mathbb{C}, \rho_{i,v} \to +\infty \), and \( \text{Im} Z_{i,v} \to +\infty \);
(ii) Denoting the translation $z \mapsto z + Z_{i,v}$ to be $\psi_{i,v}$ and viewing it as a diffeomorphism from $C \times \Sigma$ to itself, then for all sufficiently big $R > 0$, the sequence of pullbacks $\psi^*_{i,v}(a|_{\varphi_{i,v}(B_R) \times \Sigma})$, which are instantons over $B_{\rho_{i,v}R} \times \Sigma$, converge to $\tilde{u}_{v}|_{B_R}$ along with $\{\rho_{i,v}\}$ in the sense of Definition 3.4.

(e) The reparametrizations $\varphi_{i,v}$ mentioned above can all be viewed as Möbius transformations on the complex plane. One can also view all $S_v$ (being either $H$ or $C$) as subsets of $C$. For each edge $e = e'_{v'} \in \Gamma$, the map $\varphi_{i,v}^{-1} \circ \varphi_{i,v'}$ converges uniformly with all derivatives on compact sets to the constant $w_{e} \in S_v \subset C$.

(f) There is no energy lost, i.e.,

$$\lim_{i \to \infty} E(A_i) = \sum_{v \in V_i^c} E(u_v) + \sum_{v \in V_i^d} \left( E(a_v) + |m_v| \right).$$

Now we state the precise form of the main theorem (Theorem 1.4) of this paper.

**Theorem 6.4.** Given a sequence of ASD instantons on $R \times M$ with uniformly bounded energy, there exists a subsequence which converges to a stable scaled instanton in the sense of Definition 6.3.

The rest of this paper provides a proof of Theorem 6.4.

**Remark 6.5.** It is easy to extend our definitions and our compactness theorem to the case of instantons over $C \times \Sigma$. For example, when $\Gamma$ is a scaled tree with empty base, the object defined in Definition 6.2 is a possible limiting configuration of a sequence of instantons over $C \times \Sigma$. One can then prove the corresponding compactness theorem (Theorem 1.5) for such instantons in a routine way.

**7. Proof of the compactness theorem**

Now we start to prove the compactness theorem (Theorem 6.4). The basic strategy is similar to the proof of Gromov compactness for pseudoholomorphic curves used in [MS04]: starting from the root component, we construct each component of the limiting object inductively. The induction is based on the “soft rescaling” argument in which one can verify the stability condition of the limiting object. Then one uses the annulus lemma to show that the limit object satisfies the matching condition.

Before we start, we remark on certain conventions and simplifications which we will follow in our proof.

**Remark 7.1.**

(a) Reset the value of $\hbar > 0$ such that it is smaller than the $\epsilon$ of the annulus lemma (Proposition 4.5 and Corollary 4.6), the threshold of energy concentration given by Lemma 5.4, the minimal energy of nonconstant holomorphic spheres in $R_\Sigma$, the minimal energy of nonconstant holomorphic disks in $R_\Sigma$ with boundary in $\iota(L_M)$ having at most two switching points, the minimal energy of nonconstant ASD instanton over $C \times \Sigma$ (the existence of the minimal energy is proved in [DS94, Proof of Theorem 9.1] and [Weh06]), and the minimal energy of a nontrivial ASD instanton over $M$ given by Theorem 4.15.

(b) From now on we fix a sequence of ASD instantons $A_i \in A(P)$ satisfying

$$\sup_i E(A_i) < +\infty.$$ 

By Theorem 4.15 and the above item of this remark, we may assume that

$$E(A_i) \geq \hbar \quad (\forall i \geq 1). \quad (7.1)$$
(c) It is also convenient to fix the ambiguity caused by the translation invariance. For each \( i \geq 1 \) and \( Z \in \mathbb{R} \), let \( R = R(Z) \) be the minimal real number satisfying

\[
E(A_i; M \setminus M_{R(Z)}) = \frac{\hbar^2}{2},
\]

By the decay of energy, when \( Z \) approaches to \( \pm \infty \), the \( R(Z) \) satisfying the above condition approaches to \( \infty \). It is also easy to see that \( R(Z) \) depends on \( Z \) continuously. Then one can choose for each \( i \) a number \( Z_i \) (which might not be unique) such that \( R_i = R(Z_i) \) is smallest possible. Then by using a translation of \( M \) in the \( \mathbb{R} \)-direction, we may assume \( Z_i = 0 \) for all \( i \). Hence one has

\[
E(A_i; M \setminus M_{R_i}) = \frac{\hbar^2}{2}. \tag{7.2}
\]

Now we start our construction of the limiting object. By taking a subsequence, we may assume that either \( R_i \) stays bounded or diverges to infinity. We first prove that there is no energy concentrating in the area further away from \( M_{R_i} \).

