THE WILD WORLD OF 4–MANIFOLDS

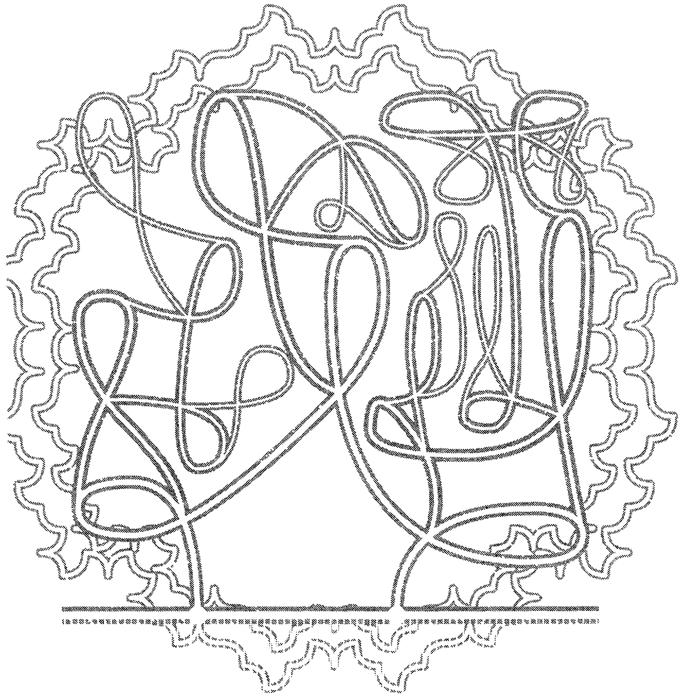
ALEXANDRU SCORPAN

AMERICAN MATHEMATICAL SOCIETY

THE WILD WORLD OF 4–MANIFOLDS

THE WILD WORLD OF 4–MANIFOLDS

ALEXANDRU SCORPAN



AMERICAN MATHEMATICAL SOCIETY
PROVIDENCE, RHODE ISLAND

2000 *Mathematics Subject Classification*. Primary 57–02, 57N13; Secondary 14J80, 32Q55, 57R17, 57R57, 57R65.

For additional information and updates on this book, visit
www.ams.org/bookpages/fourman

Library of Congress Cataloging-in-Publication Data

Scorpan, Alexandru, 1974–
The wild world of 4-manifolds / Alexandru Scorpan.
p. cm.
Includes bibliographical references and index.
ISBN 0-8218-3749-4 (alk. paper)
1. Four-manifolds (Topology) I. Title.

QA613.2.S36 2005
514′.3—dc22 2004062768

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 10 09 08 07 06 05

To the city of Berkeley in California,
to its campus and math department,
and especially to Rob Kirby

Preview

Wild at heart

Dimension 4 is unlike any other dimension. Consider the following facts:

Let M^n be a closed topological n–manifold. Then:

- *If $n \leq 3$, there is exactly one smooth structure on M.*
- *If $n \geq 5$, there are at most finitely-many smooth structures on M.*
- *If $n = 4$, there are many simply-connected closed 4–manifolds that admit infinitely-many distinct smooth structures; there are no smooth 4–manifolds known to have only finitely-many smooth structures.*

One is thus easily tempted to conjecture that all smooth closed 4–manifolds admit countably-many distinct smooth structures.

For open manifolds, things get even worse:

For $n \neq 4$, the topological manifold $\mathbb{R}^n$ admits a unique smooth structure. However, the topological 4–manifold $\mathbb{R}^4$ admits uncountably-many distinct smooth structures.

Many other open 4–manifolds admit uncountably-many distinct structures, and it is an unanswered question whether all do.

A way to think about this might be that dimension 4 is an unstable boundary case: the dimension is big enough to have room for wild things to happen, but the dimension is too small to allow room to tame and undo the wildness.

The goals

This is not a textbook.[1] We wish to offer a comfortable overview of the differential topology of dimension 4 (as it appears in 2004), a pleasant-to-read global picture, presenting the main results while suggesting some of the techniques and attempting to convey the flavor of the subject. It can be used as a travel guide, as a supplemental course material, or as a companion reader in 4–manifolds.

The audience we are trying to address includes graduate students who are learning the subject and want a panorama to be used in parallel or before other more thorough sources, or who need a bibliographical guide to the literature.

We also address mathematicians of various backgrounds, from the merely curious outsider to those genuinely interested in 4–manifolds.

The contents

This is a book about 4–dimensional topology. We restrict attention to simply-connected 4–manifolds, since that provides more than enough ground to cover. The eventual focus is on smooth 4–manifolds, but we cannot avoid discussion of topological 4–manifolds, and even a bit of higher-dimensional manifolds.

1. As motivation and context, we start with a presentation of the main technique used for dealing with higher-dimensional manifolds, namely the h–cobordism theorem. We point out how and why it fails in dimension 4: it is all about embedding disks. Then, we look at the attempt to overcome this difficulty, and review Freedman's success for topological 4–manifolds.

2. After that, we will focus on the main invariant of a 4–manifold, its intersection form. After defining it and describing in detail the fundamental example of the $K3$ complex surface, we proceed to relate intersection forms with the topology of the 4–manifold, culminating with Freedman's classification of topological 4–manifolds. We immediately counter with Donaldson's first startling result about smooth 4–manifolds, which in particular leads to exotic $\mathbb{R}^4$'s.

3. As a good source of examples of smooth 4–manifolds, we take a rapid trip through the geometry of complex surfaces, review the Enriques–Kodaira classification of complex surfaces, compare the complex point-of-view with the smooth point-of-view, and conclude by presenting in some detail the important class of (simply-connected) elliptic surfaces. As a side-effect,

---

[1] *Ceci n'est pas une pipe.*

we stumble upon infinite families of distinct smooth 4–manifolds that are homeomorphic.

4. Finally, in last part of the book we look at the explosion of results coming from the application of gauge theory to 4–manifolds. After a quick glimpse at Donaldson theory, we focus on its more recent replacement, Seiberg–Witten theory. After discussing general results and Taubes' interpretation on symplectic manifolds, we survey the problem of determining the minimum genus needed to represent a fixed homology class by embedded surfaces in a 4–manifold. At the end we describe a construction of R. Fintushel and R. Stern that yields even more infinite families of homeomorphic but non-diffeomorphic 4–manifolds.

The second and fourth parts of the book are more in depth than the expository first and third parts.

What is omitted. Everything non-simply-connected is essentially ignored. There is little discussion of Kirby calculus, which is already well-covered in the literature. The interaction between 4–manifolds and 3–manifolds, even though fundamental to the subject, is only incidentally present. In particular, $3+1$ topological quantum field theories, such as the newly-emerging Ozsváth–Szabó Heegard–Floer homologies, are not discussed.

There is little discussion of the symplectic geometry of 4–manifolds, even though there are a lot of results appearing from that area that might eventually become of fundamental importance for general smooth 4–manifolds. In particular, Lefschetz pencils/fibrations and their extension to general 4–manifolds are only briefly mentioned. Complex geometry is merely surveyed, while Riemannian geometry only plays occasional auxiliary roles.

Errata and other inevitable comments or updates that will arise after printing will be maintained on the `arXiv`[2] and also at the online AMS Bookstore.[3] The readers are encouraged to inform the author of any items that need to be included.

Travel guide

Alongside the *main text*, this volume contains several other layers: there are *footnotes*, *inserted notes* (indented paragraphs with smaller type), *end-notes* at the closing of each chapter, and of course *proofs*. The structure of this book is not linear, and varied readings will offer panoramas ranging from

---

[2] Search on `http://arXiv.org/archive/math`, or its friendlier front-end `http://front.math.ucdavis.edu`.

[3] See `http://www.ams.org/bookstore-getitem/item=fourman`.

a rapid survey to almost a textbook. The extensive index at the end of the volume should not be overlooked as a navigation tool.

The *main text* is what a minimalist tourist might want to read at a first visit, skipping all of the other layers, including the proofs. We have strived to make such a trip quite reasonable and somewhat scenic with the almost three hundred pictures on the side of the road. The main text, devoid of pictures and cleaned of the other layers, would occupy about two hundred pages.

Other itineraries are of course possible, and the reader should choose the one that fits best. The drawback of facilitating such alternative trajectories, as well as addressing as wide an audience as possible, is that repetitions must be made.

The *end-notes* of each chapter contain side developments or central arguments whose exclusion from the main text seemed to help streamline the latter—often the end-notes contain detailed proofs of statements made in the main text. At times, the end-notes to a chapter are quite extensive. The end-notes are cut into titled parts and could be viewed as appendices to their respective chapters. The end-notes also contain a section of commented bibliographical references for each chapter.

The *inserted notes* are, in a way, smaller notes or comments that we felt did not interfere with the main track of the volume, but could still be skipped by a fast traveller. Sometimes, when such a note seemed to halt the main tour, it was exiled into a footnote.

The *footnotes* also contain cross-references. Also, to make this volume readable to a wider audience, we used footnotes to recall certain definitions and technical details. Whether you should let your gaze descend upon a footnote depends on your background and interest on the marked words.[4]

The *proofs* of various results range from a mere mention of the ideas involved to fully detailed arguments, and in any case they can be safely skipped. To help with such a jump, their text was indented.

As with all tourism, a judicious choice of what to visit and what to skip is part of making for a pleasant journey, especially when revisiting only involves picking up the book again.

---

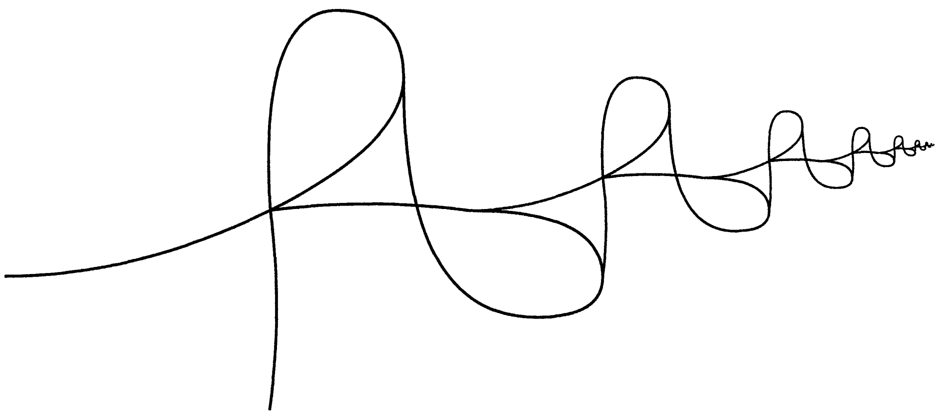
[4] We have consistently avoided marking with footnotes any mathematical formulae or symbols. When a footnote refers to a math item, the footnote mark appears on the English word preceding it.

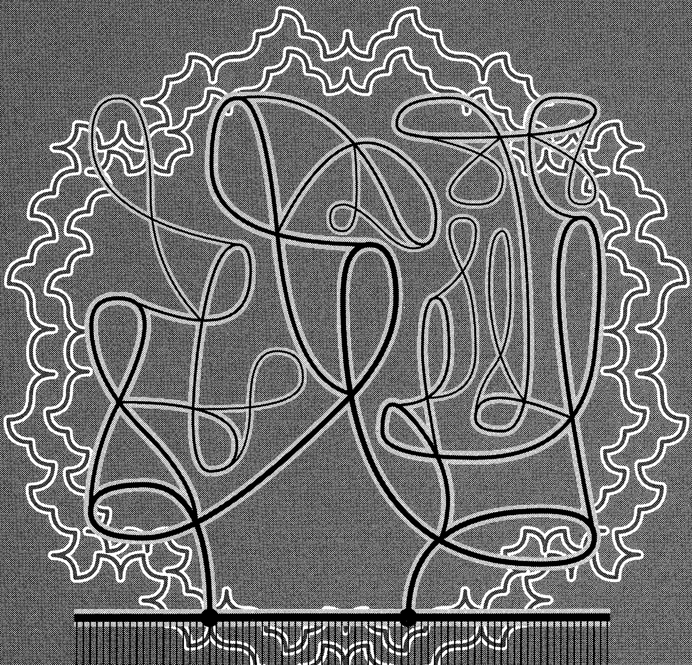
Contents

Part IV. Gauge Theory on 4–Manifolds

Contents of the Notes

THE WILD WORLD OF 4–MANIFOLDS





Introduction

THIS book is concerned with *simply-connected,* closed oriented 4–manifolds and is focused on *smooth* 4–manifolds. However, for a better understanding of smooth 4–manifolds one needs to have some perspective on their mathematical context.

In their immediate neighborhood, smooth 4–manifolds are included in the much wider class of topological 4–manifolds: we will need to see how the world of topological manifolds is different. At this moment, the topological realm is in fact much better understood than the smooth realm, while the latter has just started to unveil its wildness. Contrasting the two territories is necessary for gaining the proper perspective. A first remark is that, by softening our outlook from differential to topological, we make many smooth manifolds look topologically the same.[5] A second remark is that most topological manifolds do not admit smooth structures whatsoever.

In the opposite direction, not by weakening the structure but by strengthening its rigidity, lies the realm of 4–manifolds that admit complex structures, namely the empire of complex surfaces. These are also better understood than smooth 4–manifolds and are an excellent source of examples. The extra rigidity of the complex realm ensures that many complex surfaces that look the same as smooth 4–manifolds are different as complex manifolds. And, of course, most smooth 4–manifolds do not admit any complex structures whatsoever.

---

[5] You should probably think of this as analogous to considering topological spaces only up to homotopy equivalence: many quite different spaces look homotopically the same.

In order to gain the proper perspective on 4–manifolds, it is also inevitable to peek at what happens in other dimensions. In lower dimensions, manifolds of dimension 1 are a bore, manifolds of dimension 2 have been well-understood for quite a while, while manifolds of dimension 3 (modulo the Poincaré conjecture) are essentially governed by their fundamental groups (but of course are far from being completely deciphered). In any case, the distinction between smooth and topological manifolds (or complex manifolds, for dimension 2) does not exist in lower dimensions, and simply-connected manifolds are uninteresting.

On the other hand, in dimensions 5 and higher, a theory of a different flavor has been developed, taking advantage of the extra room available. For simply-connected high-dimensional manifolds, the main technical tool is the h–cobordism theorem, discovered in the 1960s. Its power in helping clear the waters in high dimensions cannot be understated, and its author, S. Smale, received a Fields Medal for discovering it.

Such a powerful tool available from just one dimension higher than the realm of 4–manifolds can only tempt one to extend it to our land as well. An examination of its high-dimensional proof reveals that it hinges on embedding 2–dimensional disks in the manifold, which is easy in dimension 5 or more, but not in dimension 4. Eventually, M. Freedman was able to prove in 1981 the h–cobordism theorem for dimension 4, but at the price of dropping differentiability and softening to the more flexible domain of topological manifolds. This enabled him to quickly obtain a complete classification of simply-connected topological 4–manifolds, and earned him a Fields Medal.

In contrast, just one year later S.K. Donaldson showed that the realm of smooth 4–manifolds is not yet understood. Making use of differential-geometric methods, he showed that most topological 4–manifolds do not admit any smooth structures. Later, he exhibited smoothly-distinct 4–manifolds that look the same topologically, and even infinite families of such. These results led to a Fields Medal as well.

After about ten more years, in 1994, N. Seiberg and E. Witten came up with a different approach to Donaldson's insights, which was much easier to use and thus proved to be quite more powerful.[6] While Donaldson's methods worked best on complex surfaces, the Seiberg–Witten techniques are more flexible, and led to new striking results. Among them is a method (due to R. Fintushel and R. Stern) for modifying many 4–manifolds in a manner that alters their smooth structure but does not change their topological type.

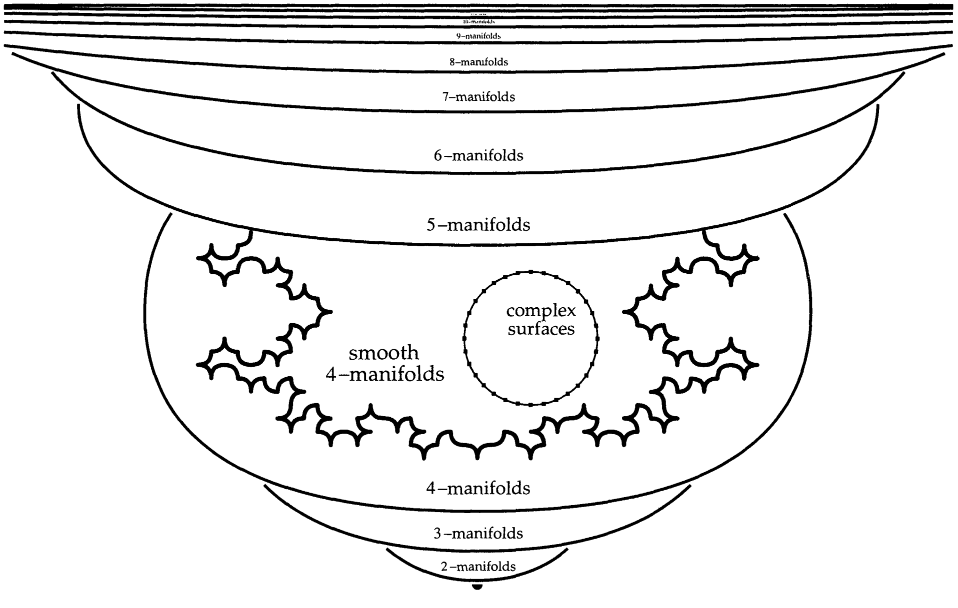
---

[6] Witten already held a Fields Medal for previous work.

The paradoxical result of all these advances is that they just made more and more obvious the level of our current ignorance, opening windows toward vast fields of unsuspected phenomena for which we presently do not have powerful enough methods of exploration. As a simple example, currently there are no tools for studying smooth manifolds homeomorphic to the 4–dimensional sphere: there might be infinitely many distinct such creatures, or just good old S^4.

It's a wide and wild world out there.

There goes the neighborhood

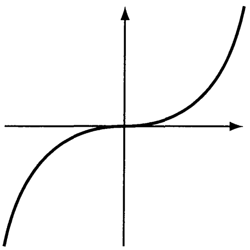


Front matter

The quickest review

Manifolds. A topological **manifold** of dimension m (or m–manifold) is a separable metrizable[7] topological space X that is locally $\mathbb{R}^m$; that is to say, for every $x \in X$, there is an open neighborhood U of x and a homeomorphism $\Phi\colon U \simeq U' \subset \mathbb{R}^m$. Two homeomorphic topological manifolds are considered identical. A **homeomorphism** is an invertible continuous map whose inverse is also continuous.

An m–manifold X is said to admit a **smooth structure** if one can find homeomorphisms $\Phi_\alpha\colon U_\alpha \simeq U'_\alpha \subset \mathbb{R}^m$ with the U_α's covering all X and with all the overlaps $\Phi_\alpha \circ \Phi_\beta^{-1}$ differentiable ($\mathcal{C}^\infty$). In general, **smooth** will stand for "$\mathcal{C}^\infty$–differentiable". Two smooth manifolds that are diffeomorphic are considered identical. A **diffeomorphism** is a smooth invertible map whose inverse is also smooth.



A. Smooth homeomorphism $\mathbb{R} \to \mathbb{R}$ that is not a diffeomorphism

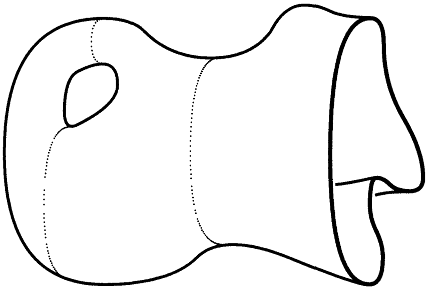
---

[7] Asking a locally Euclidean space to be separable and metrizable is equivalent to several other sets of conditions, such as Hausdorff and separable, Hausdorff and paracompact, *etc.* They are all meant to exclude pathological spaces, such as the *line with two origins* or the *long line,* see for example [**Hir94**, sec 1.1, exercise 9 & 10].

A $2m$-manifold X is said to admit a **complex structure** if one can find homeomorphisms $\Phi_\alpha\colon U_\alpha \simeq U'_\alpha \subset \mathbb{C}^m$ with the U_α's covering all X and with all the overlaps $\Phi_\alpha \circ \Phi_\beta^{-1}$ holomorphic. A map is called **holomorphic** if it is smooth and its differential is $\mathbb{C}$-linear. Two complex manifolds are considered identical if they are biholomorphic. A **biholomorphism** is an invertible holomorphic map whose inverse is also holomorphic.

> *Note on PL structures. Besides topological and smooth structures on a manifold, there is a third intermediate type of structure:* **piecewise-linear** *(or* **PL***) structure (an atlas of compatible local triangulations or a nice global triangulation*[8]*). In dimension* 4 *though (as in all dimensions below* 7*), a piecewise-linear structure is essentially equivalent to a smooth structure: A topological* 4*-manifold admits a piecewise-linear structure if and only if it admits a smooth structure; even more, if two* 4*-manifolds are piecewise-linearly homeomorphic, then they must be diffeomorphic.*[9] *As a consequence, we will not concern ourselves with PL structures any further.*[10]

Boundaries. An m-manifold X is said to have boundary if it is locally modeled not only on $\mathbb{R}^m$, but also has regions modeled on $\mathbb{R}^m_+ = \{(x_1, \ldots, x_m) \mid x_m \geq 0\}$. The **boundary** ∂X of X is the $(m-1)$-manifold coming from the portions of X mapped to $\partial \mathbb{R}^m_+ = \{(x_1, \ldots, x_{m-1}, 0)\}$. See figure B. Structures on X usually induce corresponding structures on ∂X. The manifold $X \setminus \partial X$ is called the **interior** of X and is denoted by $\operatorname{Int} X$.



B. A manifold with boundary

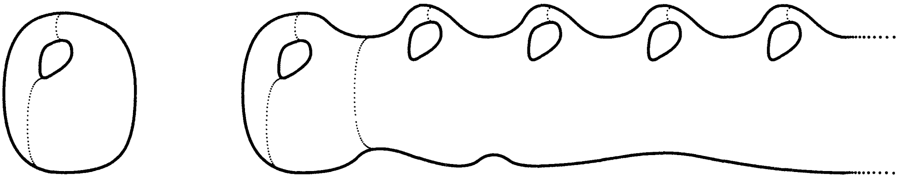
---

[8] The word "nice" is of the essence: many topological manifolds admit triangulations without being piecewise-linear. A triangulation of M is piecewise-linear if M is a **combinatorial manifold**, *i.e.*, if the link of every vertex is simplicially-homeomorphic to a sphere. This is equivalent to the existence of an atlas of charts $\Phi_\alpha\colon U_\alpha \to \mathbb{R}^m$ to $\mathbb{R}^m$, with the overlaps $\Phi_\alpha \circ \Phi_\beta^{-1}$ piecewise-linear homeomorphisms. A map $f\colon A \to B$ between two triangulated spaces is **piecewise-linear** (**PL**) if there are subdivisions of the triangulations of A and B so that f maps the resulting simplices linearly (affinely) to simplices.

[9] This follows essentially from **J. Cerf**'s ***Sur les difféomorphismes de la sphère de dimension trois*** **($\Gamma_4 = 0$)** **[Cer68a]**; see also the end-notes of chapter 4 (smoothing topological manifolds, page 207).

[10] See **C. Rourke** and **B. Sanderson**'s ***Introduction to piecewise-linear topology*** **[RS72]** for an introduction to the PL world.

A manifold is called **closed** if it is compact and without boundary. A manifold is called **open** if it is non-compact and without boundary. See also figure C. We usually assume our manifolds to be connected. Most of our manifolds will be closed or at least compact.[11]



C. A closed manifold and an open manifold

Bundles. A **vector bundle** of rank k (or a **k–plane bundle**) over an m–manifold X is an open $(m+k)$–manifold E together with a map $p\colon E \to X$ such that each fiber $p^{-1}[x]$ is a smoothly-varying copy of the vector space $\mathbb{R}^k$. We always think of the base X as embedded in E as its zero-section. We denote by $E|_Y$ the *restriction* of E to any $Y \subset X$, defined as the bundle $p\colon p^{-1}[Y] \to Y$. In particular, the fiber $p^{-1}[x]$ will be denoted by $E|_x$.

A k–plane bundle E is said to be **trivial** if there is some bundle-isomorphism of $E \to X$ with $X \times \mathbb{R}^k \to X$. A bundle E is said to be **trivialized** if a specific bundle isomorphism $E \approx X \times \mathbb{R}^k$ has been chosen (with the actual isomorphism usually considered only up to homotopy). Notice that a trivial bundle can have many non-equivalent trivializations.[12] The word **frame** will be used synonymous with basis or with linearly independent set of vectors. Thus, a **j–frame field** in E is a field of j linearly independent sections, while a **frame field** is a field of bases in E. A **framed bundle** is the same as a trivialized bundle.

Given a bundle $p\colon E \to X$ and a map $f\colon Y \to X$, we can build the **pull-back** $f^*E = \big\{(y,e) \in Y \times E \mid f(y) = p(e)\big\}$, which becomes a vector bundle over Y when endowed with the natural projection $\mathrm{pr}_1\colon f^*E \to Y$, and then fits in the diagram

$$\begin{tikzcd}
f^*E \arrow[r, "\mathrm{pr}_2"'] \arrow[d, "\mathrm{pr}_1"] & E \arrow[d, "p"] \\
Y \arrow[r, "f"] & X .
\end{tikzcd}$$

Essentially, over each y is brought back the fiber of E over $f(y)$, and together these fibers make up the bundle f^*E; notice that the projection

---

[11] When we will call a manifold "compact" rather than "closed", it is probably safe to assume that the manifold has non-empty boundary.

[12] For example, the trivializations of the bundle $\mathbb{S}^1 \times \mathbb{R}^2 \to \mathbb{S}^1$, up to homotopy, are in bijective correspondence with $\pi_1 SO(2) = \mathbb{Z}$.

$\mathrm{pr}_2\colon f^*E \to E$ is an isomorphism on fibers. Often, restrictions $E|_A$ are better understood as pull-backs ι^*E though the inclusion $\iota\colon A \subset X$. That is especially relevant when we allow A to be merely immersed in X, for example for $E|_{curve}$.

Given vector bundles $E \to X$ and $F \to X$, one can build other bundles, such as $\mathrm{Hom}(E, F)$, or $E \oplus F$, or $E \otimes F$, or the **dual** $E^* = \mathrm{Hom}(E, \mathbb{R})$. These appear by letting the corresponding algebraic operation act fiberwise on the bundles.

After picking a random inner-product in the fibers of a vector bundle, we can define its associated **disk bundle** $\mathbb{D}E \to X$, whose fiber $\mathbb{D}E|_x$ is the disk of unit radius inside $E|_x$, and we can define its **sphere bundle** $\mathbb{S}E \to X$, whose fiber $\mathbb{S}E|_x$ is the sphere of unit radius inside $E|_x$.

Tangent bundles. Every smooth m–manifold X admits a natural m–plane bundle called the **tangent bundle** T_X, that is, an m–plane bundle whose fiber over each $x \in X$ is a vector space that best approximates X around x, offering the infinitesimal picture of a neighborhood of x in X. The tangent bundle T_X is built from pieces $U_\alpha \times \mathbb{R}^m$ glued by identifying (x, v_α) from $U_\alpha \times \mathbb{R}^m$ with (x, v_β) from $U_\beta \times \mathbb{R}^m$ if and only if $v_\alpha = d(\Phi_\alpha \Phi_\beta^{-1})|_x \cdot v_\beta$.

Every smooth map $f\colon X \to Y$ has a **differential** (or **derivative**) $df\colon T_X \to T_Y$ that sends linearly the fiber over $x \in X$ to the fiber over $f(x) \in Y$, offering the linear-infinitesimal picture of f. It fits in the diagram

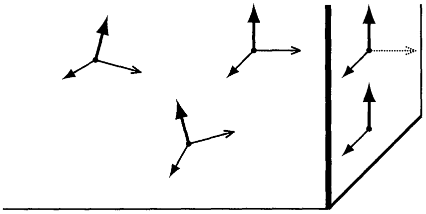
$$\begin{tikzcd}
T_X \arrow[r, "df"'] \arrow[d] & T_Y \arrow[d] \\
X \arrow[r, "f"] & Y .
\end{tikzcd}$$

In particular, any function $f\colon X \to \mathbb{R}$ has a differential $df\colon T_X \to \mathbb{R}$, which is best thought of as a section in the dual bundle T_X^*, and is often called a **1–form.**

From T_X many other wondrous objects appear, such as the cotangent bundle T_X^* and the ***p*–forms** from $\Lambda^p(T_X^*)$. The latter are the skew-symmetric part of $T_X^* \otimes \cdots \otimes T_X^*$.

Orientations. The space of bases of any vector space V has two connected components called **orientations** of V. A change-of-basis matrix A preserves orientation if $\det A > 0$ and changes it if $\det A < 0$.

A smooth manifold is called **oriented** if a coherent choice of orientation has been made in all fibers of T_X. This is equivalent to saying that there is a choice of Φ_α's such that the overlaps $\Phi_\alpha \circ \Phi_\beta^{-1}$ have Jacobian determinant $\det d(\Phi_\alpha \Phi_\beta^{-1})$ everywhere-positive. All our manifolds will be orientable, and we usually assume them to have a fixed favorite orientation. Notice that an orientation of X induces a natural orientation[13] on ∂X, as suggested in figure D.



D. Inducing an orientation on the boundary: outer first

In homology, an orientation of an m–dimensional vector space V is a choice of generator for $H^m(V,\, V \setminus 0;\, \mathbb{Z}) \approx \mathbb{Z}$. An orientation of an m–manifold X is a class in $H^m(T_X,\, T_X \setminus X;\, \mathbb{Z})$ which restricts to generators for each $H^m(T_X|_x,\, T_X|_x \setminus x;\, \mathbb{Z})$. Through the duality isomorphism $H^m(T_X,\, T_X \setminus X;\, \mathbb{Z}) \approx H_m(X; \mathbb{Z})$, this corresponds to a choice of a generator[14] $[X] \in H_m(X; \mathbb{Z}) \approx \mathbb{Z}$ called the **fundamental cycle** of X. Every closed oriented m–manifold satisfies the **Poincaré duality:** $H^k(X; \mathbb{Z}) \approx H_{m-k}(X; \mathbb{Z})$ through $\alpha \mapsto \alpha \cap [X]$.

Technology

Special types of maps. A map $f\colon X \to Y$ between two smooth manifolds is called an **immersion** if its differential $df\colon T_X \to T_Y$ is fiberwise one-to-one (injective). This means that locally f looks like the linear inclusion of $\mathbb{R}^k$ into $\mathbb{R}^m$. A map $f\colon X \to Y$ is called a **submersion** if df is fiberwise onto (surjective on) $T_Y|_{f[X]}$. This means that locally f looks like the linear projection of $\mathbb{R}^m$ onto $\mathbb{R}^k$.

For any map $f\colon X^m \to Y^n$, a point $y \in Y$ is called a **regular value** of f if either $y \notin f[X]$ or $df|_x$ is surjective for all $x \in f^{-1}[y]$. Most y's from Y are regular values[15] of f, and therefore for most y's the set $f^{-1}[y]$ is either empty or a manifold of dimension $m - n$.

If X is compact, then an injective immersion is a diffeomorphism onto its image and is thus called an **embedding.** (A *topological embedding* is merely a homeomorphism onto its image.) A **submanifold** Y^k of an m–manifold X^m is a manifold embedded in X. Locally, it looks like $\mathbb{R}^k$ inside $\mathbb{R}^n$. (If X has

---

[13] Remember Stokes' theorem?

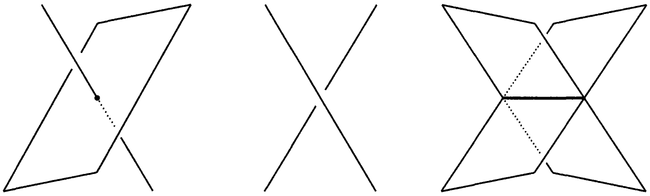
[14] Or, for manifolds with non-empty boundary, to a generator $[X, \partial X]$ of $H_m(X, \partial X;\, \mathbb{Z})$.

[15] This is a weak version of the celebrated *Sard theorem,* proved in **A. Sard**'s ***The measure of the critical values of differentiable maps*** [**Sar42**].

boundary, then we automatically assume that $Y \cap \partial X = \partial Y$ and that Y is transverse to ∂X.)

Deformations. A **homotopy** of $f\colon X \to Y$ is a path of maps starting with f, that is to say, a map $F\colon X \times [0,1] \to Y$, usually written $f_t = F(\,\cdot\,,t)$, with $f_0 = f$. The end-functions f_0 and f_1 are called **homotopic** and we write $f_0 \sim f_1$. A homotopy so that each f_t is an embedding is called an **isotopy.** An isotopy of a submanifold $Y \subset X$ is an isotopy of its inclusion map. An **ambient isotopy** of an embedding $f\colon Y \subset X$ is an isotopy f_t realized through an isotopy h_t of the identity map $id\colon X \to X$, *i.e.*, $f_t = h_t \circ f$ with $h_0 = id$ and each h_t a self-diffeomorphism of X. A typical method for obtaining an ambient isotopy is to integrate a vector field $\vartheta \in \Gamma(T_X)$ and take h_t to be the flow of ϑ. Furthermore, every isotopy of a compact submanifold can be realized by an ambient isotopy. We will freely use such happy words as *"deformation"*, *"perturbation"* and *"approximation"* to mean isotopy or ambient isotopy.

Transversality. Two linear subspaces V' and V'' of $\mathbb{R}^m$ are called **transverse** if $V' + V'' = \mathbb{R}^m$. Two submanifolds Y' and Y'' of a manifold X are called **transverse** (or **in general position**) if around any $x \in Y' \cap Y''$ they look like two transverse subspaces of $\mathbb{R}^m$. In other words, if $T_{Y'} + T_{Y''}|_{Y' \cap Y''} = T_X|_{Y' \cap Y''}$. See figure E.



E. Transverse submanifolds in $\mathbb{R}^3$

If dimensions do not add up to allow that $T_{Y'} + T_{Y''}|_{Y' \cap Y''} = T_X|_{Y' \cap Y''}$, then the transversality of Y' and Y'' must mean that $Y' \cap Y'' = \varnothing$.

If dimensions add up perfectly and transversality means $T_{Y'} \oplus T_{Y''}|_{Y' \cap Y''} = T_X|_{Y' \cap Y''}$, then Y' and Y'' meet in isolated points. Further, choosing orientations for X, Y' and Y'' and comparing them while summing will assign a sign ± to each $x \in Y' \cap Y''$, which can then be added up to yield the **intersection number** $Y' \cdot Y''$ (depending only on the homology classes $[Y']$ and $[Y'']$ inside X).

Given any two submanifolds Y' and Y'' of X, any one of them, say Y', can be isotoped to an arbitrarily-close submanifold Y'_0 that is transverse to Y''. Therefore, whenever we mention submanifolds, they should be assumed to be transverse, without our commenting further.

In particular, a curve and a surface inside a 4–manifold can be perturbed so that they do not touch at all. Two surfaces S', S'' inside a 4–manifold M can be made to meet only at isolated points x where $T_{S'} \oplus T_{S''}|_x = T_M|_x$, and an intersection number $S' \cdot S''$ can thus be defined.

Perturb into niceness. Every continuous map $f\colon X \to Z$ between smooth manifolds is homotopic to an arbitrarily-close *smooth* map $X \to Z$. Therefore, whenever we mention a map between smooth objects, it should be assumed smoothed, without our commenting further.

If $2\dim X \leq \dim Z$, then any map $f\colon X \to Z$ is homotopic to an arbitrarily-close *immersion*. Further, if $2\dim X + 1 \leq \dim Z$ and X is compact, then any map $f\colon X \to Z$ is homotopic to an arbitrarily-close *embedding*. In the borderline case $2\dim X = \dim Z$, any map $f\colon X \to Z$ can be approximated by a self-transverse immersion, *i.e.*, an immersion with isolated double-points $f(x') = f(x'')$ where the two branches of $f[X]$ meet transversely, in other words, with $df[T_X|_{x'}] \oplus df[T_X|_{x''}] = T_Z|_{f(x')}$.

In particular, any map from a closed surface into a 4–manifold can be deformed to an immersion with isolated transverse double-points, while any map of a surface into a 5–manifold can be perturbed to an embedding.

Normal bundles. The inclusion $Y \subset X$ of a submanifold Y into a manifold X induces a fiberwise-linear inclusion $T_Y \subset T_X|_Y$, and thus defines a quotient bundle $T_X|_Y / T_Y$ called the **normal bundle** of Y in X and denoted by $N_{Y/X}$. It fits in the exact sequence[16]

$$0 \longrightarrow T_Y \longrightarrow T_X|_Y \longrightarrow N_{Y/X} \longrightarrow 0\,.$$

The normal bundle $N_{Y/X}$ can be embedded around Y as a **tubular neighborhood** of Y in X, *i.e.*, as an open set U around Y together with a retraction[17] $p\colon U \to Y$ that organizes U as a vector bundle isomorphic to $N_{Y/X}$.[18] See figure F on the following page.

While there is no unique way of embedding $N_{Y/X}$ around Y, nonetheless, any two tubular neighborhoods of Y are isotopic (as embedded vector bundles[19]). Thus, for every submanifold Y in X we will automatically assume

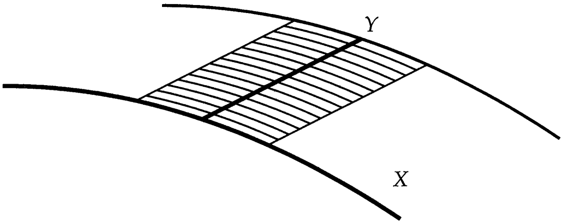
---

[16] A sequence of morphisms is called **exact** if at each node we have $\mathrm{Im}(\mathit{ending}) = \mathrm{Ker}(\mathit{beginning})$.

[17] If $A \subset B$, then a **retraction** $r\colon B \to A$ is any continuous map so that $r|_A = id$.

[18] A typical method for obtaining a tubular neighborhood is to pick a random Riemannian metric on X and use its exponential map to send a neighborhood of Y in $N_{Y/X} = (T_Y)^{\perp} \subset T_X|_Y$ to a neighborhood of Y in X.

[19] Given two embeddings $f_0\colon N_{Y/X} \subset X$ and $f_2\colon N_{Y/X} \subset X$ as tubular neighborhoods of Y in X, there is an isotopy f_t connecting f_0 to an embedding $f_1\colon N_{Y/X} \subset X$ so that $f_1^{-1} \circ f_2$ is a well-defined vector bundle isomorphism $N_{Y/X} \to N_{Y/X}$, and so that the isotopy f_t fixes the zero section $Y \subset N_{Y/X}$.



F. Tubular neighborhood of Y in X

that it is surrounded by a (essentially unique) copy of $N_{Y/X}$ as a tubular neighborhood in X.

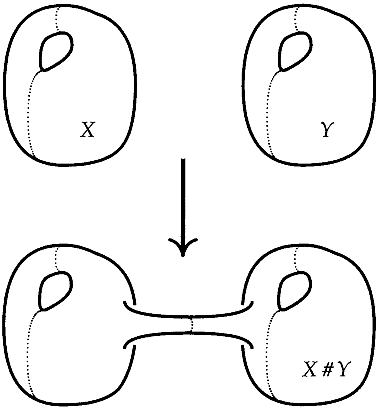
The boundary ∂X of X admits a neighborhood called a **collar**, *i.e.*, an open set U around ∂X together with a retraction $p \colon U \to \partial X$ that is isomorphic to $\partial X \times [0,1) \to \partial X$ (with $\partial X \subset X$ corresponding to $\partial X \times 0$). One should think of a collar as half a tubular neighborhood, realizing half the normal bundle of ∂X.

Cut-and-paste

Connected sums. The simplest of many cut-and-paste methods in manifold topology is the **connected sum**. Given two m–manifolds X and Y, we can build a new manifold $X\#Y$ as follows: we cut a small m–ball out of X and another out of Y; both results have an $(m-1)$–sphere as boundary; we identify these two spheres to obtain a new connected m–manifold, denoted by

$$X\#Y\,.$$

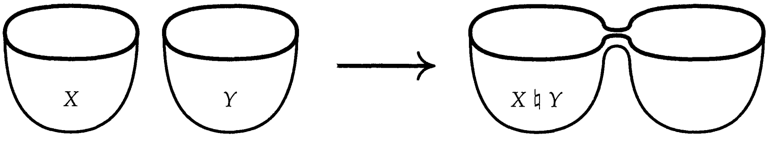
The connected sum can be pictured as connecting X and Y by a (hollow) tube, as in figure G.



G. Connected sum

A typical problem in topology is to split manifolds into connected sums of simpler manifolds. If a manifold Z^m cannot be written as $Z = X\#Y$, where neither X nor Y is a homotopy m–sphere, then Z is called **irreducible.** (A homotopy m–sphere Σ is any m–manifold that is homotopy-equivalent to a sphere; the preceding condition simply requires that parts of the homology or fundamental group of Z be taken by each of X and Y.)

> ***Boundary sums.*** *For manifolds with non-empty boundary, there is a another construction, which we mention only for the sake of completeness (it will appear in this text only marginally). Given two manifolds X and Y with non-empty boundaries, one can perform a connected sum on their boundaries. The result is denoted by*
>
> $$X \natural Y$$
>
> *and is called the* **boundary sum** *of X and Y; it is a manifold whose boundary is $\partial(X \natural Y) = (\partial X) \# (\partial Y)$. See figure* H.



H. Boundary sum

Gluing technology. Given two manifolds X and Y with homeomorphic boundaries $\partial X \simeq \partial Y$, one can identify the two boundaries and obtain a new manifold, denoted by

$$X \cup_\partial Y\,.$$

The result in general depends on the choice of homeomorphism of the boundaries.

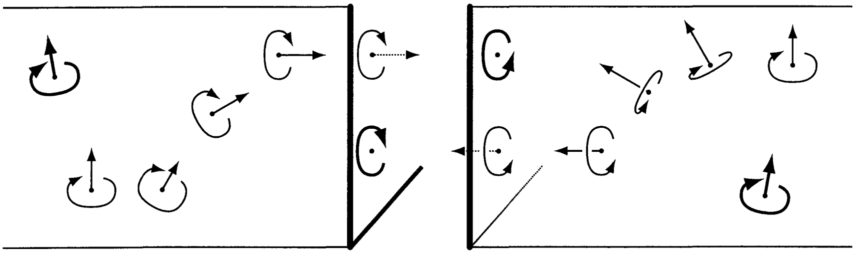
If X and Y are *oriented,* then the new manifold $X \cup_\partial Y$ could inherit an induced natural orientation, agreeing with the chosen orientations of X and Y. However, that requires that the identification of ∂X with ∂Y be made by *reversing* their inherited orientations, as suggested in figure I on the following page. In other words, we need a homeomorphism $\partial X \simeq \overline{\partial Y}$ in order to glue $X \cup_\partial Y$ as a nice oriented manifold.[20]

Further, if X and Y are *smooth,* then, to ensure that $X \cup_\partial Y$ is a smooth manifold as well, such a gluing should be made by identifying whole *collars* of the boundaries, as suggested in figure J on the next page.

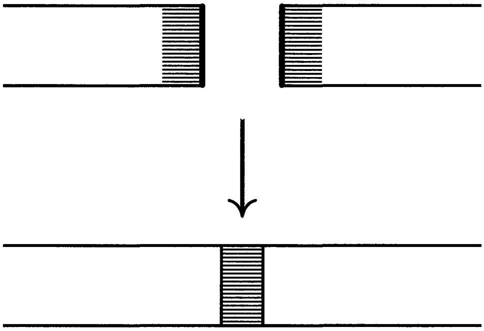
Cut-and-paste methods will be used throughout this volume, and, even without our mentioning it again, the above technology should always be understood as lying behind them.

---

[20] This of course is never a problem for the spheres that appear in a connected sum: any reflection will do. Nonetheless, there are plenty of manifolds that do not admit orientation-reversing self-homeomorphisms.

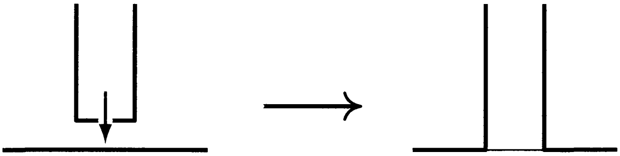


I. Orientation-reversal for gluing

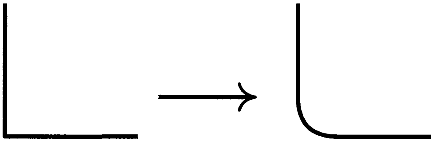


J. Gluing with collars

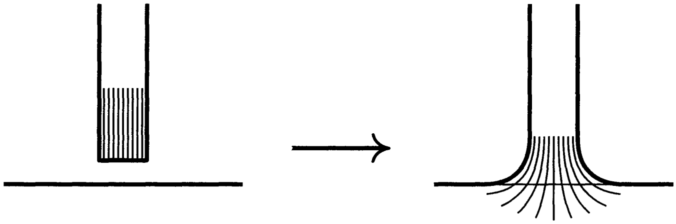
Rounding corners. Further, in many constructions that we will use (*e.g.*, attaching handles), the cut-and-paste procedure yields at first an object with *corners*, as in figure K. To obtain a smooth manifold, either these corners can be rounded (smoothed), as suggested in figure L, or the gluing can be done with more care, as in figure M on the facing page. Again, this issue will not be mentioned again.



K. Apparition of a corner



L. Rounding a corner



M. Round attachment

Homology of manifolds

The homology and cohomology of a manifold are governed by the universal coefficients theorems and by Poincaré duality. In what follows we quickly review these.

We start by recalling that we have

$$H_1(M;\mathbb{Z}) = \pi_1(M) \;/\; \big[\pi_1(M),\, \pi_1(M)\big]$$

(with H_1 written additively, but π_1 written multiplicatively). Here, $[G,G]$ is $\{aba^{-1}b^{-1} \mid a,b \in G\}$; thus $H_1(M;\mathbb{Z})$ is the Abelianization of $\pi_1(M)$.

Universal coefficients. The universal coefficient theorem for *homology* is the exact sequence

$$0 \longrightarrow H_k(M;\mathbb{Z}) \otimes_{\mathbb{Z}} G \longrightarrow H_k(M;G) \longrightarrow \operatorname{Tor}\big(H_{k-1}(M;\mathbb{Z}),\, G\big) \longrightarrow 0\,,$$

where G is any Abelian group (viewed as a $\mathbb{Z}$–module). The operator Tor is described by the properties: $\operatorname{Tor}(\mathit{Free}, G) = 0$, $\operatorname{Tor}(\mathbb{Z}_n, G) = \operatorname{Ker}(G \xrightarrow{n} G)$, $\operatorname{Tor}(A \oplus B,\, G) = \operatorname{Tor}(A,G) \oplus \operatorname{Tor}(B,G)$, and $\operatorname{Tor}(A,B) = \operatorname{Tor}(B,A)$; it essentially detects the common torsion of its arguments. While there are isomorphisms $H_k(M;G) \approx \big(H_k(M;\mathbb{Z}) \otimes_{\mathbb{Z}} G\big) \oplus \operatorname{Tor}\big(H_{k-1}(M;\mathbb{Z}),\, G\big)$, none of them is canonical.