**Proposition 7.2.** For any \( \rho_i \) with \( \rho_i \gg R_i \), one has

\[
\lim_{i \to \infty} E(A_i; M \setminus M_{\rho_i}) = 0. \tag{7.3}
\]

**Proof.** Follows from the annulus lemma (Proposition 4.5).

\( \square \)

### 7.1. Constructing the root component

Now we are going to construct the root component of the limiting object. A simple situation is when \( R_i \) is bounded. Then by Theorem 2.3, a subsequence of \( A_i \) converges to an ASD instanton \( A_\infty \) with a bubbling measure \( m_\infty \) in the Uhlenbeck sense. Then \( ([A_\infty], m_\infty) \) is the limiting object. Indeed, the scaled tree underlying this limiting object has only one vertex, and Proposition 7.2 is equivalent to the no-energy-lost condition of Definition 6.3.

From now on we assume that \( R_i \) is unbounded and we take a subsequence such that \( R_i \) diverges to infinity. Then by applying Theorem 5.1 for \( R_i \) and \( A_i \), one obtains a subsequence (still indexed by \( i \)) and a holomorphic curve with mass \( \tilde{u}_\infty = (u_\infty, m_\infty) \) from \( H \) to \( R_\Sigma \) such that \( A_i \) converges to \( \tilde{u}_\infty \) along with \( t \) in the sense of Definition 5.2. Proposition 7.2 and (5.2) imply that

\[
\lim_{i \to \infty} E(A_i) = E(u_\infty) + |m_\infty|. \tag{7.4}
\]

We would like to show that this limiting component is stable.

**Lemma 7.3.** If \( u_\infty \) is a constant map, the support of \( m_\infty \) contains either an interior point of \( H \), or contains at least two points in the boundary of \( H \).

**Proof.** The assumption (7.1) and the equality (7.4) imply that if \( u_\infty \) is constant, then \( W_\infty = \text{Supp} m_\infty \neq \emptyset \). Suppose \( W_\infty \subset \partial H \) and contains only one element \( w_0 \). Then all energy is concentrated at \( w_0 \), which means that

\[
\lim_{r \to 0} \lim_{i \to \infty} E(A_i; M \setminus M_{\varphi_{R_i}(B_r^+(w_0))}) = 0.
\]

Since the total energy is at least \( \hbar \), there exists a sequence of \( r_i \to 0 \) such that for all \( i \) sufficiently large,

\[
E(A_i; M \setminus M_{\varphi_{R_i}(B_r^+(w_0))}) = E(A_i; M \setminus M_{B_{r_i R_i}(R, w_0)}) = \frac{\hbar^2}{2}.
\]

Since \( r_i R_i \ll R_i \), this contradicts the choice of \( R_i \) (see (c) of Remark 7.1). \( \square \)
7.2. **Soft rescaling.** What we have done is constructing the root component of the limiting stable scaled instanton. Next we inductively construct all other components using the “soft-rescaling” method. The soft-rescaling method is similar to that of [MS04, Section 4.7], which was also used in [Fra08] for the compactness problem of holomorphic disks, in [Zil14] for the compactness problem of affine vortices, and in [WX17] for the adiabatic limit of disk vortices. Let \( r_\infty \) denote the root component and \( m_\infty \) the bubble measure. Suppose

\[
W_\infty = \text{Supp} m_\infty = \{ w_k \mid k = 1, \ldots, k_\infty \}.
\]

For each \( w_k \in \text{Supp} m_\infty \), one has that

\[
\lim_{r \to 0} \lim_{i \to \infty} E \left( A_i; \varphi_{R_i}(M_{B_i^+(w_k)}) \right) = m_\infty(w_k) \geq \epsilon.
\]  \((7.5)\)

Then choose \( r_0 > 0 \) such that \( B_{2r_0}(w_k) \cap W_\infty = \{ w_k \} \) and for sufficiently big \( i \) there holds

\[
m_\infty(w_k) - \frac{\epsilon}{4} \leq E \left( A_i; \varphi_{R_i}(M_{B_i^+(w_k)}) \right) = E \left( A_i; M_{B_{R_i}^+(R_iw_k)} \right) \leq m_\infty(w_k) + \frac{\epsilon}{4}.
\]

Then for each \( w \in B_{r_0}(w_k) \) and \( i \), one there exists \( r_i(w) > 0 \) such that

\[
E \left( A_i; \varphi_{R_i}(M_{B_i^+(w)}) \right) = m_\infty(w_k) - \frac{\epsilon}{2}.
\]

Notice that the minimal \( r_i(w) \) depends continuously on \( w \). Therefore, one chooses \( w_{i,k} \in B_{r_0}(w_k) \) such that

\[
r_{i,k} = r_i(w_{i,k}) = \min_{w \in B_{r_0}(w_k)} r_i(w).
\]  \((7.6)\)

Since energy concentrates at \( w_0 \), \( w_{i,k} \) converges to \( w_k \) and \( r_i(w_{i,k}) \) converges to zero. The point \( w_{i,k} \) can be viewed as the point near \( w_k \) where energy concentrates in the fastest rate (an analogue of the local maximum of the energy density).

The following argument bifurcates in different cases. Suppose\n
\[
w_{i,k} = s_{i,k} + it_{i,k}.
\]

By choosing a suitable subsequence, we are in one of the following situations.

**Case I.** Suppose we are in the situation where

\[
w_k \in \text{Int} H, \quad \lim_{i \to \infty} r_{i,k} R_i < \infty.
\]  \((7.7)\)

Then choose a sequence \( \rho_i \to \infty \) but grows slower than \( R_i \). Then when \( i \) is large, \( B_{\rho_i}(R_iw_{i,k}) \subset C \) is contained in the upper half plane. Define the sequence of translations on \( C \) by

\[
\varphi_{i,k}(z) = z + R_iw_k.
\]

Via \( \varphi_{i,k} \), the sequence of restrictions of \( A_i \) to \( B_{\rho_i}(R_iw_{i,k}) \times \Sigma \) can be identified with a sequence \( A'_{i,k} \) of solutions to the ASD equation over \( B_{\rho_i} \times \Sigma \). The total energy of \( A'_{i,k} \) is uniformly bounded, hence a subsequence (still indexed by \( i \)) converges in the Uhlenbeck sense to a limiting ASD instanton \( A_k \) over \( M_k := C \times \Sigma \) with a bubbling measure \( m_k \) on \( M_k \).

We prove that no energy is lost.

**Lemma 7.4.** There holds

\[
m_\infty(w_k) = E(A_k) + \int_{M_k} m_k.
\]
Proof. Choose \( \epsilon > 0 \). By assumption, there exists \( r_{i} > 0 \) such that for \( i \) sufficiently large, there holds

\[
E(A'_{i,k}; B_{R_{i}r_{i}} \times \Sigma) \leq m_{\infty}(w_{k}) + \epsilon.
\]

Then the Uhlenbeck convergence \( A'_{i,k} \to (A_{k}, m_{k}) \) implies that

\[
E(A_{k}) + |m_{k}| = \lim_{R \to \infty} \lim_{i \to \infty} E(A'_{i,k}; B_{R} \times \Sigma) \leq \lim_{i \to \infty} E(A_{i,k}; B_{R_{i}r_{i}} \times \Sigma) \leq m_{\infty}(w_{k}) + \epsilon.
\]

On the other hand, we claim that when \( \epsilon \) is small enough, there exists \( \tau_{\epsilon} > 0 \) such that

\[
E(A'_{i,k}; B_{\tau_{\epsilon}} \times \Sigma) \geq m_{\infty}(w_{k}) - 2\epsilon.
\]

(7.8)

If this is not true, then there exist a sequence \( \tau_{i} \to \infty \) such that

\[
E(A'_{i,k}; B_{\tau_{i}} \times \Sigma) = m_{\infty}(w_{k}) - 2\epsilon.
\]

There are also a sequence \( \tau'_{i} > \tau_{i} \) which grows slower than \( R_{i} \) such that

\[
E(A'_{i,k}; B_{\tau'_{i}} \times \Sigma) = m_{\infty}(w_{k}) - \epsilon.
\]

Hence

\[
E(A'_{i,k}; (B_{\tau'_{i}} \setminus B_{\tau_{i}}) \times \Sigma) = \epsilon.
\]

(7.9)