The universal coefficient theorem for *cohomology* is the exact sequence

$$0 \longrightarrow \operatorname{Ext}\big(H_{k-1}(M;\mathbb{Z}),\, G\big) \longrightarrow H^k(M;G) \longrightarrow \operatorname{Hom}_{\mathbb{Z}}\big(H_k(M;\mathbb{Z}),\, G\big) \longrightarrow 0\,.$$

The operator Ext is the "dual" of Tor and satisfies analogous properties: $\operatorname{Ext}(\mathit{Free}, G) = 0$, $\operatorname{Ext}(\mathbb{Z}_n, G) = G/nG$, and $\operatorname{Ext}(A \oplus B,\, G) = \operatorname{Ext}(A,G) \oplus \operatorname{Ext}(B,G)$; it also detects torsion. And again, while there are isomorphisms $H^k(M;G) \approx \operatorname{Hom}_{\mathbb{Z}}\big(H_k(M;\mathbb{Z}),\, G\big) \oplus \operatorname{Ext}\big(H_{k-1}(M;\mathbb{Z}),\, G\big)$, they are not canonical.

Let us denote by T_k the torsion submodule of $H_k(M;\mathbb{Z})$, that is to say, $T_k = \{\alpha \in H_k(M;\mathbb{Z}) \mid m\alpha = 0 \text{ for some } m \in \mathbb{Z}\}$. Choose your favorite complement F_k of T_k in $H_k(M;\mathbb{Z})$, that is, a free submodule of $H_k(M;\mathbb{Z})$ so that

$$H_k(M;\mathbb{Z}) \approx F_k \oplus T_k\,.$$

Applying the universal coefficient theorem for cohomology with $G = \mathbb{Z}$ yields isomorphisms

$$H^k(M;\mathbb{Z}) \approx F_k \oplus T_{k-1} .$$

Notice how torsion is retarded by one. See also table I. These isomorphisms, of course, are not canonical.

I. Integral co/homology

k	$\cdots$	$n-1$	n	$n+1$	$\cdots$
$H^k(X;\mathbb{Z})$	$\cdots$	$F_{n-1} \oplus T_{n-2}$	$F_n \oplus T_{n-1}$	$F_{n+1} \oplus T_n$	$\cdots$
$H_k(X;\mathbb{Z})$	$\cdots$	$F_{n-1} \oplus T_{n-1}$	$F_n \oplus T_n$	$F_{n+1} \oplus T_{n+1}$	$\cdots$

Enter Poincaré duality. While the above are true on all spaces with finitely-generated homology, the case of oriented m–manifolds is further enriched by Poincaré duality. This is a canonical isomorphism

$$H_k(M;\mathbb{Z}) = H^{m-k}(M;\mathbb{Z}) .$$

Combining with the above symmetries, it yields isomorphisms

$$F_k \approx F_{m-k} \qquad \text{and} \qquad T_k \approx T_{m-k-1} .$$

In the particular case of an oriented 4–manifold, the picture that emerges is that from table II. One should notice that the only torsion that floats around is the torsion of $H_1(M;\mathbb{Z})$, which has its origins in $\pi_1(M)$; of course, it vanishes in the simply-connected case.

II. Integral co/homology of an oriented 4–manifold

k	0	1	2	3	4
$H^k(M^4;\,\mathbb{Z})$	$\mathbb{Z}$	F_1	$F_2 \oplus T_1$	$F_1 \oplus T_1$	$\mathbb{Z}$
$H_k(M^4;\,\mathbb{Z})$	$\mathbb{Z}$	$F_1 \oplus T_1$	$F_2 \oplus T_1$	F_1	$\mathbb{Z}$

Finally, in the case of modulo 2 coefficients, the picture looks somewhat different, as in table III. Here we denoted by F_k'' the modulo 2 reduction of the free part F_k, namely $F_k'' = F_k/2F_k$; and we denoted by T_1'' the 2–torsion of $H_1(M;\mathbb{Z})$, that is, $T_1'' = \{\alpha \in H_1 \mid 2\alpha = 0\}$. The 2–torsion originates in $\pi_1(M)$ and now pollutes everything.

III. Modulo 2 co/homology of an oriented 4–manifold

k	0	1	2	3	4
$H^k(M^4;\,\mathbb{Z}_2)$	$\mathbb{Z}_2$	$F_1'' \oplus T_1''$	$F_2'' \oplus T_1''$	$F_1'' \oplus T_1''$	$\mathbb{Z}_2$
$H_k(M^4;\,\mathbb{Z}_2)$	$\mathbb{Z}_2$	$F_1'' \oplus T_1''$	$F_2'' \oplus T_1''$	$F_1'' \oplus T_1''$	$\mathbb{Z}_2$

Conventions, notations, abbreviations

Unless otherwise specified, everything is orientable and endowed with a chosen preferred orientation. All maps will be continuous and, most of them, differentiable. Most manifolds will be connected and closed. Most of them, 4-dimensional. Most of them, simply-connected. Most of them, smooth.

Therefore, should we write "M" without any comments, it is safe to assume that it represents a random smooth simply-connected closed oriented 4–manifold.

More. While the value of a function $f\colon X \to Y$ at a point x will of course be denoted by $f(x)$, the image of a subset $A \subset X$ will be written $f[A]$, while the preimage of $y \in Y$, as long as f is not invertible, will be denoted by $f^{-1}[y]$. We do this to better distinguish between sets and points.[21] On the other hand, we will often use x in notations like $A \times x \subset A \times B$, even though $\{x\}$ might be more rigorous.

We denote the integers by $\mathbb{Z}$ and their modulo m residue classes $\mathbb{Z}/m\mathbb{Z}$ by $\mathbb{Z}_m$. The rational numbers are denoted by $\mathbb{Q}$, the real numbers by $\mathbb{R}$, the complex numbers by $\mathbb{C}$ (with imaginary unit $i = \sqrt{-1}$), and the quaternions by $\mathbb{H}$.

Further, $\mathbb{R}^m$ denotes the Euclidean real m–space, with its preferred basis, orientation and metric, while $\mathbb{C}^m$ denotes complex m–space (of real dimension $2m$). Further, $\mathbb{S}^m$ denotes the standard m–dimensional sphere $\{x \in \mathbb{R}^{m+1} \mid |x| = 1\}$, and $\mathbb{D}^m$ denotes the standard m–dimensional disk (or ball) $\{x \in \mathbb{R}^m \mid |x| \leq 1\}$. In low dimensions, we take $\mathbb{D}^0 = \{point\}$, $\mathbb{D}^1 = [-1, +1]$ and $\mathbb{S}^0 = \{-1, +1\}$. Further, $\mathbb{RP}^m$ denotes real projective m–space $\mathbb{S}^m/\pm 1$, while $\mathbb{CP}^m$ denotes complex projective m–space $\mathbb{S}^{2m+1}/\mathbb{S}^1$ (of real dimension $2m$). Notice that $\mathbb{RP}^1 = \mathbb{S}^1$ and $\mathbb{CP}^1 = \mathbb{S}^2$. Finally, $\mathbb{T}^m$ will denote the m–torus $\mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ (m factors).

We will denote by $\langle \cdot, \cdot \rangle$ inner-products or Riemannian metrics; we will denote by $|\alpha|$ the (pointwise) length of α; we will denote by $\|\alpha\|$ an integral norm of α, specifically the L^2–norm. The span of $a, b, c, \ldots$ over a ring $\mathbb{F}$ will be denoted by $\mathbb{F}\{a, b, c, \ldots\}$; for example, the (x, y)–plane in $\mathbb{R}^3$ will be denoted by $\mathbb{R}\{e_1, e_2\}$ (instead of, say, $\operatorname{span}_{\mathbb{R}}\{e_1, e_2\}$).

An m–dimensional manifold X will sometimes be written X^m to emphasize its dimension m.

We will use $\overline{X}$ to denote the manifold X with the opposite orientation.

---

[21] We learned these notations from **J. Kelley**'s classic ***General topology*** [**Kel55, Kel75**].

Writing $X \sim Y$ will mean "X is homotopy-equivalent[22] to Y"; writing $X \simeq Y$ will mean "X is homeomorphic to Y"; writing $X \cong Y$ will mean "X is diffeomorphic to Y". To avoid confusion, in most cases we will support such notations with hints in English.

A manifold X will be called a **fake** $\mathbb{X}$ (for whatever typical manifold $\mathbb{X}$ might be) if X is homotopy-equivalent to $\mathbb{X}$, but not homeomorphic to it. A manifold X will be called an **exotic** $\mathbb{X}$ if X is homeomorphic to $\mathbb{X}$, but not diffeomorphic to it.

As already mentioned, we use $X \# Y$ to denote the connected sum of the manifolds X and Y. A writing like $X \# n Y$ represents the connected sum of X with n copies of Y, instead of, say, $X \# \left(\#_{k=1}^{n} Y\right)$. We use a similar convention in algebra, where we write, say, $\mathbb{Z} \oplus n\,\mathbb{Z}_2$ instead of $\mathbb{Z} \oplus \left(\bigoplus_{k=1}^{n} \mathbb{Z}_2\right)$. For multiplicative operations, we write $\mathbb{Z} \otimes \mathbb{Z}_2^{\otimes n}$ instead of $\mathbb{Z} \otimes \left(\bigotimes_{k=1}^{n} \mathbb{Z}_2\right)$.

If E is a vector bundle, then $\Gamma(E)$will denote the space of its global sections. For example, $\alpha \in \Gamma\left(\Lambda^2(T_X^*)\right)$ means that α is an exterior 2–form on X. If E has some extra structure, then the sections are assumed to respect that structure. In most cases that merely means that the sections are differentiable, but if the bundle is holomorphic, then $\Gamma(E)$ is automatically assumed to contain only holomorphic sections.

Thickenings. Often, we will need to take a k–dimensional object Σ^k and "thicken" it into a $(k+n)$–dimensional creature. Such a thickening is made by multiplying Σ by a disk $\mathbb{D}^n$. To emphasize that the product $\Sigma \times \mathbb{D}^n$ is to be thought merely as a thickened Σ, we will write the thickening-factor in smaller type, as in $\Sigma^k \times \mathbb{D}^n$. Related objects will inherit this convention, as for example in $\partial(\Sigma \times \mathbb{D}^n) = \partial\Sigma \times \mathbb{D}^n \cup \Sigma \times S^{n-1}$.

Sloppy. We will be pretty careless with notations. We will denote by the same letter a submanifold, the homology class it determines, and even its Poincaré-dual cohomology class. In general, Poincaré duality will be used blindly and tacitly. When talking about complex surfaces, we will denote by the same letter a complex-line bundle and its Chern class. Also, using the orientation, we identify without comment the top (co)homology of a manifold with the integers, $H^m(X^m;\mathbb{Z}) = \mathbb{Z}$ and $H_m(X^m;\mathbb{Z}) = \mathbb{Z}$, by identifying $[X]$ with $+1$.

As a few examples, if S' and S'' are oriented surfaces inside M^4, we will comfortably write $S' \cdot S''$ to denote their intersection number, instead of, say, $\left(PD([S']) \cup PD([S''])\right) \cap [M]$. Or, for $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$ we will denote by

---

[22] Two topological spaces A and B are called **homotopy-equivalent** if there are continuous maps $f\colon A \to B$ and $g\colon B \to A$ so that both $f \circ g$ and $g \circ f$ are homotopic to the identity.

$(3,1)$ the 2–homology class of $3\,[\mathbb{CP}^1]$ from $\mathbb{CP}^2$ added to $[\overline{\mathbb{CP}}^1]$ from $\overline{\mathbb{CP}}^2$. Or, the Pontryagin class $p_1(T_M) \in H^4(M;\mathbb{Z})$ of a 4–manifold will not be distinguished from the Pontryagin number $p_1(T_M)[M] \in \mathbb{Z}$.

Finally. In what follows, many statements will be named *conjectures.* Unless their names are followed by the word "open", it should be understood that they are in fact proved theorems, but have gained notoriety while being unsolved questions.

It is also worth noticing the double use of the words *topology / topological.* On one hand, they will be used to refer to the realm of topological manifolds, as opposed to that of smooth manifolds. On the other hand, they will be used to emphasize a topological point-of-view (as in "differential topology" or "the smooth topology of M"), as opposed to, say, a differential-geometric or algebraic-geometric one. Thus, while this book is devoted to the topology of 4–manifolds, it is not focused on topological 4–manifolds.

As the reader has probably noticed, we use sans-serif bold to emphasize a **notion being defined.** In bibliographical references we use bold for **author names,** slanted bold for ***titles of books,*** and italic bold for ***titles of articles.***

Requisites

The reader needs a reasonable understanding of manifolds, preferably with a view toward differential topology, as can be gained for example from **M. Hirsch**'s ***Differential topology*** **[Hir76, Hir94]** (skim through chapter 2 and 3 at a first reading, but do read section 8.2 on gluing manifolds), or **J. Milnor**'s little gem ***Topology from the differentiable viewpoint*** **[Mil65b, Mil97]**. **V. Guillemin** and **A. Pollack**'s ***Differential topology*** **[GP74]** is another possible introduction to smooth manifolds, and so is **A. Kosinski**'s ***Differential manifolds*** **[Kos93]**, which leads into geometric topology and includes some topics that will be sketched in our first chapter.

Some algebraic topology is of course also needed: Poincaré duality, characteristic classes, *etc.* Sources are manifold. A nice one is **A. Hatcher**'s ***Algebraic topology*** **[Hat02]**.

A smattering of differential geometry and of algebraic geometry cannot hurt, but their absence will not make the book unreadable.

Further reading

For an in-depth perspective on 4–manifold topology, see **R. Gompf** and **A. Stipsicz**'s textbook ***4–Manifolds and Kirby calculus*** **[GS99]**. For topological 4–manifolds, the standard reference is **M. Freedman** and **F. Quinn**'s ***Topology of 4–manifolds*** **[FQ90]**.

For classical (pre-gauge) developments and techniques, read **R. Kirby**'s beautiful ***The topology of* 4*–manifolds*** [Kir89]. Also for classical developments, the collection ***À la recherche de la topologie perdue*** [GM86a], edited by **L. Guillou** and **A. Marin,** contains many rare gems.[23] A historical perspective on 4–dimensional topology, as it stood in the late 1970s (pre-Freedman), can be gained from **R. Mandelbaum**'s survey ***Four-dimensional topology: an introduction*** [Man80]; reading it puts the later revolutions and developments quite in perspective.

For reading on Donaldson theory, the standard reference is **S.K. Donaldson** and **P. Kronheimer**'s ***The geometry of four-manifolds*** [DK90]. For Seiberg–Witten theory, **J. Morgan**'s ***The Seiberg–Witten equations and applications to the topology of smooth four-manifolds*** [Mor96] is a good start, and **L. Nicolaescu**'s ***Notes on Seiberg–Witten theory*** [Nic00] is comprehensive while unfriendly. For a first contact with the Seiberg–Witten invariants, we recommend **S.K. Donaldson**'s survey ***The Seiberg–Witten equations and* 4*–manifold topology*** [Don96a].

In directions less central to this volume, complex algebraic geometry needs **P. Griffiths** and **J. Harris**'s ***Principles of algebraic geometry*** [GH78, GH94], just as differential geometry needs **S. Kobayashi** and **K. Nomizu**'s ***Foundations of differential geometry*** [KN69, KN96].

For focusing on complex surfaces, the expert bible is **W. Barth, C. Peters** and **A. Van de Ven**'s ***Compact complex surfaces*** [BPVdV84], or its second enlarged edition (with **K. Hulek**) [BHPVdV04].

For a good understanding of the spin$^{\mathbb{C}}$ structures that underlie Seiberg–Witten theory, it helps to understand spin structures. **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** [LM89] is the unavoidable reference and it also contains a lot of differential geometry, including the Atiyah–Singer index theorem.

For symplectic geometry, start with **D. McDuff** and **D. Salamon**'s ***Introduction to symplectic topology*** [MS95, MS98] and continue with their ***J–holomorphic curves and symplectic topology*** [MS04].

As a classic on bundles, homotopy groups, and obstructions, **N. Steenrod**'s ***The topology of fibre bundles*** [Ste51, Ste99] is a book that every topologist or geometer should know and love. A very condensed comprehensive introduction to advanced algebraic topology is **E. Spanier**'s ***Algebraic topology*** [Spa66, Spa81], but one should probably first try **J. Davis** and **P. Kirk**'s ***Lecture notes in algebraic topology*** [DK01].

---

[23] Translations and commentary of papers of V. Rokhlin's, notes of A. Casson's lecture on Casson handles, *etc.* The collection's title could not be more appropriate.

Quite a lot of Riemannian geometry is included in **A. Besse**'s ***Einstein manifolds*** **[Bes87]**.

For further bibliographical comments, please refer to the notes at the end of each chapter.

Acknowledgments

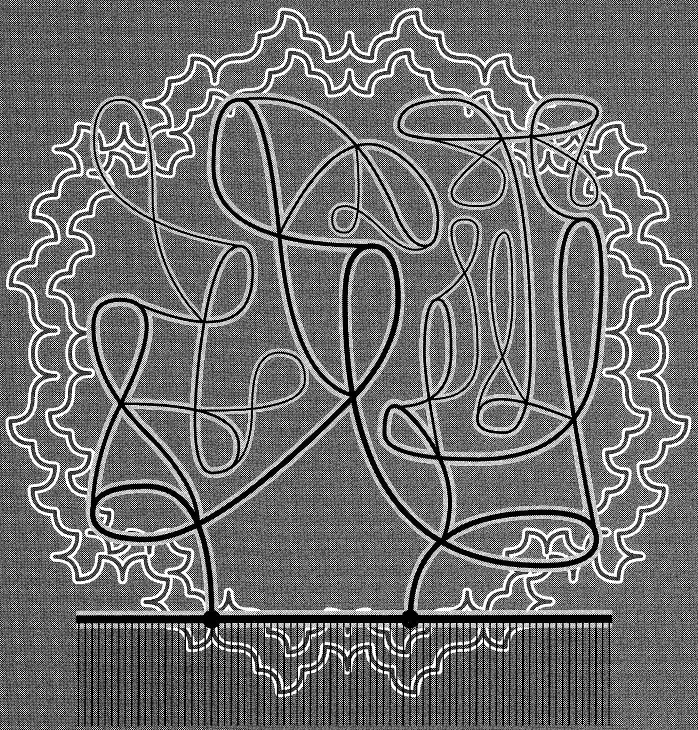
This book would not have existed without **R. Kirby**'s lectures at U.C. Berkeley or without such wonderful books as his ***The topology of 4–manifolds*** **[Kir89]** and **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** **[GS99]**. Many other works helped shape this volume, and we hope all of them are properly acknowledged in the notes at the end of each chapter.

We warmly thank **R. Gompf** for his invaluable comments on an early draft of this book, which helped us clean it of many idiocies. We are sure many have survived and many others have appeared since; they all are, of course, the author's sole responsibility. For reviewing a good part of the final manuscript, we are further indebted to **R. Kirby**.

Thanks are also due to **R. Crew** and **Y. Rudyak** for an inspection of the third part of this manuscript, and a few end-notes of chapter 4, respectively. We also wish to thank the audience of our series of talks at the University of Florida, whose questions and reactions helped the early development of this manuscript.

Finally, we also wish to thank our AMS editor **Sergei Gelfand** for a very pleasant, productive and fruitful collaboration on this project.

Production note. *The pictures were programmed directly in LaTeX 2ε; a list of figures appears toward the end of the volume. The text is typed in H. Zapf's Palatino font, combined with D. Puga's MathPazo font for mathematics, as implemented through D. Puga and W. Schmidt's* `mathpazo` *package. The general layout is built on top of AMS's* `gsm` *class file. The LaTeX code was processed using the* `pdfTeX` *engine, with Hàn Thế Thành's margin kerning (protruding characters) enabled through C. Schurig's* `pdfcprot` *package. About fifteen other packages were also quite useful. Of great help in quickly assembling the bibliography with* `bibTeX` *was the* `MathSciNet` *service of the AMS. Made on an Apple PowerMac computer.*



Part I

Background Scenery

We start our journey by taking a trip into the high-dimensional world, where we review the proof of the h–cobordism theorem. The reason for this outward look is two-fold. On one hand, it will help set 4–manifolds in their natural context among higher-dimensional manifolds, pinpointing where the split between dimension 4 and higher dimensions occurs—indeed, the h–cobordism theorem fails for *smooth* manifolds of dimension 4. On the other hand, since the theory of higher-dimensional manifolds has been so successful, it has inspired attempts to replicate its techniques in the realm of 4–manifolds; thus, inspecting the former will provide motivation for the methods that have been deployed in the study of *topological* 4–manifolds.

To the latter is devoted chapter 2 (starting on page 69), where we review Casson handles—the main technical tool for the classification of topological 4–manifolds—and culminates with M. Freedman's h–cobordism theorem for topological 4–manifolds.

This first part of the book is intended as a superficial tour of the horizon: we first look at higher-dimensional manifolds, then we focus on techniques successful on topological 4–manifolds, while throughout following h–cobordisms as a unifying thread. (This neighborhood tour ends with a short visit with dimension 3, in the end-note on page 101.) Keep in mind that our eventual goal is the realm of smooth 4–manifolds, and thus the material of this part is merely intended as a backdrop against which to later set smooth 4–manifolds.

Contents of Part I

Chapter 1

Higher Dimensions and the h–Cobordism Theorem

THIS first chapter is devoted to an outline of the proof of the h–cobordism theorem for smooth manifolds of dimensions 5 or higher. Roughly, the theorem states that if two simply-connected m–manifolds can be connected by a simply-connected $(m+1)$–manifold so that homologically nothing happens in between, then nothing can happen smoothly either, and our two m–manifolds must be diffeomorphic.

After properly stating this remarkable result that forms a bridge from the homological to the differentiable, we start in section 1.2 (page 32) to outline its proof by explaining handle decompositions of manifolds. In section 1.3 (page 40) we explain how handle decompositions can be suitably modified using so called handle moves.

The main line of argument of the h–cobordism theorem is sketched in section 1.4 (page 43), while its principal technical detail—the Whitney trick—is explained in section 1.5 (page 45), with further details in a note at the end of the chapter (page 54). The Whitney trick is the crucial point where high-dimensions and low-dimensions part ways: it hinges on embedding disks.

Section 1.6 comments on some remaining details needed to complete the proof. Finally, in the notes at the end of the chapter (page 58) is outlined the non-simply-connected version of the h–cobordism theorem, known as the s–cobordism theorem.

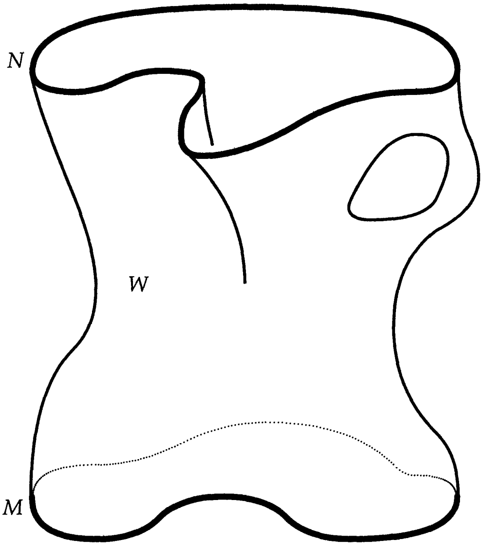
Note that further raids in the high-dimensional realm will be made in the end-notes of the next chapter (exotic spheres, page 97), and in the end-notes of chapter 4 (smoothing topological manifolds, page 207).

1.1. The statement of the theorem

Meet cobordisms. A **cobordism** between two oriented m-manifolds M and N is any oriented $(m+1)$-manifold W such that its boundary is

$$\partial W = \overline{M} \cup N ,$$

as in figure 1.1. When such a W linking M and N exists, the manifolds M and N are called **cobordant.** You should think of W as some sort of "homology-without-ambient" between M and N.



1.1. A cobordism

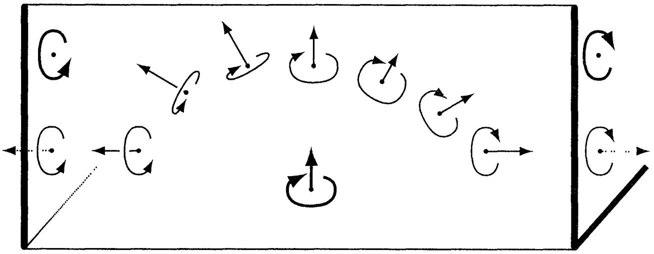
The reason M appears with reversed orientation in $\partial W = \overline{M} \cup N$ is that we think of M and N as lying at opposite ends of W. It is similar to what happens with $M \times [0,1]$, where we have

$$\partial(M \times [0,1]) = \overline{M} \times 0 \cup M \times 1 .$$

This latter equality is a consequence of how an orientation of a manifold induces an orientation on its boundary,[1] see figure 1.2 on the next page. The cobordism $M \times [0,1]$ is called the **trivial cobordism.**

---

[1] Everyone has encountered an instance of this in the fundamental theorem of calculus, where $\int_{[a,b]} df = -f(a) + f(b)$. The minus sign in $-f(a)$ and the overline in $\overline{M}$ have the same root.



1.2. Orientations on boundaries

While two manifolds being cobordant certainly points to a similarity between them[2] (and the study of cobordisms earned its creator, R. Thom, a Fields Medal), it is not a very strong similarity. For example, every 4–manifold is cobordant to a connected sum of $\mathbb{CP}^2$'s or $\overline{\mathbb{CP}}^2$'s.

Stronger cobordisms. To strengthen the cobordism relation, we will further ask that *homotopically* nothing happen between M and N. Thus, a cobordism W between M and N is called an **h–cobordism** if W is homotopically like the trivial cobordism $M \times [0, 1]$.

Specifically, we ask that the inclusion of M into W (or, equivalently, of N into W) be a homotopy-equivalence; equivalently, that W deformation retracts to M (or to N). If W, M and N are all simply-connected, then this is equivalent to merely requiring that

$$H_*(W, M; \mathbb{Z}) = 0 .$$

The theorem. Remarkably, in high dimensions a homotopically-trivial cobordism must in fact be smoothly-trivial:

h–Cobordism Theorem. *Let M^m and N^m be compact simply-connected oriented m–manifolds that are h–cobordant through the simply-connected $(m+1)$–manifold W^{m+1}. If $m \geq 5$, then there is a diffeomorphism*

$$W \cong M \times [0, 1] ,$$

which can be chosen to be the identity from $M \subset W$ to $M \times 0 \subset M \times [0, 1]$. In particular, M and N must be diffeomorphic.

The above statement is due to S. Smale in the early 1960s and was awarded a Fields Medal. The theorem's generalization to the non-simply-connected case is stated on page 65, inside the end-notes of this chapter (s–cobordisms, page 58).

---

[2] See also the end-notes of chapter 4 (cobordism groups, page 227).

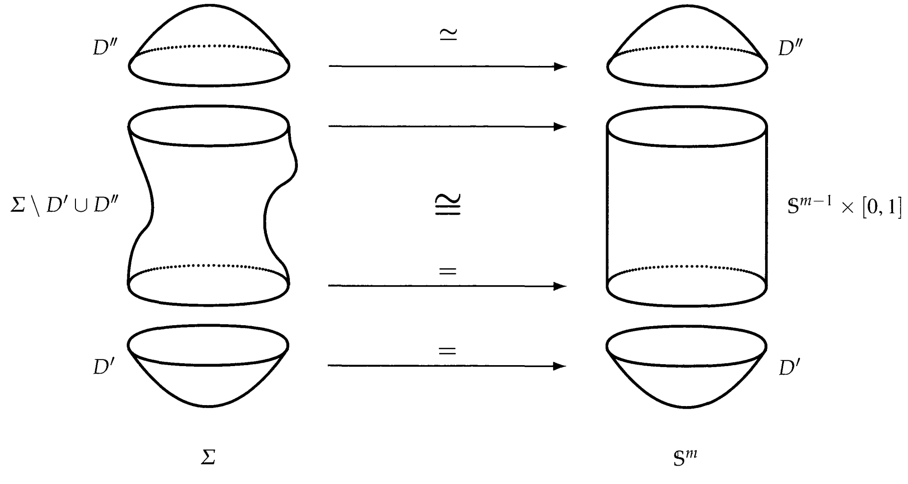
The Poincaré conjectures

An example of the power of the h–cobordism theorem is the characterization of spheres. Remember that an m–manifold Σ^m is homotopy-equivalent to $\mathbb{S}^m$ if and only if it is simply-connected and its only nontrivial homology groups are $H_0(\Sigma;\mathbb{Z}) = \mathbb{Z}$ and $H_m(\Sigma;\mathbb{Z}) = \mathbb{Z}$.

High-Dimensional Poincaré Conjecture. *If a smooth m–manifold Σ^m is homotopy-equivalent to $\mathbb{S}^m$ and $m \geq 5$, then Σ^m and $\mathbb{S}^m$ must be homeomorphic.*

Trying to obtain a *diffeomorphism* between Σ^m and $\mathbb{S}^m$ will fail in dimensions 7 or more. Nonetheless, for dimensions $m = 5$ and $m = 6$, the statement can be strengthened to offer a diffeomorphism $\Sigma^m \cong \mathbb{S}^m$. For more on the gap homeomorphisms/diffeomorphisms between Σ^m and $\mathbb{S}^m$, see the end-notes of the next chapter (exotic spheres, page 97).

Proof. For $m \geq 6$, we proceed as follows: We cut out two small m–disks D' and D'' from Σ, as in figure 1.3. The leftover $\Sigma \setminus D' \cup D''$ is an h–cobordism between two copies of $\mathbb{S}^{m-1}$. By the h–cobordism theorem, it must be a trivial cobordism: there exists a diffeomorphism $\Sigma \setminus D' \cup D'' \cong \mathbb{S}^{m-1} \times [0,1]$, which can be chosen to restrict to the identity on the lower $\mathbb{S}^{m-1}$.



1.3. Cutting up a proof of the Poincaré conjecture

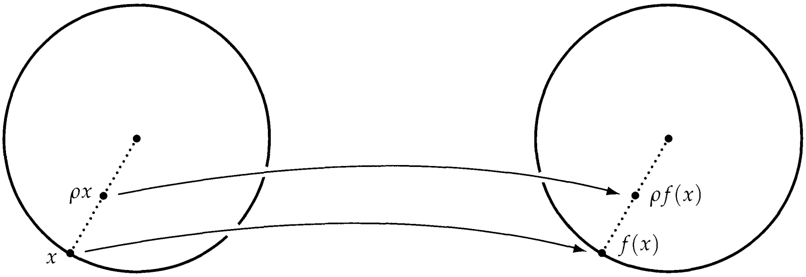
In what follows, we will rebuild Σ. As we will glue the disks D', then D'', back to $\Sigma \setminus D' \cup D''$, we will also glue copies of D', then D'', to the cylinder $\mathbb{S}^{m-1} \times [0,1]$, and then try to extend the diffeomorphism $\Sigma \setminus D' \cup D'' \cong \mathbb{S}^{m-1} \times [0,1]$ across these disks, till we obtain $\Sigma \simeq \mathbb{S}^m$.

The diffeomorphism $\Sigma \setminus D' \cup D'' \cong \mathbb{S}^{m-1} \times [0,1]$ is the identity on the bottom $\mathbb{S}^{m-1}$. Thus, after we glue D' to both sides, we can extend the diffeomorphism $\Sigma \setminus D' \cup D'' \cong \mathbb{S}^{m-1} \times [0,1]$ across the added copies of D' by the identity map and obtain a diffeomorphism $\Sigma \setminus D'' \cong \mathbb{S}^{m-1} \times [0,1] \cup D'$. The latter manifold is, of course, merely an m-ball $\mathbb{D}^m$. Hence, we can now view Σ as obtained from two m–balls, $\mathbb{D}^m$ and D'', glued along some diffeomorphism of their (upper) boundary-spheres $\mathbb{S}^{m-1}$.

Next, when we glue the disk D'' to both sides, on one hand we obtain Σ, on the other $\mathbb{S}^m$ (by gluing D'' in the standard manner to $\mathbb{D}^m$).

Our diffeomorphism $\Sigma \setminus D'' \cong \mathbb{D}^m$ induces a diffeomorphism of the (upper) boundary-spheres $\mathbb{S}^{m-1}$, which can be transported through the glue to a diffeomorphism $\partial D'' \cong \partial D''$ between the two copies of the boundary-sphere of D''. To extend to a diffeomorphism $\Sigma \cong \mathbb{S}^m$, we need to extend it across the disk D''.

Any diffeomorphism of the boundary-sphere $\mathbb{S}^{m-1}$ of an m–disk can be extended radially to the whole disk, as in figure 1.4, but only as a homeomorphism of D''. Indeed, this radial extension has a good chance to fail from being differentiable at the center, and hence homeomorphism is all we get. The theorem is proved for the case $m \geq 6$.



1.4. Extending a diffeomorphism of $\mathbb{S}^{m-1}$ to a homeomorphism of $\mathbb{D}^m$

The case $m = 5$ is a bit trickier: one first uses the fact that Σ^5 must bound a contractible 6–manifold V. If we cut out a standard 6–ball from V, the leftover is an h–cobordism from Σ^5 to $\mathbb{S}^5$, which must be trivial and establishes that Σ^5 is actually *diffeomorphic* to $\mathbb{S}^5$ (thus, there are no exotic 5–spheres). □

***Dimension* 4.** The Poincaré conjecture is also true for 4–manifolds:

Topological 4–Dimensional Poincaré Conjecture. *If a 4–manifold Σ^4 is homotopy-equivalent to $\mathbb{S}^4$, then Σ^4 and $\mathbb{S}^4$ must be homeomorphic.* □

The proof of this statement will require the whole machinery of A. Casson and M. Freedman's work, passing through a 4–dimensional topological

h–cobordism theorem from which the above is deduced. This will be discussed in the next chapter.

Encouraged by the fact that in the neighboring dimensions 5 and 6 the Poincaré conjecture can be strengthened to an actual diffeomorphism between Σ^m and S^m, we also state:

Smooth 4–Dimensional Poincaré Conjecture (*open*). *If a smooth 4–manifold Σ^4 is homotopy-equivalent to S^4, then Σ^4 and S^4 must be diffeomorphic.*

This conjecture is wide open: we do not know whether there are any exotic 4–spheres, and there are no methods in sight for either proving or disproving it. It is possible that the conjecture is true, just as it is possible that there are infinitely-many distinct smooth structures on S^4.

***Dimension* 3.** Finally, the statement that in 1904 started it all:

Poincaré Conjecture (*open?*). *If a 3–manifold Σ^3 is homotopy-equivalent to S^3, then Σ^3 and S^3 must be diffeomorphic.*

In this case, the homotopy-equivalence hypothesis reduces to merely requiring that Σ^3 be simply-connected; and homeomorphisms in dimension 3 are equivalent to diffeomorphisms.

It is possible that the conjecture has been proved[3] in 2003, a hundred years after it was stated. In other words, it is possible that dimension 4 is the only dimension left where spheres are not yet understood.

The rest of this chapter is devoted to exploring the proof of the h–cobordism theorem in dimensions 5 or more.

1.2. Handle decompositions

The strategy for proving the h–cobordism theorem is the following: we will translate the algebraic triviality $H_*(W, M; \mathbb{Z}) = 0$ into the geometric triviality of W relative to M.

For that, we must first express the homology of W by using geometric elements, which will be done by using a *handle decomposition* of W. This is the equivalent, in the realm of manifolds, of a cellular decomposition, but with the cells "thickened" so that every skeleton will still be a manifold.

A natural method for making handle decompositions appear is through the use of Morse functions:

---

[3] By G. Perelman [**Per02, Per03b, Per03a**], who might have proved Thurston's geometrization conjecture by using the (Riemannian-geometric) Ricci-flow method pioneered by R. Hamilton. The proof is still under scrutiny at the time of this writing (January 2005).

Morse functions

A **Morse function** on W is a differentiable function

$$f\colon W \longrightarrow [0,1] ,$$

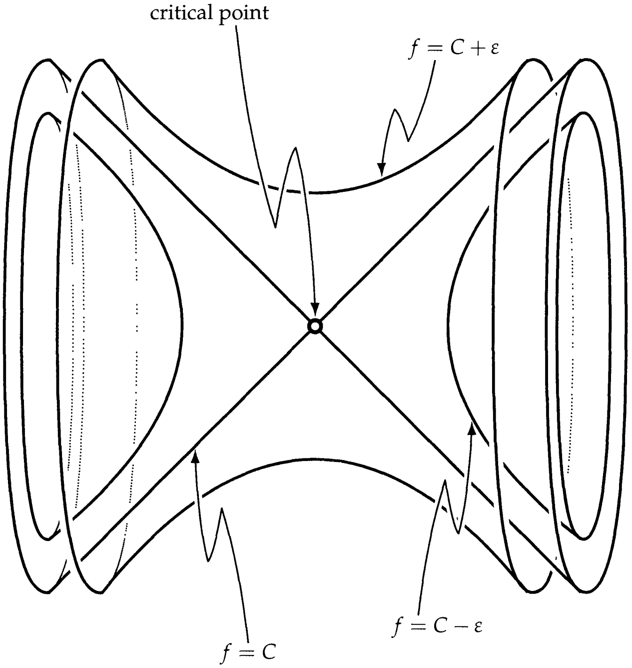
with $M = f^{-1}[0]$ and $N = f^{-1}[1]$, so that its differential $df \in \Gamma(T^*_M)$ is transverse to the zero-section of T^*_M.

Specifically, this means that around any critical point p (where $df|_p = 0$) the function f can be written locally as

$$f(x_1, \dots, x_{m+1}) = -x_1^2 - \dots - x_k^2 + x_{k+1}^2 + \dots + x_{m+1}^2 + \mathit{constant} ,$$

for a suitable k and some suitable local coordinates $(x_1, \dots, x_{m+1})$ centered at p. The critical point is then called a **critical point of index k.**

The levels of f near a critical point look as suggested in figure 1.5. As we will see, the critical points of a Morse function exhibit a lot of information about its domain W.



1.5. Levels of a Morse function around a critical point

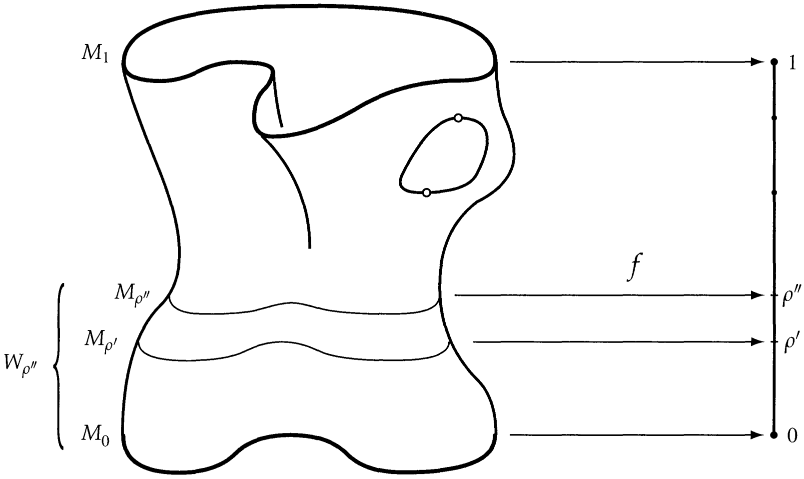
No critical, no gain. Away from the critical points, nothing much happens in W. More precisely, the *ascending cobordism*

$$W_\rho = f^{-1}[0, \rho]$$

is topologically unchanged as ρ grows without encountering critical values of f. Indeed, if there are no critical values between ρ' and ρ'', then the gradient vector field of f integrates to yield diffeomorphisms $W_{\rho'} \cong W_{\rho''}$. This restricts to diffeomorphisms between the various upper boundaries

$$M_\rho = f^{-1}[\rho] ,$$

so that $M_{\rho'} \cong M_{\rho''}$. See figure 1.6.



1.6. Ascending cobordisms

Handles

On the other hand, when running across a critical value of f, the topology of the ascending cobordism changes.

The result of passing a critical point of index k is the same as gluing a thickened k–disk to the ascending cobordism. Indeed, $W_{C+\varepsilon}$ is diffeomorphic to the result of attaching to $W_{C-\varepsilon}$ a copy of[4]

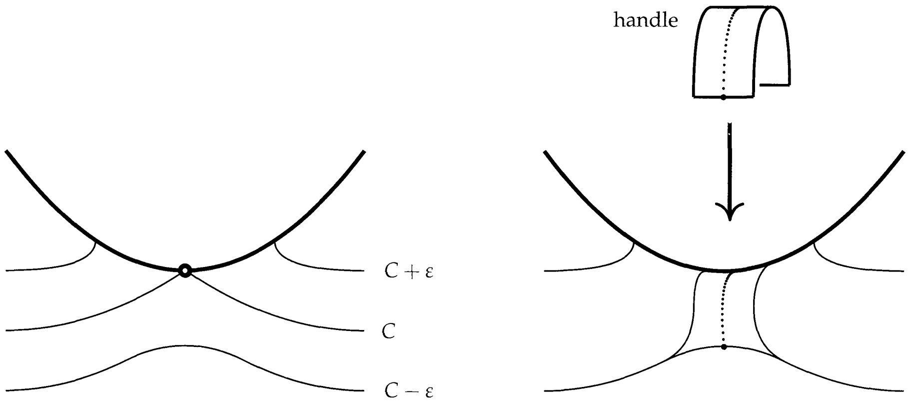
$$\mathbb{D}^k \times \mathbb{D}^{m+1-k} ,$$

along its thickened boundary-sphere $\mathbb{S}^{k-1} \times \mathbb{D}^{m+1-k}$. See figures 1.7 to 1.9 on the facing page. This is known as attaching a **k–handle** (or a *handle of order*[5] k) to $W_{C-\varepsilon}$.

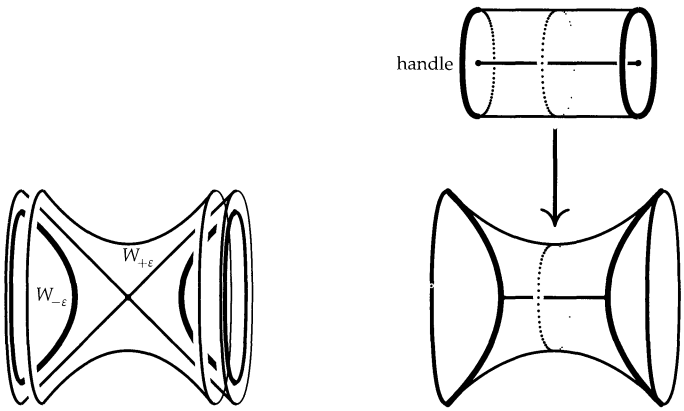
---

[4] Recall (from page 18) that, when we "thicken" something (by crossing with a disk, say), we type the "thickening" factor with smaller characters.

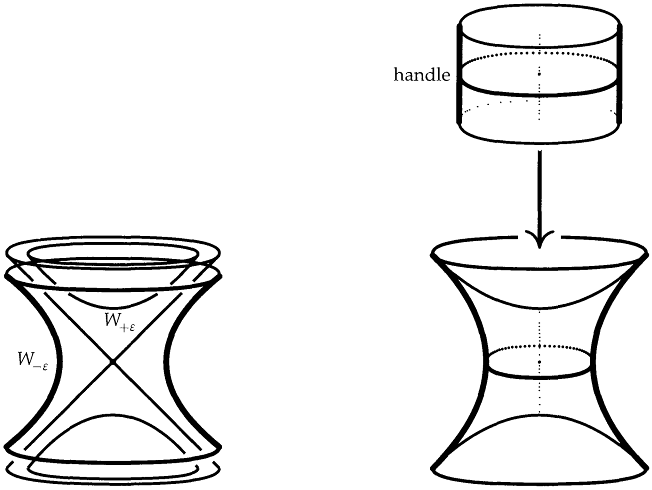
[5] We will call k the **order** of the handle $\mathbb{D}^k \times \mathbb{D}^{m+1-k}$, and $m+1$ the **dimension** of the handle. Hence, a handle of order n and dimension N is a thickening to dimension N of a n–dimensional cell.



1.7. Passing a critical point *versus* attaching a handle, I



1.8. Passing a critical point *versus* attaching a handle, II

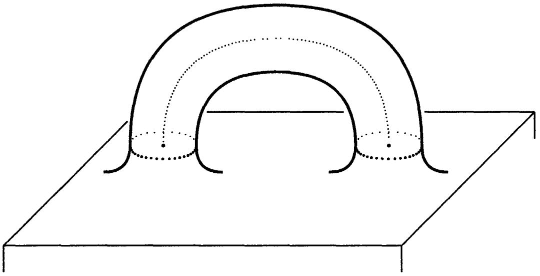


1.9. Passing a critical point *versus* attaching a handle, III

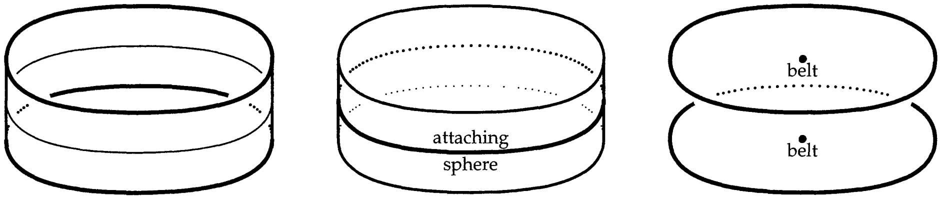
Effect of attaching a handle. Looking closer at a handle, since

$$\partial\left(\mathbb{D}^k \times \mathbb{D}^{m+1-k}\right) = \mathbb{S}^{k-1} \times \mathbb{D}^{m+1-k} \cup \mathbb{D}^k \times \mathbb{S}^{m-k},$$

we see that attaching a k–handle to $W_{C-\varepsilon}$ will delete from $M_{C-\varepsilon}$ a copy of $\mathbb{S}^{k-1} \times \mathbb{D}^{m+1-k}$ while leaving behind its border $\mathbb{S}^{k-1} \times \mathbb{S}^{m-k}$. To this border is glued as replacement a copy of $\mathbb{D}^k \times \mathbb{S}^{m-k}$, thus filling the hole and creating $M_{C+\varepsilon}$. See figures[6] 1.10 and 1.11.



1.10. Attaching a k–handle ($k = 1$)



part covered when attaching

k–handle

part left after attaching

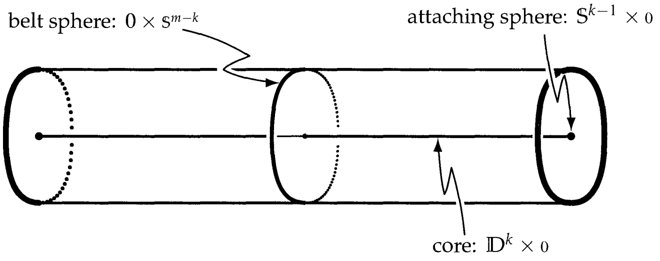
1.11. Effect of attaching a k–handle ($k = 2$)

These small steps are all that is needed to transform $M = M_0$ into $N = M_1$ by climbing on the levels of f in W. The Morse function exhibits the actual recipe for changing one into the other.

Anatomy of a handle. The disk $\mathbb{D}^k \times 0$ inside a k–handle $\mathbb{D}^k \times \mathbb{D}^{m+1-k}$ is called the **core** of the handle. The boundary $\mathbb{S}^{k-1} \times 0$ of this core is called the **attaching sphere** of the handle. Finally, the sphere $0 \times \mathbb{S}^{m-k}$ (wrapped around the "thickening" of the handle) is called the **belt sphere** (or **co-sphere**) of the handle. See figure 1.12 on the facing page, or back at figure 1.11 on this page.

---

[6] Remember that $\mathbb{S}^0 = \{-1, +1\}$

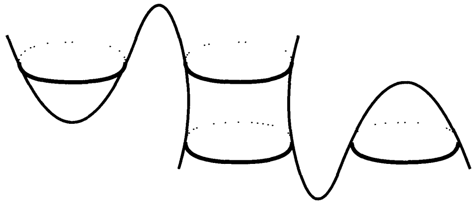


1.12. Anatomy of a k–handle $\mathbb{D}^k \times \mathbb{D}^{m+1-k}$ $(k = 1)$

Extreme handles. The case of 0– and $(m+1)$–handles is somewhat special. They correspond to local minima and maxima of f.

A 0–handle is a copy of $\mathbb{D}^0 \times \mathbb{D}^{m+1}$, where $\mathbb{D}^0 = \{point\}$; hence a 0–handle is an $(m+1)$–ball $\mathbb{D}^{m+1}$. It is "attached" through $\mathbb{S}^{-1} \times \mathbb{D}^{m+1}$, where $\mathbb{S}^{-1}$ should be understood as the empty set. In other words, attaching a 0–handle simply means setting an $(m+1)$–ball alongside the rest of our creature; this creates an m–sphere in the resulting upper boundary. See the left of figure 1.13.

An $(m+1)$–handle is a copy of $\mathbb{D}^{m+1} \times \mathbb{D}^0$ attached along $\mathbb{S}^m \times 0$. In other words, an $(m+1)$–handle is an $(m+1)$–ball as well, but attaching it means filling a spherical hole in the previous upper boundary. See the right of figure 1.13.



1.13. Minima and 0–handles, maxima and $(m+1)$–handles

As we will see later,[7] all 0– and $(m+1)$–handles can in fact be eliminated, so we will not worry about them till then.

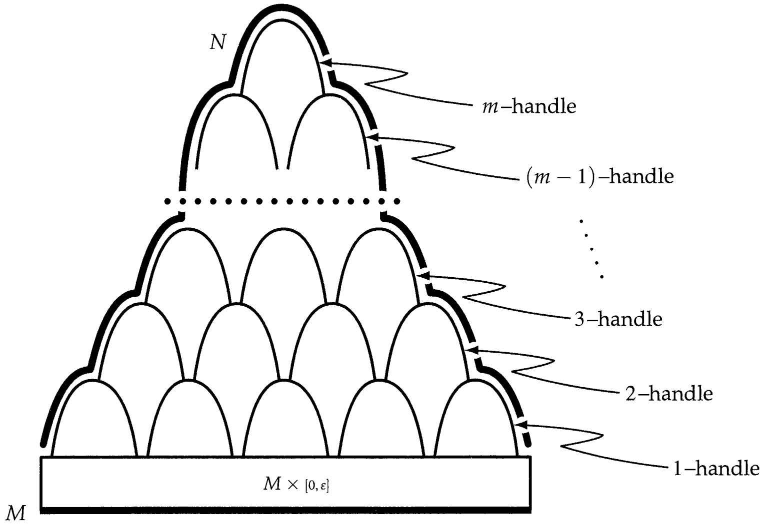
Handle decompositions. On every smooth W there exist Morse functions.[8] This implies that every cobordism W can be exhibited as a series of handle attachments to $M \times [0, \varepsilon]$, as pictured in figure 1.14 on the following page. This is called a **handle decomposition**[9] (or **handlebody structure**) of W.

---

[7] See ahead section 1.6 (page 47).

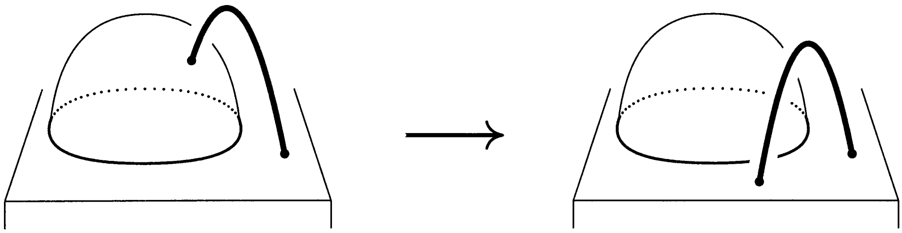
[8] Morse functions form an open and dense subset of the space of all functions on W, see for example **M. Hirsch**'s ***Differential topology*** [**Hir76, Hir94**, ch 6].

[9] In fact, it is called a handle decomposition of W **relative** to M, since one starts with M as given. Later we will also encounter handle decompositions of closed manifolds, where one starts with nothing, picks a 0–handle (a ball) and starts gluing other handles to it.



1.14. Cobordism as a tower of handles

Re-ordering handles. A dimension-counting and transversality argument shows that all k–handles can be slid off higher-order handles, as suggested in figure 1.15, so that in effect one attaches handles in stages of their increasing order.

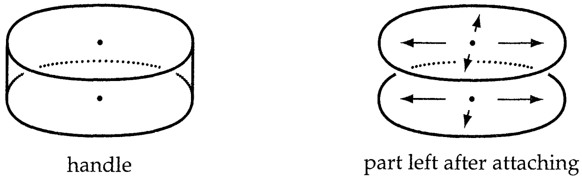


1.15. Sliding a 1–handle off a 2–handle

Indeed, let $h^i = \mathbb{D}^i \times \mathbb{D}^{m+1-i}$ be any i–handle attached after the j–handle $h^j = \mathbb{D}^j \times \mathbb{D}^{m+1-j}$. If the attaching sphere of h^i can be arranged to miss the belt sphere of h^j, then h^i can be slid off h^j, meaning that it can be viewed as attached to a level before h^j was attached.

After being attached, h^j leaves in the upper boundary a copy of $\mathbb{D}^j \times \mathbb{S}^{m-j}$. If the attaching sphere of h^i avoids the belt sphere $0 \times \mathbb{S}^{m-j}$ of h^j, then whatever part of it might sit in the remaining part $(\mathbb{D}^j \setminus 0) \times \mathbb{S}^{m-j}$ of h^j can be pushed off radially to the border $\mathbb{S}^j \times \mathbb{S}^{m-j}$, as suggested in figure 1.16 on the facing page. This border is part of what the upper boundaries M_p before and after the attachment of h^j have in common. Therefore, since sliding the attaching sphere of h^i implies that we can slide the whole attaching region of h^i (which *is just a thickening of the former), this means that h^i can be entirely pushed off h^j and viewed as attached before h^j.*

Being sure that the attaching $(i-1)$-sphere of h^i can be perturbed away from the belt $(m-j)$-sphere of h^j is a matter of transversality. These spheres both live in some m-manifold M_ρ, and thus their generic intersection will have dimension $(i-1)+(m-j)-m = i-j-1$. Therefore, if $i \leq j$, then this dimension is negative, meaning that generically the two spheres do not meet. Hence, lower-order handles can be slid off higher-order handles.



1.16. Miss the belt sphere, miss it all

From now on, we will tacitly assume that the handles of any handle decomposition at which we might be looking have already been re-ordered so as to appear attached in waves of increasing orders.

Homology from handles

Since a k-handle is merely a thickened k-cell, it should be no surprise that the homology $H_*(W, M; \mathbb{Z})$ can be retrieved directly from the handle decomposition of W.

Namely, we translate the handle decomposition into the following algebraic data: a chain complex with groups

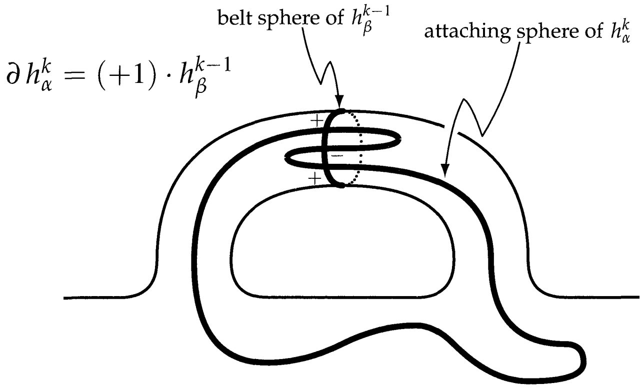
$$C_k = \mathbb{Z}\{k\text{-handles } h^k_\alpha\}$$

and boundary maps $\partial_k \colon C_k \to C_{k-1}$, given by

$$\partial_k(h^k_\alpha) = \sum \langle h^k_\alpha \mid h^{k-1}_\beta \rangle \cdot h^{k-1}_\beta ,$$

where $\langle h^k_\alpha \mid h^{k-1}_\beta \rangle$ is the **incidence number** of h^k_α with h^{k-1}_β. This coefficient is defined as the *intersection number* of the attaching sphere of h^k_α with the belt sphere of h^{k-1}_β.

Observe that the attaching sphere of h^k_α is a $(k-1)$-sphere, while the belt sphere of h^{k-1}_β is an $(m-k+1)$-sphere; both are living in the m-dimensional upper boundary M_ρ of the ascending cobordism. If assumed transverse, their intersection is in isolated points; these points can then be counted with signs to yield the coefficients $\langle h^k_\alpha \mid h^{k-1}_\beta \rangle$. See figure 1.17 on the next page.



1.17. Defining the boundary operator

It should not be hard to believe that the resulting homology groups

$$H_k(\mathcal{C}_*) = \operatorname{Ker} \partial_k \,/\, \operatorname{Im} \partial_{k+1}$$

of the complex $\{\mathcal{C}_k,\ \partial_k\}$ are naturally identical to $H_k(W, M;\ \mathbb{Z})$.

1.3. Handle moves

The handle decomposition of W that we obtain at the outset from some random Morse function is probably not the best one. To obtain more suitable ones, we will want to modify this decomposition.

There are two fundamental modifications of a handle decomposition: *handle cancellation / creation* and *handle sliding*.[10]

Handle cancellation, handle creation

If the hole created by adding a $(k-1)$–handle h_β^{k-1} is filled by the later addition of some k–handle h_α^k, then this pair of handles can be eliminated, as suggested in figures 1.18 and 1.19 on the facing page. This is called a **handle cancellation.**

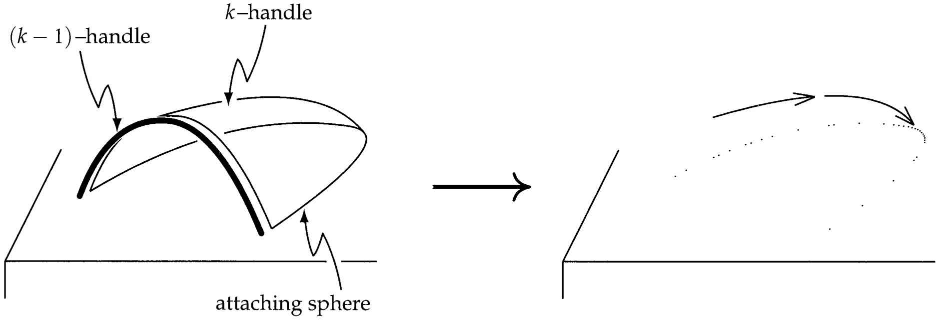
A necessary condition for canceling a pair of handles h_α^k and h_β^{k-1} is that

$$\partial h_\alpha^k = \pm h_\beta^{k-1} .$$

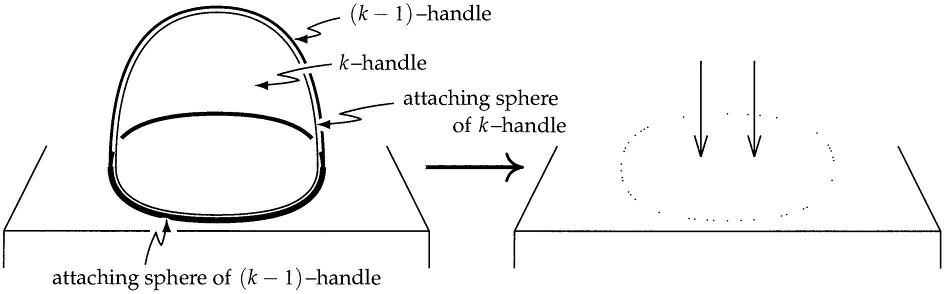
This means that the attaching sphere of h_α^k has ±1 *algebraic* intersection with the belt sphere of h_β^{k-1}, and zero *algebraic* intersection with the belt spheres of all other $(k-1)$–handles. In other words, the k–handle h_α^k passes algebraically-once over h_β^{k-1}, and only over h_β^{k-1}.

---

[10] There is also a third one, *handle trading*, which will be explained in section 1.6 (page 47) ahead.

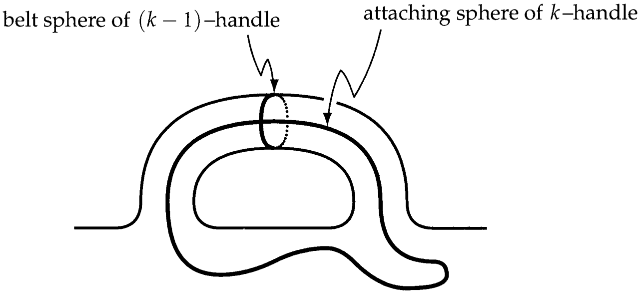


1.18. Canceling a pair of handles $(k = 2)$



1.19. Canceling a pair of handles $(k = 3)$

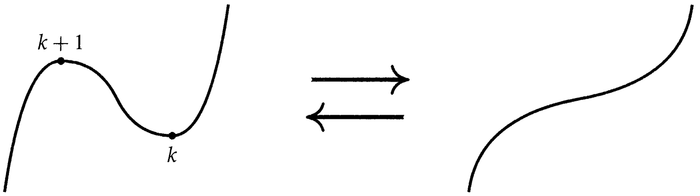
However, to actually cancel, one needs more: the algebraic intersection needs to be realized *geometrically*—the attaching sphere of h^k_α must cross exactly *once* the belt sphere of h^{k-1}_β, as in figure 1.20.



1.20. Attaching-sphere/belt-sphere position for canceling handles

The process of canceling two handles can be reversed and it is then called **handle creation** (or **handle birth**). Specifically, a pair of canceling handles can be created out of thin air, by "blistering"—just go in reverse in figures 1.18 and 1.19.

The cancellation/creation of handles can also be seen at the level of Morse functions,[11] *as in figure 1.21.*

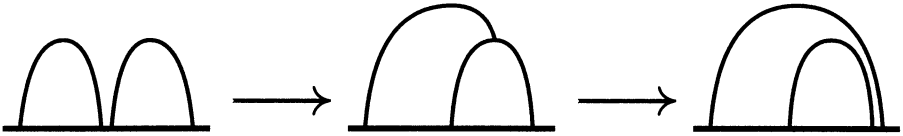


1.21. Canceling/creating a pair of critical points of a Morse function

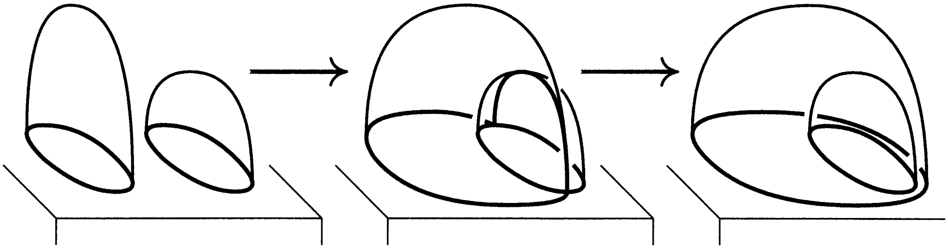
Handle sliding

The second way of modifying a handle decomposition is to change the way handles are attached.

Namely, we can slide a k–handle h^k_α over another k–handle h^k_β, as suggested in figure 1.22. What happens is that the attaching sphere of the sliding handle h^k_α travels across the core disk[12] of the other handle, as in figure 1.23.



1.22. Sliding handles, I



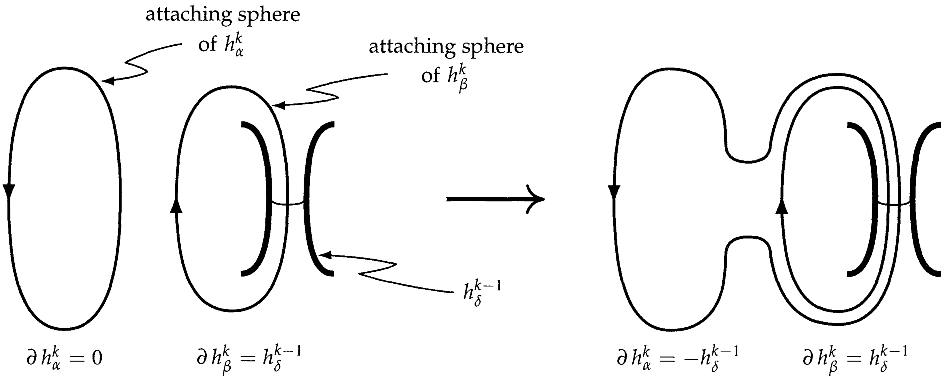
1.23. Sliding handles, II

---

[11] It is known that every two Morse functions can be linked by a homotopy that only passes through Morse functions, except for moments of creation/cancellation of critical points. See **J. Cerf's** ***La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*** [**Cer70**].

[12] Actually, the attaching sphere of h^k_α travels across a parallel copy $\mathbb{D}^k \times p$ (for some $p \in S^{m-k}$) of the core of h^k_β, which lives on the upper boundary $\mathbb{D}^k \times S^k$ of the "ridden" handle h^k_β.

The algebraic effect of sliding is that it changes the boundary operator $\partial_k\colon C_k \to C_{k-1}$. Specifically, sliding h^k_α over h^k_β modifies ∂_k the same way as would changing the basis of C_k by replacing h^k_α by $h^k_\alpha + h^k_\beta$ or $h^k_\alpha - h^k_\beta$ (depending on how one slides and orientations). See figure 1.24.



1.24. Sliding handles changes the boundary operator

1.4. Outline of proof

Remember the statement of the h–cobordism theorem: *Let M^m and N^m be compact simply-connected oriented manifolds, and let W^{m+1} be a simply-connected cobordism between them; assume that $H_*(W, M; \mathbb{Z}) = 0$ and m is at least 5; then W is diffeomorphic to $M \times [0,1]$.*

Take a handle decomposition of W. Because $H_*(W, M; \mathbb{Z}) = 0$, an algebraic reasoning shows that, by sliding handles and adding pairs of canceling handles, one can change the boundary operators till they all look like

$$\partial_k = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{or} \quad \partial_k = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 \end{bmatrix}.$$

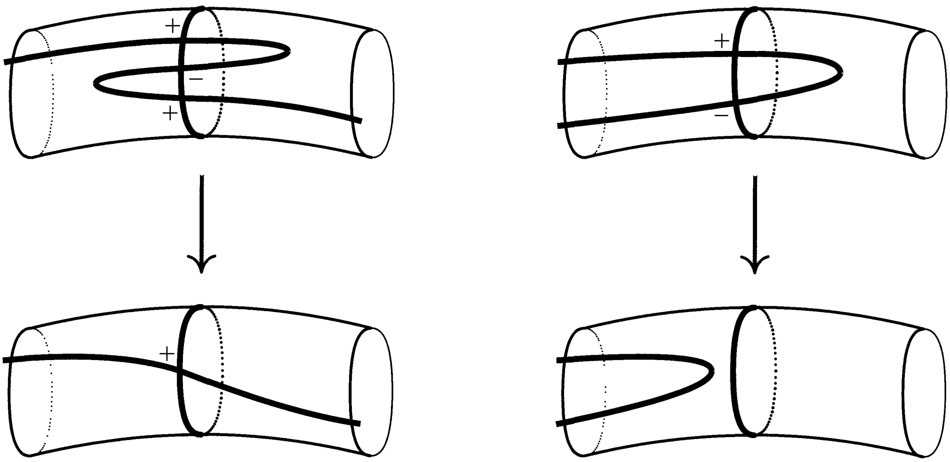
This follows since: (1) all the C_k's are free; (2) handle slides correspond to elementary row/column operations on ∂_k's matrix; and (3) creation of a pair of handles enlarges the matrix of ∂_k by adjoining an extra row and an extra column with just a 1 in the new corner.[13] Finally, if we end up with -1's in some ∂_k's matrix, then we can switch the orientation of a corresponding handle, which changes the sign to $+1$.

---

[13] An algebraic extension of this argument is presented in the end-notes of this chapter (page 58).

After performing the above modifications, and since $\partial_k \partial_{k+1} = 0$, we deduce that all the handles are now paired by 1's: that means that for every k-handle h^k_α either there exists a unique $(k-1)$-handle h^{k-1}_β such that $\partial h^k_\alpha = h^{k-1}_\beta$, or there exists a unique $(k+1)$-handle h^{k+1}_δ such that $\partial h^{k+1}_\delta = h^k_\alpha$. The handles are "$\partial$-paired".

Next, for every pair of handles with $\partial h^k_\alpha = h^{k-1}_\beta$, we want to arrange that the algebraic intersection $+1$ of the corresponding attaching and belt spheres be realized as a geometric intersection so that we can then cancel the two handles. We do that by eliminating pairs of intersection points with opposite signs, as suggested in figure 1.25.



1.25. Eliminating algebraically-canceling intersection points

Be optimistic and assume for a moment that we can somehow do that. It follows that, whenever $\partial h^k_\alpha = h^{k-1}_\beta$, we can actually cancel the pair of handles h^k_α and h^k_β. Since all handles are paired this way, it follows that all handles can be made to disappear. We are left with a handle-less decomposition of W. This means that W must be the trivial cobordism $M \times [0,1]$ and, in particular, that M and N must be diffeomorphic. The proof concludes. □

The 0–, 1–, m–, and $(m+1)$–handles must be dealt with separately, and we will do that a bit later. Besides that, there remains, of course, the crucial point of how one actually cancels those extraneous intersection points:

1.5. The Whitney trick

Consider two submanifolds P^k and Q^{m-k} of complementary dimensions inside some M^m. Then P and Q meet in isolated points, and each intersection point has a sign from comparing orientations. We will strive to geometrically eliminate algebraically-canceling pairs of intersection points.

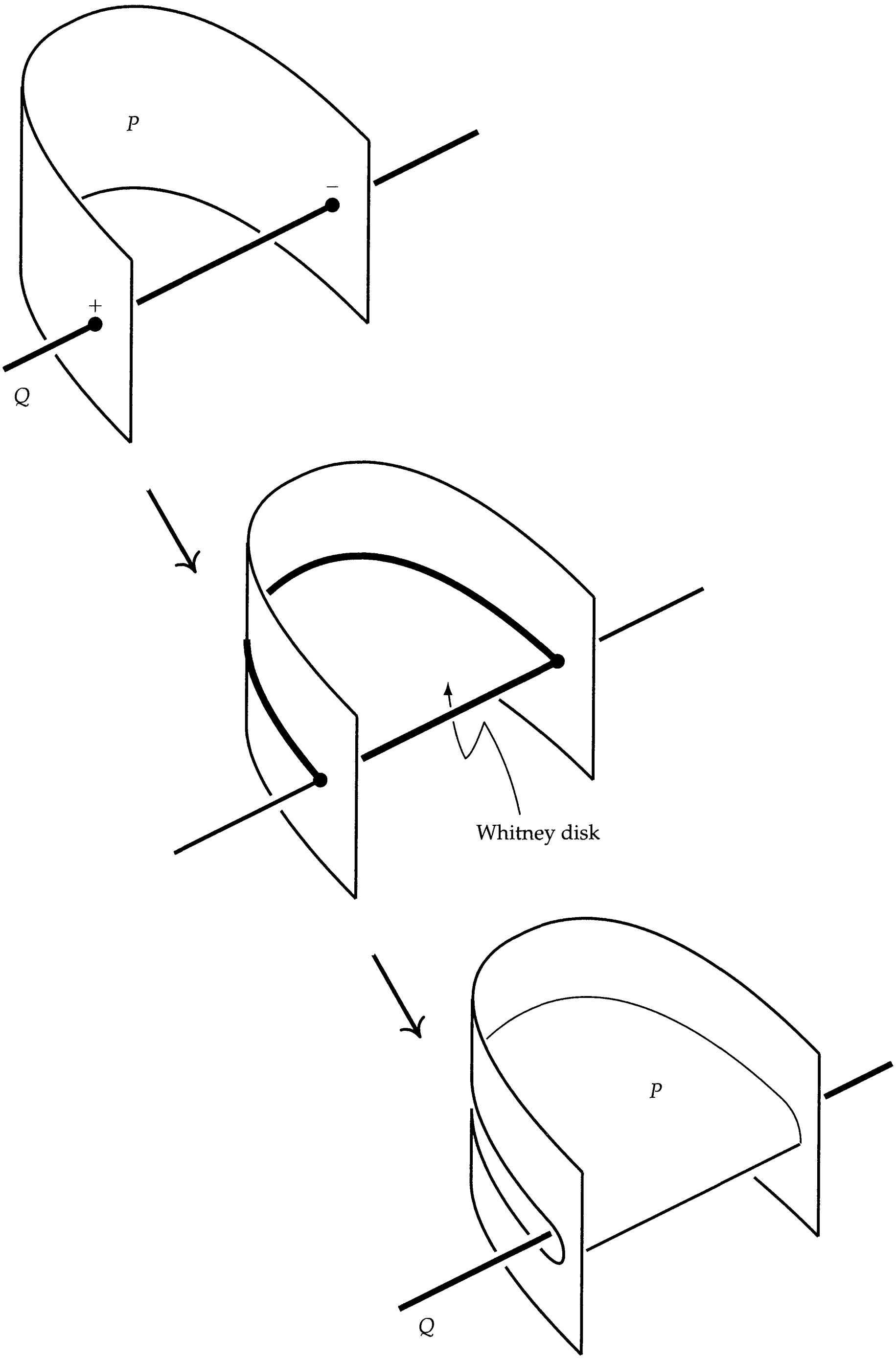
The case that concerns us is, of course, when P and Q are the attaching/belt spheres of some handles, and when M is simply-connected. After some maneuvers that will be presented in the next section, we can safely assume that the complement $M \setminus P \cup Q$ of such P's and Q's is still simply-connected. Indeed, if, on one hand, both P and Q have codimension at least 3, then their complement $M \setminus P \cup Q$ is automatically simply-connected. This can still be shown to happen for the codimension–2 cases that appear between 2– and 3– and between $(m-2)$– and $(m-1)$–handles. Finally, all 0–, 1–, m– and $(m+1)$–handles can be eliminated.

Accepting this for the moment, pick any pair of intersection points of P and Q with *opposite signs,* as suggested in figure 1.26 on the following page. Choose a path that links the two intersection points inside P and pick another path linking the two points inside Q: together, these two paths draw a circle. Owing to simple-connectedness, this circle must be homotopically-trivial in the complement $M \setminus P \cup Q$. Therefore it bounds some immersed disk in the complement of P and Q.

However, it is known that embeddings are always dense in the space of all maps $A^n \to B^{2n+1}$. In particular, this implies that immersions of disks in manifolds of dimension *at least* 5 can always be approximated by embeddings. In our case, this results in an embedded disk bounded by our circle. Such a disk, with boundary contained in $P \cup Q$ and touching the two intersection points of P and Q, is called a **Whitney disk**; its boundary is called a **Whitney circle.**

By using this Whitney disk as a guide, we can now push P past Q, till the intersection points disappear, as pictured in figure 1.26 on the next page. We have, in effect, eliminated a algebraically-canceling pair of intersection points of P and Q.

> ***Signs matter.*** *The above overview might seem to suggest that starting with intersection points of opposite signs is not really necessary. However, it is essential that, when we push P past Q along the disk, we do not inadvertently create more intersections; in other words, it is essential that P and Q be kept in separate realms during the push. For that to work, one in fact needs the signs of the two intersection points to be opposed. This fundamental technical point is explained in detail in the end-notes of this chapter (page 54).*



1.26. The Whitney trick

By applying this procedure to the attaching and belt spheres of our cobordism W, we eliminate all the algebraically-canceling intersections and thus finish the proof of the h–cobordism theorem when $m \geq 5$.

> ***Dimension* 4.** *This maneuver is the point where one sees why the proof of the h–cobordism theorem fails for dimension $m = 4$: the essential Whitney trick cannot be performed in dimension 4. The failure is owing to the impossibility of making an immersed disk into a smoothly embedded one: embeddings are* not *dense among the maps $\mathbb{D}^2 \to M^4_\rho$. Thus, in 4–dimensional topology any progress along the lines of the h–cobordism strategy must hinge on the problem of embedding disks in 4–manifolds. That will be the topic of the next chapter.*

1.6. Low and high handles; handle trading

We will now explain a few details that were glossed over in the previous outline of the h–cobordism theorem. Specifically, we will explain how, before anything else, one must deal with 0–, 1–, m–, and $(m+1)$–handles.

First, since $H_0(W, M; \mathbb{Z}) = 0$, all 0–handles can be paired with 1–handles and cancelled. Further, since W^{m+1} is simply-connected, every 1–handle bounds an embedded disk (again, we use that $m \geq 5$), which can be thickened into a canceling pair of a 2– and a 3–handle, so that the 2–handle kills the 1–handle, but leaves the 3–handle behind. In effect, we are trading all 1–handles for 3–handles, and thus a good name for this maneuver is **handle trading.** Afterwards, the whole handle decomposition can be "turned upside-down", thus making the m– and $(m+1)$–handles look like 1– and 0–handles, and we can then apply the preceding methods to eliminate them as well.

Finally, we will comment on a technical detail for applying the Whitney trick when either P or Q has codimension 2, specifically the seeming failure of simple-connectedness that appears between 2– and 3–handles and between $(m-2)$– and $(m-1)$–handles.

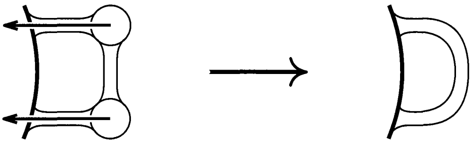
The reader should feel free to skip the remainder of this section, to the start of the end-notes on page 54, or to the start of the next chapter on page 69.

Canceling* 0*–handles. The manifold W^{m+1} is connected and is obtained from $M \times [0, \varepsilon]$ by adding various handles. A 0–handle is simply a thickened point, *i.e.*, an $(m+1)$–ball. Adding a 0–handle simply means setting such a ball alongside $M \times [0, \varepsilon]$. As such, it is disconnected from the rest and it is the job of higher-order handles to achieve connectedness for W.

However, the only handles that can link two distinct connected components are the 1–handles. Therefore, each 0–handle must be linked by a

1–handle to $M \times [0,\varepsilon]$ (or to another 0–handle). Then, as suggested in figures 1.27 and 1.28, we can cancel pairs of 0– and 1–handles till there are no 0–handles left.

1.27. Canceling 0–handles, I



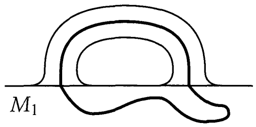
1.28. Canceling 0–handles, II

Trading* 1*–handles for* 3*–handles. Just as 0–handles have the potential to ruin the connectedness of W, 1–handles could ruin the simple-connectedness of W. Since W^{m+1} is assumed simply-connected, we will be able to eliminate all 1–handles.

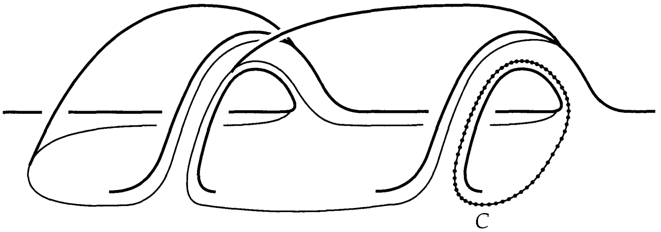
We assume as usual that, when building W^{m+1}, handles are attached in stages of their increasing order. We also assume there are no 0–handles. Thus, we start with a thickening of M into $M \times [0,\varepsilon]$. To its upper boundary $M \times \varepsilon$ we attach all the 1–handles. Call the resulting upper boundary M_1. Then we attach the 2–handles, with their attaching regions $S^1 \times D^{m-1}$ glued to M_1; call the new upper boundary M_2. To it we attach the 3–handles. And so on.

Now, consider some random 1–handle, *i.e.*, a copy of $[-1,+1] \times D^m$, attached to $M \times \varepsilon$ through $\{-1\} \times D^m$ and $\{+1\} \times D^m$ and living between $M \times \varepsilon$ and M_1. Take a parallel copy ℓ of the core $[-1,+1] \times 0$ inside the boundary $[-1,+1] \times S^{m-1}$ (for example, take $\ell = [-1,+1] \times p$ with $p \in S^{m-1}$). The endpoints of ℓ are attached to $M \times \varepsilon$, and, since M is connected, these endpoints can also be linked by some path ℓ' inside $M \times \varepsilon$. By putting ℓ and ℓ' together we get a circle C, as in figure 1.29 on the next page. This circle travels once across our 1–handle. If we could make appear a 2–handle h^2_C attached to C, then this h^2_C would cancel our 1–handle.

All 2–handles are glued along their (thickened) attaching circles to M_1. Since circles can be pushed away from circles, we can then assume that C misses all the attaching regions of the 2–handles and thus that it survives untouched into the next level M_2, as in figure 1.30 on the facing page.

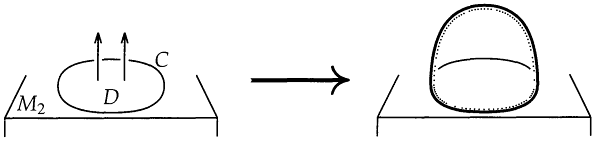


1.29. Circle across 1-handle



1.30. Preparation for handle trading

Since W is simply-connected, C is a homotopically-trivial loop and hence there must be some map $\mathbb{D}^2 \to W$ that sends $\partial\,\mathbb{D}^2$ to C. Since the dimension of W is at least 5, we can in fact get a disk D embedded in W and bounded by C. A little dimension counting also shows that in fact D can be pushed below all handles of order 3 or more and pushed on top of all handles of order 2 or less, so that we end up with a disk D, embedded in M_2 and bounded by C. Now we can thicken this disk D into a canceling pair of a 2- and a 3-handle, as in figure 1.31. Thus, we get a 2-handle attached along C, which cancels our 1-handle, and then we are left with the new 3-handle instead: the 1-handle was traded for the 3-handle. See figure[14] 1.32 on the following page.



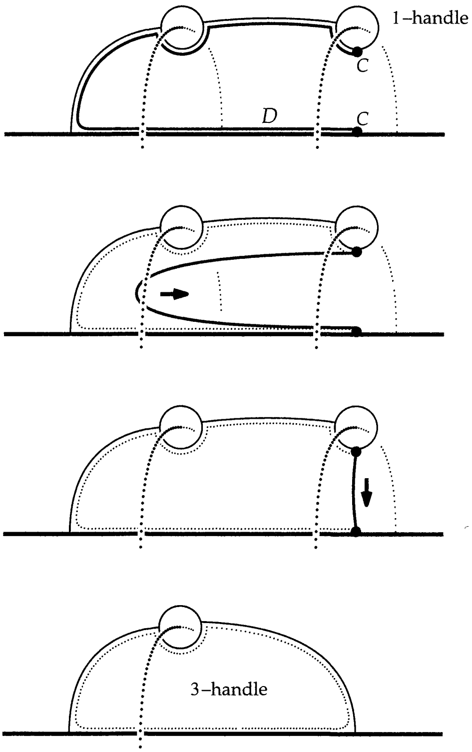
1.31. Creating a pair of 2- and 3-handles

High tradings. *Handle trading can also be done on higher-order handles. Let* h^k_α *be a* k*-handle* $\mathbb{D}^k \times \mathbb{D}^{m-k+1}$ *in* W^{m+1}*, glued along its attaching sphere* $\mathbb{S}^{k-1} \times 0$*. A parallel copy* $\mathbb{D}^k \times p$ *(with* $p \in \mathbb{S}^{m-k}$*) of the core of* h^k_α *determines an element in*[15] $\pi_k(W, M)$*. However, since we assumed* W *to be an* h*-cobordism, all the groups* $\pi_k(W, M)$ *are trivial. Therefore there must be a map* $\mathbb{D}^{k+1} \to W$ *that sends a hemisphere of* $\partial\,\mathbb{D}^{k+1}$ *to cover* $\mathbb{D}^k \times p$*, while*

---

[14] Think of figure 1.32 as a continuation of figure 1.30 on this page.

[15] Remember that $\pi_n(A, B)$ is the group of homotopy classes of all maps $f\colon \mathbb{D}^n \to A$ with $\partial\,\mathbb{D}^n$ sent into B. Such a map f represents the identity 0 if and only if there is a map $\mathbb{D}^{n+1} \to A$ such that a hemisphere of $\partial\,\mathbb{D}^{n+1}$ is mapped as $f\colon \mathbb{D}^n \to A$, while the rest is sent into B.



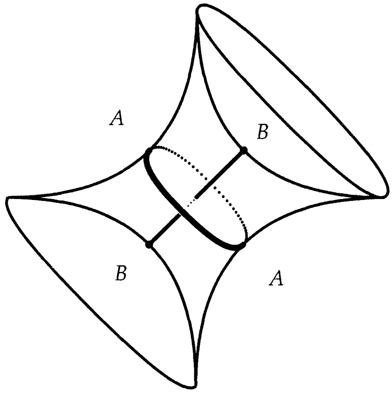
1.32. Handle trading

the rest is sent to M. With a bit of care we can arrange that this $\mathbb{D}^{k+1}$ is sent entirely into a level M_ρ in between the k– and the $(k+1)$–handles. Further, if $m \geq 2k+3$, then $\mathbb{D}^{k+1}$ can be assumed embedded *in M_ρ. The disk $\mathbb{D}^{k+1}$ can then be thickened into a canceling pair of a $(k+1)$– and a $(k+2)$–handle, the $(k+1)$–handle will cancel the k–handle h^k_α, and we are left with the $(k+2)$–handle. Thus, if $m \geq 2k+3$, then we can trade k–handles for $(k+2)$–handles.*

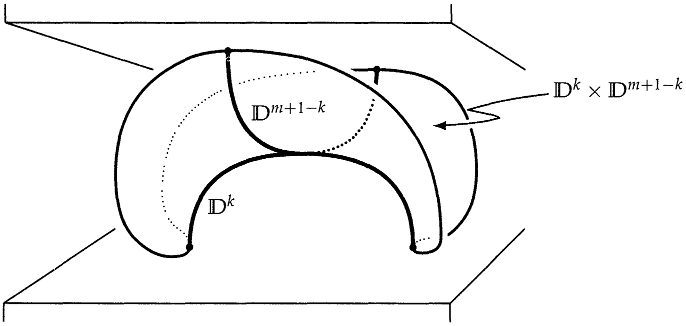
Turning the decomposition upside-down. After eliminating all 0– and 1–handles, we now eliminate the m– and $(m+1)$–handles. For that, we turn the cobordism W^{m+1} upside-down.

Turning W upside-down is equivalent, in terms of Morse functions, to switching from $f\colon W \to [0,1]$ to $-f\colon W \to [-1,0]$, which transforms a critical point of index k into a critical point of index $m+1-k$.

In terms of handles, a same product $\mathbb{D}^k \times \mathbb{D}^{m-k+1}$ can be viewed either as a k–handle $\mathbb{D}^k \times \mathbb{D}^{m+1-k}$, attached "downwards", or as an $(m+1-k)$–handle $\mathbb{D}^k \times \mathbb{D}^{m+1-k}$, attached "upwards". Thus, the belt sphere $0 \times S^{m-k}$ of the k–handle is viewed as the attaching sphere $0 \times S^{m-k}$ of an $(m+1-k)$–handle, while the attaching sphere $S^{k-1} \times 0$ of the k–handle works as the belt sphere $S^{k-1} \times 0$ of an $(m+1-k)$–handle, *etc.* See figures 1.33–1.35.



1.33. Attaching a 2–handle to A, or attaching a 1–handle to B?

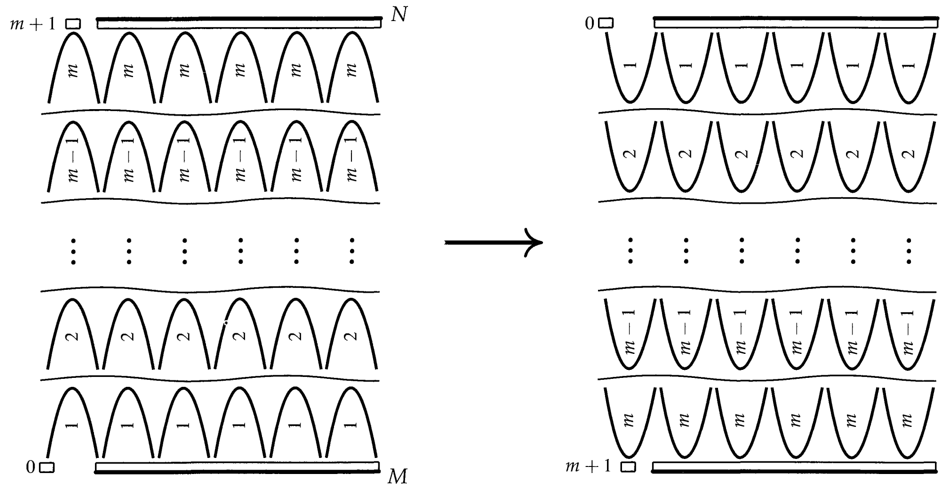


1.34. Viewing a handle upside-down

Since turning the decomposition upside-down transforms $(m+1)$–handles into 0–handles and m–handles into 1–handles, we apply the techniques presented before and eliminate all $(m+1)$–handles as well as trade all m–handles for $(m-2)$–handles.

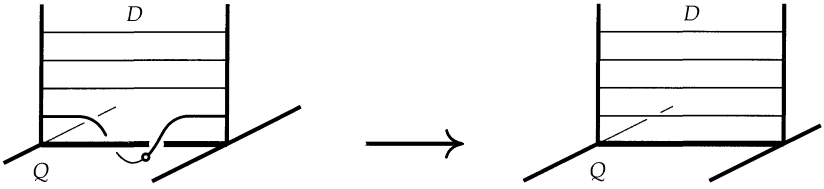
The result of all the above maneuvers is that the handle decomposition of W has now been modified so that it contains no 0–, 1–, m– or $(m+1)$–handles.

***Whitney trick in codimension* 2.** Finally, there is another detail, appearing when we apply the Whitney trick between 2– and 3–handles and between $(m-1)$– and $(m-2)$–handles. Namely, belt spheres of 2–handles and attaching spheres of $(m-1)$–handles have codimension 2, and their complements *a priori* seem like they might be non-simply-connected.

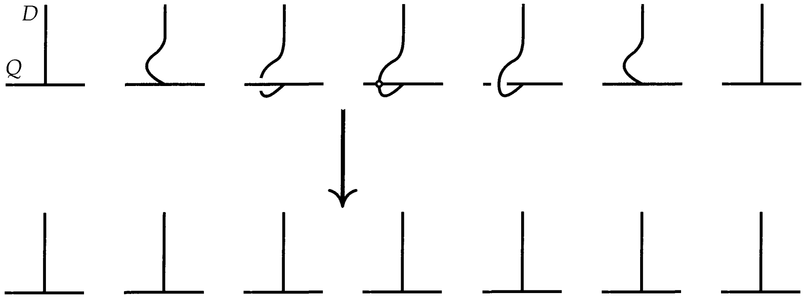


1.35. Turning a handle decomposition upside-down

Spinning. *In general, when one tries to apply the Whitney trick to submanifolds P^2 and Q^{m-2}, one cannot just assume that the disk D will be disjoint from Q^{m-2}, since it is not clear that $M^m \setminus Q$ is simply-connected. Thus, one might need to untangle D by spinning it around Q, as is suggested in figures*[16] *1.36 and 1.37.*



1.36. Untangling the Whitney disk



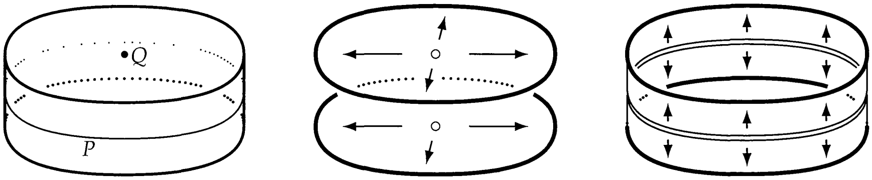
1.37. Untangling the Whitney disk: the movie

---

[16] Figure 1.37 should be understood as having time as one of its dimensions. One can then imagine the surface D twisting around the surface Q and crossing it once, at mid-time.

In the cases actually needed for the h–cobordism theorem, when $m \geq 5$ and P and Q are attaching spheres or belt spheres, this situation does not actually arise: Assume for example that Q^{m-2} is the belt sphere of a 2–handle; then the complement of Q is always simply-connected. Indeed, call M_1 the upper boundary just before this 2–handle was attached and M_2 the one just after the attachment. To obtain M_2 from M_1, one deletes the attaching region $\mathbb{S}^1 \times \mathbb{D}^{m-1}$ from M_1, and glues $\mathbb{D}^2 \times \mathbb{S}^{m-2}$ instead, along their common boundary $\mathbb{S}^1 \times \mathbb{S}^{m-2}$. Our Q is the new $0 \times \mathbb{S}^{m-2}$ in M_2.

If we remove the belt sphere $Q = 0 \times \mathbb{S}^{m-2}$ from M_2, we can collapse the remaining $(\mathbb{D}^2 \setminus 0) \times \mathbb{S}^{m-2}$ to its boundary $\mathbb{S}^1 \times \mathbb{S}^{m-2}$ in $M_1 \cap M_2$. On the other hand, if we remove the attaching circle $\mathbb{S}^1 \times 0$ of the 2–handle from M_1, we can collapse the remaining $\mathbb{S}^1 \times (\mathbb{D}^{m-1} \setminus 0)$ to the same boundary $\mathbb{S}^1 \times \mathbb{S}^{m-2}$. See figure 1.38. What becomes apparent is that $M_2 \setminus Q$ is diffeomorphic to $M_1 \setminus \mathbb{S}^1 \times 0$, but the simple-connectedness of a manifold of dimension at least 5 (such as M_1) cannot be ruined by removing a circle. Thus, the complement $M_2 \setminus Q$ of Q is simply-connected.