On the other hand, choose \( R_{0} > 1 \) which is bigger than the limit of \( r_{i,k}R_{i} \). Then one has

\[
E(A'_{i,k}; (B_{R_{i}r_{i}} \setminus B_{R_{0}}) \times \Sigma) = E(A'_{i,k}; B_{R_{i}r_{i}} \times \Sigma) - E(A'_{i,k}; B_{R_{0}} \times \Sigma)
\]

\[
\leq E(A'_{i,k}; B_{R_{i}r_{i}} \times \Sigma) - E(A'_{i,k}; B_{R_{i}r_{i}} \times \Sigma) \leq m_{\infty}(w_{k}) + \epsilon - (m_{\infty}(w_{k}) - \frac{\hbar}{2}) = \frac{\hbar}{2} + \epsilon.
\]

The conformal radius of \( B_{R_{i}r_{i}} \setminus B_{R_{0}} \) diverges to infinity; one also has \( \tau_{i}/R_{0} \to \infty \), \( \tau'_{i}/R_{i}r_{i} \to 0 \). Hence when \( \epsilon \) is small enough, (7.9) contradicts Corollary 4.6. Therefore the claim of the existence of \( \tau_{\epsilon} \) satisfying (7.8) is true. Then one has

\[
E(A_{k}) + |m_{k}| \geq \lim_{i \to \infty} E(A'_{i,k}; B_{\tau_{i}} \times \Sigma) \geq m_{v}(w_{k}) - 2\epsilon.
\]

Since \( \epsilon \) can be arbitrary small number, this lemma is proved. \( \square \)

The following lemma follows directly from Lemma 7.4.

Lemma 7.5. If \( A_{k} \) is a trivial instanton, then \( m_{k} \neq 0 \).

Case II. Suppose we are in the situation where

\[
w_{k} \in \partial H, \quad \lim_{i \to \infty} (r_{i,k} + t_{i,k})R_{i} < \infty.
\]

Then choose a sequence \( \rho_{i} \to \infty \) but grows slower than \( R_{i} \). Define the sequence of real translations by (recall \( s_{i,k} = Re w_{i,k} \))

\[
\varphi_{i,k}(z) = z + R_{i}s_{i,k}.
\]

Via \( \varphi_{i,k} \), the sequence of restrictions of \( A_{i} \) to \( M_{B_{R_{i}}^{\pm}(r_{i,k},s_{i,k})} \) can be identified with a sequence \( A'_{i,k} \) of solutions to the ASD equation over \( M_{B_{R_{i}}^{\pm}} \). The energy of \( A'_{i,k} \) is uniformly bounded. Then a subsequence of \( A'_{i,k} \) converges in the Uhlenbeck sense to an ASD instanton \( A'_{k} \) over \( M_{k} = M = R \times M \) with a bubbling measure \( m_{k} \). Similar to Lemma 7.4, one use the annulus lemma (Proposition 4.5) for the case \( N = M^{\text{double}} \) to prove that no energy is lost, i.e.,

\[
m_{\infty}(w_{k}) = E(A_{k}) + \int_{M_{k}} m_{k}.
\]
Similar to the above case, one has the following facts.

**Lemma 7.6.** If $A_k$ is trivial, then $m_k \neq 0$.

**Case III.** Suppose we are in the situation where

$$ w_k \in \text{Int} H, \quad \lim_{i \to \infty} r_{i,k} := \lim_{i \to \infty} r_{i,k} R_i = \infty. $$

In this case, choose a sequence $p_i \to \infty$ which grows faster than $r_{i,k}$ but slower than $R_i$. Then when $i$ is large, the disk $B_{p_i}(R_i w_{i,k}) \subset C$ is contained in the upper half plane. Define a sequence of affine transformations $\phi_{i,k}$ and a sequence of translations $\psi_{i,k}$ by

$$ \phi_{i,k}(z) = r_{i,k} \left( z + \frac{w_{i,k}}{r_{i,k}} \right), \quad \psi_{i,k}(z) = z + \frac{w_{i,k}}{r_{i,k}}. $$

Via $\psi_{i,k}$, the sequence of restrictions of $A_i$ to $B_{p_i}(R_i w_{i,k}) \times \Sigma$ can be identified with a sequence $A_{i,k}'$ of solutions to the ASD equation over $B_{\delta_i} \times \Sigma$. By Theorem 3.3 a subsequence of $A_{i,k}'$ converges modulo bubbling to a holomorphic map $u_k = (u_k, m_k)$ with mass from $C$ to $R_\Sigma$ along with the sequence $\{\tau_{i,k}\}$. Using similar method as proving Lemma 7.4 one can prove that

$$ m_{\mathcal{X}}(w_k) = E(u_k) + \int_C m_k. $$

Moreover, we verify the stability condition as follows.