2–handle part in $M_2 \setminus Q$ part in M_1

1.38. Complement of codimension–2 belt sphere is simply-connected

Since P is here the attaching sphere $\mathbb{S}^2 \times 0$ of a 3–handle $\mathbb{D}^3 \times \mathbb{D}^{m-2}$, removing P as well from our M_2 (of dimension ≥ 5) will also preserve simple-connectedness, and we can now safely start hunting for Whitney disks embedded in the simply-connected complement of both P and Q.

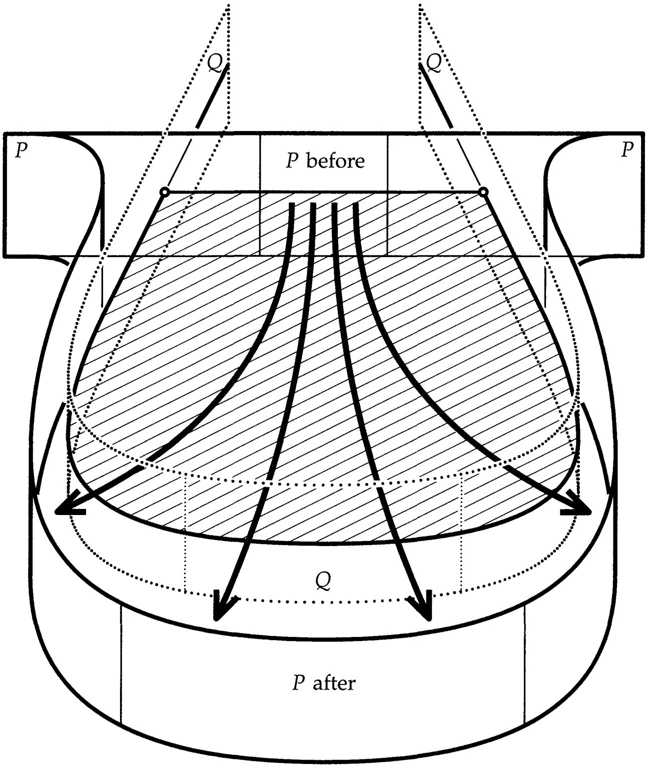
For the other codimension–2 case, when Q is the attaching sphere $\mathbb{S}^{m-2} \times 0$ of a $(m-1)$–handle $\mathbb{D}^{m-1} \times \mathbb{D}^2$ and P is the belt sphere $0 \times \mathbb{S}^2$ of a $(m-2)$–handle $\mathbb{D}^{m-2} \times \mathbb{D}^3$, all one needs to do is turn the decomposition upside-down and argue as above.

***Dimension* 4.** *This reasoning fails in the happy case of dimension $m = 4$. When P is the attaching 2–sphere of a 3–handle, and Q is the belt 2–sphere of a 2–handle, then both have codimension 2 inside their 4–dimensional level M_ρ; and even though both $M_\rho \setminus P$ and $M_\rho \setminus Q$ are simply-connected, the full complement $M \setminus P \cup Q$ has a good chance to not be. Even if we somehow deal with this problem, and then even manage to embed the Whitney disk, in dimension 4 the headaches are not all gone, since the resulting disk might have the wrong* framing; *this last issue is another peculiarity of the 4–dimensional case and is explained on page 57 ahead.*

1.7. Notes

Note: Whitney trick—the technicality

In what follows, we will explain why the Whitney trick can only be applied when the Whitney disk links intersection points of *opposite* signs. Incidentally, the discussion will also make visible an extra obstruction for the 4–dimensional case.



1.39. Pushing P along a Whitney disk while avoiding Q

Let us take a closer look at the Whitney push, as in figure[1] 1.39. The push of P along the Whitney disk must be made so that, during the movement, we do not

---

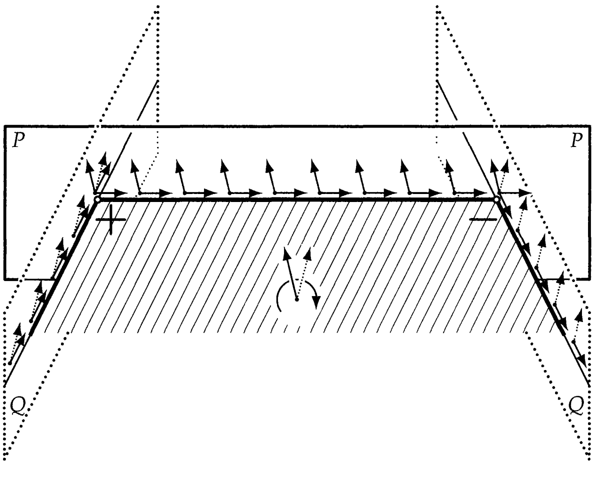
[1] In figure 1.39, as in many others to come, when we run into trouble with the at-most-three dimensions that we can easily represent in a picture, we use dotted lines to suggest objects that exist in other dimensions. Thus, in figure 1.39, we used dotted lines to suggest that Q is a surface, but does not live in the same 3–dimensional slice as P (apart from the Whitney circle, where Q crosses through our slice).

bump into Q, since that would create undesired new intersections. In other words, we want to keep P and Q in separate realms while we push. (Of course, they will necessarily meet along the Whitney circle when P actually passes Q, but that is not what we worry about.)

It turns out that it is all a matter of arranging and fitting bundles. Indeed, what is needed for such a push is a splitting of the normal bundle $N_{D/M}$ of the Whitney disk D into complementary parts, one tangent to P and normal to Q, the other normal to P and tangent to Q; the existence of such a split allows us to keep P normal to Q all through the Whitney push. As we will see, the existence of such a split for $N_{D/M}$ is governed by $\pi_1\big(\mathscr{G}r_k(\mathbb{R}^n)\big) = \mathbb{Z}_2$, corresponding to the two cases when the intersection points have opposite or same signs.

Playing with bundles. Denote by p the dimension of P, by q the dimension of Q, and by $m = p + q$ the dimension of their ambient M (usually the upper boundary M_p of the ascending cobordism W_p). Let x' and x'' be the two intersection points of P and Q on which we are focused. Denote the Whitney disk by D, and denote its normal bundle in M by $N_{D/M}$. As usual, think of $N_{D/M}$ both as an $(m-2)$-plane bundle on D and as a tubular neighborhood of D in M. This neighborhood, slightly enlarged, will contain the support of the isotopy that pushes P beyond Q: it is the place of change. For definiteness, pick a Riemannian metric on M, and arrange that P and Q are orthogonal at their intersection points, that $N_{D/M}$ is orthogonal to T_D, and so on.

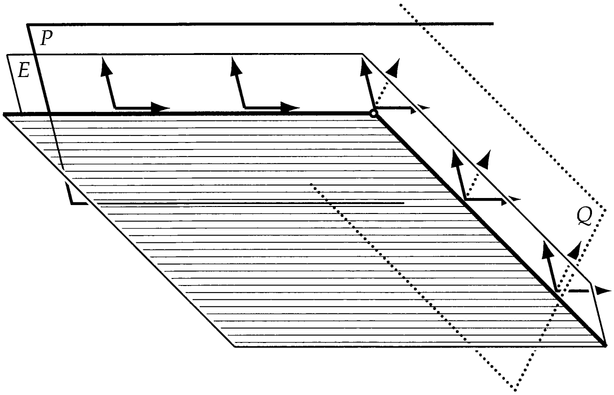
Denote by $\partial_P D$ the half-circle of ∂D contained in P, and by $\partial_Q D$, the one contained in Q; clearly, $\partial_P D \cap \partial_Q D = \{x', x''\}$. The tangent bundle T_P to P splits over $\partial_P D$ naturally into $T_{\partial_P D}$ and a part entirely contained in $N_{D/M}|_{\partial_P D}$. Denote the latter $(p-1)$-subbundle of $N_{D/M}|_{\partial_P D}$ by E_P. Thus, E_P is tangent to P and normal to D. See also figure 1.40.



1.40. Fitting of bundles for the Whitney trick, I

On the other hand, the normal bundle $N_{Q/M}$ of Q splits over $\partial_Q D$ into $N_{\partial_Q D\,/\,D}$ (contained in T_D) and a part entirely contained in $N_{D/M}|_{\partial_Q D}$. Denote the latter $(p-1)$-subbundle of $N_{D/M}|_{\partial_Q D}$ by E_Q. Thus, E_Q is normal to both Q and D.

At the intersection points x' and x'' the two subbundles E_P and E_Q match, and thus define a subbundle E of $N_{D/M}$, defined only along the whole Whitney circle ∂D. The bundle E is tangent to P and normal to Q. See figure 1.41.



1.41. Fitting of bundles for the Whitney trick, II

Extension issues. We wish to extend E to a subbundle of $N_{D/M}$ over the whole disk D. For that, we think of E as a $(p-1)$-plane field inside $N_{D/M}|_{\partial D}$. Since D is contractible, the entire bundle $N_{D/M}$ must be trivial, and thus we can think of E as field of $(p-1)$-planes in $\partial D \times \mathbb{R}^{m-2}$, that is to say, as a map $\partial D \to \mathscr{G}r_{p-1}(\mathbb{R}^{m-2})$, where $\mathscr{G}r$ denotes the Graßmann space of all $(p-1)$-planes inside $\mathbb{R}^{m-2}$. In this light, extending E across D means extending the plane field $E\colon \partial D \to \mathscr{G}r_{p-1}(\mathbb{R}^{m-2})$ to a map $\widetilde{E}\colon D \to \mathscr{G}r_{p-1}(\mathbb{R}^{m-2})$. Therefore, E extends if and only if it determines the trivial element of $\pi_1\big(\mathscr{G}r_{p-1}(\mathbb{R}^{m-2})\big)$.

It is known that, either if $n-k \geq 2$ or if $k > n/2$ and $k \geq 2$, then we have

$$\pi_1\big(\mathscr{G}r_k(\mathbb{R}^n)\big) = \mathbb{Z}_2 \,.$$

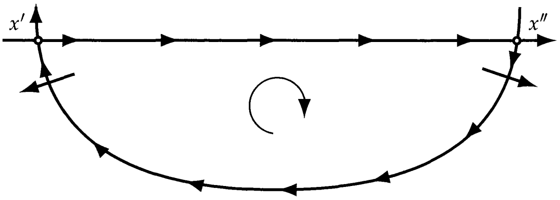
It is generated by any closed path drawn by a k-plane traveling inside $\mathbb{R}^n$ and coming back to its initial position with reversed orientation.[2] (Of course, the k-planes are not oriented; here "reversed orientation" means that we pick a random orientation, we preserve it along the loop, and we compare at the end; if it flipped, we have a winner.)

Therefore, to extend E over D, all we need is that, as we travel along the Whitney circle, E does not reverse orientation. In other words, the bundle E needs to be orientable. This happens exactly when the intersection points x' and x'' of P and Q have opposite signs, as we will argue next.

---

[2] See for example **N. Steenrod**'s ***The topology of fibre bundles*** **[Ste51, Ste99**, ch 22].

Orientations. Remember that P, Q and M are oriented; thus both T_P and $N_{Q/M}$ are oriented bundles. Choose an orientation for D, inducing an orientation of ∂D; then both $T_{\partial D}$ and $N_{\partial D/D}$ get an orientation. Notice that, at the intersection points x' and x'', since we assumed P and Q to be orthogonal, the Whitney circle has "corners", and $T_{\partial_P D}|_{\{x',x''\}}$ coincides with $N_{\partial_Q D\,/\,D}|_{\{x',x''\}}$. Further, at one intersection point their orientations must fit, while at the other they must be opposite. See figure 1.42.



1.42. Along the boundary of the Whitney disk

On one hand, over $\partial_P D$ the bundle E_P fits in the split

$$T_P|_{\partial_P D} = E_P \oplus T_{\partial_P D} \,.$$

On the other hand, over $\partial_Q D$, the bundle E_Q fits in the split

$$N_{Q/M}|_{\partial_Q D} = E_Q \oplus N_{\partial_Q D\,/\,D} \,.$$

At the intersection points x', x'' the normal bundle $N_{Q/M}$ of Q coincides with the tangent bundle T_P of P; correspondingly, E_Q and E_P are identified, and so are $T_{\partial_P D}$ and $N_{\partial_Q D\,/\,D}$.

The bundle E can be extended across the disk D if and only if E is orientable, that is, if and only if E_P and E_Q can be given orientations that induce a well-defined orientation of E. In other words, if there exist orientations for E_P and E_Q that fit at both x' and x''.

To test, orient E_P as induced from the splitting $T_P|_{\partial_P D} = E_P \oplus T_{\partial_P D}$, and orient E_Q as induced from the splitting $N_{Q/M}|_{\partial_Q D} = E_Q \oplus N_{\partial_Q D\,/\,D}$. The bundle E is orientable if and only if these orientations of E_P and E_Q either fit at both x' and x'', or are opposite at both.

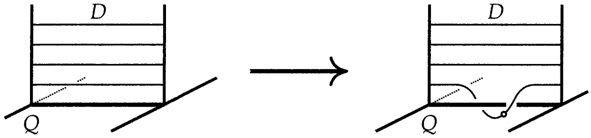
On one hand, the intersection point x' has positive sign if and only if the orientations of $T_P|_{x'}$ and $N_{Q/M}|_{x'}$ fit; similarly for x''. On the other hand, the orientations of $T_{\partial_P D}$ and $N_{\partial_Q D\,/\,D}$ fit at exactly one of x' or x'', while at the other one they are opposite. Therefore, either E is not orientable, or the intersection points x' and x'' must have opposite signs.

***Framing obstruction in dimension* 4.** In the case of dimension 4, besides the inherent difficulty of finding embedded Whitney disks, there appears an extra problem: If P and Q are surfaces inside a 4–manifold, then the bundle-extension problem lives in

$$\pi_1(\mathscr{G}r_1(\mathbb{R}^2)) = \mathbb{Z} \,.$$

Indeed, $\mathscr{G}r_1(\mathbb{R}^2) = \mathbb{RP}^1 = S^1$. Thus, the obstruction to extending E across a Whitney disk is now an *integer*, which we can call the *framing* obstruction. Its modulo 2 reduction vanishes if the intersection points are of opposite signs.

Therefore, the h–cobordism program in dimension 4 runs into three problems: (1) the problem of embedding disks; (2) finding a disk disjoint from P and Q; (3) the right framing for that disk. And, indeed, there are cases when Whitney disks exist, but they all appear with the wrong framing. The spinning procedure, explained earlier and recalled in figure 1.43, can be used to repair the framing, but at the cost of introducing intersection points between the Whitney disk and P or Q...



1.43. Spinning the disk

In dimension 4, there is no sidestepping these three difficulties while remaining in the realm of differentiability.

Note: Diagonalizing chain complexes, and s–cobordisms

In this note, we will explain in some detail how the diagonalization of the boundary operators ∂_k in the proof of the h–cobordism theorem is achieved. We will set the problem in enough generality so that afterwards we can comment on the non-simply-connected version of the h–cobordism theorem, known as the s–cobordism theorem.

Algebra: K–theory. We claimed earlier that, by sliding handles and creating handle pairs, we can eventually make the boundary operators $\partial_k \colon C_k \to C_{k-1}$ appear as a diagonal of 1's. Sliding handles has the same effect as changing basis from h^k_α to $h^k_\alpha \pm h^k_\beta$. In terms of a matrix for ∂_k, this means that we can change the matrix by adding/subtracting rows/columns. In other words, we can multiply by elementary matrices $I \pm E_{ij}$; or more generally, for repeated slides, by $I \pm mE_{ij}$. Creating a pair of canceling handles is the same as extending the matrix of ∂_k by a row and a column, with entries zero but for the new corner. We now generalize all this into an algebraic machinery.

Let R be a ring with unit. Consider $GL_R(n)$ the group of $n \times n$ *invertible* matrices[3] over R, and its subgroup $E_R(n)$ generated by the elementary matrices $I \pm aE_{ij}$, with $a \in R$.

We can map $GL_R(n)$ into $GL_R(n+1)$ by

$$GL_R(n) \longrightarrow GL_R(n+1) \colon \qquad A \longmapsto \begin{bmatrix} A & \\ & 1 \end{bmatrix} .$$

---

[3] An R–valued matrix A is invertible if and only if its determinant $\det A$ is a unit (= invertible element) in R. For matrices with integer values, this means $\det A = \pm 1$, while for real matrices, it means $\det A \neq 0$.

This is called **stabilization.** Taking the direct limit, we obtain

$$GL_{\mathbb{R}} = \varinjlim GL_{\mathbb{R}}(n) \qquad \text{and} \qquad E_{\mathbb{R}} = \varinjlim E_{\mathbb{R}}(n) ,$$

and $E_{\mathbb{R}}$ turns out to be a normal subgroup in $GL_{\mathbb{R}}$. Even more, $E_{\mathbb{R}}$ is exactly the commutator[4] $[GL_{\mathbb{R}}, GL_{\mathbb{R}}]$. Therefore we can define the quotient

$$K_1(R) = GL_{\mathbb{R}} \,/\, E_{\mathbb{R}} ,$$

which is thus the Abelianization of $GL_{\mathbb{R}}$.

An interesting detail is that matrix multiplication induces the same operation in $K_1(R)$ as "block addition":[5]

$$A \cdot B = \begin{bmatrix} A & \\ & B \end{bmatrix} \qquad \text{in } K_1(R) .$$

Since $K_1(R)$ is Abelian, the above operation is written additively as $A + B$, and the class of the identity matrix is denoted by 0.

Higher K's. *On the side, note that $K_1(R)$ is just one of many K_n's from algebraic K-theory. Its origins are in J.H.C. Whitehead's torsion and groups* Wh(G)*, used in the study of simple homotopy,*[6] *and in Grothendieck's groups of projective modules. These ideas were extended by M. Atiyah and F. Hirzebruch's* topological K-theory *into a cohomology theory built with vector bundles; see* **M. Atiyah***'s* ***K-theory*** **[Ati67, Ati89]***. Then H. Bass algebraized it into* algebraic K-theory *by defining $K_0(R)$ and $K_1(R)$ of a ring*[7] *R; the reference is* **H. Bass***'s monograph* ***Algebraic K-theory***[8] **[Bas68]***. The next step was taken by* **J. Milnor***'s definition of $K_2(R)$ in his* ***Introduction to algebraic K-theory*** **[Mil71]***. Finally* **D. Quillen***'s* ***Higher algebraic K-theory*** **[Qui73]** *defined all higher*[9] *$K_n(R)$'s. A recent introduction to algebraic K-theory is* **J. Rosenberg***'s* ***Algebraic K-theory and its applications*** **[Ros94]**.

---

[4] For every group G, its **commutator** is the subgroup $[G, G] = \{aba^{-1}b^{-1} \mid a, b \in G\}$; the quotient $G \,/\, [G, G]$ is a commutative group, called the **Abelianization** of G.

[5] This should suggest links with the direct sum $E \oplus F$ of two vector bundles, just as the stabilization above is linked to the bundle-stabilization $E \oplus \underline{\mathbb{R}}$. We will not pursue this, but it points to the reason why both topological and algebraic K-theory are called "K-theory".

[6] See the pages ahead.

[7] $K_0(R)$ measures the stable obstruction to the existence of bases in projective modules over a ring R, while $K_1(R)$ measures stable obstructions to the uniqueness of bases, up to automorphisms.

[8] Bass's book also includes a nice study of $K_1(\mathbb{Z}[G])$ for finite groups G; these, as we will see shortly, are important in topology.

[9] $K_n(R)$ is defined as follows: First, for a space A and for G a perfect subgroup (*i.e.*, $G = [G, G]$) of $\pi_1 A$, there is a technique called the **plus-construction** that kills G without altering the rest of the homology of A. Roughly, one adds 2-cells that kill the elements of G, then keeps on adding higher cells to kill every new homology class that is created by the preceding additions. The resulting space is denoted by A^+_G. (A mild version of the plus-construction will be used in section 2.3, on page 83.) For $K_n(R)$, one starts with the *classifying space* $\mathscr{B}GL_{\mathbb{R}}$ of $GL_{\mathbb{R}}$; its fundamental group is $\pi_1(\mathscr{B}GL_{\mathbb{R}}) = GL_{\mathbb{R}}$, and thus $E_{\mathbb{R}}$ is a perfect subgroup of $\pi_1\mathscr{B}$. Thus we can build the space $(\mathscr{B}GL_{\mathbb{R}})^+_{E_{\mathbb{R}}}$. The K-groups of R are defined as the homotopy groups of this space, specifically as $K_n(R) = \pi_n((\mathscr{B}GL_{\mathbb{R}})^+_{E_{\mathbb{R}}})$. For a short review of classifying spaces, see the end-notes of chapter 4 (page 204).

Back to cobordisms. It is not too hard to see that, for $R = \mathbb{Z}$, the group $K_1(\mathbb{Z})$ is fully covered by the classes of matrices like

$$\begin{bmatrix} \pm 1 & & \\ & \ddots & \\ & & \pm 1 \end{bmatrix} .$$

In other words, every integral *invertible* matrix can be diagonalized by using elementary operations and by increasing the matrix. In geometric terms, this means that, by sliding handles and by adding canceling pairs of handles, we can manage to diagonalize the boundary operators ∂_k. (The signs of the 1's are irrelevant, since we can always switch the orientations of our handles.)

Specifically (if, say, there are no 0–, 1–, m– and $(m+1)$–handles), we have the chain complex

$$0 \longrightarrow C_{m-1} \xrightarrow{\partial_{m-1}} \cdots \xrightarrow{\partial_5} C_4 \xrightarrow{\partial_4} C_3 \xrightarrow{\partial_3} C_2 \longrightarrow 0 .$$

Since $H_*(W, M; \mathbb{Z}) = 0$, we must have $\operatorname{Im} \partial_{k+1} = \operatorname{Ker} \partial_k$; in other words, the above sequence is exact.

Since ∂_3 is onto and splits (its image $\operatorname{Im} \partial_3 = C_2$ being free), there is a decomposition of C_3 as $\operatorname{Im} \partial_4 \oplus \operatorname{Coker} \partial_4 = \operatorname{Ker} \partial_3 \oplus \operatorname{Coker} \partial_4$; then ∂_3 restricts to an invertible morphism $\operatorname{Coker} \partial_4 \to \operatorname{Im} \partial_3$. Further, by the above remark on $K_1(\mathbb{Z})$, this invertible ∂_3 can be diagonalized *via* handle slides and handle creations.

Going upwards, ∂_4 has free image and thus splits, and restricts to an invertible $\partial_4 \colon \operatorname{Coker} \partial_5 \to \operatorname{Im} \partial_4$ which can be diagonalized, and so on upwards, till all the ∂_k's have been diagonalized; look also at the diagram

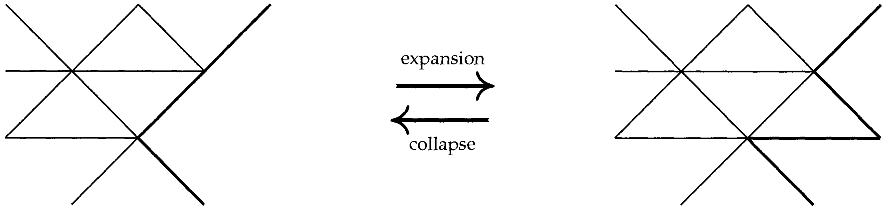
$$\begin{tikzcd}
\cdots \arrow[r, "\partial_5"] & C_4 \arrow[r, "\partial_4"] \arrow[ddd, equal] & C_3 \arrow[r, "\partial_3"] \arrow[d, equal] & C_2 \arrow[r] \arrow[d, equal] & 0 . \\
 & & \operatorname{Coker} \partial_4 \arrow[r, "\partial_3", "\approx"'] & \operatorname{Im} \partial_3 & \\
 & & \oplus & & \\
 & \operatorname{Coker} \partial_5 \arrow[r, "\partial_4", "\approx"'] & \operatorname{Im} \partial_4 & & \\
 & \oplus & & & \\
\cdots \arrow[r, "\partial_5", "\approx"'] & \operatorname{Im} \partial_5 & & &
\end{tikzcd}$$

In the end, all handles are ∂–paired, and the proof of the h–cobordism theorem can proceed as outlined in the main text.

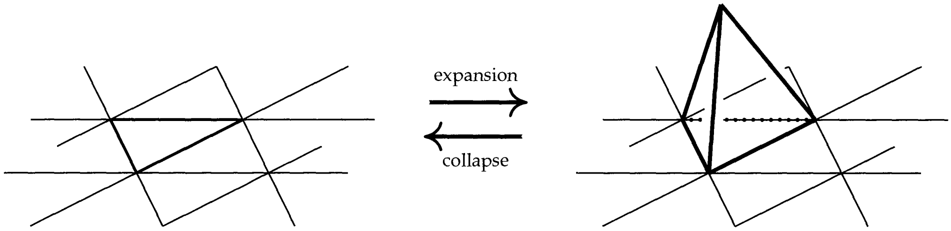
Simple-homotopy type. The argument above—using $K_1(\mathbb{Z})$ in the proof of the h–cobordism theorem—is somewhat of an overkill. Nonetheless, there is a powerful related statement that needs all this general abstract machinery, namely the *non-simply-connected* analogue of the h–cobordism theorem. That result is known as the *s–cobordism theorem,* with the "*s*" coming from "*s*imple-homotopy".

For a statement analogous to the h–cobordism theorem to hold for non-simply-connected manifolds, one needs to strengthen the hypotheses: instead of merely asking that the inclusion $M \subset W$ be a homotopy equivalence, we must now ask that it be a *simple-homotopy equivalence.* We start by explaining this notion:

Let A and B be two polyhedra (*i.e.*, A and B are triangulated, they are simplicial complexes[10]). Then A and B are called **simple-homotopy equivalent** if A can be transformed into B by a finite sequence of elementary expansions and collapses, as in figures 1.44 and 1.45. (Of course, for this to have any chance of happening, A and B must first be homotopy-equivalent in the usual sense.) For convenience, we can also allow "internal" expansions or collapses, with simplices being created or crushed in the "insides" of A, not merely at free faces as suggested in the pictures. (One should think of an elementary expansion or collapse as a simplicial analogue of handle creation/cancellation.) It is known that this notion of simple-homotopy equivalence does not depend on the chosen triangulations; and for simply-connected spaces it actually coincides with homotopy equivalence.



1.44. Elementary collapse / elementary expansion, I



1.45. Elementary collapse / elementary expansion, II

A good start for understanding simple-homotopy equivalence is to think about the statement: A and B are simple-homotopy equivalent if and only if there are simplicial embeddings of A and B into some $\mathbb{R}^N$ (with $N \geq 2\max\{\dim A, \dim B\} + 1$) so that A and B have simplicially-homeomorphic tubular neighborhoods in $\mathbb{R}^N$. Roughly, A and B are simple-homotopy equivalent if they can be "nicely thickened" into being homeomorphic.

Simple-homotopy for h-cobordisms. Consider an h-cobordism W^{m+1} between M^m and N^m. We do not assume M to be simply-connected, but do assume that the inclusion $M \subset W$ is a homotopy equivalence. In particular, $\pi_1 W$ is naturally isomorphic to $\pi_1 M$. We wish to determine whether M and W are simple-homotopy equivalent.

---

[10] We use simplicial complexes instead of cell complexes for simplicity. If one wishes, one can substitute "compact CW-complex" throughout. Simplicial complexes will be recalled in footnote 5 on page 182.

For that, we pick a handle decomposition of W that describes W as a series of handles h^k_α attached to $M \times [0,\varepsilon]$. A little preparation cancels all 0- and $(m+1)$-handles, in the usual manner explained in section 1.6 (page 47).

Threaded handles. For an argument similar to the h-cobordism theorem to proceed, we need to take into account the fundamental groups; that is to say, we must be careful with base-points and loops. Thus, we choose a random base-point $x_0 \in M$. Further, we link every handle h^k_α to the base-point by choosing a path

$$c^k_\alpha \colon [0,1] \longrightarrow W$$

from x_0 to the "center" $0 \times 0 \in \mathbb{D}^k \times \mathbb{D}^{m+1-k}$ of the handle h^k_α. We call *threaded handle* such a handle h^k_α endowed with a path c^k_α linking the base point to its center. The path c^k_α is, of course, considered only up to homotopy.

Notice that there is an obvious action of $\pi_1 M$ on threaded handles: for every $\ell \in \pi_1 M \approx \pi_1 W$ and every threaded handle h^k_α, the handle $\ell \cdot h^k_\alpha$ is just the result of adding an ℓ-loop at the start of c^k_α.

Incidence numbers. Given a threaded k-handle h^k_α and a threaded $(k-1)$-handle h^{k-1}_β, we now consider their *incidence number* $\langle h^k_\alpha \mid h^{k-1}_\beta \rangle$. This is still defined as the intersection number of the attaching sphere of h^k_α with the belt sphere of h^{k-1}_β, but we construe this intersection "number" not merely as an integer from $\mathbb{Z}$, but as an element of $\mathbb{Z}[\pi_1 M]$. This is achieved by using the threads to twist the usual integer intersection numbers with $\pi_1 M$-data.

Specifically, given any intersection point p where the attaching sphere of h^k_α meets the belt sphere of h^{k-1}_β, we pick a random path inside h^k_α from the center of h^k_α to the point p, and then continue inside h^{k-1}_β to the center of h^{k-1}_β. (Up to homotopy, there is only one such path.) Using this path, we can now join c^k_α with the reverse of c^{k-1}_β, and obtain a *loop*

$$c^k_\alpha *_p \bar{c}^{k-1}_\beta .$$

The loop $c^k_\alpha *_p \bar{c}^{k-1}_\beta$ starts at x_0, goes along c^k_α to the center of h^k_α, then descends touching p to the center of h^{k-1}_β, and finally comes back to x_0 by going backwards along c^{k-1}_β; see figure 1.46 on the next page. This loop determines an element in $\pi_1 W$, but since $M \subset W$ is a homotopy-equivalence, we can think of it as an element of $\pi_1 M$.

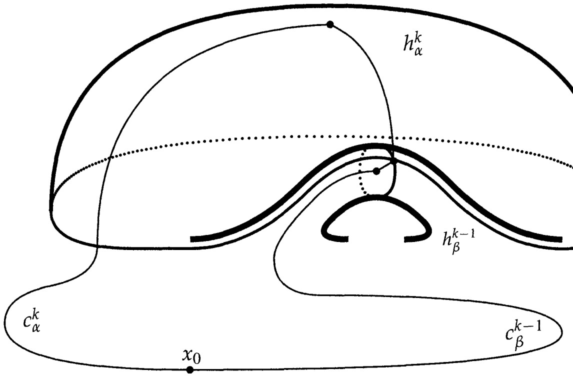
We can now redefine the **incidence number** of the handles h^k_α and h^{k-1}_β as

$$\langle h^k_\alpha \mid h^{k-1}_\beta \rangle = \sum \pm \left[c^k_\alpha *_p \bar{c}^{k-1}_\beta \right] ,$$

with summation over all points p that the attaching sphere of h^k_β and the belt sphere of h^{k-1}_α have in common. The signs are the usual intersection-signs coming from orientations; the addition is performed inside the group-ring $\mathbb{Z}[\pi_1 M]$.

Chain complex. Once endowed with these twisted incidence numbers, we define a chain complex $\{\mathcal{C}_k, \partial_k\}$ with

$$\mathcal{C}_k = \mathbb{Z}[\pi_1 M]\{k\text{-handles}\} .$$



1.46. Incidence numbers in the non-simply-connected case

In other words, the module C_k is freely generated over the group-ring $\mathbb{Z}[\pi_1 M]$ by the k–handles of W. We define the boundary maps $\partial_k \colon C_k \longrightarrow C_{k-1}$ by

$$\partial_k(h^k_\alpha) = \sum \langle h^k_\alpha \mid h^{k-1}_\beta \rangle \cdot h^{k-1}_\beta ,$$

using the $\pi_1 M$–twisted incidence numbers. The resulting homology groups

$$H_k(C_*) = \operatorname{Ker} \partial_k \,/\, \operatorname{Im} \partial_{k+1}$$

of this complex are, of course, modules over $\mathbb{Z}[\pi_1 M]$.

A good way to think about these $H_k(C_*)$ is as the usual integral homology groups $H_k(\widetilde{W}, \widetilde{M}; \mathbb{Z})$ of the universal covers $\widetilde{W}$ and $\widetilde{M}$. Indeed, we have

$$H_k(C_*) = H_k(\widetilde{W}, \widetilde{M}; \mathbb{Z}) .$$

The action of $\mathbb{Z}[\pi_1 M] = \mathbb{Z}[\pi_1 W]$ on the latter appears from the deck transformations of the universal cover $\widetilde{W} \longrightarrow W$.

Since $M \subset W$ is a homotopy equivalence, so will be $\widetilde{M} \subset \widetilde{W}$, and thus we must have $H_*(\widetilde{W}, \widetilde{M}; \mathbb{Z}) = 0$. Hence

$$H_*(C_*) = 0 .$$

Diagonalizing. The natural thing to do now is to try to diagonalize the ∂_k's. This is a problem in $K_1(\mathbb{Z}[\pi_1 M])$. Indeed, creating/canceling pairs of handles and sliding handles will modify the ∂_k's almost in the usual manner but for the use of coefficients from $\mathbb{Z}[\pi_1 M]$.

As a bit of algebra, we notice that inside every $K_1(\mathbb{Z}[G])$ there is a not-so-interesting part represented by the classes of diagonal matrices with entries $\pm g$ with $g \in G$. We denote this subgroup by U_G. The remainder

$$\operatorname{Wh}(G) = K_1(\mathbb{Z}[G]) \,/\, U_G$$

is called the **Whitehead group** of G, and we have an exact sequence

$$0 \longrightarrow U_G \longrightarrow K_1(\mathbb{Z}[G]) \longrightarrow \operatorname{Wh}(G) \longrightarrow 0 .$$

In particular, any $\mathbb{Z}[G]$–valued matrix A determines a unique class $[A] \in \operatorname{Wh}(G)$.

Note that $\mathrm{Wh}(G) = 0$ for all finitely-generated *free Abelian* groups $G = \oplus\, m\mathbb{Z}$. Also, $\mathrm{Wh}(G) = 0$ for all $G = \pi_1 S$ fundamental groups of *surfaces*, orientable or non-orientable.

Back to our manifolds, assume now that some operator $\partial_k \colon C_k \to C_{k-1}$ happens to be an isomorphism. Then the element

$$[\partial_k] \in \mathrm{Wh}(\pi_1 M)$$

is the complete obstruction to modifying ∂_k through handle creations and handle slides so that its matrix become a diagonal with entries $\pm g$ with $g \in \pi_1 M$.

If indeed we have $[\partial_k] = 0$ and we do modify its matrix to such a diagonal, then modifying it further into a diagonal of 1's only amounts to changing the base-loops c^k_α (to kill the g's) and changing orientations (to eliminate minuses).

In general, of course, the ∂_k's are not isomorphisms. Nonetheless, one can climb up in the complex

$$0 \longrightarrow C_m \xrightarrow{\partial_m} \cdots \xrightarrow{\partial_4} C_3 \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \longrightarrow 0$$

by restricting and splitting the various ∂_k to get their nontrivial isomorphism-parts

$$\partial_k \colon \operatorname{Coker} \partial_{k+1} \approx \operatorname{Im} \partial_k$$

in a *coherent* fashion throughout the chain complex, as we suggested earlier with the diagram on page 60. These isomorphism-parts determine elements $[\partial_k] \in \mathrm{Wh}(\pi_1 M)$ that can be combined into the **Whitehead torsion**

$$\tau(W, M) = \Sigma(-1)^k\, [\partial_k]$$

of the chain complex C_*.

As the notation already suggests, it turns out that this quantity does not depend on the choices made along the way, but only on the topology of the h–cobordism W. Moreover, we have:

Lemma. *The boundary operators ∂_k of W can be diagonalized if and only if the torsion $\tau(W, M)$ vanishes.* □

The s–cobordism theorem. Remember that a cobordism W^{m+1} between some M^m and N^m is called an h–cobordism if $M \subset W$ is a homotopy equivalence. Going further, an h–cobordism W is called an **s–cobordism** if its torsion $\tau(W, M)$ vanishes. The name comes from "simple-homotopy", and will be justified later on.

Given an s–cobordism W^{m+1}, we can diagonalize all the ∂_k's as above. After doing that, if the dimension m of M is bigger than 5, then we can apply the Whitney trick to transform algebraic intersections into geometric intersections, and thus eventually cancel all handles.

Of course, since M_ρ is no longer simply-connected, the Whitney trick will now be applied only to intersection points with opposite "threaded" intersection numbers; that is, two intersection points x' and x'' can be cancelled if the intersection number at x' is $+g$, while at x'' it is $-g$, for some $g \in \pi_1 M$. The $\pm$–sign assures that the bundles fit, as explained earlier (page 54), while the g–part assures that the resulting Whitney circle is null-homotopic and thus bounds a disk. A similar type of care must be taken when trading threaded 1–handles for 3–handles, but it works. Eventually, all this yields:

s–Cobordism Theorem. *Assume W^{m+1} is an s–cobordism between M^m and N^m, with $m \geq 5$. Then W is diffeomorphic to the trivial cobordism $M \times [0,1]$, and hence M and N are diffeomorphic.* □

More, when $m \geq 6$, every element $\tau \in \mathrm{Wh}(\pi_1 M)$ can actually be realized as the torsion of some h–cobordism W built on top of M. In particular, all the h–cobordisms that sit on top of M are classified up to diffeomorphisms by the elements of $\mathrm{Wh}(\pi_1 M)$.

A bit more on simple-homotopy. The same procedure used for defining the torsion of a cobordism W can be used to define the torsion of any simplicial inclusion $A \subset B$. Indeed, if something can be done with handles, it should be doable with simplices. Assume that $A \subset B$ is a homotopy equivalence (and so $\pi_1 A = \pi_1 B$), think of B as obtained by adding simplices to A, and define the simplicial chain complex

$$C_k = \mathbb{Z}[\pi_1 A]\{\,k\text{–simplices}\}\ ,$$

with corresponding boundary operators ∂_k defined similarly to what we did previously. We then define the torsion

$$\tau(B, A) \in \mathrm{Wh}(\pi_1 A)$$

in an analogous manner.

The algorithm followed in the proof of the s–cobordism theorem can be used here as well, thinking of an elementary expansion as analogous to a handle creation, and of an elementary collapse as analogous to a handle cancellation. And there is, of course, a simplicial analogue of handle sliding (which is easiest to describe using cellular decompositions instead of simplicial complexes). Further, since we are dealing with simplicial complexes (not manifolds) and simple-homotopy allows thickening to increase dimensions, the dimensional restriction for performing the Whitney trick does *not* appear, and we end up with:

Lemma. *The inclusion $A \subset B$ is a simple-homotopy equivalence if and only if it is a homotopy equivalence and the torsion $\tau(B, A)$ vanishes.* □

In particular, this explains the name of "s–cobordism": a cobordism W between M and N is an s–cobordism if and only if the inclusion $M \subset W$ is a simple-homotopy equivalence.

Furthermore, one can extend the above considerations from inclusion maps $A \subset B$ to general maps $A \to B$. Indeed, any map $f\colon A \to B$ can be viewed as an inclusion by using its **mapping cylinder**

$$\mathcal{M}_f = A \times [0,1] \cup B \,/\, \{(x,1) = f(x)\}$$

and replacing $f\colon A \to B$ with the inclusion $A \times 0 \subset \mathcal{M}_f$ in order to define its **Whitehead torsion**[11]

$$\tau(f) = \tau(\mathcal{M}_f,\, A)\ .$$

A homotopy equivalence $f\colon A \to B$ is called a **simple-homotopy equivalence** if it is homotopic to a composition of elementary collapses and expansions. We then have

---

[11] Of course, first f needs to be homotoped to a simplicial map.

Lemma. *Let $f\colon A \to B$ be a homotopy equivalence. Then f is a simple-homotopy equivalence if and only if its torsion $\tau(f)$ vanishes.* □

Finally, we should mention that simple-homotopy type is in fact a *topological invariant,* and therefore does not depend on the choice of triangulations (or cell decompositions).[12]

References. The s-cobordism theorem was proved independently around 1963 by **D. Barden** in his thesis ***The structure of manifolds*** **[Bar63]**, as well as in **J. Stallings'** ***On infinite processes leading to differentiability in the complement of a point*** **[Sta65]** (see also his ***Lectures on polyhedral topology*** **[Sta67]**), and **B. Mazur**'s ***Relative neighborhoods and the theorems of Smale*** **[Maz63]**.

The notion of simple homotopy type and Whitehead torsion were introduced by **J.H.C. Whitehead**'s ***Simple homotopy types*** **[Whi50]**. A nice introduction to simple-homotopy theory is **M. Cohen**'s ***A course in simple homotopy theory*** **[Coh73]**. For an excellent discussion of torsions, read **J. Milnor**'s ***Whitehead torsion*** **[Mil66]**, where other versions (Reidemeister torsion) are also described and applied for example on 3-manifolds. A proof of the s-cobordism theorem can also be found in **C. Rourke** and **B. Sanderson**'s ***Introduction to piecewise-linear topology*** **[RS72]**.

Bibliography

As a general reference for differentiable manifolds as seen from topology, the best reference is **M. Hirsch**'s ***Differential topology*** **[Hir76]** or its corrected reprint **[Hir94]** (with the recommendation to skim through the technical chapters 2 and 3 at first lecture; nonetheless, those chapters are one of the few places to see in detail why, say, immersed 2-disks in 5-manifolds can easily be made embedded, *etc.*) For a milder start, one can take delight in **J. Milnor**'s ***Topology from the differentiable viewpoint*** **[Mil65b, Mil97]**. With a view toward more advanced topology, read **A. Kosinski**'s ***Differential manifolds*** **[Kos93]**. Another introduction to smooth manifolds is **V. Guillemin** and **A. Pollack**'s ***Differential topology*** **[GP74]**.

Cobordisms were studied by **R. Thom** in his celebrated ***Quelques propriétés globales des variétés différentiables*** **[Tho54]** (results first announced in **[Tho53b]**). He classified manifolds up to cobordism by using tools from algebraic topology, work that earned him a Fields Medal. A mild introduction can be found in **M. Hirsch**'s ***Differential topology*** **[Hir76, Hir94**, ch 7**]**, with the actual theory developed in **R. Stong**'s ***Notes on cobordism theory*** **[Sto68]**.

The generalized Poincaré conjecture was proved using combinatorial methods by **J. Stallings** for dimensions $m \geq 7$ in ***Polyhedral homotopy-spheres*** **[Sta60b]** (extended to $m \geq 5$ by **E. Zeeman** **[Zee61]**), then in the smooth case by **S. Smale**'s ***Generalized Poincaré's conjecture in dimensions greater than four*** **[Sma61]**. The h-cobordism theorem is due to **S. Smale**'s ***On the structure of manifolds*** **[Sma62]**,

---

[12] This was proved in **R. Kirby** and **L. Siebenmann**'s ***Foundational essays on topological manifolds, smoothings, and triangulations*** [KS77] for topological manifolds, and in **T. Chapman**'s ***Topological invariance of Whitehead torsion*** [Cha74] for general CW-complexes. For contrast, look also at **J. Milnor**'s ***Two complexes which are homeomorphic but combinatorially distinct*** [Mil61].

and earned him a Fields Medal. For a proof, one can also look at **C. Rourke** and **B. Sanderson**'s ***Introduction to piecewise-linear topology*** **[RS72**, ch 6] (even though they discuss it in the piecewise-linear setting, it is easy to translate the main argument to the smooth setting). Another source is **J. Milnor**'s ***Lectures on the h-cobordism theorem*** **[Mil65a]**, but he uses Morse functions and no handles, which makes the proof less geometric and visual than it should be. Yet another account of the h-cobordism theorem is contained in **A. Kosinski**'s ***Differential manifolds*** **[Kos93]**.

The Whitney trick was first imagined by **H. Whitney** in ***The self-intersections of a smooth n-manifold in 2n-space*** **[Whi44]**. He used it to eliminate self-intersections of immersions, and prove that every m-manifold can be embedded in $\mathbb{R}^{2m}$.

Morse functions were devised by **M. Morse**'s ***The critical points of a function of n variables*** **[Mor31]**. They can be found nicely explained in **M. Hirsch**'s ***Differential topology*** **[Hir76, Hir94**, ch 6]. **J. Milnor**'s book ***Morse theory*** **[Mil63a]** is devoted to the subject, and goes on to apply it to Riemannian geometry (Morse theory on spaces of geodesics). See also **Y. Matsumoto**'s ***An introduction to Morse theory*** **[Mat02]**.

> *It is worth noting that Morse theory admits a formulation that can be used to obtain the whole homology of the manifold directly from Morse data.*[13] *This formulation allows itself for generalizations to infinite-dimensional settings, for example in various versions of Floer homology. (The latter are gaining a lot of prominence on 3-manifolds, from where they have strong ramifications in the 4-dimensional realm.*[14]*) The finite-dimensional case is explained in this spirit—with a view toward infinite-dimensional applications—in* **M. Schwarz***'s* ***Morse homology*** **[Sch93]***, or in the third edition* **[Jos02**, ch 6] *of* **J. Jost***'s* ***Riemannian geometry and geometric analysis****.*

Closer to our subject, we should mention that the theory of handles for 4- (and 3-) manifolds has become a powerful symbolic calculus with diagrams, known as *Kirby calculus*. As a preview, look ahead at the end-notes of the next chapter (page 91). An exhaustive exposition can be found in **R. Gompf** and **A. Stipsicz**'s monograph ***4-Manifolds and Kirby calculus*** **[GS99]**.

Topology of manifolds. The theory of high-dimensional manifolds, with the help of the h- and s-cobordism theorems, grew into surgery theory. Surgery theory traces its origins to **J. Milnor**'s ***Differentiable structures on spheres*** **[Mil59]**, to **M. Kervaire** and **J. Milnor**'s ***Groups of homotopy spheres*** **[KM63]**, and to **A. Wallace**'s ***Modifications and cobounding manifolds*** **[Wal60]**. A nice place to start is **W. Browder**'s monograph ***Surgery on simply-connected manifolds*** **[Bro72]**. The non-simply-connected case gains a very heavily-algebraic flavor, and is presented by one of the main surgeons, **C.T.C. Wall**, in his hard-to-read ***Surgery on compact manifolds*** **[Wal70, Wal99]**. A recent treatment is **A. Ranicki**'s ***Algebraic and geometric surgery*** **[Ran02]**. Some surgery can also be found at the end of **A. Kosinski**'s ***Differential manifolds*** **[Kos93]**.

---

[13] One defines $C_k = \mathbb{Z}\{\text{critical points of index } k\}$, and lets ∂_k count the number of gradient flow-lines that go from a critical point of index k to a critical point of index $k-1$. The difficult part is to assign the appropriate sign to each flow-line.

[14] See also the references on page 475 at the end of chapter 10.

The issue of the various smooth structures on a topological (or PL) manifold rose its head with **J. Milnor**'s ***On manifolds homeomorphic to the 7–sphere*** **[Mil56b]**, where some of the twenty-eight distinct smooth structures on S^7 were uncovered. More exotic spheres were uncovered by **J. Milnor** in ***Differentiable structures on spheres*** **[Mil59]**; then **M. Kervaire**'s ***A manifold which does not admit any differentiable structure*** **[Ker60]** found a (triangulated) 10–manifold with no smooth structures whatsoever.

Since the study of purely topological manifolds appeared unapproachable, extra structure was added, such as a nice triangulation[15] (called **piecewise-linear** or **PL structure**). The problem of uniqueness for smooth structures was first studied along the gap between PL manifolds and smooth ones, was started by **S. Cairns**' ***The manifold smoothing problem*** **[Cai61]**, then strengthened by **R. Lashof** and **M. Rothenberg**'s ***Microbundles and smoothing*** **[LR65]** and by **M. Hirsch** and **B. Mazur** in ***Smoothings of piecewise-linear manifolds*** **[HM74]**. The crowning achievement of this phase was the so-called Cairns–Hirsch theorem. A consequence was that there is no difference between PL and smooth up to dimension 7. See also the references inside the end-notes of chapter 4 (page 207).

The general existence issue for smooth structures on topological manifolds was breached by **J. Milnor** in ***Microbundles*** **[Mil64]**, where he used microbundles as analogues of tangent bundles for topological manifolds.

When the time was ripe, the naked homeomorphism finally opened itself to study through **R. Kirby**'s ***Stable homeomorphism and the annulus conjecture*** **[Kir69]**, and then grew into **R. Kirby** and **L. Siebenmann**'s theory of smoothing topological manifolds of dimension at least 5, explained in their ***Foundational essays on topological manifolds, smoothings, and triangulations*** **[KS77]**. For example, up to dimension 7, the existence of a smooth structure on a topological m–manifold X depends solely on the vanishing of the *Kirby–Siebenmann invariant* $\mathrm{ks}(X) \in H^4(X;\mathbb{Z}_2)$. If one smooth structure exists, then all others are classified by the elements of $H^3(X;\mathbb{Z}_2)$. For manifolds of dimension 7 or higher, new obstructions start to appear from the PL/smooth gap. Of course, the theory fails in dimension 4.

For more details on the smoothing theory of high-dimensional manifolds, see the end-notes of chapter 4 (page 207). A few exotic high-dimensional spheres are discussed in the end-notes of the next chapter (page 97).

---

[15] A "nice" triangulation here means a triangulation in which the link of every vertex is simplicially equivalent to a sphere.

Chapter 2

Topological 4–Manifolds and h–Cobordisms

ANY straightforward attempt to use the high-dimensional proof of the h–cobordism theorem in the case of manifolds of dimension 4 fails, and thus a virtual juggernaut for the classification problem is lost. The failure is owing to the difficulty of embedding 2–disks in 4–manifolds, or, more to the point, to the problem of eliminating self-intersections of immersed disks. If we could solve that, we would modify any immersed disk to an embedded disk and hence proceed with the h–cobordism program as outlined in the preceding chapter.

In this chapter we will see what has been done to embed disks and thus prove a 4–dimensional h–cobordism theorem. The inevitable price to pay was dropping differentiability, and thus weakening the conclusion from diffeomorphisms to homeomorphisms. Thus, while the preceding chapter could be viewed as a glance into higher dimensions, one can think of this chapter as a quick visit with topological 4–manifolds. As far as rigor is concerned, this chapter is even more of a fairy tale than the previous.

The chapter starts by presenting Casson handles, which are the main tool used for topologically embedding disks in 4–manifolds; indeed, the fundamental technical result of M. Freedman proves that these wild creatures are topologically standard handles. Using this, the h–cobordism program can be followed through and Freedman's topological h–cobordism theorem for dimension 4 follows in section 2.2 (page 80). As is explained in the endnotes (page 96), Casson handles are rather easy to embed, justifying their earlier name of "flexible handles".

Embeddings of Casson handles are also used in section 2.3 (page 83) to show that every homology 3–sphere bounds a contractible topological 4–manifold—a fake 4-ball. An important example of homology 3–sphere is the Poincaré sphere, which is used to build the so-called E_8–manifold. The higher-dimensional analogues of the Poincaré sphere are actual exotic spheres and are described in the end-note on page 97.

The chapter concludes with section 2.4 (page 89), in which the failure of the smooth h–cobordism theorem in dimension 4 is located around some twisted contractible sub–h–cobordisms.

An end-note on page 91 briefly outlines the description of handle decompositions of 4–manifolds known as Kirby calculus. Completing the tour of horizon, another end-note (page 101) takes a peek at the realm of 3–manifolds.

2.1. Casson handles

Here is the plan: we try to transplant the high-dimensional proof of the h–cobordism theorem to the case of dimension 4; we fail, then we push our difficulties away to infinity, and thus we eventually succeed.

Starting the h–cobordism program

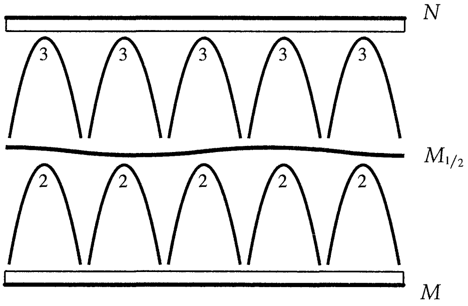
Imagine we are pursuing the h–cobordism program on a 5–dimensional cobordism W between two 4–manifolds M and N, all simply-connected.

Preparing the cobordism. As explained in section 1.6 (page 47), we start by eliminating any 0– or 5–handles that might appear in the handle decomposition of W. Then we trade all the 1–handles for 3–handles and all the 4–handles for 2–handles. We are left with a handle decomposition of W that contains only 2– and 3–handles. The corresponding chain complex is now simply

$$0 \longrightarrow C_3 \xrightarrow{\partial} C_2 \longrightarrow 0 .$$

Since $H_*(W, M; \mathbb{Z}) = 0$, the boundary operator ∂ is an isomorphism and can be diagonalized (by handle slides and handle creations) to a diagonal of 1's. This means that all 2– and 3–handles are paired by ∂ into pairs h^2_α, h^3_β so that $\partial h^3_\beta = h^2_\alpha$.

Let $1/2$ be the level in W that appears immediately after all 2–handles have been attached, but before any 3–handle is attached, as sketched in figure 2.1 on the facing page. Thus, the ascending cobordism $W_{1/2}$ contains just M and all 2–handles. In its (4–dimensional) upper boundary $M_{1/2}$ are located both the belt spheres of the 2–handles and the attaching spheres of the 3–handles.



2.1. The middle level of the cobordism W

The condition $\partial h^3_\beta = h^2_\alpha$ means that, algebraically, the attaching sphere of h^3_β has a nontrivial intersection number only with the belt sphere of h^2_α and that number is $+1$.

> *Keep in mind that in our context a* 2*–handle is now a copy of* $\mathbb{D}^2 \times \mathbb{D}^3$*, attached to* $M \times [0,\varepsilon]$ *along* $\mathbb{S}^1 \times \mathbb{D}^3$*; its belt sphere is* $0 \times \mathbb{S}^2$*. A* 3*–handle is a copy of* $\mathbb{D}^3 \times \mathbb{D}^2$*, attached along* $\mathbb{S}^2 \times \mathbb{D}^3$*; its attaching sphere is* $\mathbb{S}^2 \times 0$*. Thus,* $\partial h^3_\beta = h^2_\alpha$ *means that the corresponding spheres* $0 \times \mathbb{S}^2$ *and* $\mathbb{S}^2 \times 0$ *have intersection number* $+1$ *inside the* 4*–manifold* $M_{1/2}$*.*

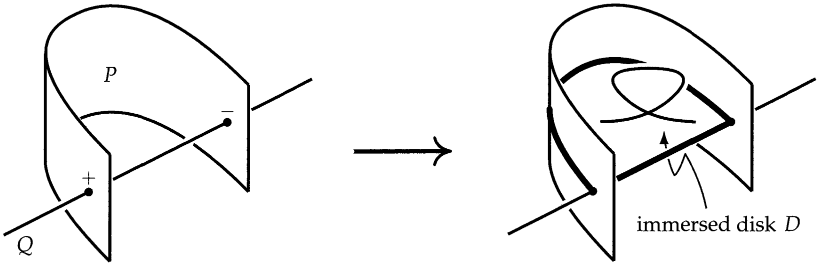
For the h–cobordism program to proceed, we now need to realize the algebraic intersection numbers of the attaching 2–spheres with the belt 2–spheres as geometric intersections. That is, we need to eliminate all algebraically-canceling intersections.

I want to do the Whitney. For this purpose, as in the Whitney trick, we choose an attaching 2–sphere P of some 3–handle and a belt 2–sphere Q of some 2–handle, both living in $M_{1/2}$. We pick a pair of intersection points of opposite signs, then choose a path linking the two points inside P and a path linking the two inside Q. These two paths draw a circle, which must bound an immersed disk D inside $M_{1/2}$. See figure 2.2 on the next page. Assume for now that the complement $M_{1/2} \setminus P \cup Q$ is still simply-connected, so that D can in fact be immersed while avoiding both P and Q.

If D were an embedded disk,[1] then we could use the Whitney trick to eliminate the two intersection points. However, modifying D to an embedded disk is no longer guaranteed by a simple dimension count: embeddings are no longer open and dense in the space of maps $D \longrightarrow M_{1/2}$. Nonetheless, we can safely assume that D fails to be embedded only owing to transverse double-point singularities.

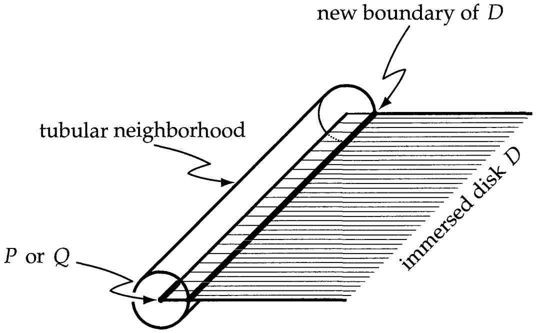
---

[1] Throughout this discussion we gloss over the framing problem that was detailed back in the endnotes of the preceding chapter (page 57). A more complete discussion would say "if D were an embedded disk with the right framing..."



2.2. Immersed disk, wants to be a Whitney disk

Clean-up. To simplify the setting, we will cut tubular neighborhoods of P and Q out of $M_{1/2}$ and think of D as immersed in the resulting 4–manifold with boundary, with ∂D sent to that boundary, as suggested in figure 2.3.



2.3. Cutting out P and Q

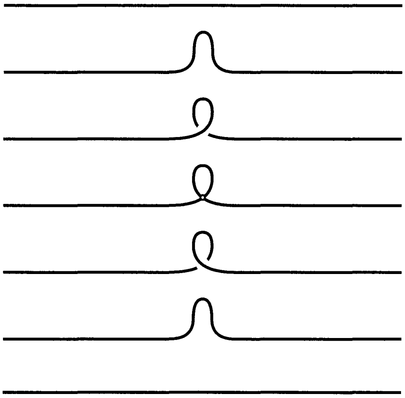
Even more, by cleaning away all the context, we will just think about a disk D immersed in some random simply-connected 4–manifold M with non-empty boundary, with ∂D being sent into ∂M. The immersion of D fails from being an embedding only through the existence of transverse double-points.

These double-points are the enemy.

A few tricks

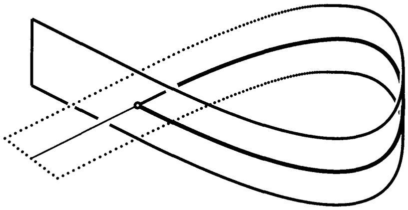
Before we can actually focus on the immersed disk D, we must first gather a few techniques.

Creating self-intersections. To start things off, we observe that, for any surface S immersed in a 4–manifold, we can create more self-intersections at will. This can be done by pinching and twisting a kink, as suggested in figure[2] 2.4. The sign of these self-intersections can be adjusted as needed (in the figure, run time upwards or downwards to get opposite signs).



2.4. Creating a self-intersection

Eliminating self-intersections. Imagine now that we have some immersed surface S in a 4–manifold, and that it has a self-intersection point. Choose a loop in S that is based at that double-point, leaves along one branch, and returns along the other branch, as in figure[3] 2.5. Be optimistic and assume for a moment that our loop bounds an embedded disk in the complement of S.



2.5. Loop at a self-intersection of S

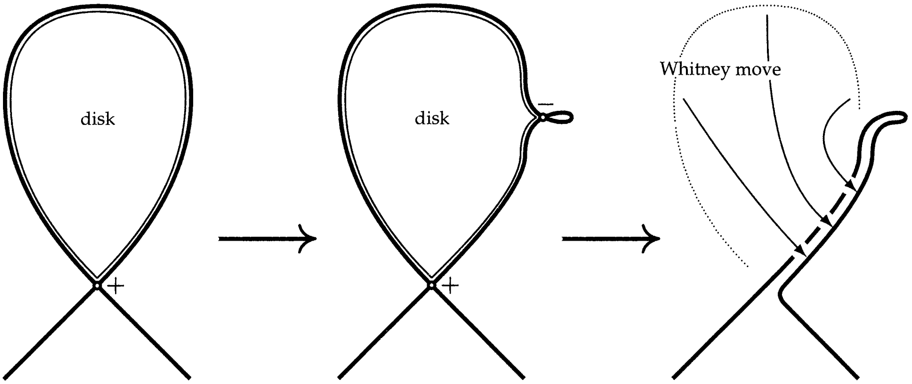
Then we could create another self-intersection (of opposite sign) of S right on the boundary of the disk and end up with a Whitney situation:[4] we can now push our surface along the disk and eliminate both self-intersections, as in figure 2.6 on the next page.

---

[2] In figure 2.4 we represent a surface in 4–dimensional space as the movie of a curve, with time running vertically.

[3] In figure 2.5 we represent creatures that escape in the fourth dimension by dotted lines.

[4] Again, assume the right framing of the disk.



2.6. Eliminating a self-intersection

Therefore, if we wish to eliminate the self-intersections of S, it would be enough to hunt for disks that are bounded by such loops in S, based at the double-points.

Finger moves. For a loop as above to have any chance of bounding an embedded disk in the complement of our surface S, that complement better be simply-connected.

A method of reducing the fundamental group of the complement of S is the **finger move** suggested in figure 2.7 on the facing page. The name comes from imagining that we push our finger through a rubbery surface.

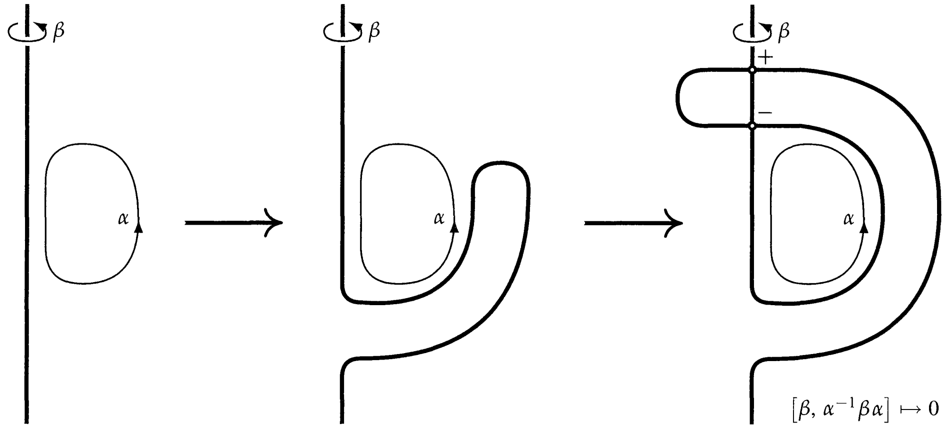
If a finger move is made following a loop α in the complement, then it results in killing the commutator[5] $[\beta, \alpha^{-1}\beta\alpha]$ in π_1 of the complement, where β is a loop around S.

> ***Commuting by a torus.*** *That β and $\alpha^{-1}\beta\alpha$ commute in the complement of the fingered surface is not hard to see. First, model a small neighborhood of a self-intersection point as $\mathbb{D}^2 \times \mathbb{D}^2$, with $\mathbb{D}^2 \times 0$ and $0 \times \mathbb{D}^2$ representing the two branches of S. In the complement of S lives the torus $\mathbb{S}^1 \times \mathbb{S}^1$. One generating circle of the torus is $\mathbb{S}^1 \times 1$ and represents β, while the other generating circle $1 \times \mathbb{S}^1$ represents $\alpha^{-1}\beta\alpha$, since the loop around the fingered branch is simply the translate of β along α. But the fundamental group of a torus is commutative, and therefore β and $\alpha^{-1}\beta\alpha$ must commute. See figure 2.8 on the next page.*

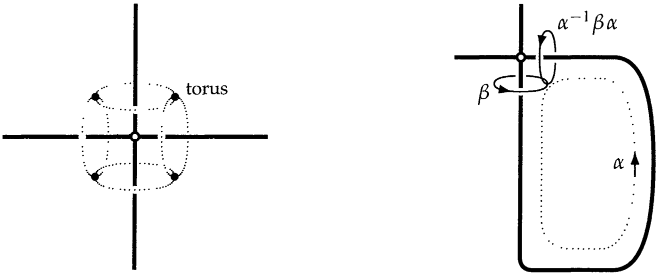
By killing commutators, a finger move thus reduces the fundamental group. The price to pay, though, is that with each finger move we create a new pair of self-intersection points (see figure 2.9 on the facing page).

---

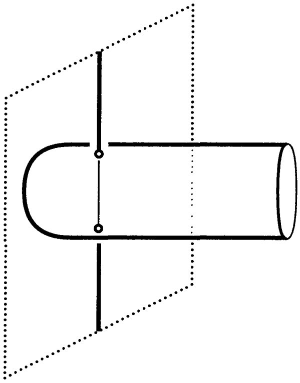
[5] The **commutator** $[x, y]$ of two elements in a group is $[x, y] = xyx^{-1}y^{-1}$. It vanishes when x and y commute.



2.7. A finger move



2.8. Torus around an intersection point and its generators

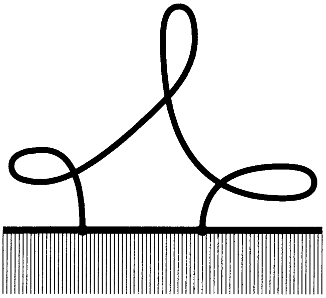


2.9. Fingering creates crossings

Introducing Casson handles

We start with a "thin" version of the construction, then we "thicken" to the actual Casson handles.

Growing a tower of failures. Consider an immersed disk D (a wannabe Whitney disk), inside a simply-connected 4–manifold M with non-empty boundary (with ∂D embedded in ∂M). The disk D has singularities as transverse double-points, as in figure 2.10.



2.10. An immersed disk

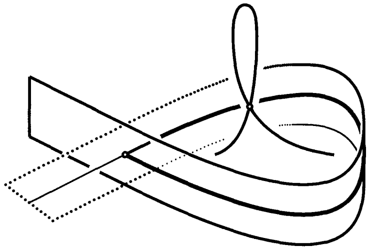
In all cases that actually concern us here, the fundamental group of the complement of D in M is a perfect group;[6] if we destroy enough commutators in $\pi_1(M \setminus D)$, then the fundamental group will disappear. Hence we can use finger moves to make the complement of D simply-connected.

Now, for each self-intersection of D, choose a loop in D, based at the self-intersection, leaving along one branch and returning along the other. If one of these loops actually bounds an embedded disk in the complement, then the corresponding self-intersection can be eliminated, as we saw earlier. In general, all we can find is merely an immersed disk, with its own self-intersections. Still, for each chosen loop we pick such an immersed disk, as in figures 2.11 and 2.12 on the facing page. If we can eliminate the self-intersections of these new immersed disks, then the self-intersections of D would disappear as well.

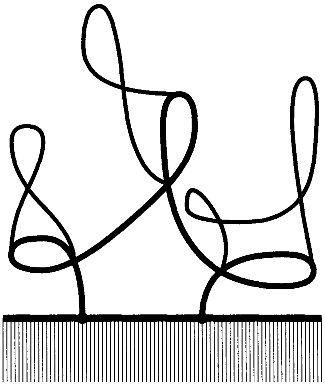
Think of this procedure as "pushing the problem away" a little: instead of having to deal with the self-intersections of the initial disk D, now we have to deal with the self-intersections of the new disks. While that seems only to make things worse, the wonderful idea of A. Casson was to keep repeating the procedure indefinitely, and thus "push the problem away to infinity".

---

[6] A group G is called **perfect** if $G = [G, G]$, where $[G, G] = \{[x, y] \mid x, y \in G\}$.



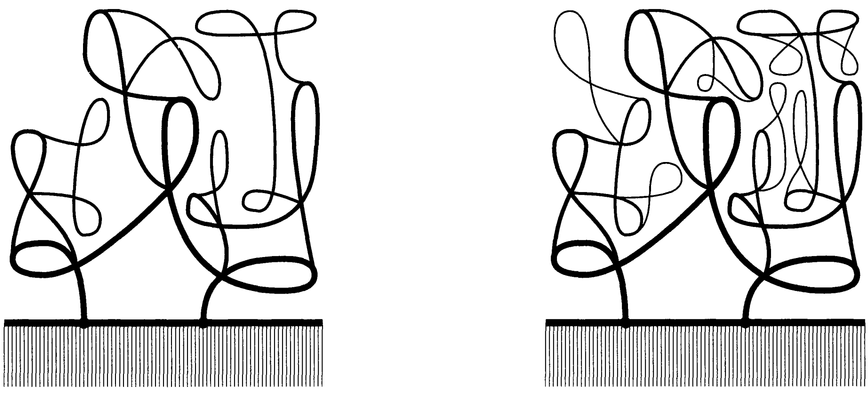
2.11. Immersed disk bounded by loop at double-point



2.12. Growing a Casson handle: first stage

Therefore, we continue: we perform finger moves on the second generation of immersed disks until the complement of the whole thing is simply-connected, then we pick loops at each self-intersection of the new disks, and choose for each an immersed disk, as in figure 2.13.

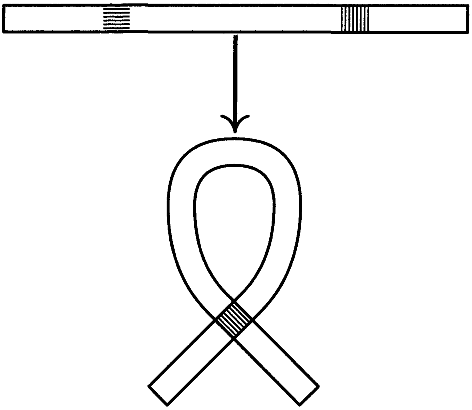
Lather, rinse, repeat. Infinitely many times.



2.13. Growing a Casson handle: second and third stages

Thicken to a Casson handle. Now imagine that the above process is carried out not with simple disks (copies of $\mathbb{D}^2$), but with thickened disks (copies of $\mathbb{D}^2 \times \mathbb{D}^2$, attached along $\mathbb{S}^1 \times \mathbb{D}^2$): each of them is like a handle that is allowed to intersect itself.

The model for such a self-intersection is suggested in figure 2.14: Start with $\mathbb{D}^2 \times \mathbb{D}^2$*, and choose two small disks* D' *and* D'' *inside* $\mathbb{D}^2$*. Then identify* $D' \times \mathbb{D}^2$ *with* $D'' \times \mathbb{D}^2$ *by flipping factors: namely, identify each slice* $D' \times p$ *from* $D' \times \mathbb{D}^2$ *with* $p \times \mathbb{D}^2$ *from* $D'' \times \mathbb{D}^2$*. This is called a* **self-plumbing.**[7]



2.14. Self-plumbing a handle

Away from such self-intersections, our "thickened immersed disk" looks very much like a 2–handle and is attached along its boundary region $\mathbb{S}^1 \times \mathbb{D}^2$. Of course, we must also choose a suitable[8] way to attach the thickening of $\mathbb{S}^1 \times 0$. We will not dwell on this here: have faith that there is a good way of attaching each of these thickened disks, so that what follows below will actually work.

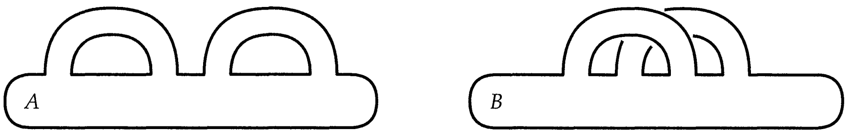
One reason behind this thickening procedure is that it creates more room to clean up. Think of the simple example in figure 2.15 on the next page—two distinct spaces that become identical after thickening. Another reason is, of course, trying to develop directly a handle theory better suited to the peculiarities of 4–manifolds.

We now follow the infinite procedure described before and grow a tower of such thickened disks, each disk attempting to eliminate a double-point

---

[7] This is a particular instance of a slightly more general procedure called *plumbing*, which will be described a bit later, on page 86.

[8] The framing obstruction, explained back in the end-notes of the preceding chapter (page 57) must vanish.



2.15. $A \not\simeq B$, but $A \times [0,1] \simeq B \times [0,1]$.

of the preceding generation, but itself adding new self-intersections, which the next generation will try to repair, and so on, to infinity.

We want to end up with an object which, away from its attaching part to ∂M, does not have a boundary. In other words, we want the end-result to be an open set in $\operatorname{Int} M$. The only part of boundary that we wish to keep is the "attaching part", where the whole construction anchors itself to the boundary of M. Thus, for each immersed thickened disk $\mathbb{D}^2 \times \mathbb{D}^2$ that we add to our tower, we immediately discard $\mathbb{D}^2 \times \mathbb{S}^1$ from its boundary (the other part of the boundary, $\mathbb{S}^1 \times \mathbb{D}^2$, is used up when attaching it either to ∂M or to the preceding generation of disks).

We carry out the above process for infinitely many steps, then take the union of all these thickened disks. The resulting monster is called a **Casson handle.**[9]

The miracle

What we did is that we pushed our problems away toward infinity. The miracle is that, when working with thickened disks as above, the procedure actually succeeds and our problems vanish.

First, A. Casson proved around 1973 that every Casson handle is proper homotopic, relative to its attaching boundary, to $\mathbb{D}^2 \times \mathbb{R}^2$ (think $\mathbb{R}^2 = \operatorname{Int} \mathbb{D}^2$). For example, one can notice that all Casson handles are simply-connected: any loop must be contained in the tower obtained after finitely-many stages; and attaching the next stage will kill it.

And then, in 1981, came the revolution:

Freedman's Theorem on Casson Handles. *Any Casson handle is homeomorphic to a thickened disk* $\mathbb{D}^2 \times \mathbb{R}^2$ *and thus is a genuine (open)* 2*–handle, having as core a genuine topologically embedded* 2*–disk.* □

This very hard technical result led (comparatively) pretty quickly to a complete classification of *topological* 4–manifolds, which we will present later.[10]

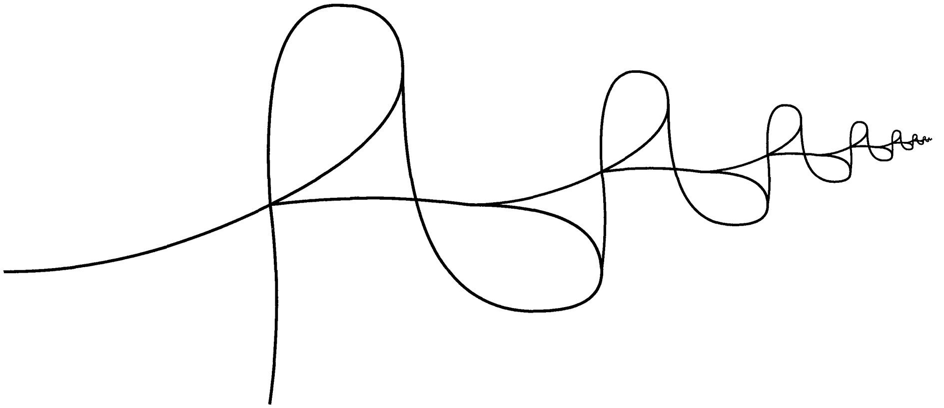
---

[9] "Handle" because if the initial disk D were embedded, then its thickening would be a genuine 2–handle; and a Casson handle tries to make up for the failure of D to provide a genuine handle. A. Casson himself called his creation *flexible handles* because they were easier to embed than actual handles; for a discussion of embedding Casson handles, look ahead at the the end-notes of this chapter (page 96).

[10] See section 5.2 (page 239) ahead.

Smooth sadness. Unfortunately, this type of wild botany is bound to fail in the smooth case.

The simplest Casson handle known to be exotic (*i.e.*, homeomorphic to $\mathbb{D}^2 \times \mathbb{R}^2$, but *not* diffeomorphic to it) is the one in figure 2.16 (with all self-intersections of the same sign).[11] There are infinitely-many non-diffeomorphic Casson handles.



2.16. The simplest exotic Casson handle.

Keep in mind that a Casson handle is exotic only *relative to its boundary:* if we discard the attaching boundary, we are left with a very standard open 4–ball.

2.2. The topological *h*–cobordism theorem

Putting the above result to work, we can now obtain topologically-embedded Whitney disks as easily as in high-dimensions and thus apply the Whitney trick to undo intersections. Therefore, the high-dimensional strategy for proving the *h*–cobordism theorem can be followed through and yields M. Freedman's 4–dimensional version of the theorem:

Topological 4–Dimensional h–Cobordism Theorem. *Let W^5 be an h–cobordism between M^4 and N^4, with everybody simply-connected. Then we have a homeomorphism $W \simeq M \times [0,1]$, and in particular M and N are homeomorphic.*

Note that the *smooth* version of the 4–dimensional *h*–cobordism theorem is false: W does not need be diffeomorphic to $M \times [0,1]$. The first counterexample was brought to light by S.K. Donaldson, and many others followed.

---

[11] If the self-intersections in figure 2.16 have varying signs, it is not known whether the resulting Casson handles are exotic. If all such were exotic, then it would follow that all Casson handles whatsoever are exotic.

Outline of proof. Before merely applying the h–cobordism program, we must first show that the 5–dimensional *topological* manifold W admits a topological handle decomposition.[12] This is far from trivial (so far we obtained handle decompositions from Morse functions, an inherently differentiable object) and was proved in **F. Quinn**'s ***Ends of maps. III. Dimensions* 4 *and* 5** **[Qui82]**.

Of course, if we start with smooth M and N, connected through a smooth W, then the point is moot. For our goals in this volume, since our focus is on smooth 4–manifolds, that is indeed quite enough. □

Smoothness and handle decompositions. *While we are talking about it, it is worth saying a few more words about handle decompositions. If the manifold is smooth, then a handle decomposition can be obtained from any Morse function, as was explained in the preceding chapter. If the manifold is piecewise-linear (i.e., nicely triangulated), then one can use polyhedral regular neighborhoods to obtain one. If the manifold is merely topological, we can seek topological handle decompositions (i.e., decompositions into handles that are attached by homeomorphisms). If the dimension of the manifold is* not *4, then it is known that it will always admit a topological handle decomposition.*[13] *In dimension 4 though, any such handle decomposition would have the handles attached* via *homeomorphisms of 3–manifolds, but these can always be deformed to diffeomorphisms, and thus exhibit our 4–manifold as a smooth manifold. Therefore, dimension 4 is the* only *dimension where the existence of a handle decomposition is equivalent to the existence of a smooth structure.*[14]

***Spheres in dimension* 4.** A corollary of the topological h–cobordism theorem is, as one might expect:

Topological* 4*–Dimensional Poincaré Conjecture. *If a topological 4–manifold Σ^4 is homotopy-equivalent to S^4, then Σ^4 is homeomorphic to S^4.*

Sketch of proof. One builds the cone on Σ, then argues that this cone is actually a topological 5–manifold, including at its vertex. After that,

---

[12] The original proof of **M. Freedman** in ***The topology of four-dimensional manifolds*** **[Fre82]** was more complicated: he took a 1–complex out of W, added a smooth structure on its complement (and thus obtained a handle decomposition), then used a non-compact version of the h–cobordism theorem due to L. Siebenmann.

[13] Indeed, if dim ≤ 3, everything is smooth; if dim ≥ 6, it was shown in **R. Kirby** and **L. Siebenmann**'s ***Foundational essays on topological manifolds, smoothings, and triangulations*** **[KS77]**; if dim $= 5$, it was done, as mentioned, in **F. Quinn**'s in ***Ends of maps. III. Dimensions* 4 *and* 5** **[Qui82]**. Compare with the end-notes of chapter 4 (smoothing of topological manifolds, page 221).

[14] This statement does not contradict what we said earlier: for the h–cobordism theorem one needs topological handle decompositions of 5–manifolds, and these always exist.

one cuts out a 5–ball from the cone, thus obtaining an h–cobordism between Σ and $\mathbb{S}^4$.

On the other hand, if Σ is assumed smooth, then one can instead use Wall's theorem on h–cobordisms (which we will prove later[15]) to get a smooth cobordism between Σ and $\mathbb{S}^4$. □

In contrast, the smooth version of the 4–dimensional Poincaré conjecture is very much wide open to this day, and there are not even methods in sight that one might hope would lead to a solution. We do not know whether there exist exotic 4–spheres; we are not even able to make an educated guess.

Gluck twists. *A series of possible counter-examples to the smooth Poincaré conjecture in dimension 4 is obtained from the following surgeries on $\mathbb{S}^4$: Let S be a 2–sphere embedded in $\mathbb{S}^4$. A tubular neighborhood of S is a copy of $S \times \mathbb{D}^2$. We cut this neighborhood out of $\mathbb{S}^4$ and then glue it back in by using a certain self-diffeomorphism of its boundary $S \times \mathbb{S}^1$.*

The only interesting automorphism of $\mathbb{S}^2 \times \mathbb{S}^1$ is the following spinning: send $(s, \vartheta) \in \mathbb{S}^2 \times \mathbb{S}^1$ to[16] $(e^{i\vartheta}s, \vartheta)$; that is, as one travels around the $\mathbb{S}^1$–factor, one simultaneously rotates the sphere-factor.[17]

Cutting the neighborhood $S \times \mathbb{D}^2$ out of $\mathbb{S}^4$ and gluing it back in by using the above automorphism is known as performing a **Gluck twist**[18] *on S. The result is proved to always be homotopy-equivalent to $\mathbb{S}^4$, and thus homeomorphic to $\mathbb{S}^4$. However, in the vast majority of cases it is unknown whether the result is smoothly $\mathbb{S}^4$ or is an exotic 4–sphere.*

The only 2–spheres in $\mathbb{S}^4$ on which it is worth performing Gluck twists are the knotted spheres. Indeed, if S bounds a 3–ball in $\mathbb{S}^4$, then the spinning of the Gluck twist can be extended across the bounded 3–ball to yield a diffeomorphism between the twisted 4–sphere and $\mathbb{S}^4$. Nonetheless, 2–spheres in $\mathbb{S}^4$ can exhibit knotting phenomena. (A sphere S in $\mathbb{S}^4$ is called **knotted** *if it does not bound an embedded 3–ball in $\mathbb{S}^4$.) Simple examples of knotted spheres can be built by spinning a 1–dimensional knotted thread in the fourth dimension, as in figure 2.17 on the next page. However, Gluck twists on this type of knotted spheres are known to never lead to exotic 4–spheres, since they also bound "spinnable" 3–submanifolds. Besides these simple cases, little is known.*

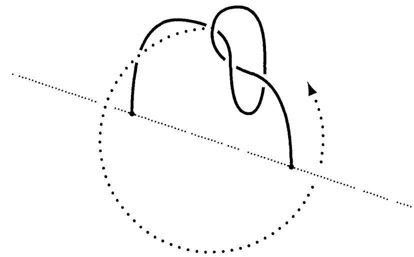
---

[15] See section 4.2 (page 155).

[16] Here, think of $\vartheta \in \mathbb{S}^1$ as being represented as an angle $\vartheta \in [0, 2\pi]$, and view $\mathbb{S}^2$ as the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Multiplying by $e^{i\vartheta}$ merely means rotating the sphere by an angle of ϑ around the axis 0-∞.

[17] The group of self-diffeomorphisms of $\mathbb{S}^2 \times \mathbb{S}^1$ up to isotopy is isomorphic to $\oplus 3\,\mathbb{Z}_2$, generated by the antipodal of $\mathbb{S}^2$, the reflection of $\mathbb{S}^1$, and the above spinning.

[18] This construction is due to **H. Gluck**'s ***The embedding of two-spheres in the four-sphere*** [**Glu62**].



2.17. Building a knotted sphere in $\mathbb{R}^4$

2.3. Homology 3–spheres bound fake 4–balls

Another fundamental result in the theory of topological 4–manifolds is the following:

Freedman's Theorem on Fake Balls. *Every* 3*–manifold* Σ^3 *with the same homology as the* 3*–sphere* $\mathbb{S}^3$ *must bound a contractible*[19] *topological* 4*–manifold, in other words, a fake* 4*–ball.*[20]

A 3–manifold Σ^3 has the same homology as $\mathbb{S}^3$ if and only if its only nontrivial homology groups are $H_0(\Sigma; \mathbb{Z}) = \mathbb{Z}$ and $H_3(\Sigma; \mathbb{Z}) = \mathbb{Z}$. If Σ^3 were also simply-connected, then it would be homotopy-equivalent to $\mathbb{S}^3$ and thus most likely homeomorphic to $\mathbb{S}^3$.

Sketch of proof. The construction of a contractible Δ^4 with $\partial \Delta = \Sigma$ proceeds as follows: we take the product

$$\Sigma \times [0, 1]$$

and we modify it by surgery until it becomes simply-connected, but without altering its boundary or its homology (such a procedure is known as the **plus-construction**). In other words, we modify $\Sigma \times [0, 1]$ to obtain not merely a homological copy of $\mathbb{S}^3 \times [0, 1]$, but a *homotopy* copy of $\mathbb{S}^3 \times [0, 1]$. The resulting 4–manifold $\mathcal{S}$ will have boundary $\overline{\Sigma} \cup \Sigma$ and be homotopy-equivalent to $\mathbb{S}^3$. The procedure for obtaining $\mathcal{S}$ is to add disks to $\Sigma \times [0, 1]$ in order to kill π_1, then kill the homology that we created by adding those disks. It is worth noting that the construction must make use of Casson handles. After building $\mathcal{S}$, we will stack end-to-end infinitely-many copies of this creature, add a point, and the result will be Δ.

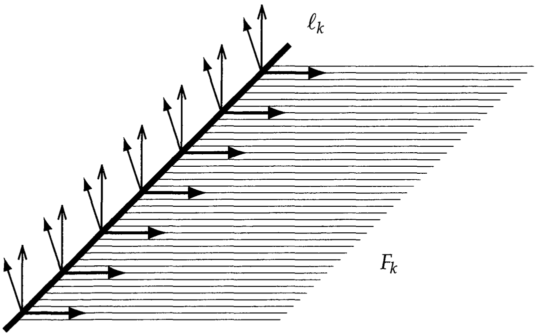
Make it simply-connected. Pick generators for $\pi_1(\Sigma \times [0, 1])$ and represent them as embedded disjoint circles $\ell_1, \ldots, \ell_n$. Since the homology group $H_1(\Sigma \times [0, 1]; \mathbb{Z})$ vanishes, these circles must bound embedded

---

[19] Remember that **contractible** means homotopy-equivalent to a point.

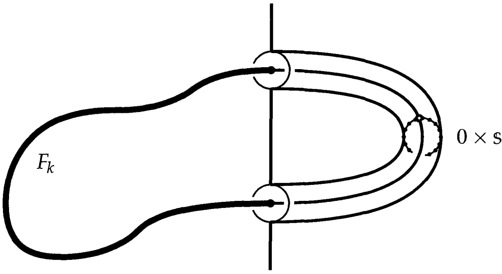
[20] A **fake 4–ball** is a 4–manifold homotopy-equivalent to a 4–ball $\mathbb{D}^4$; we certainly do *not* take this to include that its boundary be homotopy-equivalent to a sphere $\mathbb{S}^3$.

surfaces $F_1, \ldots, F_n$ in $\Sigma \times [0,1]$. Each such surface F_k induces a trivialization of the normal bundle of its boundary-circle ℓ_k (see figure 2.18) and thus prescribes an embedding (up to isotopy) of $S^1 \times D^3$ into $\Sigma \times [0,1]$ as a tubular neighborhood of ℓ_k. The boundary of $S^1 \times D^3$ is $S^1 \times S^2$, the same as the boundary of $D^2 \times S^2$. We can therefore cut this $S^1 \times D^3$ out of $\Sigma \times [0,1]$ and replace it with a copy[21] of $D^2 \times S^2$.



2.18. Framing from bounded surface

The class $[\ell_k]$ is now homotopically-trivial in the surgered manifold, since it bounds the newly added disk $D^2 \times 0$. Repeating this for all ℓ_k's, we end up with a simply-connected manifold. The problem is that in the process of killing π_1 we created some new 2–homology: each of our glued-in $D^2 \times S^2$'s brings in two new homology classes, one from $F_k \cup (D^2 \times 0)$, and one from $0 \times S^2$, as is suggested in figure 2.19. The class $[F_k \cup (D^2 \times 0)]$ has zero self-intersection (insured from the way we trivialized N_{ℓ_k}).



2.19. New homology classes created

---

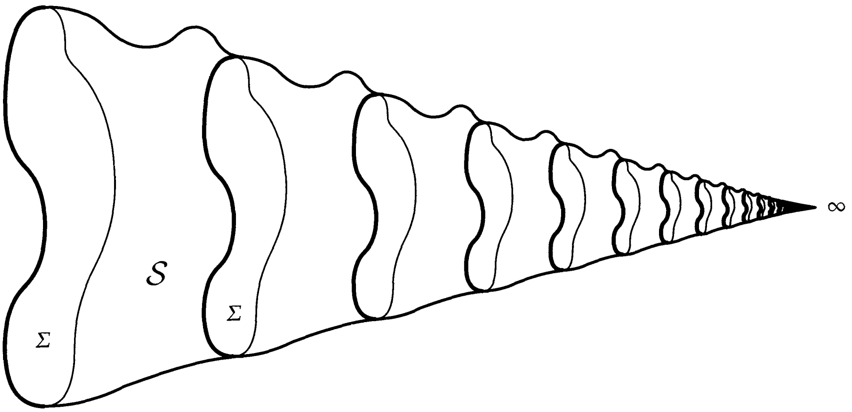
[21] An alternative way of seeing this modification is as follows: we are modifying $\Sigma \times [0,1]$ by a 5–dimensional cobordism; specifically, we are attaching a 5–dimensional 2–handle $D^2 \times D^3$ along ℓ_k to a thickening $\Sigma \times [0,1] \times [0,\varepsilon]$; this modifies the upper boundary by deleting the attaching region $S^1 \times D^3$ and replacing it with $D^2 \times S^2$.

Eliminate the 2-homology. Using Casson handles, one can show that each of the classes $[F_k \cup (\mathbb{D}^2 \times 0)]$ can be represented by topologically-embedded *spheres* S_k, all disjoint and whose complements are still simply-connected.[22]

Each sphere S_k has zero self-intersection, and so its normal bundle is trivial. Therefore each S_k admits an embedding of $\mathbb{S}^2 \times \mathbb{D}^2$ around it. However, $\mathbb{S}^2 \times \mathbb{D}^2$ has the same boundary as $\mathbb{D}^3 \times \mathbb{S}^1$, and thus we can cut the former out and glue the latter in.[23] Repeating this for all S_k's results in the destruction of all 2-homology. Since the complement of the S_k's was simply-connected, the result will still be simply-connected.

Therefore, we have finally obtained a simply-connected 4-manifold with the same homology and boundary as $\Sigma \times [0, 1]$. We will denote this creature by $\mathcal{S}$.

Stack'em. To build Δ, we attach one after the other countably-many copies of $\mathcal{S}$ and compactify the result by adding one point "at infinity", as in figure 2.20. Since each $\mathcal{S}$ is homotopy-equivalent to $\mathbb{S}^3$, so will be any finite stacking of $\mathcal{S}$'s, as it can be retracted into the right-most copy of $\mathcal{S}$. It follows that Δ deformation-retracts to its added point ∞, in other words, that Δ is *contractible.*



2.20. Building a contractible Δ bounded by Σ.

All we still need to argue is that Δ is in fact a manifold. The only region where this is an issue is, of course, around ∞. Somehow the

---

[22] Technically, to embed the spheres we must use Casson's embedding theorem stated in the endnotes of this chapter (page 96). It is important there that the classes $[F_k \cup (\mathbb{D}^2 \times 0)]$ and $[0 \times \mathbb{S}^2]$ have zero self-intersections, but cross each other once.

[23] Alternatively, this can be viewed as attaching a 5-dimensional 3-handle $\mathbb{D}^3 \times \mathbb{D}^2$, which deletes the attaching region $\mathbb{S}^2 \times \mathbb{D}^2$ and replaces it by $\mathbb{D}^3 \times \mathbb{S}^1$.

homotopy-equivalence of all the $\mathcal{S}'$s that accumulate at ∞ with S^3, together with local simply-connectedness and local contractibility at ∞, make a neighborhood of the ∞–point look like a cone on S^3 (instead of, say, a cone on the homology 3–sphere Σ). We have thus built a topological 4–manifold, contractible and bounded by Σ. □

The Poincaré homology 3–sphere and the E_8–manifold

A closed manifold with the same homology as a sphere is called a **homology sphere.** A closed 3–manifold Σ is a homology sphere if and only if its first homology $H_1(\Sigma;\mathbb{Z})$ vanishes. Since $H_1 = \pi_1/[\pi_1, \pi_1]$, this ultimately depends on the fundamental group of Σ: a 3–manifold Σ is a homology sphere if and only if its fundamental group $\pi_1(\Sigma)$ is a perfect group.

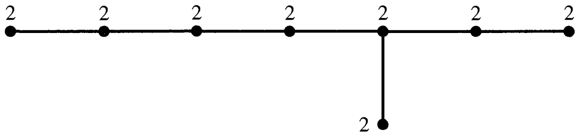
In what follows, we will build an example of such a creature Σ_P, called the *Poincaré homology sphere.*

Historically, H. Poincaré first conjectured that every homology 3–sphere must be S^3; then he discovered this nontrivial homology 3–sphere Σ_P, and thus had to strengthen his conjecture by requiring simple-connectedness. In any case, the homology 3–sphere Σ_P upon which Poincaré stumbled turned out to be ubiquitous in low-dimensional topology and proudly carries his name.

The Poincaré 3–sphere will appear as the boundary of a 4–manifold. Afterwards, we will cap this 4–manifold with one of Freedman's contractible Δ's to obtain a special *closed* 4–manifold, called the *E_8–manifold.*

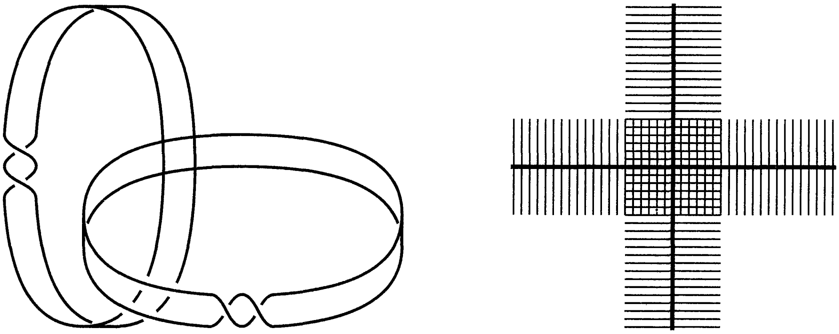
Plumbing. To build Σ_P, start with eight copies of S^2. Build on each a disk bundle with Euler class $+2$, for example the unit-disk bundle $\mathbb{D}T_{S^2}$ of its tangent bundle. We now have eight 4–manifolds with boundary, each containing a sphere of self-intersection $+2$.