**Lemma 7.7.** When $u_k$ is a constant, the support of $m_k$ contains at least two points.

**Proof.** If this is not the case, then $m_k$ is supported at a single point $z_0 \in C$ with $m_k(z_0) = m_{\mathcal{X}}(w_k)$. This implies that

$$ \lim_{r \to \infty} \lim_{i \to \infty} E\left( A_{i,k}' ; B_{r\tau_{i,k}}(z_0) \times \Sigma \right) = m_k(z_0) = m_{\mathcal{X}}(w_k). $$

Since $m_k(z_0) \geq h$, there is a sequence $\delta_i \to 0$ such that for large $i$

$$ E\left( A_{i,k}' ; B_{\delta_i\tau_{i,k}}(z_0) \times \Sigma \right) = E\left( A_i ; \phi_{\delta_i}(M_{\delta_i\tau_{i,k}}(w_{i,k} + z_0)) \right) = m_{\mathcal{X}}(w_k) - \frac{h}{2}. $$

Here

$$ r_i' = \frac{\delta_i\tau_{i,k}}{R_i} = \delta_i r_{i,k} $$

which is smaller than $r_{i,k}$. This contradicts the choice of $w_{i,k}$ and $r_{i,k}$ (see (7.6)). Hence the support of $m_k$ contains at least two points.

**Case IV.** Suppose we are in the situation where

$$ w_k \in \partial H, \quad \lim_{i \to \infty} r_{i,k} := \lim_{i \to \infty} R_i(r_{i,k} + t_{i,k}) = \infty. \quad (7.10) $$

Define the sequence of affine transformations $\phi_{i,k}$ and translations $\psi_{i,k}$ by

$$ \phi_{i,k}(z) = r_{i,k}(z + R_is_{i,k}), \quad \psi_{i,k}(z) = z + R_is_{i,k} $$

Denote the pull-back of $A_k$ via $\psi_{i,k}$ by $A_{i,k}'$, which are still ASD instantons over $M$. Without loss of generality, we may assume that $R_i s_{i,k} = 0$ for all $i$. Then by Theorem 5.1, a subsequence of $A_{i,k}'$ (still indexed by $i$) converges to a holomorphic map with mass $\bar{u}_k = (u_k, m_k)$ from $H$ to $R_\Sigma$ along with the sequence $\{\tau_{i,k}\}$. Moreover, one has

$$ \lim_{i \to \infty} \lim_{R \to \infty} E\left( A_{i,k}' ; M_{\phi_{i,k}(B_{R_i})} \right) = E(u_k) + \int_H m_k. $$
As before one can prove that the left hand side above is \( m_\mathcal{X}(w_k) \), hence

\[
m_\mathcal{X}(w_k) = E(u_k) + \int_H m_k.
\]

We verify the stability condition as follows.

**Lemma 7.8.** Suppose \( u_k \) is a constant map.

(a) If \( \lim_{i \to \infty} t_{i,k}/r_{i,k} < \infty \), then \( \text{Supp} m_k \) contains at least two points;

(b) If \( \lim_{i \to \infty} t_{i,k}/r_{i,k} = \infty \), then \( \text{Supp} m_k \) contains at least one point in the interior of \( H \).

**Proof.** Suppose \( \lim_{i \to \infty} t_{i,k}/r_{i,k} < \infty \) and \( \text{Supp} m_k \) contains only one point \( z_0 \in H \) with \( m_k(z_0) = m_\mathcal{X}(w_k) \). This implies that

\[
\lim_{r \to 0} \lim_{i \to \infty} E\left( A_{i,k}'; M^{\phi_{\tau_{i,k}}(B_{\tau_{i,k}}^+(z_0))} \right) = m_\mathcal{X}(w_k).
\]

Then there exists a sequence \( \delta_i \rightarrow 0 \) satisfying

\[
E\left( A_{i,k}'; M^{\phi_{\tau_{i,k}}(B_{\tau_{i,k}}^+(z_0))} \right) = E\left( A_{i,k}'; M^{\phi_{\tau_{i,k}}(B_{\tau_{i,k}}^+(z_0))} \right) = m_\mathcal{X}(w_k) - \frac{h}{2}.
\]

However we have

\[
\frac{\tau_{i,k} \delta_i}{\alpha_{i,k}} = \delta_i(r_{i,k} + t_{i,k}) < r_{i,k},
\]

which contradicts the choice of \( r_{i,k} \) (see (7.6)).