We "plumb" these according to the E_8 Dynkin diagram from figure 2.21. Each dot in the diagram stands for one of our eight disk bundles (labelled by the self-intersection of its zero-section), while each edge stands for a plumbing connection, as we explain next.



2.21. The E_8 diagram

A **plumbing** is obtained through the identification suggested in figure 2.22 on the facing page: We pick a small disk D' of center p' in one sphere and a disk D'' of center p'' in another. Then locally the corresponding disk bundles over them look respectively like $D' \times \mathbb{D}^2$ and $D'' \times \mathbb{D}^2$. We prefer to write the latter as $\mathbb{D}^2 \times D''$ because we identify it with $D' \times \mathbb{D}^2$ factor-by-factor as just re-written: fiber-factor with basis-factor and vice-versa (*i.e.*, $D' \times 0$ is sent to $\mathbb{D}^2 \times p''$, and $p' \times \mathbb{D}^2$ to $0 \times D''$, *etc.*). Further, when plumbing we take care of orientations so that the intersection of D' and D'' is positive.



2.22. Plumbing

After plumbing our eight disk-bundles following the recipe from the E_8–diagram (and after rounding corners), we obtain a smooth 4–manifold that we denote by

$$P_{E_8} .$$

It has non-empty boundary and is called the 4–dimensional **E_8–plumbing.**

Homology. The manifold P_{E_8} contains eight spheres, each with self-intersection $+2$ and intersecting the other spheres either 0 or $+1$. All this intersection data is gathered in the 8×8 matrix

$$E_8 = \begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix} ,$$

which is called the **intersection form**[24] of P_{E_8}. One can compute its determinant to be $\det E_8 = +1$, and thus deduce that the E_8–matrix is invertible over $\mathbb{Z}$.

---

[24] Intersection forms will be discussed in detail and generality in the next chapter, which starts on page 106.

Since our eight spheres generate $H_2(P_{E_8};\mathbb{Z})$, it follows that the E_8–matrix in fact governs the whole 2–homology of P_{E_8}. The invertibility of the intersection form can then be used to show that the boundary of P_{E_8} has trivial 1–homology, and thus must in fact be a homology 3–sphere.[25] (As an alternative, one could of course figure out directly that the fundamental group of ∂P_{E_8} is a perfect group.)

This boundary 3–manifold, which we denote by

$$\Sigma_P = \partial P_{E_8} \, ,$$

is the **Poincaré homology sphere** that we set out to build.

> ***Alternatives.*** *Another possible description of P_{E_8} is as follows: Consider the open 4–manifold Q described by the equation $z_1^5 + z_2^3 + z_3^2 = \varepsilon$ in $\mathbb{C}^3$. It is diffeomorphic*[26] *to the interior* $\operatorname{Int} P_{E_8}$ *of P_{E_8}. If we take a small ball $\mathbb{D}^6$ in $\mathbb{C}^3$, then $\mathbb{D}^6 \cap Q \cong P_{E_8}$, with boundary $\mathbb{S}^5 \cap Q \cong \Sigma_P$.*
>
> *A different description of Σ_P is as the quotient $\Sigma_P = SO(3)/A_5$, where A_5 is the group of the sixty symmetries of the dodecahedron. In fact, Σ_P can be obtained directly by identifying with a twist the opposite faces of a dodecahedron.*[27] *Thus, Σ_P is sometimes called the* **dodecahedral space.**
>
> *We should also mention that P_{E_8} is just the lowest-dimensional version of a series of E_8–plumbings, whose boundaries are exotic $(4k-1)$–spheres.*[28]

Many more homology 3–spheres can be built as boundaries of plumbings.[29]

The E_8–manifold. Reversing orientation, $\overline{\Sigma}_P$ is still a homology sphere. By Freedman's result above, $\overline{\Sigma}_P$ must bound some contractible topological 4–manifold, a fake 4–ball Δ. Then we can glue[30] P_{E_8} and Δ along their common boundary Σ_P, and thus obtain a closed manifold

$$\mathcal{M}_{E_8} = P_{E_8} \cup_{\Sigma_P} \Delta \, .$$

---

[25] Such an argument works in general. For every compact 4–manifold M, its boundary ∂M is a homology sphere if and only if the intersection form $Q_M \colon H^2(M;\mathbb{Z}) \times H^2(M;\mathbb{Z}) \to \mathbb{Z}$ is unimodular (= has invertible matrix); the intersection form being unimodular establishes an isomorphism $H^2(M;\mathbb{Z}) \approx H^2(M,\partial M;\mathbb{Z})$, which forces the homology of ∂M to be trivial in dimensions 1 and 2. The full argument will be presented in the end-notes of chapter 5 (page 261).

[26] For a more general discussion of complex singularities (like $z_1^5 + z_2^3 + z_3^2 = 0$) and their topology, see **J. Milnor**'s ***Singular points of complex hypersurfaces*** **[Mil68]**. See also a few comments on page 318.

[27] This, in fact, is the original construction of H. Poincaré. For several descriptions of Σ_P and elegant proofs of their equivalence, read **R. Kirby** and **M. Scharleman**'s ***Eight faces of the Poincaré homology 3–sphere*** **[KS79]**.

[28] See the end-notes of this chapter (page 97).

[29] See the end-notes of chapter 5 (unimodular forms and homology spheres, page 261).

[30] We must take $\partial\Delta$ to be $\overline{\Sigma}_P$, with opposite orientation to Σ_P, because all boundary-gluings must be made through orientation-reversing homeomorphisms, as was explained on page 13.

This simply-connected topological 4–manifold is known as the **E_8–manifold.** The striking fact is that $\mathcal{M}_{E_8}$ *cannot* admit any smooth structures, as we will argue later.[31]

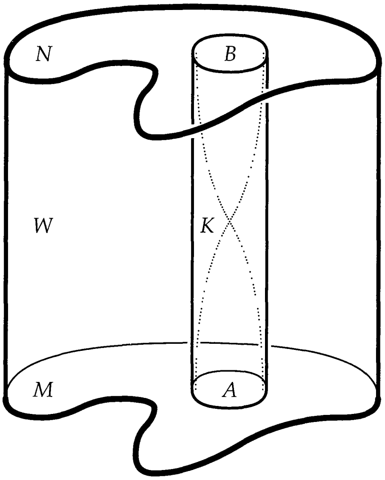
2.4. Smooth failure: the twisted cork

In contrast with the case of topological 4–manifolds, the *smooth* case of the 4–dimensional h–cobordism theorem fails. Some insight on where this failure occurs is provided by the following rather startling result:

Theorem (*C. Curtis, M. Freedman, W. Hsiang and R. Stong*). *Let W^5 be any smooth h–cobordism between M^4 and N^4. Then inside W there exists a compact contractible sub–h–cobordism (with non-empty boundary) K^5, between some compact contractible submanifolds $A^4 \subset M^4$ and $B^4 \subset N^4$, such that W is a trivial cobordism outside K. That is, we have a diffeomorphism*

$$W \setminus \operatorname{Int} K \;\cong\; (M \setminus \operatorname{Int} A) \times [0,1]\,.$$

Further, K can be chosen so that it is diffeomorphic to the 5–ball $\mathbb{D}^5$, that $W \setminus K$ is simply-connected, and furthermore so that A and B are diffeomorphic through a diffeomorphism that, when restricted to the boundary $\partial A = \partial B$, is an involution.[32] □



2.23. Failure of the 4–dimensional h–cobordism theorem

In other words, if M and N are h–cobordant, then N can be obtained from M by cutting out a compact contractible submanifold A and gluing it back in by using an involution of ∂A. The h–cobordism K somehow connects

---

[31] See section 4.4 (page 170).

[32] An **involution** of Z is any map $f\colon Z \to Z$ so that $f \circ f = id$.

A and its "reversed" version B. The fake 4–ball A is called an **Akbulut cork.** Several of these can be described quite explicitly in terms of handle diagrams.

Even more, each Akbulut cork is surrounded by an exotic $\mathbb{R}^4$. We will revisit this result later[33] during our discussion of exotic $\mathbb{R}^4$'s.

For us, the main point of this theorem is that smooth 4–manifolds (and the gap between the smooth and topological realm) are wrought with much subtlety, of which today we have a rather poor understanding.

---

[33] In section 5.4 (page 253).

2.5. Notes

Note: Kirby calculus

One method of describing smooth 4–manifolds is through their handle decompositions. This has been transformed into a calculus with link diagrams which is called **Kirby calculus.** An extensive discussion can be found in **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** [GS99]. Here we will just sketch its rudiments.

Handles. A 0–handle is simply a 4–ball $\mathbb{D}^4$, with boundary a 3–sphere $\mathbb{S}^3$. If we deal with connected closed manifolds, then a single 0–handle is all we need.[1] To this single 0–handle are attached 1–, 2–, 3– and 4–handles. Since the relevant attachments are done mostly to the boundary $\mathbb{S}^3 = \mathbb{R}^3 \cup \{\infty\}$ of the 0–handle, we can picture these by diagrams in $\mathbb{R}^3$.

A *1–handle* is a copy of $[-1, +1] \times \mathbb{D}^3$, to be attached to the 0–handle by gluing $\{-1\} \times \mathbb{D}^3$ and $\{+1\} \times \mathbb{D}^3$ to $\mathbb{S}^3$ (the "feet" of the 1–handle). If we are building an oriented manifold, then once we specify two 3–balls in $\mathbb{S}^3$ there is a unique way (up to isotopy) to glue a 1–handle to them (without ruining orientability). Thus, we can picture a 1–handle as in figure 2.24.

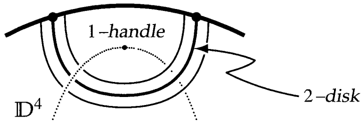
2.24. Diagram for attaching a 1–handle

Alternative. The same result as attaching a 1–handle can be obtained by removing a neighborhood of a disk. Namely, draw an unknotted circle in $\mathbb{S}^3 = \partial\,\mathbb{D}^4$ and imagine it bounding a disk; push the interior of that disk inside the 4–ball, then remove a neighborhood of it. See figure 2.25 on the following page.

That the result is the same as adding a 1–handle can be argued as follows: If to a genuine 1–handle we add a canceling 2–handle, then the whole thing can be collapsed back to the 4–ball;

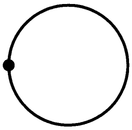
---

[1] This is a handle decomposition of a closed manifold, not of a cobordism. Thus, one needs *one* 0–handle, unlike in the case of a cobordism. In the cobordism case, all 0–handles were eliminated by absorbing them into the lower boundary. Here, a similar argument shows that all 0–handles can be absorbed into a single 0–handle. However, one needs at least one minimum for the Morse function, and thus one 0–handle must remain.



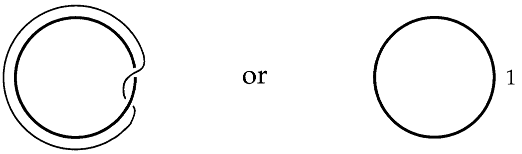
2.25. Removing a 2–handle, same as adding a 1–handle

not adding the 2-handle is then equivalent to removing the collapsed 2-handle from the 4-ball. That is, adding a 1-handle is the same as removing the 2-handle that cancels it, which is exactly the disk mentioned above. A consequence is that we can now represent a 1-handle also by specifying the circle whose bounded disk is to be removed, as in figure 2.26. It is marked with a dot to distinguish it from later objects.



2.26. Alternative diagram for attaching a 1–handle

A *2-handle* is a copy of $\mathbb{D}^2 \times \mathbb{D}^2$, attached by gluing $S^1 \times \mathbb{D}^2$ to S^3. To specify such an attachment, a first thing to describe is, of course, the image of $S^1 \times 0$, specifically an embedded knot K in S^3. Second, we need to provide instructions for the way the "thickening" is to be glued around this knot K. This is done by specifying a **framing**[2] of the knot, that is to say, a trivialization of the normal bundle N_{K/S^3} of K, drawn as a parallel curve that might twist around K, and which determines how $S^1 \times \mathbb{D}^2$ is to be identified with N_{K/S^3}, by showing where, say, $S^1 \times 1$ is to be attached, as on the left of figure 2.27.



2.27. Diagram for attaching a 2–handle: framed knot

In fact, the framing is completely described (up to isotopy) by an integer, for example the linking number of K with its framing curve, or, equivalently, the intersection number of the framing curve with a Seifert surface[3] for K. Thus, one can specify the attachment of a 2–handle by drawing a knot and labeling it with an integer, as on the right of figure 2.27.

---

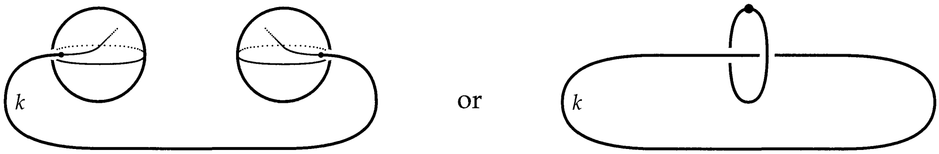
[2] Framings of knots are discussed in more detail in section 4.1 (page 147).

[3] A Seifert surface for a knot K is any oriented surface in S^3 that is bounded by K. It can be used to define the linking number of two knots K and K' as the intersection number of K with a Seifert surface of K'. See page 147 for a better discussion.

It remains to attach the 3*–handles* and the 4*–handles.* If we are dealing with closed 4–manifolds, then a single 4–handle is enough.[4] It is not hard to see that the union of this 4–handle with the 3–handles is a copy of either $\mathbb{D}^4$ (when there are no 3–handles) or $\#k\,\mathbb{S}^1 \times \mathbb{D}^3$ (when there are k 3–handles). This means that, if we try to build a closed 4–manifold, then, after attaching all our 1– and 2–handles to the 0–handle, we better end up with a 4–manifold with boundary either $\mathbb{S}^3$ or $\#k\,\mathbb{S}^1 \times \mathbb{S}^2$.

If the boundary does turn out to be as needed, then it is known that at this stage it does not matter how the 3– and 4–handles are attached: the result will be the same closed 4–manifold.[5] Therefore, to describe a closed 4–manifold, it is enough to describe the attachment of the 1– and 2–handles. (Keep in mind that a random Kirby diagram in general will not describe any closed 4–manifold: to be able to close it up, one needs that the resulting boundary be $\mathbb{S}^3$ or $\#k\,\mathbb{S}^1 \times \mathbb{S}^2$.)

Handle moves. A pair of canceling 1– and 2–handles can be represented as in figure 2.28. Canceling the pair means deleting the couple from the diagram, while creating the pair means adding such a couple.



2.28. Diagrams for a pair of canceling 1– and 2–handles

Sliding a 2–handle over another 2–handle is achieved by band-summing the attaching knots (and taking into account the framing of the handle over which we slide), as in the simple example from figure 2.29 on the next page. There are many ways to execute such a slide, and an alternative can be seen in figure 2.30 on the following page. If h_α is the 2–handle being slid over h_β, then the slide in 2.29 changes ∂ the same way as changing basis in $H_2(M;\mathbb{Z})$ from h_α to $h_\alpha + h_\beta$, while the slide in 2.30 changes h_α to $h_\alpha - h_\beta$.

For determining the framing of the slid handle, one should use the formula:

$$\begin{aligned}\text{framing}(\text{new } h_\alpha) &= (\alpha \pm \beta)\cdot(\alpha \pm \beta) = \alpha\cdot\alpha + \beta\cdot\beta \pm 2\alpha\cdot\beta \\ &= \text{framing}(h_\alpha) + \text{framing}(h_\beta) \pm 2\,\text{linking}(\text{knot } h_\alpha,\ \text{knot } h_\beta)\ .\end{aligned}$$

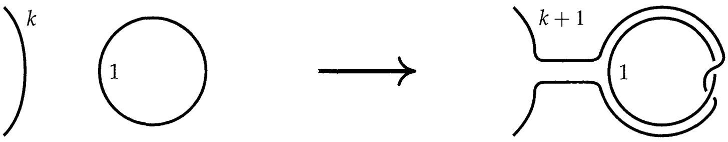
(In our pictures, the linking was zero.)

It is worth noting that, in the absence of 1– and 3–handles, each 2–handle represents a generator of $H_2(M;\mathbb{Z})$, and that the intersection form of M (see the next chapter) is exactly the linking matrix of the diagram (with self-linking given by the framing). Compare with the discussion of Whitehead's theorem, in section 4.1

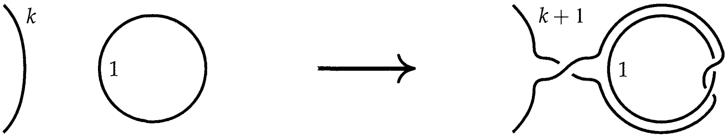
---

[4] "Single 4–handle is enough" is the upside-down version of "single 0–handle is enough".

[5] This follows from the fact that every self-diffeomorphism of $\#k\,\mathbb{S}^1 \times \mathbb{S}^2$ extends to a self-diffeomorphism of $\#k\,\mathbb{S}^1 \times \mathbb{D}^3$, as was proved by **F. Laudenbach** and **V. Poénaru** in ***A note on 4–dimensional handlebodies*** [**LP72**].



2.29. Diagram for sliding a 2–handle over another, I; $h_\alpha \mapsto h_\alpha + h_\beta$

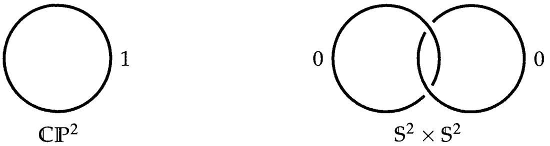


2.30. Diagram for sliding a 2–handle over another, II; $h_\alpha \mapsto h_\alpha - h_\beta$

(page 140). It is unknown whether every simply-connected 4–manifold admits handle decompositions without 1– or 3–handles.

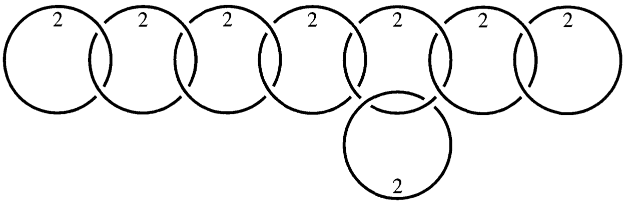
An example of an actual Kirby calculus computation will appear in figure 4.10 on page 151, where it will be proved that $\mathbb{CP}^2 \# S^2 \times S^2 \cong \mathbb{CP}^2 \# S^2 \tilde{\times} S^2$.

Examples, and* 3*–manifolds. As examples, the diagrams in figure 2.31 represent $\mathbb{CP}^2$ and $S^2 \times S^2$. Since the connected sum of two 4–manifolds is represented by simply putting together (untangled) the Kirby diagrams of the two summands, this whole figure can be thought of as the diagram of $\mathbb{CP}^2 \# S^2 \times S^2$.



2.31. The Kirby diagrams of $\mathbb{CP}^2$ and $S^2 \times S^2$

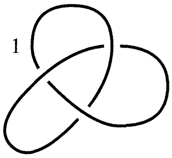
The diagram in figure 2.32 represents the P_{E_8} plumbing of eight spheres. It is a 4–manifold with boundary, and its boundary is the Poincaré homology 3–sphere Σ_P. While it could be closed-up to the closed topological manifold $\mathcal{M}_{E_8}$ by using Freedman's mysterious fake ball Δ, such a capping cannot be done smoothly; thus, the diagram does not represent any smooth closed 4–manifold.



2.32. The Kirby diagram of P_{E_8} (or of Σ_P)

Nonetheless, this last example points us in another direction for using Kirby calculus: Since every 3–manifold is the boundary of some 4–manifold, we can use a diagram of such a 4–manifold to describe its boundary 3–manifold (just as the diagram of P_{E_8} above can be thought of as describing Σ_P). Further, all 3–manifolds are boundaries of 4–manifolds without 1– or 3–handles. (This follows from an intrinsically 3–dimensional interpretation of a Kirby diagram as describing a sequence of *Dehn surgeries* on $\mathbb{S}^3$, that is, removals of solid tori—around the components of the diagram—from $\mathbb{S}^3$ and gluing them back using a twist of the boundary torus specified by the respective framing coefficient.)

Of course, many 4–manifolds have the same boundary 3–manifold. For example, the diagram in figure 2.33 represents a simpler 4–manifold whose boundary is also the Poincaré homology 3–sphere.



2.33. Another diagram for Σ_P

Observe that the 4–manifold in figure 2.33 can also be closed-up using one of Freedman's fake 4–balls. This yields a closed topological manifold that is homotopy-equivalent to $\mathbb{CP}^2$ but not homeomorphic to it; the result is denoted by $*\,\mathbb{CP}^2$ and admits no smooth structures either. We will encounter it again in section 5.2 (page 241).

While two diagrams that represent a same 4–manifold must be related by the handle moves discussed above, if we want merely to preserve the boundary 3–manifold, then we can also allow blow-ups and blow-downs (complex and anti-complex). This means that we are allowed to connect sum to our 4–manifold copies of $\mathbb{CP}^2$'s and $\overline{\mathbb{CP}}^2$'s or to split off such copies; indeed, this does not change the boundaries. Diagrammatically, this simply means adding or removing an unknotted circle with framing ± 1 that is separated from the rest of the diagram.

In other words, the set of all 3–manifolds coincides with the set of all Kirby diagrams (with only 2–handles) modulo handle moves and blow-ups/blow-downs. This point of view has led to the so-called *quantum invariants* for 3–manifolds obtained from suitable invariants of knots and links.

References. Kirby calculus (both 3– and 4–dimensional) found its origin in **R. Kirby**'s ***A calculus for framed links in*** $\mathbb{S}^3$ **[Kir78]** and was then used extensively by R. Kirby, S. Akbulut, and R. Gompf, as well as many others. Kirby calculus, along with Casson's ideas, was used by **M. Freedman** when he proved his results in ***The topology of four-dimensional manifolds*** **[Fre82]** (later, in the book ***Topology of 4–manifolds*** **[FQ90]**, Casson handles and Kirby calculus were dropped in favor of gropes).

For a fuller discussion and applications of Kirby calculus, see **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89]**, and, for diagrams of complex surfaces, **J. Harer**, **A. Kas** and **R. Kirby**'s ***Handlebody decompositions of complex surfaces*** **[HKK86]**. Most of the material from the latter is discussed in **R. Gompf** and **A. Stipsicz**'s

4–Manifolds and Kirby calculus [GS99], which is currently the best reference for Kirby calculus.

That all 3–manifolds appear as results of Dehn surgeries on $\mathbb{S}^3$ was proved by **A. Wallace**'s ***Modifications and cobounding manifolds*** **[Wal60]** and, with an intrinsic 3–dimensional argument, by **R. Lickorish**'s ***A representation of orientable combinatorial 3–manifolds*** [Lic62b]. A simplification of the needed set of diagram-modifications for the 3–dimensional case appeared in **R. Fenn** and **C. Rourke**'s ***On Kirby's calculus of links*** [FR79]. A beautiful intrinsic 3–dimensional discussion of diagrams and surgeries, including a nice proof of Lickorish's result, is contained in **D. Rolfsen**'s classic ***Knots and links*** **[Rol76, Rol90, Rol03]**.

The quantum invariants for 3–manifolds have their origin in **E. Witten**'s ***Quantum field theory and the Jones polynomial*** [Wit89], with a heavily-algebraized but mathematically-sound version appearing in **N. Reshetikhin** and **V. Turaev**'s ***Invariants of 3–manifolds via link polynomials and quantum groups*** **[RT91]**. A simplification and nice presentation is **R. Kirby** and **P. Melvin**'s ***The 3–manifold invariants of Witten and Reshetikhin–Turaev for*** $\mathfrak{sl}(2,\mathbb{C})$ **[KM91]**. The literature on this topic is ever-growing, and the topic is certainly closer to knot theory and non-commutative algebra than to 4–manifolds.

Note: Embedding Casson handles

Casson handles are quite useful because they are easy to embed (and A. Casson himself named them "flexible handles"). It is thus worth stating Casson's main result:

Casson's Embedding Theorem. *Let M be a simply-connected 4–manifold, with non-empty boundary. Let $f_1, \dots, f_n$ be immersions $\mathbb{D}^2 \to M$ such that $f_k|_{\partial \mathbb{D}^2}$ are disjoint embeddings into ∂M.*

Assume that, when $i \neq j$, we have intersection numbers[6] *$f_i \cdot f_j = 0$. Further, assume there are classes $\alpha_1, \dots, \alpha_n \in H_2(M; \mathbb{Z})$ such that all $\alpha_k \cdot f_k = 1$ but $\alpha_i \cdot f_j = 0$ if $i \neq j$, and so that all self-intersections $\alpha_k \cdot \alpha_k$ are even.*

Then there must exist disjoint open sets (Casson handles) $C_1, \dots, C_n$ such that:

- *we have proper*[7] *homotopy equivalences $\left(C_k,\ C_k \cap \partial M\right) \sim \left(\mathbb{D}^2 \times \mathbb{R}^2,\ \partial \mathbb{D}^2 \times \mathbb{R}^2\right)$;*
- *$C_k \cap \partial M$ are open tubular neighborhoods of the circles $f_k\left[\partial \mathbb{D}^2\right]$ in ∂M;*
- *f_k is homotopic, relative to its boundary $\mathbb{S}^1$, to a map into C_k.* □

The role of the classes α_k is to help untangle various constructions along the way, somewhat similar to the use of transverse spheres in section 4.2 (page 149) ahead.

---

[6] These intersection numbers are well-defined when $i \neq j$: keep the disjoint boundaries $f_i[\partial \mathbb{D}^2]$ and $f_j[\partial \mathbb{D}^2]$ fixed, and perturb the interiors of the immersed disks until they become transverse; then, count.

[7] A map $f \colon A \to B$ is called **proper** if $f^{-1}[\mathit{compact}]$ is compact. If A and B are compact, the condition is automatically satisfied. A homotopy $f_t \colon A \to B$ is called proper if the map $f_\bullet \colon A \times [0,1] \to B$ is proper, or if all f_t's are proper.

Furthermore, we in fact have homeomorphisms

$$\left(C_k,\ C_k \cap \partial M\right) \ \simeq\ \left(\mathbb{D}^2 \times \mathbb{R}^2,\ \partial\mathbb{D}^2 \times \mathbb{R}^2\right),$$

as M. Freedman later proved.

A typical use of the embedding theorem is as follows: one starts with a closed 4–manifold M, removes a small 4–ball, then tries to embed Casson handles in the remaining 4–manifold with boundary. Topologically, the set made of the 4–ball together with the Casson handles is homeomorphic to the result of adding standard 2–handles to the 4–ball. If the Casson handles managed to exhaust all the homology of M, then the leftovers (M without the 4–ball and all Casson handles) must be somewhat simple. Thus appears a topological decomposition of M.

As an example, the embedding theorem leads to:

Corollary. *Let M be any simply-connected closed 4–manifold. For every 2–class β from $H_2(M;\mathbb{Z})$ with $\beta \cdot \beta = 0$ and for which we can find a class α with $\alpha \cdot \beta = 1$ and $\alpha \cdot \alpha$ even, there exists a sphere topologically-embedded in M that represents β.* □

This has already been used in section 2.3 (page 83) to build Freedman's Δ's.

Corollary. *Let M be a simply-connected closed 4–manifold. Assume there are classes α, $\overline{\alpha} \in H_2(M;\mathbb{Z})$ such that $\alpha \cdot \alpha = 0$ and $\overline{\alpha} \cdot \overline{\alpha} = 0$ but $\alpha \cdot \overline{\alpha} = 1$. Then there is an open set U in M such that U is homeomorphic to the complement of a topological 4–ball inside $\mathbb{S}^2 \times \mathbb{S}^2$, and so that α and $\overline{\alpha}$ belong to the image of $H_2(U;\mathbb{Z})$ into $H_2(M;\mathbb{Z})$. Moreover, α and $\overline{\alpha}$ can be realized as embedded topological spheres inside U, with only one crossing.* □

The above two constructions, and the conflict between their topological success but smooth failure, will be used in section 5.4 (page 250) to exhibit exotic $\mathbb{R}^4$'s.

Notes **[Cas86**, lecture I**]** from **A. Casson**'s 1974 lecture (where the above ideas were presented and proved) can be found in the volume ***À la recherche de la topologie perdue*** **[GM86a]**, edited by **L. Guillou** and **A. Marin.**

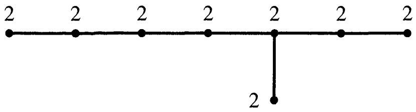
Note: Milnor plumbing and high-dimensional exotic spheres

The plumbing procedure we have described for building the Poincaré homology 3–sphere has higher-dimensional analogues. There, instead of yielding a homology sphere, it actually yields smooth manifolds homeomorphic to $\mathbb{S}^m$ but not diffeomorphic to it. In other words, it exhibits *exotic spheres.* Further, the set of all high-dimensional exotic spheres can be organized as a group Θ_m that can be explicitly determined. These groups play an essential role in the theory of smooth structures on high-dimensional topological manifolds and will be encountered again in the end-notes of chapter 4 (page 207), where this topic is discussed.

Plumbing along E_8. Start by taking eight copies of $\mathbb{S}^{2k}$ and consider the unit-disk bundles $\mathbb{D}T_{\mathbb{S}^{2k}}$ of their tangent bundles. Plumb these according to the E_8–diagram. The result will be a $4k$–manifold $P^{4k}_{E_8}$, called a *(Milnor)* **E_8–plumbing.** Its boundary

$$\Sigma_P^{4k-1} = \partial P^{4k}_{E_8}$$

is a homology sphere,[8] but when $k > 1$ it is also simply-connected. Therefore Σ_P^{4k-1} is a homotopy $(4k-1)$–sphere, and, by using the generalized Poincaré conjecture,[9] it follows that it must be homeomorphic to $\mathbb{S}^{4k-1}$.



2.34. The E_8 diagram, again

Assume that Σ_P^{4k-1} is diffeomorphic to $\mathbb{S}^{4k-1}$. Then we could *smoothly* glue $\mathbb{D}^{4k}$ to $P_{E_8}^{4k}$ to obtain a smooth closed $4k$–manifold of signature 8, but that is prohibited in all dimensions[10] $4k$. Therefore Σ_P^{4k-1} must be smoothly distinct from the standard sphere; it is an *exotic* $(4k-1)$*–sphere.*

Since Σ_P^{4k-1} is homeomorphic to $\mathbb{S}^{4k-1}$, we can still *topologically* attach $\mathbb{D}^{4k}$ to $P_{E_8}^{4k}$ and obtain a topological $4k$–manifold[11]

$$\mathcal{M}_{E_8}^{4k} = P_{E_8}^{4k} \cup_{\Sigma_P} \mathbb{D}^{4k} .$$

This manifold does not admit *any* smooth structures.

Plumbing along A_m and Kervaire spheres. *An even simpler exotic sphere can be obtained in dimensions $4k+1$ as follows: Start with two copies of the disk bundle $\mathbb{D}T_{\mathbb{S}^{2k+1}}$ of the tangent bundle to $\mathbb{S}^{2k+1}$ and plumb them exactly once. The resulting $(4k+2)$–manifold $P_{A_2}^{4k+2}$ has as boundary a smooth $(4k+1)$–manifold that can be proved to be homeomorphic to a sphere. It is called the* **Kervaire sphere** *and will be denoted by*

$$\Sigma_K^{4k+1} .$$

When k is even, *the $(4k+1)$–dimensional sphere Σ_K^{4k+1} is never diffeomorphic to $\mathbb{S}^{4k+1}$; it is an exotic sphere. When k is odd, Σ_K^{4k+1} is sometimes a standard sphere (for example for $k=1$).*

More generally, one can plumb m copies of $\mathbb{D}T_{\mathbb{S}^{2k+1}}$ following the simple A_m diagram in figure 2.35 on the facing page, obtaining the manifold $P_{A_m}^{4k+2}$. Its boundary $\partial P_{A_m}^{4k+2}$ is not always homeomorphic to a sphere. On one hand, when $m+1 = 0 \pmod 8$, it is known that $\partial P_{A_m}^{4k+2}$ is diffeomorphic to $\mathbb{S}^{4k+1}$, while, on the other hand, when $m+1 = 3$ or $5 \pmod 8$, it is proved that $\partial P_{A_m}^{4k+2}$ is diffeomorphic to the Kervaire sphere Σ_K^{4k+1}.

The first Kervaire sphere to appear was Σ_K^9, in **M. Kervaire***'s* ***A manifold which does not admit any differentiable structure*** *[***Ker60***], where it was used in building a non-smoothable 10–manifold (the first non-smoothable manifold ever created).*

---

[8] Compare with the proof in the end-notes of chapter 5 (page 261). See also **W. Browder**'s ***Surgery on simply-connected manifolds*** [**Bro72**, ch V].

[9] Stated back on page 30.

[10] The reason for this prohibition is similar to the one behind the non-smoothability of Freedman's E_8–manifold, $\mathcal{M}_{E_8}$. Intersection forms can be defined for all $4k$–manifolds. A higher-dimensional analogue of Rokhlin's theorem, uncovered in **M. Kervaire** and **J. Milnor**'s ***Bernoulli numbers, homotopy groups, and a theorem of Rohlin*** [**KM60**], excludes the existence of smooth $4k$–manifolds with intersection forms of signature 8, and in particular excludes the E_8–form and prohibits $\mathcal{M}_{E_8}^{4k}$ from supporting any smooth structures.

[11] In fact, $\mathcal{M}_{E_8}^{4k}$ is a PL manifold; compare with the end-notes of chapter 4 (smoothing topological manifolds, page 207), especially page 220.

2.35. The A_m diagram

The group of exotic spheres. High-dimensional exotic spheres can be organized as a group. For any dimension $m \geq 5$, start with the set

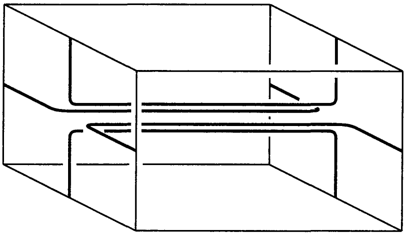
$$\Theta_m$$

of all smooth m–manifolds Σ^m that are homotopy-equivalent to S^m, considered up to smooth h–cobordisms. Since $m \geq 5$, the generalized Poincaré conjecture implies all these manifolds are in fact homeomorphic to S^m. Thus, they represent exotic smooth structures on S^m.

Two exotic smooth structures (S^m, ζ') and (S^m, ζ'') determine a same element of Θ_m if and only if they are smoothly h–cobordant, but the h–cobordism theorem implies that (S^m, ζ') and (S^m, ζ'') are diffeomorphic. Further, the h–cobordism linking them must be trivial, and hence exhibit a smooth structure on $S^m \times [0,1]$ that coincides with ζ' on $S^m \times 0$ and with ζ'' on $S^m \times 1$. The smooth structures ζ' and ζ'' are called **concordant.**

Thus, Θ_m can be called either the set of homotopy-spheres or the set of smooth structures on S^m up to concordance. (Notice that there exist smooth structures that are diffeomorphic but *not* concordant.)

Together with connected sums, the set Θ_m becomes an Abelian group. Its identity element is the standard smooth sphere S^m. The inverse of any $\Sigma \in \Theta_m$ is $\overline{\Sigma}$. Indeed, it is not hard to prove that, since $\Sigma \# \overline{\Sigma}$ can be realized as the boundary of $(\Sigma \setminus \textit{ball}) \times [0,1]$ as suggested in figure 2.36, and the latter can be shown to be contractible, the h–cobordism theorem implies that $\Sigma \# \overline{\Sigma}$ must be diffeomorphic to S^m.



2.36. $\Sigma \# \overline{\Sigma}$ is the boundary of $(\Sigma \setminus \textit{ball}) \times [0,1]$

For example, since the Kervaire spheres Σ_K^{4k+1} can be shown to have $\Sigma_K \# \Sigma_K \cong S^{4k+1}$, it follows that Σ_K is either standard or it generates an order 2 subgroup in Θ_{4k+1} (certainly the latter when k is even).

The group Θ_m is usually called the **group of homotopy spheres** in dimension m. One might also call it the *group of exotic spheres* or the *group of smooth structures* on S^m. The groups Θ_m are always *finite.* The orders of some of these groups are listed in table IV on the following page, while a few actual groups are listed in table V on the next page.

IV. Orders for groups of smooth structures on $\mathbb{S}^m$

m	5	6	7	8	9	10	11	12	13	14	15	16	17
$\lvert\Theta_m\rvert$	1	1	28	2	8	6	992	1	3	2	16256	2	16

V. Groups of smooth structures on $\mathbb{S}^m$

m	5	6	7	8	9	10	11
Θ_m	0	0	$\mathbb{Z}_{28}$	$\mathbb{Z}_2$	$\oplus 4\,\mathbb{Z}_2$	$\mathbb{Z}_6$	$\mathbb{Z}_{992}$

Low dimensions. *When $m \leq 4$, the set Θ_m can still be defined as the collection of smooth m–manifolds homotopy-equivalent to $\mathbb{S}^m$, considered up to h–cobordisms. Clearly, we have $\Theta_1 = 0$ and $\Theta_2 = 0$. Since all topological manifolds of dimensions up to 3 admit unique smooth structures, there are no exotic 3–spheres, but one would need to be sure that the Poincaré conjecture has indeed been proved to conclude that there are no fake 3–spheres and thus that $\Theta_3 = 0$. Finally, since the topological Poincaré conjecture holds in dimension 4, any 4–manifold Σ^4 homotopy-equivalent to $\mathbb{S}^4$ must be homeomorphic to it. Since Σ^4 and $\mathbb{S}^4$ are linked by a smooth h–cobordism, we have that $\Theta_4 = 0$ as well, regardless of whether there are exotic 4–spheres hiding around.*

***Example: dimension* 7.** Going upwards past $\Theta_5 = 0$ and $\Theta_6 = 0$, the first dimension where Θ_m is nontrivial is $m = 7$; we have

$$\Theta_7 = \mathbb{Z}_{28}\,.$$

As it turns out, the generator of Θ_7 is exactly the exotic sphere Σ^7_P obtained earlier as boundary of the 8–dimensional E_8–plumbing $P^8_{E_8}$. In other words,

$$\mathbb{S}^7,\ \Sigma^7_P,\ \Sigma^7_P \# \Sigma^7_P,\ \ldots\ ,\ \#27\,\Sigma^7_P$$

is the complete list of smooth structures on $\mathbb{S}^7$, *up to concordance.* We have diffeomorphisms $\#28\,\Sigma^7_P \cong \mathbb{S}^7$, as well as $\overline{\#k\,\Sigma^7_P} \cong \#(28-k)\,\Sigma^7_P$.

Notice that reversing orientations establishes diffeomorphisms between non-concordant smooth structures on $\mathbb{S}^7$. Thus, the complete list of smooth structures on $\mathbb{S}^7$, considered *up to diffeomorphisms,* counts exactly 15 elements:

$$\mathbb{S}^7,\ \Sigma^7_P,\ \ldots\ ,\ \#14\,\Sigma^7_P\,.$$

The only smooth structures on $\mathbb{S}^7$ that admit orientation-reversing self-diffeomorphisms are $\#14\,\Sigma^7_P$ and standard $\mathbb{S}^7$.

References. Plumbings were created by **J. Milnor** and presented in his mimeographed notes ***Differentiable manifolds which are homotopy spheres*** **[Mil58a]**, as well as in his ***Differentiable structures on spheres*** **[Mil59]**. A nice brief treatment is in **W. Browder**'s ***Surgery on simply-connected manifolds*** [**Bro72**, ch V].

A different source of exotic spheres (Brieskorn spheres, from singularities of complex hypersurfaces) will be mentioned in the end-notes of chapter 8 (page 318). In particular, we will notice there that all the plumbing examples above can be described by explicit equations in $\mathbb{C}^m$.

The groups Θ_m were defined and studied in **M. Kervaire** and **J. Milnor**'s ***Groups of homotopy spheres*** **[KM63]**, which is also considered as one of the founding papers of surgery theory.

Note: The world of 3–manifolds

After visiting in chapter 1 the realm of high-dimensional manifolds, dominion of the h–cobordism theorem, and in this chapter the world of topological 4–manifolds, it is now worth also taking a peek downwards, at the land of 3–manifolds. There too, embeddings of disks play a fundamental role in the theory.

First of all, as we mentioned earlier, in dimension 3 there is no distinction between topological and smooth manifolds. This was proved[12] by **E. Moise** in his series ***Affine structures in* 3*–manifolds. I–VIII*** [Moi54], or in his book ***Geometric topology in dimensions* 2 *and* 3** [Moi77a]; the proof was simplified by **R.H. Bing**'s ***An alternative proof that* 3*–manifolds can be triangulated*** [Bin59].

Second, just as surfaces play an essential part in the topology of 4–manifolds, the middle dimension is essential for 3–manifolds as well. The middle dimension, though, happens to be $1\frac{1}{2}$, which gives the fundamental group quite a prominence, but keeps surfaces at the forefront as well. The fundamental group is usually non-commutative, and that gives a strong group-theoretic flavor to the theory, as you can sample from **J. Stallings**' ***Group theory and three-dimensional manifolds*** [Sta71]. Another standard reference for the classical theory is **J. Hempel**'s **3*–Manifolds*** [Hem76].

As a show of strength of the fundamental group, the inevitably-known Poincaré conjecture (1904) claims that, without fundamental group, nothing much happens in dimension 3:

Poincaré Conjecture *(open?). The only simply-connected closed 3–manifold is* S^3.

A more general conjecture on the structure of 3–manifolds is W. Thurston's *geometrization conjecture*, which claims that all closed oriented 3–manifolds split into pieces that admit certain few Riemannian geometric structures. The geometrization conjecture (and thus the Poincaré conjecture as well) might have been proved in 2003 by G. Perelman, following the direction traced by R. Hamilton's program to deform Riemannian metrics along their Ricci-flow into hyperbolic metrics (*i.e.*, metrics of constant sectional curvature -1). This remarkable proof is presented in **G. Perelman**'s ***The entropy formula for the Ricci flow and its geometric applications*** [Per02], with technical developments in [**Per03b**] and a further technical detail in [**Per03a**]. This differential-geometric proof is still under intense scrutiny at the time of this writing.

The theory of geometric structures on 3–manifolds was founded by W. Thurston around 1977. One can start with **P. Scott**'s exposition ***The geometries of* 3*–manifolds*** [Sco83], then continue with **W. Thurston**'s own ***Three-dimensional geometry and topology*** [Thu97], or with **M. Kapovich**'s ***Hyperbolic manifolds and discrete groups*** [Kap01].

---

[12] Strictly speaking, Moise and Bing only proved that any topological 3–manifold admits an essentially unique triangulation. The surprising issue of triangulated manifolds having more than one smooth structure only became visible later, after J. Milnor's exotic 7–spheres [**Mil56b**]. The first proofs that 3–manifolds admit unique smooth structures are due to **J. Munkres**' ***Obstructions to the smoothing of piecewise-differentiable homeomorphisms*** [**Mun59, Mun60b**] and to **J.H.C. Whitehead**'s last paper, ***Manifolds with transverse fields in euclidean space*** [**Whi61**].

Going back into the prehistory of the subject, **M. Dehn** in his ***Über die Topologie des dreidimensionalen Raume*** **[Deh10]** thought to have proved the following:

Dehn's Lemma. *Let* N^3 *be a* 3*–manifold and* $f\colon \mathbb{D}^2 \longrightarrow N$ *be an immersion such that* $f|_{\partial \mathbb{D}^2}$ *is an embedding. Then* $f|_{\partial \mathbb{D}^2}$ *can be extended to an embedding of* $\mathbb{D}^2$ *into* N. □

Dehn's argument was incomplete though, and it took fifty years to find the correct proof. That was due to **C.D. Papakyriakopoulos** in ***On Dehn's lemma and the asphericity of knots*** **[Pap57a]**.

In the same paper was proved the following very important result:

Sphere Theorem. *Let* N *be an orientable* 3*–manifold, and assume that* $\pi_2(M)$ *is nontrivial. Then there must be a homotopically-nontrivial embedding* $\mathbb{S}^2 \longrightarrow N$. □

This theorem can be strengthened by the uniqueness result from **F. Laudenbach**'s ***Sur les 2–sphères d'une variété de dimension 3*** **[Lau73]**: *If two spheres embedded in a* 3*–manifold are homotopic, then they are in fact ambiently isotopic.*[13]

Further, **C.D. Papakyriakopoulos** proved in ***On solid tori*** **[Pap57b]** a result that is most useful in the form improved by **J. Stallings**' ***On the loop theorem*** **[Sta60a]**, and combined with Dehn's lemma:

Loop Theorem. *Let* M *be a* 3*–manifold, and let* S *be a connected component of* ∂M. *Assume that the natural morphism* $\pi_1(S) \longrightarrow \pi_1(M)$ *has nontrivial kernel. Then some nontrivial element of that kernel can be represented by a circle* C *embedded in* S. *Further, there is an embedding* $f\colon \mathbb{D}^2 \longrightarrow N$ *with* $f\left[\partial \mathbb{D}^2\right] = C$. □

It might be amusing to compare the above 3–dimensional statements to the 4–dimensional Casson embedding theorem discussed in a previous note (page 96).

In any case, the role played by these results in the theory of 3–manifolds can hardly be understated, even though, of course, since the 1960s many new methods have gained prominence in the field: foliations and laminations, and geometric structures, especially hyperbolic ones.

Apart from a few words on foliations and surfaces in 3–manifolds in section 11.3 (page 491), and some on the Rokhlin invariant in the end-notes of chapter 4 (page 224), we now leave the 3–dimensional realm.