On the other hand, suppose \( \lim_{i \to \infty} t_{i,k}/r_{i,k} = \infty \). Then \( \tau_{i,k} \gg r_{i,k}R_i \). Still assume for simplicity that \( s_{i,k} = \text{Re} w_{i,k} = 0 \). Then for any \( r > 0 \), for \( i \) sufficiently large, the disk centered at \( R_iw_{i,k} = iR_i\tau_{i,k} \) of radius \( r_{i,k}R_i \) is contained in the disk centered at \( i\tau_{i,k} \) of radius \( r_{i,k} \). This implies that \( m_k \) is supported at the single point \( i \in \text{Int} H \).

7.3. **Bubble connects.** The above discussion of four different cases allows us to inductively construct the limiting object. The process stops after finitely many steps due to the energy quantization phenomenon (see item (a) of Remark 7.1). This finishes the construction of a collection of ASD instantons with measures \( ([A_i], m_i) \), a collection of holomorphic maps \( u_v \) from \( S_v \) to \( R \) with boundary lying in \( \iota(L_M) \). They satisfy

\[
\lim_{i \to \infty} E(A_i) = \sum_{v \in V_1^i} \left( E(A_v) + |m_v| \right) + \sum_{v \in V_2^i} E(u_v).
\]

Lemma 7.5, 7.6, 7.7, and 7.8 imply that the collection satisfies the stability condition of Definition 6.2. To show that the limiting object satisfies the definition of stable scaled instantons (Definition 6.2), it remains to prove that this collection form a stable scaled instanton, i.e., proving “bubbles connect” or more precisely, the matching condition of Definition 6.2. We only prove this condition for a boundary edge/node. The case for interior edge/node can be proved in a similar way. Let \( e_{i,v} \rightarrow v \in E_T \) be a boundary edge of the limiting graph \( \Gamma \). Then by the construction of the limit, there are two sequences of Möbius transformations

\[
\varphi_{i,v}(z) = \rho_{i,v}(z + Z_{i,v}), \quad \varphi_{i,v'}(z) = \rho_{i,v'}(z + Z_{i,v'})
\]

with \( Z_{i,v}, Z_{i,v'} \in \mathbb{R} \) and \( \rho_{i,v} \gg \rho_{i,v'} \). By the translation invariance, one can assume that \( Z_{i,v'} \equiv 0 \). Then one has the following lemma (recall the notion of the diameter in Subsection 4.3.)
Lemma 7.9 (Bubble connects). For sufficiently large $s$ one has
\[
\lim_{i \to \infty} \sup_{z \in \text{Ann}^+ \left( e^s \rho_{i,v}, e^{-s} \rho_{i,v} \right)} \| F_{A_i} \|_{L^2(z \times \Sigma)} = 0
\] (7.11)
and
\[
\lim_{s \to \infty} \lim_{i \to \infty} \text{diam}(A_i; M e^s \rho_{i,v}, e^{-s} \rho_{i,v}) = 0.
\] (7.12)

Proof. By the inductive construction of the limiting object, there exists $s_0 > 0$ such that for all sufficiently large $i$, there holds
\[
E(A_i; M e^s \rho_{i,v}, e^{-s} \rho_{i,v}) \leq \frac{\hbar}{2}.
\]
Then the annulus lemma (Proposition 4.5), one has
\[
\lim_{s \to \infty} \lim_{i \to \infty} E(A_i; M e^s \rho_{i,v}, e^{-s} \rho_{i,v}) = 0.
\]
Then (7.11) follows from standard estimate of the ASD equation. Hence the diameter in the limit (7.12) is defined for sufficiently large $s$ and sufficiently large $i$. Then (7.12) follows from the diameter bound given by Lemma 4.9 and Lemma 4.10. \hfill \Box

The above lemma implies that the evaluations of the two sides of the node corresponding to the edge $e_{v' > v}$ (which are points in $L_N$) agree after being mapped to $\iota(L_M) \subset R$. One can use the ASD equation on $\partial \text{Ann}^+ \left( e^s \rho_{i,v'}, e^{-s} \rho_{i,v} \right) \times M_0$ and the fact that the energy of the solution restricted to this region shrinks to zero to prove that the evaluations are the same point of $L_N$ (after transposing). This finishes the proof of the matching condition for the edge $e_{v' > v}$.

We declare that we have finished the proof of Theorem 6.4 (the same as Theorem 1.4). As we have addressed previously, the case of instantons over $C \times \Sigma$ (Theorem 1.5) can be proved using a similar and more simplified method.

References

[AJ10] Manabu Akaho and Dominic Joyce, Immersed Lagrangian Floer theory, Journal of Differential Geometry 86 (2010), no. 3, 381–500.

[Ati88] Michael Atiyah, New invariants of three-dimensional and four-dimensional manifolds, Proceedings, The Mathematical Heritage of Hermann Weyl: Durham, NC, USA, May 12-16, 1987, Proceedings of Symposia in Pure Mathematics, vol. 48, 1988, pp. 285–299.

[BJSV95] Michael Bershadsky, Andrei Johansen, Vladimir Sadov, and Cumrun Vafa, Topological reduction of 4D SYM to 2D $\sigma$-models, Nuclear Physics B 448 (1995), no. 1-2, 166–186.

[BW18] Nathaniel Bottman and Katrin Wehrheim, Gromov compactness for squiggly strip shrinking in pseudoholomorphic quilts, Selecta Mathematica 24 (2018), 3381–3443.

[CGS00] Kai Cieliebak, Ana Gaio, and Dietmar Salamon, $J$-holomorphic curves, moment maps, and invariants of Hamiltonian group actions, International Mathematics Research Notices 16 (2000), 831–882.

[Che98] Jingyi Chen, Convergence of anti-self-dual connections on $SU(n)$-bundles over product of two Riemann surfaces, Communications in Mathematical Physics 196 (1998), 571–590.

[DF18] Aliakbar Daemi and Kenji Fukaya, Atiyah-Floer conjecture: a formulation, a strategy of proof and generalizations, Modern geometry: a celebration of the work of Simon Donaldson, Proceedings of Symposia in Pure Mathematics, vol. 99, American Mathematical Society, Providence, RI, 2018, pp. 23–57.

[DK90] Simon Donaldson and Peter Kronheimer, The geometry of four-manifolds, Oxford Mathematical Monographs, Clarendon Press - Oxford, 1990.

[Don02] Simon Donaldson, Floer homology groups in Yang–Mills theory, Cambridge Tracts in Mathematics, vol. 147, Cambridge University Press, Cambridge, 2002.

[DS94] Stamatis Dostoglou and Dietmar Salamon, Self-dual instantons and holomorphic curves, Annals of Mathematics 139 (1994), no. 3, 581–640.
Corrigendum: Self-dual instantons and holomorphic curves, Annals of Mathematics
165 (2007), no. 2, 665–673.

David Duncan, Compactness results for neck-stretching limits of instantons, https://arxiv.org/abs/1212.1547v5.

Daniel Archilla, Compactness results for the quilted Atiyah–Floer conjecture, Ph.D. thesis, Rutgers, the State University of New Jersey, 2013.

Andreas Floer, An instanton-invariant for 3-manifolds, Communications in Mathematical Physics
118 (1988), no. 2, 215–240.

Morse theory for Lagrangian intersections, Journal of Differential Geometry 28 (1988),
513–547.

Symplectic fixed points and holomorphic spheres, Communications in Mathematical Physics
120 (1989), 575–611.

Fukaya, Yong-Geun Oh, Hiroshi Ohta, and Kaoru Ono, Lagrangian intersection Floer theory: anomaly and obstruction, part i & ii, AMS/IP studies in advanced mathematics, vol. 46, American Mathematical Society, 2009.

Urs Frauenfelder, Gromov convergence of pseudoholomorphic disks, Journal of Fixed Point Theory and Applications 3 (2008), 215–271.

Kenji Fukaya, Floer homology for 3-manifolds with boundary 1, http://www.math.kyoto-u.ac.jp/~fukaya/bdrt1.pdf.

SO(3)-Floer homology for 3-manifolds with boundary 1, https://arxiv.org/abs/1506.01435.

Floer homology of connected sum of homology 3-spheres, Topology 35 (1996), no. 1, 89–136.

Anti-self-dual equation on 4-manifolds with degenerate metric, Geometric and Functional Analysis 8 (1998), no. 3, 466–528.

Categorification of invariants in gauge theory and symplectic geometry, Japanese Journal of Mathematics 13 (2018), no. 1, 1–65.

The symplectic nature of fundamental groups of surfaces, Advances in Mathematics 54 (1984), no. 2, 200–225.