Bibliography

Casson handles were created by **A. Casson** and presented in a lecture in 1973. Notes from that lecture were published as **[Cas86**, lecture I**]** inside the volume ***À la recherche de la topologie perdue*** **[GM86a]** edited by **L. Guillou** and **A. Marin.**

For **M. Freedman**'s theory, one can refer to the original paper ***The topology of four-dimensional manifolds*** **[Fre82]** or read the survey from **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89**, ch XIII**]**. A quick overview is also contained in **M. Freedman**

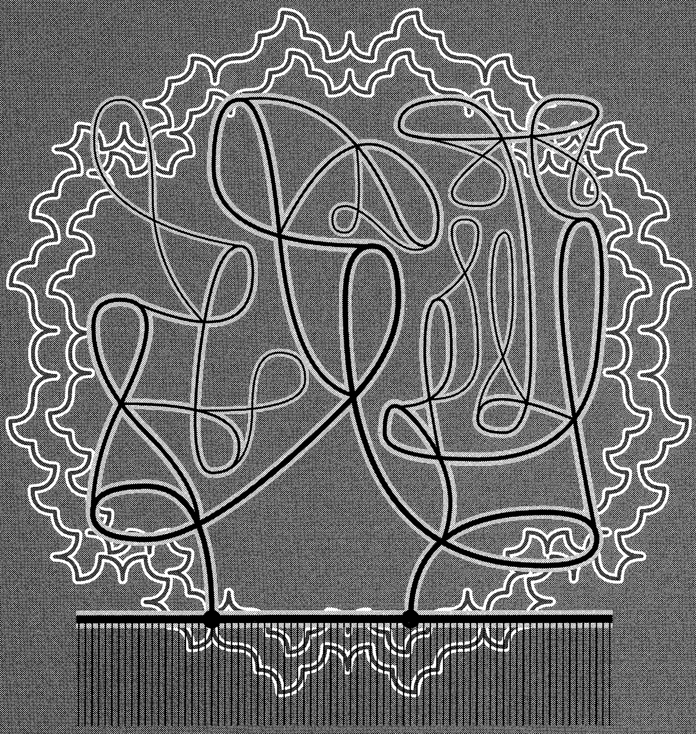
---

[13] This result is also the main technical detail that is used to show that every self-diffeomorphism of $\#k\mathbb{S}^1 \times \mathbb{S}^2$ extends to $\#k\mathbb{S}^1 \times \mathbb{D}^3$. This is relevant to 4–manifold theory since it shows that, if we know the 1– and 2–handles of 4–manifold, we need not worry about the 3–handles. See page 93 and footnote 5 there.

and F. **Luo**'s ***Selected applications of geometry to low-dimensional topology*** **[FL89]**. **M. Freedman** and **F. Quinn**'s volume ***Topology of 4–manifolds*** **[FQ90]** is the standard monograph on topological 4–manifolds. Notice that the latter substitutes *capped gropes* instead of Casson handles, which are technically better suited for dealing with non-simply-connected 4–manifolds; this use of gropes originates with **M. Freedman**'s ***The disk theorem for four-dimensional manifolds*** **[Fre84]**. The method of construction of the fake 4–balls Δ has its origins in **M. Freedman**'s earlier ***A fake*** $\mathbb{S}^3 \times \mathbb{R}$ **[Fre79]**.

The first examples of nontrivial h–cobordisms were provided by **S.K. Donaldson** in ***Irrationality and the h–cobordism conjecture*** **[Don87]**. We will see examples in section 8.4 (page 314) and section 12.4 (page 545). The simple Casson handle that we claimed is exotic was proved to be so in **Ž. Bižaca**'s ***An explicit family of exotic Casson handles*** **[Biž95]**.

A first example of a nontrivial h–cobordism supported by $\mathbb{D}^5$ appeared in **S. Akbulut**'s ***A fake compact contractible 4–manifold*** **[Akb91]**. The theorem on Akbulut corks was first proved in a preprint of **C. Curtis** and **W. Hsiang**, then improved together with **M. Freedman** and **R. Stong** in ***A decomposition theorem for h–cobordant smooth simply-connected compact 4–manifolds*** **[CFHS96]**, with further contributions by **R. Matveyev** in ***A decomposition of smooth simply-connected h–cobordant 4–manifolds*** **[Mat96]**, by **R. Kirby** in ***Akbulut's corks and h–cobordisms of smooth, simply connected 4–manifolds*** **[Kir96]**, and by **Ž. Bižaca** in ***A handle decomposition of an exotic*** $\mathbb{R}^4$ **[Biž94]**. These improvements are related to exotic $\mathbb{R}^4$'s and some will be stated in section 5.4 (page 253). A nice exposition and proof is in **R. Kirby**'s **[Kir96]**, and one can also read from **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** **[GS99**, ch 9**]**. The latter sources also contain explicit handle decompositions of Akbulut corks.



Part II

Smooth 4–Manifolds and Intersection Forms

WE are now entering the domain of smooth 4–manifolds. This part is devoted to what could be called the classic smooth topology of 4–manifolds. The discussion is centered around the main invariant of a 4–manifold, its intersection form. As the name suggests, this form describes how surfaces intersect inside the 4–manifold. After defining it in chapter 3, we look at a few simple examples, then build an elaborate but essential 4–manifold, the $K3$ complex surface.

Afterwards, in chapter 4 (starting on page 139) we set out to explore the strong interactions between the intersection form and the topology of its underlying 4–manifold. We look at the homotopy type of a 4–manifold, at h–cobordisms between 4–manifolds (Wall's theorems), at characteristic classes of 4–manifolds, and at Rokhlin's theorem.

We open chapter 5 (starting on page 237) with statements about the algebraic classification of forms. The influence of intersection forms on topology culminates with M. Freedman's classification of topological 4–manifolds, which implies that if two smooth 4–manifolds have the same intersection form, then they must be homeomorphic; nonetheless, they do not need to be diffeomorphic. The opening salvo illuminating the chasm between topological and smooth 4–manifolds was shot by S.K. Donaldson, who showed that most intersection forms cannot correspond to smooth 4–manifolds, even though they can be realized by topological 4–manifolds. As dessert, we conclude by building exotic $\mathbb{R}^4$'s.

Contents of Part II

Chapter 3

Getting Acquainted with Intersection Forms

We define the intersection form of a 4–manifold, which governs intersections of surfaces inside the manifold. We start by representing every homology 2–class by an embedded surface, then, in section 3.2 (page 115), we explore the properties of the intersection form. Among them is unimodularity, which is essentially equivalent to Poincaré duality. An important invariant of an intersection form is its signature, and we discuss how its vanishing is equivalent to the 4–manifold being a boundary of a 5–manifold. After listing a few simple examples of 4–manifolds and their intersection form, in section 3.3 (page 127) we present in some detail the important example of the $K3$ manifold.

Given any closed oriented 4-manifold M, its **intersection form** is the symmetric 2–form defined as follows:

$$\begin{gathered} Q_M \colon H^2(M;\mathbb{Z}) \times H^2(M;\mathbb{Z}) \longrightarrow \mathbb{Z} \\ Q_M(\alpha,\beta) = (\alpha \cup \beta)[M] \ . \end{gathered}$$

This form is bilinear[1] and is represented by a matrix of determinant ± 1. While over $\mathbb{R}$ this is a recipe for boredom, since this intersection form is defined over the *integers* (and thus changes of coordinates must be made only through integer-valued matrices), our Q_M is a quite far-from-trivial object.

---

[1] Notice that Q_M vanishes on any torsion element, and thus can be thought of as defined on the free part of $H^2(M;\mathbb{Z})$; since our manifolds are assumed simply-connected, torsion is not an issue.

For convenience, we will often denote $Q_M(\alpha, \beta)$ by $\alpha \cdot \beta$. Further, we will identify without comment a cohomology class $\alpha \in H^2(M; \mathbb{Z})$ with its Poincaré-dual homology class $\alpha \in H_2(M; \mathbb{Z})$.

For defining Q_M more geometrically,[2] we will represent classes α and β from $H_2(M; \mathbb{Z})$ by embedded surfaces S_α and S_β, and then equivalently define $Q_M(\alpha, \beta)$ as the intersection number of S_α and S_β:

$$Q_M(\alpha, \beta) = S_\alpha \cdot S_\beta .$$

First, though, we need to argue that any class $\alpha \in H_2(M; \mathbb{Z})$ can indeed be represented by a smoothly embedded surface S_α:

3.1. Preparation: representing homology by surfaces

It is known from general results[3] that every homology class of a 4–manifold can be represented by embedded submanifolds. Nonetheless, we present a direct argument for the case of 2–classes, owing to the useful techniques that it exhibits.

Simply-connected case. Assume first that M is simply-connected. Then by Hurewicz's theorem $\pi_2(M) \approx H_2(M; \mathbb{Z})$, and hence all homology classes of M can be represented as images of maps $f \colon S^2 \to M$. The latter can always be perturbed to yield immersed spheres, whose only failures from being embedded are transverse double-points. These double-points can be eliminated at the price of increasing the genus.

For example, by using complex coordinates, a double-point is isomorphic to the simple nodal singularity of equation $z_1 z_2 = 0$ in $\mathbb{C}^2$: the complex planes $z_1 = 0$ and $z_2 = 0$ meeting at the origin. It can be eliminated by perturbing to $z_1 z_2 = \varepsilon$, as suggested in figure 3.1 on the facing page. (A simple change of coordinates transforms the situation into perturbing $w_1^2 + w_2^2 = 0$ to $w_1^2 + w_2^2 = \varepsilon$.)

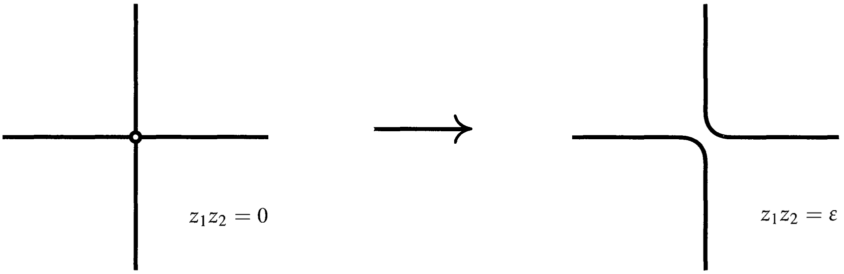
More geometrically, imagine two planes meeting orthogonally at the origin of $\mathbb{R}^4$. Their traces in the 3–sphere S^3 are two circles, linking once.[4] We can eliminate the singularity if we discard the portions contained in the open 4–ball bounded by S^3, and instead connect the two circles in S^3 by an annular

---

[2] *"Think with intersections, prove with cup-products."*

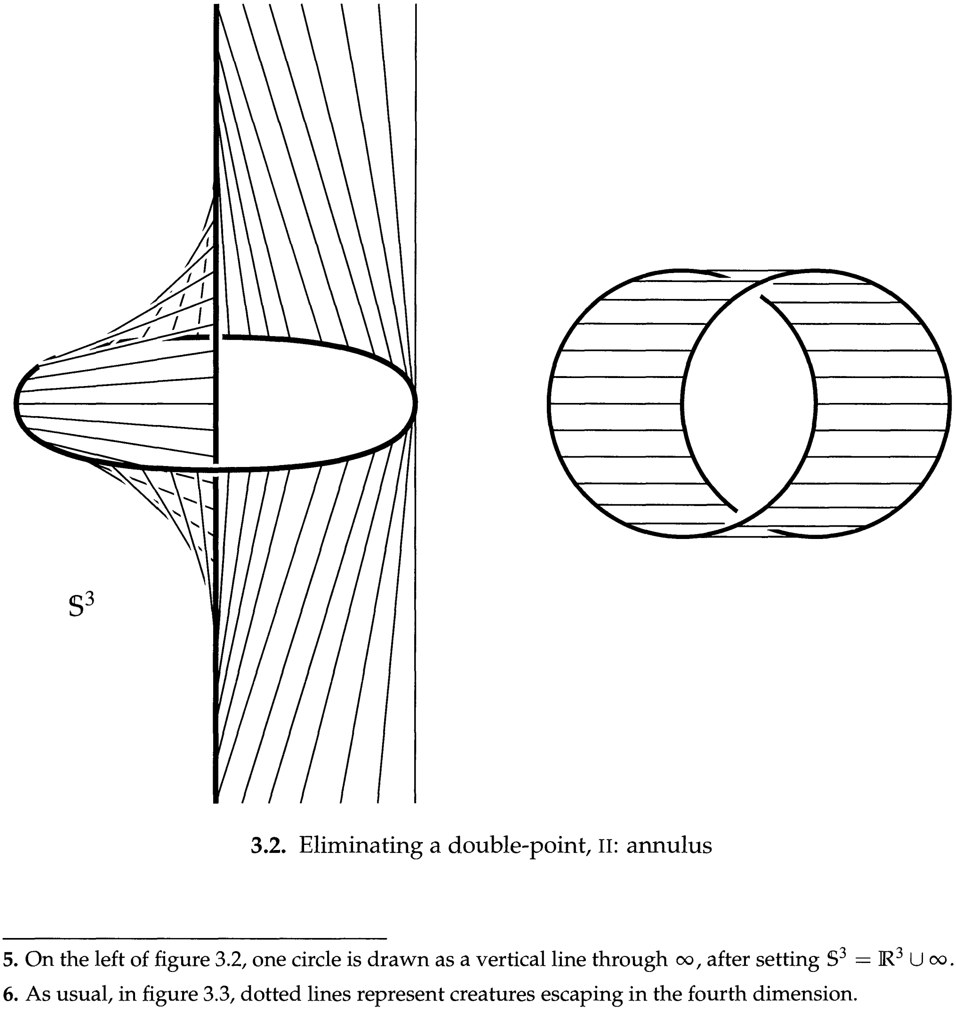
[3] For example, for any smooth oriented X^m and any $\alpha \in H^*(X; \mathbb{Z})$, there is some integer k so that $k\alpha$ can be represented by an embedded submanifold; if α has dimension at most 8 or codimension at most 2, then it can be represented directly by a submanifold; if X^m is embedded in $\mathbb{R}^{m+2}$, then X is the boundary of an oriented smooth $(m+1)$–submanifold in $\mathbb{R}^{m+2}$. These results were announced in **R. Thom**'s ***Sous-variétés et classes d'homologie des variétés différentiables*** **[Tho53a]** and proved in his celebrated ***Quelques propriétés globales des variétés différentiables*** **[Tho54]**.

[4] Think: fibers of the Hopf map $S^3 \to \mathbb{CP}^1$; the Hopf map will be recalled in footnote 34 on page 129.



3.1. Eliminating a double-point, I: complex coordinates

sheet, as suggested in figure[5] 3.2. Thus, we replaced two disks meeting at the double-point by an annulus. A 4–dimensional image is attempted in figure[6] 3.3 on the following page.

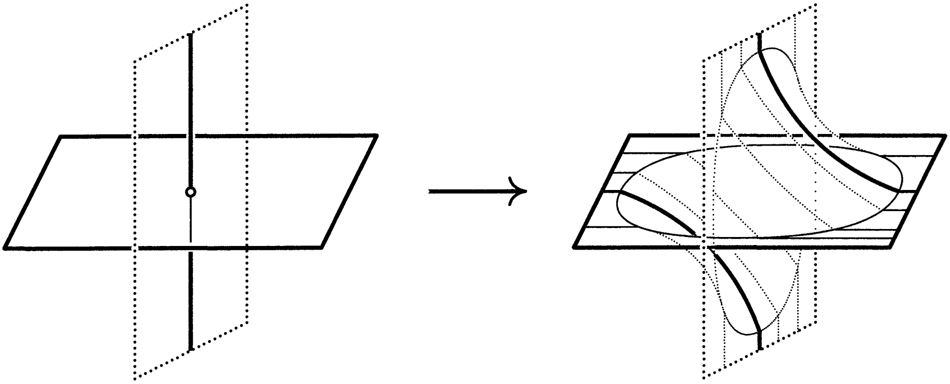


3.2. Eliminating a double-point, II: annulus

---

[5] On the left of figure 3.2, one circle is drawn as a vertical line through ∞, after setting $S^3 = \mathbb{R}^3 \cup \infty$.

[6] As usual, in figure 3.3, dotted lines represent creatures escaping in the fourth dimension.



3.3. Eliminating a double-point, III

Either way, we can eliminate all double-points of the immersed sphere, and the result is then an embedded surface representing that homology class. Thus, all homology classes can be represented by embedded surfaces, but rarely by spheres.

> *The failure to represent homology classes by smoothly embedded spheres is of course related to the failure of smoothly embedding disks. The natural question to ask is then: what is the minimum genus needed to represent a given homology class? We will come back to this question later.*[7]

In general. The method above only works for simply-connected M^4's. An argument for general 4–manifolds has two equivalent versions:

(1) Since $\mathbb{CP}^\infty$ is an Eilenberg–Maclane $K(\mathbb{Z},2)$–space,[8] it follows that the elements of $H^2(M;\mathbb{Z})$ correspond to homotopy classes of maps $M \to \mathbb{CP}^\infty$. Since M is 4–dimensional, such maps can be slid off the high-dimensional cells of $\mathbb{CP}^\infty$ and thus reduced to maps $M \to \mathbb{CP}^2$. For any class $\alpha \in H^2(M;\mathbb{Z})$, pick a corresponding $f_\alpha \colon M \to \mathbb{CP}^2$ and arrange it to be differentiable and transverse to $\mathbb{CP}^1 \subset \mathbb{CP}^2$. Then $f_\alpha^{-1}[\mathbb{CP}^1]$ is a surface Poincaré-dual to α.

(2) Equivalently, since $\mathbb{CP}^\infty$ coincides with the classifying space[9] $\mathscr{B}U(1)$ of the group $U(1)$, classes in $H^2(M;\mathbb{Z})$ correspond to complex line bundles on M, with α being paired to L_α whenever $c_1(L_\alpha) = \alpha$. If we pick a

---

[7] See ahead, chapter 11 (starting on page 481).

[8] An **Eilenberg–Maclane $K(G,m)$–space** is a space whose *only* non-zero homotopy group is $\pi_m = G$; such a space is unique up to homotopy-equivalence. It can be used to represent cohomology as $H^m(X;G) = [X;\, K(G,m)]$, where $[A;B]$ denotes the set of homotopy classes of maps $A \to B$.

[9] A **classifying space** $\mathscr{B}G$ for a topological group G is a space so that $[X;\mathscr{B}G]$ coincides with the set of isomorphisms classes of G–bundles over X. A bit more on classifying spaces is explained in the end-notes of the next chapter (page 204).

generic section σ of L_α, then its zero set $\sigma^{-1}[0]$ will be an embedded surface Poincaré-dual to α.

3.2. Intersection forms

Given a closed oriented 4-manifold M, we defined its **intersection form** as

$$Q_M \colon H_2(M;\mathbb{Z}) \times H_2(M;\mathbb{Z}) \longrightarrow \mathbb{Z} \qquad Q_M(\alpha,\beta) = S_\alpha \cdot S_\beta \, ,$$

where S_α and S_β are any two surfaces representing the classes α and β.

Notice that, if M is simply-connected, then $H_2(M;\mathbb{Z})$ is a free $\mathbb{Z}$–module and there are isomorphisms $H_2(M;\mathbb{Z}) \approx \oplus m\,\mathbb{Z}$, where $m = b_2(M)$. If M is not simply-connected, then $H_2(M;\mathbb{Z})$ inherits the torsion of $H_1(M;\mathbb{Z})$, but by linearity the intersection form will always vanish on these torsion classes; thus, when studying intersection form, we can safely pretend that $H_2(M;\mathbb{Z})$ is always free.

Lemma. *The form $Q_M(\alpha,\beta) = S_\alpha \cdot S_\beta$ on $H_2(M;\mathbb{Z})$ coincides modulo Poincaré duality with the pairing $Q_M(\alpha^*,\beta^*) = (\alpha^* \cup \beta^*)[M]$ on $H^2(M;\mathbb{Z})$.*

Proof. Given any class $\alpha \in H_2(M;\mathbb{Z})$, denote by α^* its Poincaré-dual from $H^2(M;\mathbb{Z})$; we have $\alpha^* \cap [M] = \alpha$. We wish to show that the pairing

$$Q_M(\alpha^*,\beta^*) = (\alpha^* \cup \beta^*)[M]$$

on $H^2(M;\mathbb{Z})$ defines the same bilinear form as the one defined above.

We use the general formula[10] $(\alpha^* \cup \beta^*)[M] = \alpha^*\big[\beta^* \cap [M]\big]$, from which it follows that $Q_M(\alpha^*,\beta^*) = \alpha^*[\beta]$, or

$$Q_M(\alpha^*,\beta^*) = \alpha^*[S_\beta] \, .$$

Therefore, we need to show that

$$\alpha^*[S_\beta] = S_\alpha \cdot S_\beta \, .$$

Since Q_M vanishes on torsion classes, it is enough to check the last formula by including the free part of $H^2(M;\mathbb{Z})$ into $H^2(M;\mathbb{R})$ and by interpreting the latter as the de Rham cohomology of exterior 2–forms.

Moving into de Rham cohomology translates cup products into wedge products and cohomology/homology pairings into integrations. We have, for example,

$$Q_M(\alpha^*,\beta^*) = \int_M \alpha^* \wedge \beta^* \qquad \text{and} \qquad \alpha^*[S_\beta] = \int_{S_\beta} \alpha^*$$

for all 2–forms α^*, $\beta^* \in \Gamma(\Lambda^2(T^*_M))$.

---

[10] More often written in terms of the Kronecker pairing as $\langle \alpha^* \cup \beta^*, \, [M] \rangle = \langle \alpha^*, \, \beta^* \cap [M] \rangle$.

In this setting, given a surface S_α, one can find a 2–form α^* dual to S_α so that it is non-zero only close to S_α. Further, one can choose some local oriented coordinates $\{x_1, x_2, y_1, y_2\}$ so that S_α coincides locally with the plane $\{y_1 = 0; y_2 = 0\}$, oriented by $dx_1 \wedge dx_2$. One can then choose α^* to be locally written $\alpha^* = f(x_1, x_2)\, dy_1 \wedge dy_2$, for some suitable bump-function f on $\mathbb{R}^2$, supported only around $(0,0)$ and with integral $\int_{\mathbb{R}^2} f = 1$.

If S_β is some surface transverse to S_α and we arrange that, around the intersection points of S_α and S_β, we have S_β described by $\{x_1 = 0;\ x_2 = 0\}$, then clearly

$$\int_{S_\beta} \alpha^* = S_\alpha \cdot S_\beta ,$$

with each intersection point of S_α and S_β contributing ±1 depending on whether $dy_1 \wedge dy_2$ orients S_β positively or not.[11] □

Unimodularity and dual classes

The intersection form Q_M is $\mathbb{Z}$*–bilinear* and *symmetric*. As a consequence of Poincaré duality, the form Q_M is also **unimodular**, meaning that the matrix representing Q_M is invertible over $\mathbb{Z}$. This is the same as saying that

$$\det Q_M = \pm 1 .$$

Unimodularity is further equivalent to the property that, for every $\mathbb{Z}$–linear function $f\colon H_2(M;\mathbb{Z}) \to \mathbb{Z}$, there exists a unique $\alpha \in H_2(M;\mathbb{Z})$ so that $f(x) = \alpha \cdot x$.

Lemma. *The intersection form Q_M of a 4–manifold is unimodular.*

Proof. The intersection form is unimodular if and only if the map

$$\begin{aligned} \widehat{Q}_M\colon H_2(M;\mathbb{Z}) &\longrightarrow \operatorname{Hom}_{\mathbb{Z}}(H_2(M;\mathbb{Z}),\ \mathbb{Z}) \\ \alpha &\longmapsto x \mapsto \alpha \cdot x \end{aligned}$$

is an isomorphism. We will argue that this last map coincides with the Poincaré duality morphism. Indeed, Poincaré duality is the isomorphism

$$\begin{aligned} H_2(M;\mathbb{Z}) &\xrightarrow[\approx]{} H^2(M;\mathbb{Z}) \\ \alpha &\longmapsto \alpha^* , \end{aligned}$$

with α^* characterized by $\alpha^* \cap [M] = \alpha$. Assume for simplicity that $H_2(M;\mathbb{Z})$ is free.[12] Then the universal coefficient theorem[13] shows that

---

[11] See **R. Bott** and **L. Tu**'s ***Differential forms in algebraic topology*** **[BT82]** for more such play with exterior forms.

[12] If not free, a similar argument is made on the free part $H^2(M;\mathbb{Z}) / \operatorname{Ext}(H_1(M;\mathbb{Z});\mathbb{Z})$ of $H^2(M;\mathbb{Z})$, which is all that matters since Q_M vanishes on torsion.

[13] The universal coefficient theorem was recalled on page 15.

we have an isomorphism

$$\begin{array}{ccc} H^2(M;\mathbb{Z}) & \xrightarrow[\approx]{} & \operatorname{Hom}\bigl(H_2(M;\mathbb{Z}),\, \mathbb{Z}\bigr) \\ \alpha^* & \longmapsto & x \mapsto \alpha^*[x]\,. \end{array}$$

Combining Poincaré duality with the latter yields the isomorphism

$$\begin{array}{ccc} H_2(M;\mathbb{Z}) & \xrightarrow[\approx]{} & \operatorname{Hom}\bigl(H_2(M;\mathbb{Z}),\, \mathbb{Z}\bigr) \\ \alpha & \longmapsto & x \mapsto \alpha^*[x]\,. \end{array}$$

However, as argued in the preceding subsection, we have $Q_M(\alpha, x) = \alpha^*[x]$, and therefore the above isomorphism coincides with the map $\widehat{Q}_M$. That proves that the intersection form Q_M is unimodular. □

Further, the unimodularity of Q_M is equivalent to the fact that, for every basis $\{\alpha_1, \ldots, \alpha_m\}$ of $H_2(M;\mathbb{Z})$, there is a unique **dual basis** $\{\beta_1, \ldots, \beta_m\}$ of $H_2(M;\mathbb{Z})$ so that $\alpha_k \cdot \beta_k = +1$ and $\alpha_i \cdot \beta_j = 0$ if $i \neq j$.

To see this, start with the basis $\{\alpha_1, \ldots, \alpha_m\}$ in $H_2(M;\mathbb{Z})$, pick the familiar dual basis[14] *$\{\alpha_1^*, \ldots, \alpha_m^*\}$ in the dual $\mathbb{Z}$–module $\operatorname{Hom}(H_2(M;\mathbb{Z}), \mathbb{Z})$, then transport it back to $H_2(M;\mathbb{Z})$ by using Poincaré duality (or $\widehat{Q}_M$) and hence obtain the desired basis $\{\beta_1, \ldots, \beta_m\}$.*

In particular, for every *indivisible* class α (*i.e.*, not a multiple), there exists at least one **dual class** β such that $\alpha \cdot \beta = +1$: complete α to a basis and proceed as above. (Of course, such β's are *not* unique: once you find one, you can obtain others by adding any γ with $\alpha \cdot \gamma = 0$.)

Intersection forms and connected sums

The simplest way of combining two 4–manifolds yields the the simplest way of combining two intersection forms. First, a bit of review:

Remembering connected sums. The **connected sum** of two manifolds M and N, denoted by

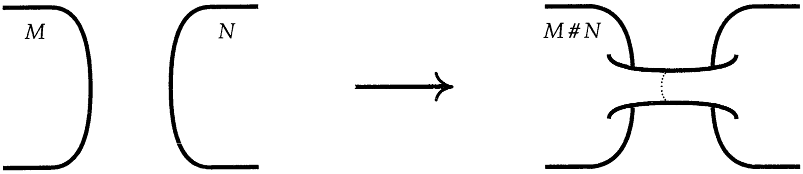
$$M \# N\,,$$

is the simplest method for combining M and N into one connected manifold, by joining them with a tube as sketched in figure 3.4 on the next page. Notice that the 4–sphere is an identity element for connected sums: $M \# S^4 \cong M$.

Connected sums are described more rigorously by choosing in each of M and N a small open 4–ball and removing it to get two manifolds M° and N°, each with a 3–sphere as boundary, then identifying these 3–spheres to obtain the closed manifold $M \# N$.

---

[14] Recall that, given a basis $\{e_1, \ldots, e_m\}$ in a module Z, the dual basis $\{e_1^*, \ldots, e_m^*\}$ in Z^* is specified by setting $e_k^*(e_k) = 1$ and $e_i^*(e_j) = 0$ for $i \neq j$.



3.4. The connected sum of two manifolds, I

More about connected sums. *The identification of the two 3-spheres must be made through an orientation-reversing diffeomorphism* $\partial M^\circ \cong \overline{\partial N^\circ}$*, as was mentioned on page 13. Indeed, if* M *and* N *are oriented, then the new boundary 3-spheres will inherit orientations. In order that the orientations of* M *and* N *be nicely compatible with an orientation of* $M \# N$*, we must identify the 3-spheres with an orientation flip.*

Furthermore, to ensure that $M \# N$ *is a smooth manifold, this gluing must be done as follows: Choose open 4-balls in* M *and* N*, then remove them. Embed copies of* $S^3 \times [0,1]$ *as collars to the new boundary 3-spheres. Take care to embed these collars so that, on the side of* M*, the sphere* $S^3 \times 1$ *be sent onto* ∂M°*, with* $S^3 \times [0,1)$ *going into the interior of* M°*. On the* N *side,* $S^3 \times 0$ *should be sent onto* ∂N° *and* $S^3 \times (0,1]$ *into the interior of* N°*. Now identify the two collars* $S^3 \times [0,1]$ *in the obvious manner and thus obtain* $M \# N$*, as in figure 3.5. This automatically forces the boundary-spheres to be identified "inside-out", reversing orientations, and further makes it clear that* $M \# N$ *is smooth.*[15] *See figure 3.6 on the next page. The equivalence of this procedure with "joining by a tube" is explained in figure 3.7 on the facing page.*

3.5. Gluing by identifying collars

Sums and forms. This connected sum operation is nicely compatible with intersection forms:

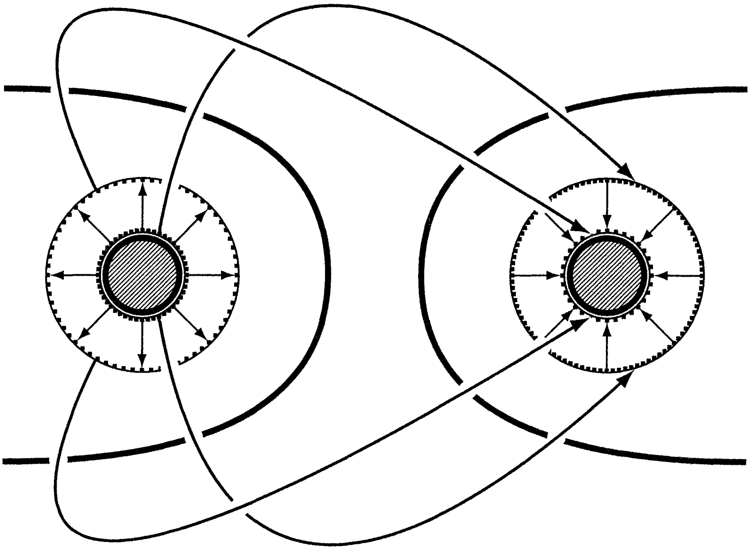
Lemma. *If* M *and* N *have intersection forms* Q_M *and* Q_N*, then their connected sum* $M \# N$ *will have intersection form*

$$Q_{M\#N} = Q_M \oplus Q_N \,.$$

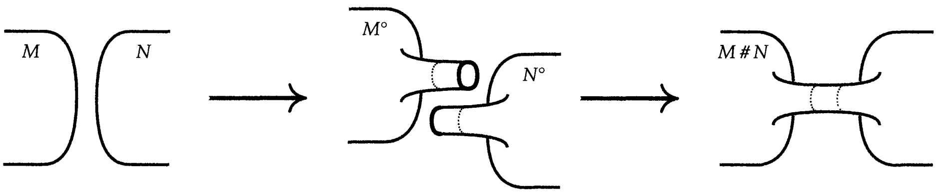
Proof. Since M° and N° can be viewed as M and N without a 4-handle (or a 4-cell), and since 2-homology is influenced only by 1-, 2- and 3-handles, it follows that the 2-homology of $M \# N$ will merely be the friendly gathering of the 2-homologies of M and N, intersections and all. □

---

[15] In fact, each time you read "*A and B both have the same boundary, so we glue A and B along it*", you should understand that the "gluing" is done *via* an orientation-reversing diffeomorphism $\partial A \cong \overline{\partial B}$, and that a collaring procedure as above is used. This was already explained on page 13. For more on the foundation of these gluings, read from **M. Hirsch**'s ***Differential topology*** [**Hir94**, sec 8.2].



3.6. The connected sum of two manifolds, II



3.7. The connected sum of two manifolds, III

Topological heaven. *For topological* 4*–manifolds a converse is true:*

Theorem (*M. Freedman*). *If M is simply-connected and Q_M splits as a direct sum $Q_M = Q' \oplus Q''$, then there exist* topological 4*–manifolds N' and N'' with intersection forms Q' and Q'' such that $M = N' \# N''$.* □

This is a direct consequence of Freedman's classification that we will present later.[16] *Such a result certainly fails in the smooth case, and its failure spawns exotic*[17] *$\mathbb{R}^4$'s.*

Invariants of intersection forms

To start to distinguish between the various possible intersection forms, we define the following simple algebraic invariants:

---

[16] See ahead section 5.2 (page 239). For a more refined topological sum-splitting result, we refer to **M. Freedman** and **F. Quinn**'s ***Topology of 4–manifolds*** [**FQ90**, ch 10].

[17] See ahead section 5.4 (page 250).

— The **rank** of Q_M:

It is the size of Q_M's domain, defined simply as

$$\operatorname{rank} Q_M = \operatorname{rank}_{\mathbb{Z}} H^2(M;\mathbb{Z}) ,$$

or $\operatorname{rank} Q_M = \dim_{\mathbb{R}} H^2(M;\mathbb{R})$. In other words, the rank is the second Betti number $b_2(M)$ of M.

— The **signature** of Q_M:

It is obtained as follows: first diagonalize Q_M as a matrix over $\mathbb{R}$ (or $\mathbb{Q}$), separate the resulting positive and negative eigenvalues, then subtract their counts; that is

$$\operatorname{sign} Q_M = \dim H^2_+(M;\mathbb{R}) - \dim H^2_-(M;\mathbb{R}) ,$$

where $H^2_\pm$ are any maximal positive/negative-definite subspaces for Q_M. We can set partial Betti numbers $b_2^\pm = \dim H^2_\pm$, and thus we can read $\operatorname{sign} Q_M = b_2^+(M) - b_2^-(M)$.

— The **definiteness** of Q_M (*definite* or *indefinite*):

If for all non-zero classes α we always have $Q_M(\alpha,\alpha) > 0$, then Q_M is called **positive definite.**

If, on the contrary, we have $Q_M(\alpha,\alpha) < 0$ for all non-zero α's, then Q_M is called **negative definite.**

Otherwise, if for some α_+ we have $Q_M(\alpha_+, \alpha_+) > 0$ and for some α_- we have $Q_M(\alpha_-, \alpha_-) < 0$, then Q_M is called **indefinite.**

— The **parity** of Q_M (*even* or *odd*):

If, for all classes α, we have that $Q_M(\alpha,\alpha)$ is even, then Q_M is called **even.** Otherwise, it is called **odd.** Notice that it is enough to have *one* class with odd self-intersection for Q_M to be called odd.

Signatures and bounding 4–manifolds

A first remark is that signatures are *additive:* $\operatorname{sign}(Q' \oplus Q'') = \operatorname{sign} Q' + \operatorname{sign} Q''$. In particular,[18]

$$\operatorname{sign}(M \# N) = \operatorname{sign} M + \operatorname{sign} N .$$

Another remark is that changing the orientation of M will change the sign of the signature:

$$\operatorname{sign} \overline{M} = -\operatorname{sign} M ,$$

since it obviously changes the sign of its intersection form: $Q_{\overline{M}} = -Q_M$.

---

[18] The additivity of signatures still holds for gluings $M \cup_\partial N$ more general than connected sums. This result *(Novikov additivity)* and an outline of its proof can be found in the the end-notes of the next chapter (page 224).

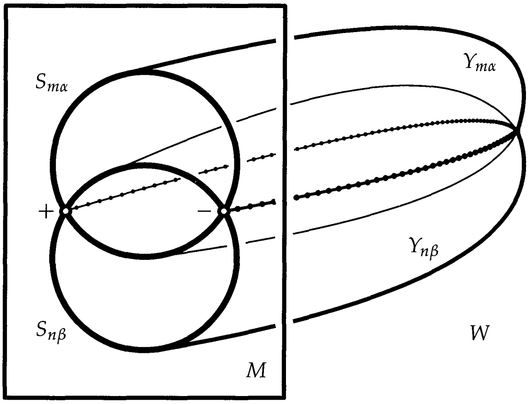
The signature vanishes for boundaries. More remarkably, the vanishing of the signature of a 4–manifold M has a direct topological interpretation:

Lemma. *If M^4 is the boundary of some oriented 5–manifold W^5, then*

$$\operatorname{sign} Q_M = 0\,.$$

Proof. Since the signature appears after diagonalizing over some field, we will work here with homology with rational coefficients. Thus, denote by $\iota\colon H_2(M;\mathbb{Q}) \to H_2(W;\mathbb{Q})$ the morphism induced from the inclusion of M^4 as the boundary of W^5.

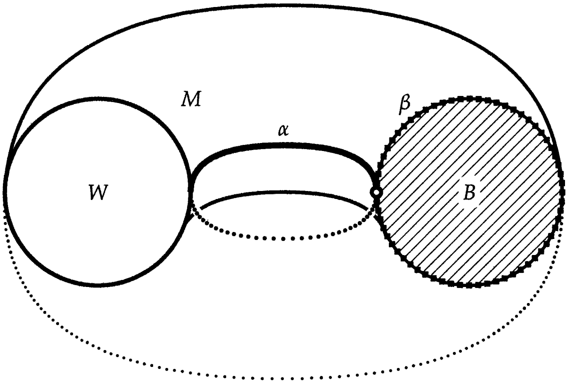
If bounding. First, we claim that if both $\alpha, \beta \in H_2(M;\mathbb{Q})$ have $\iota\alpha = 0$ and $\iota\beta = 0$ then their intersection must be $\alpha \cdot \beta = 0$. Indeed, since α and β are rational, some of their multiples $m\alpha$ and $n\beta$ will be integral. Then $m\alpha$ and $n\beta$ can be represented by two embedded surfaces $S_{m\alpha}$ and $S_{n\beta}$ in M. Since $\iota\alpha = 0$ and $\iota\beta = 0$, this implies that $S_{m\alpha}$ and $S_{n\beta}$ will bound two oriented 3–manifolds $Y_{m\alpha}$ and $Y_{n\beta}$ inside W. The intersection number $\alpha \cdot \beta$ is determined by counting the intersections of the surfaces $S_{m\alpha}$ and $S_{n\beta}$, then dividing by mn. However, the intersection of $Y^3_{m\alpha}$ and $Y^3_{n\beta}$ inside W^5 consists of arcs, which connect pairs of intersection points of $S_{m\alpha}$ and $S_{n\beta}$ with opposite signs, as pictured in figure 3.8. It follows that $S_{m\alpha} \cdot S_{n\alpha} = 0$, and therefore $\alpha \cdot \beta = 0$, as claimed.



3.8. Bounding surfaces have zero intersection

If not bounding. Second, we claim that for every $\alpha \in H_2(M;\mathbb{Q})$ with $\iota\alpha \neq 0$ there must be some $\beta \in H_2(M;\mathbb{Q})$ so that $\alpha \cdot \beta = +1$ but $\iota\beta = 0$.

To see that, we notice that, since $\iota\alpha \neq 0$ in $H_2(W;\mathbb{Q})$, there exists a 3–class $B \in H_3(W, \partial W; \mathbb{Q})$ that is *dual*[19] to our $\iota\alpha \in H_2(W;\mathbb{Q})$, *i.e.*, has $\alpha \cdot B = +1$ in W^5. Its boundary $\partial B = \beta$ is a class in $H_2(M;\mathbb{Q})$, and we have that $\alpha \cdot \beta = \iota\alpha \cdot B = +1$ and also that $\iota\beta = 0$. See figure 3.9.



3.9. A non-bounding class has a bounding dual

Unravel the form. Finally, we are ready to attack the actual intersection form of M. Any class α that bounds in W, *i.e.*, has $\iota\alpha = 0$, must have zero self-intersection $\alpha \cdot \alpha = 0$. We are thus more interested in classes α that do not bound.

Assume we choose some $\alpha \in H_2(M;\mathbb{Q})$ so that $\iota\alpha \neq 0$. Then there is some $\beta \in H_2(M;\mathbb{Q})$ so that $\alpha \cdot \beta = +1$, while $\iota\beta = 0$, and thus $\beta \cdot \beta = 0$. Therefore the part of Q_M corresponding to $\{\alpha, \beta\}$ has matrix

$$Q_{\alpha\beta} = \begin{bmatrix} * & 1 \\ 1 & 0 \end{bmatrix},$$

which has determinant -1 and diagonalizes over $\mathbb{Q}$ as $[+1] \oplus [-1]$.

Since Q_M is unimodular, this means that Q_M must actually split as a direct sum $Q_M = Q_{\alpha\beta} \oplus Q^{\perp}$ for some unimodular form $Q^{\perp}$ defined on a complement of $\mathbb{Q}\{\alpha, \beta\}$ in $H_2(M;\mathbb{Q})$. Since the signature is additive and one can see that $\operatorname{sign} Q_{\alpha\beta} = 0$, we deduce that we must have $\operatorname{sign} Q_M = \operatorname{sign} Q^{\perp}$.

We continue this procedure for $Q^{\perp}$, splitting off 2–dimensional pieces until there are no more classes α with $\iota\alpha \neq 0$ left. Then whatever is still there has to bound in W, and hence cannot contribute to the signature. Therefore $\operatorname{sign} Q_M = 0$. □

---

[19] A reasoning analogous to the one we made earlier for Q_M applies to the intersection pairing $H_2(W;\mathbb{Z}) \times H_3(W, \partial W; \mathbb{Z}) \to \mathbb{Z}$. In particular, it is unimodular, and thus we have dual classes; since we work over $\mathbb{Q}$, the indivisibility of α is not required.

A consequence of this result is that, whenever two 4–manifolds can be linked by a cobordism, they must have the same signature. Indeed, if $\partial W = \overline{M} \cup N$, then $0 = \operatorname{sign}(\overline{M} \cup N) = -\operatorname{sign} M + \operatorname{sign} N$. That is:

Corollary. *If two manifolds are cobordant, then they have the same signature. Signature is a cobordism invariant.* □

The signature vanishes only for boundaries. A result quite more difficult to prove is the following:

Theorem (*V. Rokhlin*). *If a smooth oriented* 4*–manifold M has*

$$\operatorname{sign} Q_M = 0\,,$$

then there is a smooth oriented 5*–manifold W such that* $\partial W = M$.

> ***Idea of proof.*** A classic result of Whitney assures that any manifold X^m can be immersed in $\mathbb{R}^{2m-1}$; in particular, our M^4 can be immersed in $\mathbb{R}^7$. By performing various surgery modifications, we then arrange that M be cobordant to a 4–manifold M' that embeds in $\mathbb{R}^6$. Furthermore, a result of R. Thom[20] implies that M' must bound a 5–manifold W' inside $\mathbb{R}^6$. Attaching W' to the earlier cobordism from M to M' creates the needed W^5. A few more details for such a proof will be given in an inserted note on page 167. □

Therefore, the signature of M is zero if and only if M bounds. And hence:

Corollary (*Cobordisms and signatures*). *Two* 4*–manifolds have the same signature if and only if they are cobordant. Signature is the complete cobordism invariant.* □

A consequence is that, unlike h–cobordisms, simple cobordisms are not very interesting: *Every* 4*–manifold M is cobordant to a connected sum of* $\mathbb{CP}^2$*'s or of* $\overline{\mathbb{CP}}^2$*'s or to* S^4. Indeed, assume that $\operatorname{sign} M = m > 0$; then, since $\operatorname{sign} \mathbb{CP}^2 = 1$, it follows that M and $\#m\,\mathbb{CP}^2$ must be cobordant; if $m < 0$, use $\overline{\mathbb{CP}}^2$'s instead.

Simple examples of intersection forms

Since the first example of a 4–manifold that comes to mind, namely the sphere S^4, does not have any 2–homology, it has no intersection form worth mentioning. Thus, we move on:

---

[20] The result was quoted back in footnote 3 on page 112.

The complex projective plane. The complex projective plane $\mathbb{CP}^2$ has intersection form

$$Q_{\mathbb{CP}^2} = [+1] \ .$$

Indeed, since $H_2(\mathbb{CP}^2;\mathbb{Z}) = \mathbb{Z}\{[\mathbb{CP}^1]\}$ where $[\mathbb{CP}^1]$ is the class of a projective line, and since two projective lines always meet in a point, the equality above follows.

The oppositely-oriented manifold $\overline{\mathbb{CP}^2}$ has

$$Q_{\overline{\mathbb{CP}^2}} = [-1] \ .$$

Sphere bundles. The manifold $\mathbb{S}^2 \times \mathbb{S}^2$ has intersection form

$$Q_{\mathbb{S}^2\times\mathbb{S}^2} = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} \ .$$

We will denote this matrix by H (from "hyperbolic plane").

Reversing orientation does not exhibit a new manifold: there exist orientation-preserving diffeomorphisms $\mathbb{S}^2 \times \mathbb{S}^2 \ \cong\ \overline{\mathbb{S}^2 \times \mathbb{S}^2}$, and they correspond algebraically to isomorphisms $H \approx -H$.

The twisted product $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$ is the unique nontrivial sphere-bundle[21] over $\mathbb{S}^2$. It is obtained by gluing two trivial patches (*hemisphere*) $\times\ \mathbb{S}^2$ along the equator of the base-sphere, using the identification of the $\mathbb{S}^2$–fibers that rotates them by 2π as we travel along the equator. The intersection form is

$$Q_{\mathbb{S}^2\tilde{\times}\mathbb{S}^2} = \begin{bmatrix} 1 & 1 \\ 1 & \end{bmatrix} \ .$$

A simple change of basis in $H_2(\mathbb{S}^2 \tilde{\times} \mathbb{S}^2;\mathbb{Z})$ exhibits the intersection form as

$$Q_{\mathbb{S}^2\tilde{\times}\mathbb{S}^2} = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} = [+1] \oplus [-1] \ .$$

Even more, it is not hard to argue that in fact we have a diffeomorphism[22]

$$\mathbb{S}^2 \tilde{\times} \mathbb{S}^2 \ \cong\ \mathbb{CP}^2 \# \overline{\mathbb{CP}^2} \ ,$$

and so we have not really encountered anything essentially new.

---

[21] Since an $\mathbb{S}^2$–bundle over $\mathbb{S}^2 = \mathbb{D}^2{}_1 \cup \mathbb{D}^2{}_2$ is described by an equatorial gluing map $\mathbb{S}^1 \to SO(3)$, and $\pi_1 SO(3) = \mathbb{Z}_2$, it follows that there are only two topologically-distinct sphere-bundles over a sphere.

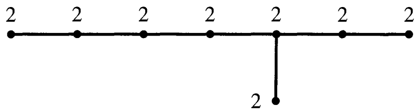
[22] Quick argument: The equatorial gluing map $\mathbb{S}^1 \to SO(3)$ of $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$ can be imagined as follows: as we travel along the equator of the base-sphere, it fixes the poles of the fiber-sphere and rotates the equator of the fiber-sphere by an angle increasing from 0 to 2π. Then these fiber-equators describe a circle-bundle of Euler number 1, which thus has to be the Hopf circle-bundle $\mathbb{S}^3 \to \mathbb{S}^2$. Hence the sphere-bundle is cut into two halves by a 3–sphere. Each of these halves is a disk-bundle of Euler number 1 and can therefore be identified with a neighborhood of $\mathbb{CP}^1$ inside $\mathbb{CP}^2$, but the complement of such a neighborhood is just a 4–ball. Taking care of orientations yields the splitting $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2 = \mathbb{CP}^2 \# \overline{\mathbb{CP}^2}$.

Connected sums. Of course, through the use of connected sums we can build a lot of boring examples, such as $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2 \# S^2 \times S^2$, whose intersection form is the sum $[+1] \oplus [-1] \oplus H$. (Incidentally, notice that this manifold has signature zero, and thus must be the boundary of some 5–manifold.)