Ana Gaio and Dietmar Salamon, Gromov–Witten invariants of symplectic quotients and adiabatic limits, Journal of Symplectic Geometry 3 (2005), no. 1, 55–159.

Sergei Ivashkovich and Vsevolod Shevchishin, Gromov compactness theorem for J-complex curves with boundary, International Mathematics Research Notices 2000 (2000), no. 22, 1167–1206.

Dusa McDuff and Dietmar Salamon, J-holomorphic curves and symplectic topology, Colloquium Publications, vol. 52, American Mathematical Society, 2004.

Ignasi Mundet i Riera, Yang–Mills–Higgs theory for symplectic fibrations, Ph.D. thesis, Universidad Autónoma de Madrid, 1999.

Hamiltonian Gromov–Witten invariants, Topology 43 (2003), no. 3, 525–553.

Florin Manolescu and Chris Woodward, Floer homology on the extended moduli space, Perspectives in analysis, geometry, and topology, Progress in Mathematics, vol. 296, Birkhäuser/Springer, Basel, Switzerland, 2012.

Takeo Nishinou, Convergence of adiabatic family of anti-self-dual connections on products of Riemann surfaces, Journal of Mathematical Physics 51 (2010), no. 2, 10 pp.

Eleonora Nezza, Giampiero Palatucci, and Enrico Valdinoci, Hitchhiker’s guide to the fractional Sobolev spaces, https://arxiv.org/abs/1104.4345.

Yong-Geun Oh, Floer cohomology of Lagrangian intersections and pseudo-holomorphic disks. I, Communications on Pure and Applied Mathematics 46 (1993), no. 7, 949–993.

Floer cohomology of Lagrangian intersections and pseudo-holomorphic disks. II, (CP^n, RP^{n+1}), Communications on Pure and Applied Mathematics 46 (1993), no. 7, 995–1012.

Peter Ozsváth and Zoltán Szabó, Holomorphic disks and topological invariants for closed three-manifolds, Annals of Mathematics 159 (2004), no. 3, 1027–1108.

Marcin Poźniak, Floer homology, Novikov rings and clean intersections, Ph.D. thesis, University of Warwick, 1994.

Dietmar Salamon, Lagrangian intersections, 3-manifolds with boundary, and the Atiyah–Floer conjecture, Proceedings of the International Congress of Mathematicians, Springer, 1995, pp. 526–536.

Dietmar Salamon and Katrin Wehrheim, Instanton Floer homology with Lagrangian boundary conditions, Geometry & Topology 12 (2008), 747–918.
[Tau90] Clifford Taubes, *Casson’s invariant and gauge theory*, Journal of Differential Geometry 31 (1990), 547–599.

[Weh03] Katrin Wehrheim, *Uhlenbeck compactness*, European Mathematical Society series of lectures in mathematics, European Mathematical Society, 2003.

[Weh05a] ______, *Anti-self-dual instantons with Lagrangian boundary conditions I: Elliptic theory*, Communications in Mathematical Physics 254 (2005), no. 1, 45–89.

[Weh05b] ______, *Anti-self-dual instantons with Lagrangian boundary conditions II: Bubbling*, Communications in Mathematical Physics 258 (2005), no. 2, 275–315.

[Weh05c] ______, *Lagrangian boundary conditions for anti-self-dual instantons and the Atiyah–Floer conjecture*, Journal of Symplectic Geometry 3 (2005), no. 4, 703–747.

[Weh06] ______, *Energy identity for anti-self-dual instantons on $\mathbb{C} \times \Sigma$*, Mathematical Research Letters 13 (2006), 161–166.

[WW15] Katrin Wehrheim and Chris Woodward, *Pseudoholomorphic quilts*, Journal of Symplectic Geometry 13 (2015), 849–904.

[WX] Chris Woodward and Guangbo Xu, *An open quantum Kirwan map*, https://arxiv.org/abs/1806.06717.

[WX17] Dongning Wang and Guangbo Xu, *A compactification of the moduli space of disk vortices in adiabatic limit*, Mathematische Zeitschrift 287 (2017), 405–459.

[Zil05] Fabian Ziltener, *Symplectic vortices on the complex plane and quantum cohomology*, Ph.D. thesis, Swiss Federal Institute of Technology Zurich, 2005.

[Zil14] ______, *A quantum Kirwan map: bubbling and Fredholm theory*, Memoirs of the American Mathematical Society 230 (2014), no. 1082, 1–129.