The E_8–manifold. More interesting, though rather exotic, is Freedman's E_8–manifold $\mathcal{M}_{E_8} = P_{E_8} \cup_{\Sigma_P} \Delta$. This topological 4–manifold was built earlier[23] by plumbing on the E_8 diagram and capping with a fake 4–ball. Its intersection form can be read from the plumbing diagram to be

$$Q_{\mathcal{M}_{E_8}} = \begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix} .$$

From now on, we will denote this matrix[24] by E_8, and succinctly write $Q_{\mathcal{M}} = E_8$. The E_8–manifold does not admit any smooth structures.[25]



3.10. The E_8 diagram, yet again

An alternative algebraic description of this most important E_8–form is the following: Consider the form $Q = [-1] \oplus 8\,[+1]$, with corresponding basis $\{e_0, e_1, \ldots, e_8\}$. The vector $\kappa = 9e_0 + e_1 + \cdots + e_8$ has $\kappa \cdot \kappa = -1$; therefore its Q–orthogonal complement must be unimodular. This complement is the E_8–form. In particular, we have $E_8 \oplus [-1] \approx [-1] \oplus 8\,[+1]$.

Lemma. *The E_8–form is positive-definite, even, and of signature 8.*

Unexpectedly, proof. We will perform elementary operations on the rows and columns of the E_8–matrix. This will be fun.

---

[23] See section 2.3 (page 86).

[24] Various people have slightly different favorite choices for their E_8–matrix, for example, the negative of the above matrix. A brief discussion is contained in the end-notes of this chapter (page 137).

[25] This is a consequence of Rokhlin's theorem, see section 4.4 (page 170) ahead.

First off, notice that these operations must be applied symmetrically, corresponding to changes of basis in $H_2(M;\mathbb{Z})$. That is to say, when for example we subtract $3/2$ times the first row from the third, we must afterwards also subtract $3/2$ times the first column from the third column. Indeed, since the matrix A of a bilinear form acts on $H_2 \times H_2$ by $(x,y) \mapsto x^t A y$, any elementary change of basis $I + \lambda E_{ij}$ on H_2 will transform A into $(I+\lambda E_{ji})A(I+\lambda E_{ij})$.

Denote by $(1),(2),(3),(4),(5),(6),(7),(8)$ the eight rows/columns of the E_8-matrix, and let us start: We write down the E_8-matrix, then subtract $1/2 \times (1)$ from (2):

$$\begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix} \quad \text{then} \quad \begin{bmatrix} 2 & & & & & & & \\ & 3/2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix}.$$

Subtract $2/3 \times (2)$ from (3), then subtract $3/4 \times (3)$ from (4):

$$\begin{bmatrix} 2 & & & & & & & \\ & 3/2 & & & & & & \\ & & 4/3 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix} \quad \text{then} \quad \begin{bmatrix} 2 & & & & & & & \\ & 3/2 & & & & & & \\ & & 4/3 & & & & & \\ & & & 5/4 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix}.$$

Subtract $4/5 \times (4)$ from (5), then subtract $1/2 \times (8)$ from (5):

$$\begin{bmatrix} 2 & & & & & & & \\ & 3/2 & & & & & & \\ & & 4/3 & & & & & \\ & & & 5/4 & & & & \\ & & & & 6/5 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix} \quad \text{then} \quad \begin{bmatrix} 2 & & & & & & & \\ & 3/2 & & & & & & \\ & & 4/3 & & & & & \\ & & & 5/4 & & & & \\ & & & & 7/10 & 1 & & \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & & & & 2 \end{bmatrix}.$$

Subtract $10/7 \times (5)$ from (6), then subtract $7/4 \times (6)$ from (7):

$$\begin{bmatrix} 2 & & & & & & & \\ & 3/2 & & & & & & \\ & & 4/3 & & & & & \\ & & & 5/4 & & & & \\ & & & & 7/10 & & & \\ & & & & & 4/7 & 1 & \\ & & & & & 1 & 2 & \\ & & & & & & & 2 \end{bmatrix} \quad \text{then} \quad \begin{bmatrix} 2 & & & & & & & \\ & 3/2 & & & & & & \\ & & 4/3 & & & & & \\ & & & 5/4 & & & & \\ & & & & 7/10 & & & \\ & & & & & 4/7 & & \\ & & & & & & 1/4 & \\ & & & & & & & 2 \end{bmatrix}.$$

We have diagonalized E_8, and its signature is 8. It is positive-definite. Its determinant is $\det E_8 = 2 \cdot 3/2 \cdot 4/3 \cdot 5/4 \cdot 7/10 \cdot 4/7 \cdot 1/4 \cdot 2 = 1$ and hence E_8 is unimodular, as claimed. □

A few more examples. *(1) The intersection form of $\mathcal{M}_{E_8} \# \overline{\mathcal{M}}_{E_8}$ is $E_8 \oplus -E_8$. Algebraically, we have $E_8 \oplus -E_8 \;\approx\; \oplus 8\, H$ through a suitable change of basis. As it turns out, this corresponds to an actual homeomorphism*[26]

$$\mathcal{M}_{E_8} \# \overline{\mathcal{M}}_{E_8} \;\simeq\; \# 8\, \mathbb{S}^2 \times \mathbb{S}^2 \, .$$

Hence the smooth manifold $\# 8\, \mathbb{S}^2 \times \mathbb{S}^2$ can be cut into two non-smoothable topological 4–manifolds, along a topologically-embedded 3–sphere.

(2) The intersection form of $\mathcal{M}_{E_8} \# \overline{\mathbb{CP}}^2$ is $[-1] \oplus 8\, [+1]$, same as the intersection form of $\overline{\mathbb{CP}}^2 \# 8\, \mathbb{CP}^2$. The two 4–manifolds, though, are not *homeomorphic, and the manifold $\mathcal{M}_{E_8} \# \overline{\mathbb{CP}}^2$ does not admit any smooth structures.*[27]

(3) The manifold $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8}$, with intersection form $E_8 \oplus E_8$, is not smooth.[28] *Neither is $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8} \# \, \mathbb{S}^2 \times \mathbb{S}^2$, nor is $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8} \# 2\, \mathbb{S}^2 \times \mathbb{S}^2$. However, suddenly $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8} \# 3\, \mathbb{S}^2 \times \mathbb{S}^2$* does *admit smooth structures, and in what follows we will display such a smooth structure:*

3.3. Essential example: the *K*3 surface

A less exotic example (than the E_8–manifold) of a 4–manifold whose intersection form contains E_8's is the remarkable $K3$ complex surface that we build next:

The Kummer construction

Take the 4-torus

$$\mathbb{T}^4 = \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$$

and think of each $\mathbb{S}^1$–factor as the unit-circle inside $\mathbb{C}$. Consider the map

$$\sigma\colon \mathbb{T}^4 \to \mathbb{T}^4 \qquad \sigma(z_1, z_2, z_3, z_4) = (\bar{z}_1, \bar{z}_2, \bar{z}_3, \bar{z}_4)$$

given by complex-conjugation in each circle-factor, as in figure 3.11 on the next page. The involution σ has exactly $16 = 2^4$ fixed points, and thus the quotient

$$\mathbb{T}^4 / \sigma$$

will have sixteen singular points where it will fail to be a manifold. Small neighborhoods of these singular points are cones[29] on $\mathbb{RP}^3$.

We wish to surger away these singular points of $\mathbb{T}^4 / \sigma$ in order to obtain an actual 4–manifold. For that, we consider the *complex cotangent* bundle $T^*_{\mathbb{S}^2}$

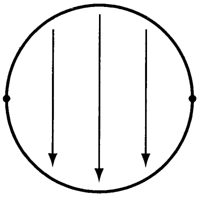
---

[26] This homeomorphism follows from Freedman's classification, see section 5.2 (page 239). A direct argument can also be made, starting with the observation that $\mathcal{M}_{E_8} \# \overline{\mathcal{M}}_{E_8}$ is the boundary of $(\mathcal{M}_{E_8} \setminus$ *ball*$) \times [0, 1]$.

[27] This follows, again, from Freedman's classification.

[28] This is a consequence of Donaldson's theorem, section 5.3 (page 243).

[29] Remember that the **cone** C_A of a space A is simply the result of taking $A \times [0, 1]$ and collapsing $A \times 1$ to a single point (the "vertex").



3.11. Conjugation, acting on $\mathbb{S}^1$

of the 2–sphere. It is the 2–plane bundle over $\mathbb{S}^2$ with Euler number -2 (it has *opposite orientation*[30] to the tangent bundle $T_{\mathbb{S}^2}$, whose Euler number is $+2$). Its unit-disk subbundle $\mathbb{D}T^*_{\mathbb{S}^2}$ is a 4–manifold bounded by $\mathbb{RP}^3$.

Since a neighborhood of a singular point in $\mathbb{T}^4/\sigma$ has the same boundary as $\mathbb{D}T^*_{\mathbb{S}^2}$, we can cut the former out of $\mathbb{T}^4/\sigma$ and replace it by a copy of $\mathbb{D}T^*_{\mathbb{S}^2}$. The result of this maneuver is essentially to remove the singular point and replace it with a sphere of self-intersection -2 (the zero-section of $\mathbb{D}T^*_{\mathbb{S}^2}$). We do this for all sixteen singular points.

Such a desingularization of $\mathbb{T}^4/\sigma$ yields a *simply-connected* smooth 4–manifold. This manifold admits a complex structure (thus it is a complex surface) and is called the ***K*3 surface.** The name comes from Kummer–Kähler–Kodaira.[31] The construction above is due to Kummer, which is why this manifold used to be known merely as the *Kummer surface.*

Homology. The $K3$ surface has homology $H_2(K3;\mathbb{Z}) = \oplus\, 22\mathbb{Z}$ (superficially, from 6 tori surviving from $\mathbb{T}^4$, plus the 16 desingularizing spheres). Its intersection form is

$$Q_{K3} = -\begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix} \oplus -\begin{bmatrix} 2 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ & 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & & \\ & & & 1 & 2 & 1 & & 1 \\ & & & & 1 & 2 & 1 & \\ & & & & & 1 & 2 & \\ & & & & 1 & & & 2 \end{bmatrix}$$

$$\oplus \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} \oplus \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} \oplus \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}$$

and clearly it is better kept abbreviated as

$$Q_{K3} = \oplus\, 2(-E_8) \oplus 3H\,.$$

---

[30] For a discussion of orientations for complex-duals, see the end-notes of this chapter (page 134).

[31] A. Weil wrote that, besides honoring Kummer, Kodaira and Kähler, the name "*K*3" was also chosen in relation to the famous K2 peak in the Himalayas: *"[Surfaces] ainsi nommées en l'honneur de Kummer, Kähler, Kodaira, et de la belle montagne K2 au Cachemire."*

Even if this manifold does not seem simple at all, it is in many ways as simple as it gets. We will see that $K3$ is indeed the *simplest*[32] simply-connected *smooth* 4–manifold that is not $\mathbb{S}^4$ nor a boring sum of $\mathbb{CP}^2$, $\overline{\mathbb{CP}}^2$ and $\mathbb{S}^2 \times \mathbb{S}^2$'s.

The desingularization, revisited. Let us take a closer look at the desingularization of $\mathbb{T}^4/\sigma$ that created $K3$ and try to better visualize it.

Consider first a neighborhood inside $\mathbb{T}^4$ of a fixed point x_0 of σ. It is merely a 4–ball, which can be viewed as a cone over its boundary 3–sphere $\mathbb{S}^3$, with vertex at x_0. The action of σ on this cone can itself be viewed as being the cone[33] of the antipodal map $\mathbb{S}^3 \to \mathbb{S}^3$ (which sends w to $-w$). Therefore, the quotient of this neighborhood of x_0 by σ must be a cone on the quotient of $\mathbb{S}^3$ by the antipodal map, in other words, a cone on $\mathbb{RP}^3$.

Furthermore, $\mathbb{S}^3$ is fibrated by the Hopf map,[34] which makes it into a bundle with fiber $\mathbb{S}^1$ and base $\mathbb{S}^2$. Then its quotient $\mathbb{RP}^3$ inherits a structure of $\mathbb{RP}^1$–bundle over $\mathbb{S}^2$:

$$\begin{tikzcd}
\mathbb{S}^1 \arrow[r, phantom, "\subset"] \arrow[d, "2"'] & \mathbb{S}^3 \arrow[r] \arrow[d, "2"'] & \mathbb{S}^2 \arrow[d, equal] \\
\mathbb{RP}^1 \arrow[r, phantom, "\subset"] & \mathbb{RP}^3 \arrow[r] & \mathbb{S}^2.
\end{tikzcd}$$

However, $\mathbb{RP}^1$ is simply a circle, so in fact we exhibited $\mathbb{RP}^3$ as an $\mathbb{S}^1$–bundle over $\mathbb{S}^2$.

Now let us look back at the neighborhood of a singular point of $\mathbb{T}^4/\sigma$. It is a cone on $\mathbb{RP}^3$, and we can think of it as being built by attaching a disk to each circle-fiber of $\mathbb{RP}^3$, and then identifying all their centers in order to obtain the vertex of the cone, the singular point. When we desingularize, we replace this cone-neighborhood in $\mathbb{T}^4/\sigma$ with a copy of $\mathbb{D}T^*_{\mathbb{S}^2}$. This can be viewed simply as *not* identifying the centers of those disks attached to the fibers of $\mathbb{RP}^3$, but keeping them disjoint. The space of the circle-fibers of $\mathbb{RP}^3$ is the base $\mathbb{S}^2$ of the fibration. Thus the space of the attached disks is $\mathbb{S}^2$ as well, and thus their centers (now distinct) will draw a new 2–sphere, which replaced the singular point.

We can thus think of our desingularization as simply replacing each of the sixteen singular points of $\mathbb{T}^4/\sigma$ by a sphere with self-intersection -2.

---

[32] We take "simple" to include "simple to describe". Smooth manifolds with simpler intersection forms already exist (*e.g.*, exotic $\#m\,\mathbb{S}^2 \times \mathbb{S}^2$'s, see page 553), and exotic $\mathbb{S}^4$'s could always appear.

[33] Remember that the **cone** C_f of a map $f\colon A \to B$ is the function $C_f\colon C_A \to C_B$ defined by first extending $f\colon A \to B$ to $f \times id\colon A \times [0,1] \to B \times [0,1]$, then collapsing $A \times 1$ to a point and $B \times 1$ to another, with the the resulting function $C_f\colon C_A \to C_B$ sending vertex to vertex.

[34] Remember that the **Hopf map** is defined to send a point $x \in \mathbb{S}^3 \subset \mathbb{C}^2$ to the point from $\mathbb{S}^2 = \mathbb{CP}^1$ that represents the complex line spanned by x inside $\mathbb{C}^2$. Topologically, the Hopf bundle $\mathbb{S}^3 \to \mathbb{S}^2$ is a circle-bundle of Euler class $+1$. Two distinct fibers will be two circles in $\mathbb{S}^3$ linked once (a so-called Hopf link, see figure 8.16 on page 318). The Hopf map $\mathbb{S}^3 \to \mathbb{S}^2$ represents the generator of $\pi_3\mathbb{S}^2 = \mathbb{Z}$.

Holomorphic construction

A complex geometer would construct the Kummer $K3$ in a way that visibly exhibits its complex structure. Specifically, she would start with $\mathbb{T}^4$ being a complex torus—for example the simplest such, the product of two copies of $\mathbb{C}/(\mathbb{Z} \oplus i\mathbb{Z})$. Such a $\mathbb{T}^4$ comes equipped with complex coordinates (w_1, w_2), and the involution σ can be described as $\sigma(w_1, w_2) = (-w_1, -w_2)$ (which is obviously holomorphic).

As before, the action of σ has sixteen fixed points, but, before taking the quotient, the complex geometer will blow-up[35] $\mathbb{T}^4$ at these sixteen points. This has the result of replacing each fixed point of σ with a sphere of self-intersection -1 (a neighborhood of which looks like a neighborhood of $\overline{\mathbb{CP}}^1$ inside $\overline{\mathbb{CP}}^2$). The map σ can be extended across this blown-up 4–torus: since she replaced the fixed points of σ by spheres, she can extend σ across the new spheres simply as the identity, thus letting the whole spheres be fixed by the resulting σ.

Only now will the complex geometer take the quotient by σ of the blown-up 4–torus. The result is the $K3$ surface. The spheres of self-intersection -1 created when blowing-up the torus will project to the quotient $K3$ as themselves (they were fixed by σ), but their neighborhoods are doubly-covered through the action of σ; thus these spheres inside $K3$ have now self-intersection -2.

Many K3's. This is the place to note that a complex geometer will in fact see a *multitude* of $K3$ surfaces. Indeed, "*K3*" is not the name of *one* complex surface, but the name of a class of surfaces.[36] Any non-singular simply-connected complex surface with $c_1 = 0$ is a ***K*3 surface.**

For example, in the construction above, if we start with a different complex structure on $\mathbb{T}^4$ (from factoring $\mathbb{C}^2$ by a different lattice), then we will end up with a different $K3$ surface. All $K3$'s that result from such a construction are called **Kummer surfaces.** However, $K3$ surfaces can be built in many other ways. One example is the hypersurface of $\mathbb{CP}^3$ given by the homogeneous equation

$$z_1^4 + z_2^4 + z_3^4 + z_4^4 = 0$$

(or any other smooth surface of degree 4). Another is the $E(2)$ elliptic surface that we will describe in chapter 8 (page 301).

This whole multitude of complex $K3$ surfaces, through the blinded eyes of the topologist, are just *one* smooth 4–manifold: any two $K3$'s are complex-deformations of each other, and thus are diffeomorphic. Hence, in this book we will carelessly be saying "*the* $K3$ surface".

---

[35] For a discussion of blow-ups, see ahead section 7.1 (page 286).

[36] For instance, the moduli space of all $K3$ surfaces has dimension 20.

$K3$ as an elliptic fibration

The $K3$ surface can be structured as a singular fibration over $\mathbb{S}^2$, with generic fiber a torus. A (singular) fibration by tori of a complex surface is called an **elliptic fibration** (because a torus in complex geometry is called an *elliptic curve*). A complex surface that admits an elliptic fibration is called an **elliptic surface.** The Kummer $K3$ is such an elliptic surface. Other examples of elliptic surfaces, as well as a different elliptic fibration on the $K3$ manifold, will be discussed later.[37]

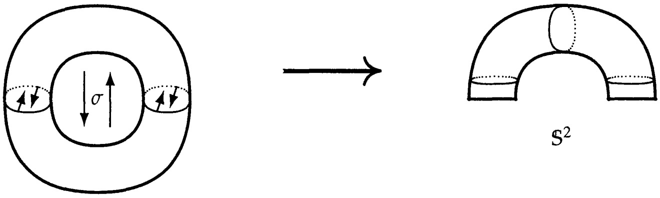
In any case, describing the elliptic fibration of $K3$ will help us better visualize this manifold. To exhibit it, we start with the projection

$$\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1 \longrightarrow \mathbb{S}^1 \times \mathbb{S}^1$$

of $\mathbb{T}^4$ onto its first two factors. After taking the quotient by the action of σ, this projection descends to a map

$$\mathbb{T}^4/\sigma \longrightarrow \mathbb{T}^2/\sigma\,.$$

Its target $\mathbb{T}^2/\sigma$ is a non-singular sphere $\mathbb{S}^2$, as suggested in figure 3.12 (it seems like it has four singular points at the corners, but these are merely metric-singular, and can be smoothed over).



3.12. Obtaining the base sphere: $\mathbb{T}^2/\sigma = \mathbb{S}^2$

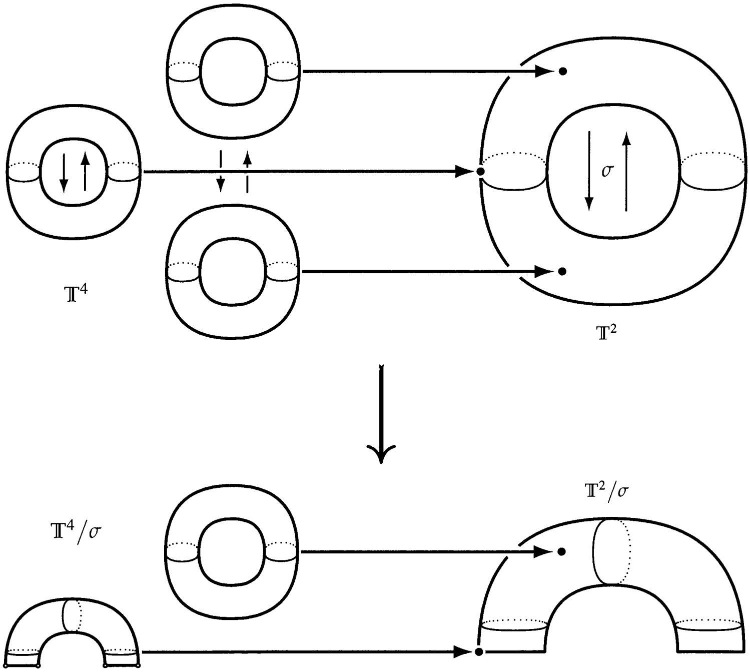
Aside from the corner-points of the base-sphere $\mathbb{T}^2/\sigma$, each of its other points comes from two distinct points (p,q) and $(\overline{p},\overline{q})$ of $\mathbb{T}^2$ identified by σ. Thus, the fiber of the map $\mathbb{T}^4/\sigma \to \mathbb{T}^2/\sigma$ over a generic point appears from σ's identifying two distinct tori $p \times q \times \mathbb{S}^1 \times \mathbb{S}^1$ and $\overline{p} \times \overline{q} \times \mathbb{S}^1 \times \mathbb{S}^1$ from $\mathbb{T}^4$. The resulting fiber will itself be a torus. This is the generic fiber of $\mathbb{T}^4/\sigma \to \mathbb{T}^2/\sigma$. See also figure 3.13 on the following page.

On the other hand, each of the four corner-points of the sphere $\mathbb{T}^2/\sigma$ comes from a single fixed point (p_0,q_0) of σ on $\mathbb{T}^2$. Thus, the fiber of $\mathbb{T}^4/\sigma \to \mathbb{T}^2/\sigma$ over such a corner appears from σ's sending a torus $p_0 \times q_0 \times \mathbb{S}^1 \times \mathbb{S}^1$ to itself. The quotient of this torus is again a cornered-sphere (just as before, in figure 3.12), but now its corners coincide with the sixteen global fixed points of σ on $\mathbb{T}^4$. In other words, each such sphere-fiber contains four

---

[37] See chapter 8 (starting on page 301), which is devoted to these creatures.

of the sixteen singular points of the quotient $\mathbb{T}^4/\sigma$, points where the latter fails to be a manifold. See again figure 3.13.



3.13. The map $\mathbb{T}^4/\sigma \to \mathbb{T}^2/\sigma$ and its fibers

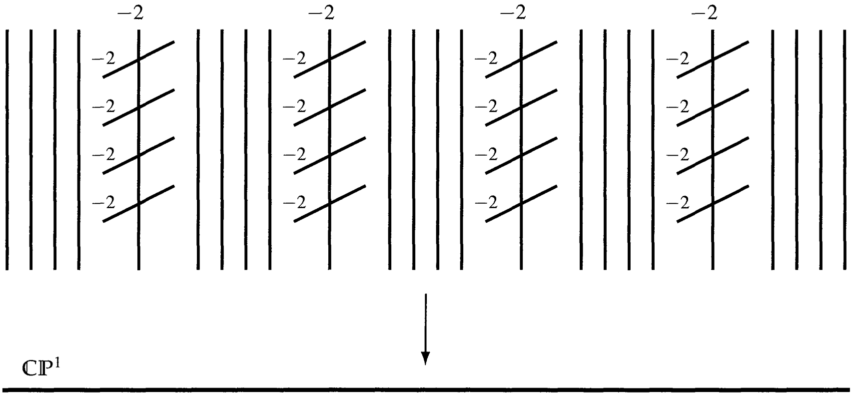
> *This might be a good moment to notice that $\mathbb{T}^4/\sigma$ is simply-connected. It fibrates over $\mathbb{S}^2$, which is simply-connected, and any loop in a generic torus fiber can be moved along to one of the singular sphere-fibers and contracted there. The desingularization of $\mathbb{T}^4/\sigma$ into $K3$ does not create any new loops, and therefore the $K3$ surface is, as claimed, simply-connected.*

As explained before, we cut neighborhoods of the singular points out of $\mathbb{T}^4/\sigma$ and glue a copy of $\mathbb{D}T^*_{\mathbb{S}^2}$ in their stead, thus replacing each singular point by a sphere; the result is the $K3$ surface. The projection $\mathbb{T}^4/\sigma \to \mathbb{T}^2/\sigma$ survives the desingularization as a map

$$K3 \longrightarrow \mathbb{S}^2 .$$

Indeed, since we only replaced sixteen points by sixteen spheres, we can send each of these spheres wherever the removed point used to go in $\mathbb{S}^2$.

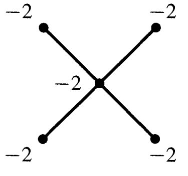
The generic fiber of $K3 \to \mathbb{S}^2$ is still a torus. However, there are now also four singular fibers, each made of five transversely-intersecting spheres: the old singular sphere-fiber of $\mathbb{T}^4/\sigma$, together with its four desingularizing spheres. A symbolic picture of this fibration is figure 3.14.



3.14. $K3$ as the Kummer elliptic fibration

Observe that the main sphere of the singular fiber must have self-intersection -2*. This can be can argued as follows: Denote by* S *the main sphere of a singular fiber and by* S_1, S_2, S_3, S_4 *the desingularizing spheres. Recall how the main sphere* S *appeared from factoring by* σ*: doubly-covered by a torus. Imagine a moving generic torus-fiber* F *of* $K3$ *approaching our singular fiber: it will wrap around the main sphere twice, covering it. Also, the approaching fiber will extend to cover the desingularizing spheres once, and so in homology we have* $F = 2S + S_1 + S_2 + S_3 + S_4$*. We know that* $F \cdot F = 0$ *(since it is a fiber), and that each* $S_k \cdot S_k = -2$*; then one can compute that we must also have* $S \cdot S = -2$*.*

Finally, note that a neighborhood of the singular fiber inside $K3$ can be obtained by plumbing five copies of $\mathbb{D}T^*_{S^2}$ following the diagram from figure 3.15.



3.15. Plumbing diagram for neighborhood of singular fiber

3.4. Notes

Note: Duals of complex bundles and orientations

The pretext for this note is to explain why the cotangent bundle $T^*_{S^2}$ (used earlier for building $K3$) has Euler class -2 rather than $+2$; that is to say, why $T^*_{S^2}$ and T_{S^2} have opposite orientations.

Let V be a real vector space, endowed with a complex structure. There are two ways to think of such a creature: (1) we can view V as a complex vector space, in other words, think of it as endowed with an action of the complex scalars $\mathbb{C} \times V \to V$ that makes V into a vector space over the field of complex numbers; or (2) we can view V as a real space endowed with an automorphism $J \colon V \to V$ with the property that $J \circ J = -id$. One should think of this J as a proxy for the multiplication by i. The two views are clearly equivalent, related by

$$J(v) = i \cdot v\,.$$

Nonetheless, they naturally lead to two different versions of a complex structure for the dual vector space.

The real version. Let us first discuss the case when we view V as a real vector space endowed with an anti-involution J. As a real vector space, the dual of V is

$$V^* = \mathrm{Hom}_{\mathbb{R}}(V;\mathbb{R})\,.$$

A vector space and its dual are isomorphic, but there is no natural choice of isomorphism. To fix a choice of such an isomorphism, we endow V with an auxiliary inner-product $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathbb{R}}$. Then V and V^* are naturally isomorphic through

$$V \xrightarrow{\approx} V^*\colon \qquad v \longmapsto v^* = \langle\,\cdot\,, v\rangle_{\mathbb{R}}\,.$$

If V is endowed with a complex structure J, then it is quite natural to restrict the choice of inner-product to those that are compatible with J. This means that we only choose inner-products that are invariant under J: we require that

$$\langle Jv,\, Jw\rangle_{\mathbb{R}} = \langle v, w\rangle_{\mathbb{R}}\,.$$

An immediate consequence is that we have $\langle Jv,\, w\rangle_{\mathbb{R}} = -\langle v,\, Jw\rangle_{\mathbb{R}}$.

We now wish to endow the dual V^* with a complex structure of its own. In other words, we want to define a natural anti-involution $J^* \colon V^* \to V^*$ induced by J. Since an isomorphism $V \approx V^*$ was already chosen, it makes sense now to simply transport J from V to V^* through that isomorphism. Namely, we define the complex structure J^* of V^* by

$$J^*(v^*) = (Jv)^*\,.$$

More explicitly, if $f \in V^*$ is given by $f(x) = \langle x, v\rangle_{\mathbb{R}}$ for some $v \in V$, then $(J^*f)(x) = \langle x, Jv\rangle_{\mathbb{R}}$. However, this means that $(J^*f)(x) = -\langle Jx, v\rangle_{\mathbb{R}}$, and so we have

$$J^*f = -f(J\cdot)\,.$$

Notice that we ended up with a formula that does *not* depend on the choice of inner-product. Hence we have defined a natural complex structure J^* on the real vector space $V^* = \mathrm{Hom}_{\mathbb{R}}(V,\mathbb{R})$.

The complex version. If, on the other hand, we think of the complex structure of V as an action of the complex scalars that makes V into a vector space $V_{\mathbb{C}}$ over the complex numbers, then a different notion of dual space comes to the fore. We must define the dual as

$$V_{\mathbb{C}}^* = \mathrm{Hom}_{\mathbb{C}}(V,\mathbb{C})\,.$$

This vector space comes from birth equipped with a complex structure, namely

$$(i\cdot f)(x) = i\,f(x)$$

for every $f \in V_{\mathbb{C}}^*$. To better grasp what this $V_{\mathbb{C}}^*$ looks like, we will endow $V_{\mathbb{C}}$ with an auxiliary inner-product. The appropriate notion of inner-product for complex vector spaces is that of **Hermitian** inner-products. This differs from the usual inner products by the facts that it is complex-valued, and it is complex-linear in its first variable, but complex anti-linear in the second. We have $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathbb{C}} : V \times V \to \mathbb{C}$ with $\langle zv,\, w\rangle_{\mathbb{C}} = z\langle v, w\rangle_{\mathbb{C}}$, but $\langle v,\, zw\rangle_{\mathbb{C}} = \bar{z}\langle v, w\rangle_{\mathbb{C}}$ for every[1] $z \in \mathbb{C}$.

Any Hermitian inner product can then be used to define a complex-isomorphism of $V_{\mathbb{C}}^*$, though not with $V_{\mathbb{C}}$, but with its **conjugate** vector space $\overline{V_{\mathbb{C}}}$. The latter is defined as being the real vector space V endowed with an action of complex scalars that is conjugate to that of $V_{\mathbb{C}}$. That is to say, in $\overline{V_{\mathbb{C}}}$ we have $i\cdot v = -iv$. The complex-isomorphism with the dual is:

$$\overline{V_{\mathbb{C}}} \xrightarrow{\approx} V_{\mathbb{C}}^*: \qquad v \longmapsto v^* = \langle\,\cdot\,, v\rangle_{\mathbb{C}}\ .$$

Notice that in the definition of v^* we must put v as the second entry in $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathbb{C}}$, so that v^* be a complex-linear function and thus indeed belong to $V_{\mathbb{C}}^*$.

If $f \in V_{\mathbb{C}}^*$ is given by $f(x) = \langle x, v\rangle_{\mathbb{C}}$ for some $v \in V$, then we have $(if)(x) = i\,f(x) = i\langle x, v\rangle_{\mathbb{C}} = \langle x,\, -iv\rangle_{\mathbb{C}}$. This means that we have

$$i\cdot v^* = (-iv)^*\,,$$

which shows that the complex-isomorphism above is indeed between the dual $V_{\mathbb{C}}^*$ and the *conjugate* vector space $\overline{V_{\mathbb{C}}}$.

Comparison. In review, if we view a complex vector space as (V, J), then its dual is (V^*, J^*) and the two are complex-isomorphic. If we view a complex vector space as $V_{\mathbb{C}}$, then its dual is $V_{\mathbb{C}}^*$, which is complex-isomorphic to $\overline{V_{\mathbb{C}}}$. To compare the two versions, it is enough to notice that $\overline{V_{\mathbb{C}}}$ translates simply as $(V, -J)$. Indeed, as *real* vector spaces (*i.e.*, ignoring the complex structures) V^* and $V_{\mathbb{C}}^*$ are

---

[1] It is worth noticing that the concept of a real inner product compatible with a complex structure, and the concept of Hermitian inner product are equivalent: one can go from one to the other by using $\langle v, w\rangle_{\mathbb{C}} = \langle v, w\rangle_{\mathbb{R}} - i\,\langle iv,\, w\rangle_{\mathbb{R}}$ and $\langle v, w\rangle_{\mathbb{R}} = \mathrm{Re}\,\langle v, w\rangle_{\mathbb{C}}$.

naturally isomorphic. Specifically, the isomorphism $\mathrm{Hom}_{\mathbb{R}}(V,\mathbb{R}) \approx \mathrm{Hom}_{\mathbb{C}}(V,\mathbb{C})$ sends $f\colon V \to \mathbb{R}$ to the function $f_{\mathbb{C}}\colon V \to \mathbb{C}$ given by

$$f_{\mathbb{C}}(x) = \tfrac{1}{2}\big(f(x) - if(Jx)\big) .$$

The duals (V^*, J^*) and $V^*_{\mathbb{C}}$ thus differ not as real vector spaces, but because their complex structures are conjugate. This could be checked directly against the isomorphism above, or, in the simplifying presence of an inner-product, we could simply write:

$$J^*(v^*) = (iv)^* \qquad \text{and} \qquad i \cdot v^* = (-iv)^* .$$

Usage. We should emphasize that, while the "complex" version of dual is certainly the most often used, nonetheless both these versions are important.

As a typical example, consider a complex manifold X, which is endowed with a tangent bundle T_X and a cotangent bundle T^*_X. Owing to the complex structure of X, the tangent bundle has a natural complex structure on its fibers. The complex structure on T^*_X is *always* taken to be dual to the one on T_X in its "complex" version: as complex bundles, we have $T^*_X \approx \overline{T_X}$. In general for vector bundles with complex structures, the dual is usually taken to be the "complex" dual.

The "real" version of dual is also used in complex geometry. Thinking now of the complex structure of T_X as $J\colon T_X \to T_X$, we let it induce its own dual complex structure J^* on T^*_X. We then extend J^* by linearity to the complexified vector space $T^*_X \otimes_{\mathbb{R}} \mathbb{C}$. The advantage of such an extension is that now J^* has eigenvalues $\pm i$, and thus splits the bundle $T^*_X \otimes \mathbb{C}$ into its $\pm i$–eigenbundles as

$$T^*_X \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1} ,$$

and hence separates complex-valued 1–forms on X into type $(1,0)$ and type $(0,1)$. This is simply a splitting into complex-linear and complex-anti-linear parts: indeed $J^*(\alpha) = -i\alpha$ if and only if $\alpha(Jx) = +i\alpha(x)$, and then $\alpha \in \Lambda^{1,0}$.

The advantage of using J lies in part with clarity of notation: for a complex-valued creature, J will denote the complex action on its arguments (living on X), while i denotes the complex action on its values (living in $\mathbb{C}$).

More on complex-valued forms. *Every complex-valued function $f\colon X \to \mathbb{C}$ has its differential $df \in \Gamma(T^*_X \otimes \mathbb{C})$ split into its $(1,0)$-part $\partial f \in \Gamma(\Lambda^{1,0})$ and its $(0,1)$-part $\bar\partial f \in \Gamma(\Lambda^{0,1})$. Hence, $\bar\partial f = 0$ means that f's derivative is complex-linear, $df(Jx) = i\,df$, and thus that f is* holomorphic.

By using local real coordinates $(x_1, y_1, \dots, x_m, y_m)$ on X such that $z_k = x_k + iy_k$ are local complex coordinates on X, we can define $dz_k = dx_k + i\,dy_k$ and $d\bar z_k = dx_k - i\,dy_k$, and write $\Lambda^{1,0} = \mathbb{C}\{dz_1, \dots, dz_m\}$ and $\Lambda^{0,1} = \mathbb{C}\{d\bar z_1, \dots, d\bar z_m\}$. Indeed, $J^*(d\bar z_k) = +i\,d\bar z_k$.

The split $\Lambda^1 \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1}$ further leads to a splitting of all complex-valued forms into (p,q)-types, as in $\Lambda^k \otimes \mathbb{C} = \Lambda^{k,0} \oplus \Lambda^{k-1,1} \oplus \cdots \oplus \Lambda^{1,k-1} \oplus \Lambda^{0,k}$. Specifically, $\Lambda^{p,q}$ is made of all complex-valued forms that can be written using p of the dz_k's and q of the $d\bar z_k$'s. For example, $\Lambda^{2,0}$ contains all complex-bilinear 2-forms.

The exterior differential $d\colon \Gamma(\Lambda^k) \to \Gamma(\Lambda^{k+1})$ splits, after complexification, as $d = \partial + \bar\partial$ with $\partial\colon \Gamma(\Lambda^{p,q}) \to \Gamma(\Lambda^{p+1,q})$ and $\bar\partial\colon \Gamma(\Lambda^{p,q}) \to \Gamma(\Lambda^{p,q+1})$. Since $\bar\partial\bar\partial = 0$, this can be used to define cohomology groups $H^{p,q}(X) = \mathrm{Ker}\,\bar\partial \,/\, \mathrm{Im}\,\bar\partial$ (called **Dolbeault cohomology***), which offer a cohomology splitting $H^k(X;\mathbb{C}) = H^{k,0}(X) \oplus H^{k-1,1}(X) \oplus \cdots \oplus H^{1,k-1}(X) \oplus H^{0,k}(X)$, with $H^{p,q}(X) \approx \overline{H^{q,p}(X)}$; further, if X is Kähler, then the Hodge duality operator[2] $*$ will take*

---

[2] The Hodge operator will be recalled in section 9.3 (page 350).

(p,q)-forms to $(m-q,\ m-p)$-forms, and lead into complex Hodge theory, to just drop some names. Any complex geometry book will explain these topics properly, for example **P. Griffiths** *and* **J. Harris***'s* ***Principles of algebraic geometry*** **[GH78, GH94]***; we ourselves will make use of (p,q)-forms for some technical points later on.*[3] *Part of this topic will be explained in more detail in the end-notes of chapter 9 (page 365).*

Orientations. Every vector space with a complex structure (defined either way) is naturally oriented by any basis like $\{e_1, ie_1, \ldots, e_k, ie_k\}$ (or $\{e_1, Je_1, \ldots, e_k, Je_k\}$). Thus its dual vector space, getting a complex structure itself, will be naturally oriented as well. However, the choice of duality matters: if our vector space V is odd-dimensional (over $\mathbb{C}$), then the two versions of dual complex structure lead to *opposite* orientations of V's dual. Specifically, the real-isomorphism $V \approx V^*_{\mathbb{C}}$ reverses orientations, while $V \approx (V^*, J^*)$ preserves them.

For complex manifolds and their tangent/cotangent bundles, as we mentioned above, one uses the "complex" version of duality. Therefore, for a complex curve C (for example, $\mathbb{S}^2$) we have that the tangent bundle T_C and the cotangent bundle T^*_C, while isomorphic as real bundles, are naturally oriented by opposite orientations. In particular, the tangent bundle $T_{\mathbb{S}^2}$ is the plane bundle of Euler class $+2$, while the cotangent bundle $T^*_{\mathbb{S}^2}$ is the plane bundle with Euler class -2.

For a complex surface M (for example, $K3$), the tangent and cotangent bundles do not have opposite orientations. Nonetheless, their complex structures are conjugate, and this leads to phenomena like $c_1(T^*_M) = -c_1(T_M)$.

Note: Positive E_8, negative E_8

In some texts, the E_8-form is sometimes described by the matrix

$$E_8 \approx \begin{bmatrix} 2 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & -1 & 2 & -1 & & & \\ & & & -1 & 2 & -1 & & -1 \\ & & & & -1 & 2 & -1 & \\ & & & & & -1 & 2 & \\ & & & & -1 & & & 2 \end{bmatrix}.$$

Correspondingly, the negative-E_8-form is sometimes written

$$-E_8 \approx \begin{bmatrix} -2 & 1 & & & & & & \\ 1 & -2 & 1 & & & & & \\ & 1 & -2 & 1 & & & & \\ & & 1 & -2 & 1 & & & \\ & & & 1 & -2 & 1 & & 1 \\ & & & & 1 & -2 & 1 & \\ & & & & & 1 & -2 & \\ & & & & 1 & & & -2 \end{bmatrix}.$$

These alternative matrices are in fact equivalent with the ones presented earlier, because one can always find an isomorphism between the two versions: simply change the sign of "every other" element of the basis. Then the self-intersections

---

[3] In section 6.2 (page 278), the end-notes of chapter 9 (connections and holomorphic bundles, page 365) and the end-notes of chapter 10 (Seiberg–Witten on Kähler and symplectic, page 457).

are preserved, but, if done properly, the intersections between distinct elements will all change signs. Peek back at the E_8 diagram for inspiration.

Complex geometers always prefer to have $+1$'s off the diagonal (thinking in terms of complex submanifolds, which always intersect positively), and so they will write $-E_8$ in the version displayed above.

More than this, certain texts prefer to switch the names of the E_8– and negative–E_8–matrices. Since what we denote here by $-E_8$ appears quite more often than E_8, calling it E_8 does save some writing.

Pick your own favorites.

Bibliography

For strong and general results about representing homology classes by submanifolds, see **R. Thom**'s celebrated paper ***Quelques propriétés globales des variétés différentiables*** **[Tho54]** (the results were first announced in **[Tho53a]**).

For a definition of intersections directly in terms of cycles (not necessarily submanifolds), see **P. Griffiths** and **J. Harris**'s ***Principles of algebraic geometry*** **[GH78, GH94, sec 0.4]**; there one can also find an intersection-based view of Poincaré duality. For the rigorous algebraic topology development of various products and pairings of co/homology, see **A. Dold**'s ***Lectures on algebraic topology*** **[Dol80, Dol95]**; the differential-forms approach is best culled from **R. Bott** and **L. Tu**'s ***Differential forms in algebraic topology*** **[BT82]**.

Intersection forms can be defined in all dimensions $4k$, and their signature is an important invariant. For example, it is the main surgery obstruction in those dimensions, see for example **W. Browder**'s ***Surgery on simply-connected manifolds*** **[Bro72]**.

That all 4–manifolds of zero signature bound was proved in **V. Rokhlin**'s ***New results in the theory of four-dimensional manifolds*** **[Rok52]**, alongside his celebrated Rokhlin's theorem that we will discuss in the next chapter. A French translation of the paper can be read as **[Rok86]** from the volume ***À la recherche de la topologie perdue*** **[GM86a]**. A geometric proof of this result can be read from **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89, ch VIII]**, and a slightly more complete outline will be presented on page 167 ahead. A different-flavored proof is contained in **R. Stong**'s ***Notes on cobordism theory*** **[Sto68]**.

The E_8–manifold was defined, alongside the fake 4–balls Δ, in **M. Freedman**'s ***The topology of four-dimensional manifolds*** **[Fre82]**; see also, of course, **M. Freedman** and **F. Quinn**'s ***Topology of 4–manifolds*** **[FQ90]**.

For the $K3$ surface, the algebro-geometric point-of-view is discussed at length in **W. Barth, C. Peters** and **A. Van de Ven**'s ***Compact complex surfaces*** **[BPVdV84]** (or the second edition **[BHPVdV04]**, with **K. Hulek**). Also, inevitably, in **P. Griffiths** and **J. Harris**'s ***Principles of algebraic geometry*** **[GH78, GH94]**. For a topological point-of-view on $K3$, see **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** **[GS99]**, or **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89]**. We will come back to the $K3$ surface ourselves in chapter 8 (starting on page 301), where we will discuss it alongside its elliptic-surface brethren.

Chapter 4

Intersection Forms and Topology

We explore in what follows the topological ramifications of a 4–manifold having a certain intersection form. The results discussed are classical, such as Whitehead's theorem, Wall's theorems, and Rokhlin's theorem. All classification results are postponed until the next chapter.

We start by showing that the intersection form determines the homotopy type of a 4–manifold. This theorem of Whitehead is argued in two ways, once by using homotopy theory and once through a Pontryagin–Thom argument. The end-notes (page 230) contain a more general discussion of the Pontryagin–Thom technique.

In section 4.2 (page 149) we explain the results of C.T.C. Wall: first, if two smooth 4–manifolds are h–cobordant, then they become diffeomorphic after summing with enough copies of $\mathbb{S}^2 \times \mathbb{S}^2$; second, if two smooth 4–manifolds have the same intersection form, then they must be h–cobordant. Notice that this last result can be combined with M. Freedman's h–cobordism theorem to show that two smooth 4–manifolds with the same intersection forms must be homeomorphic.

In section 4.3 (page 160) we discuss the characteristic classes of the tangent bundle of a 4–manifold. Most important among these is the second Stiefel–Whitney class $w_2(T_M)$. Its vanishing is equivalent, on one hand, to the intersection form being even, and on the other hand, to the existence of a spin structure on M. Various definitions of spin structures and related concepts are explained in the end-notes, and we refer to their introduction on page 173 for an outline of their contents.

Section 4.4 discusses the integral lifts of $w_2(T_M)$, called characteristic elements. These always exist, and their self-intersections are congruent modulo 8 to the signature of M. A striking result of Rokhlin's states that if $w_2(T_M)$ vanishes and M is smooth, then the signature of M is not merely a multiple of 8, but of 16; the consequences of this fact pervade all of topology. For us, an immediate consequence is that E_8 can never be the intersection form of a smooth simply-connected 4–manifold.

Finally, we should also mention that the end-notes contain a discussion of the theory of smooth structures on topological manifolds of high dimensions (page 207).

4.1. Whitehead's theorem and homotopy type

It is obvious that, if two 4–manifolds are homotopy-equivalent, then their intersection forms must be isomorphic. A first hint of the overwhelming importance that intersection forms have for 4–dimensional topology comes from the following converse:

Whitehead's Theorem. *Two simply-connected 4–manifolds are homotopy-equivalent if and only if their intersection forms are isomorphic.*

The result as stated was proved by J. Milnor, based on J.H.C. Whitehead's work. The rest of this section is devoted to a proof of this result.[1]

Start of the proof. Take a simply-connected 4–manifold M: it has homology only in dimensions 0, 2 and 4. Therefore, by Hurewicz's theorem,

$$\pi_2(M) \approx H_2(M;\mathbb{Z}) .$$

Since M is simply-connected, the latter has no torsion and thus is isomorphic to some $\oplus\, m\,\mathbb{Z}$. Hence the isomorphism $\pi_2 \approx H_2$ can be realized by a map[2]

$$f\colon S^2 \vee \cdots \vee S^2 \longrightarrow M .$$

Such f induces an isomorphism on 2–homology, and thus on all homology groups but the fourth.

To remedy this defect, we can cut out a small 4–ball from M and thus annihilate its H_4. The remainder, denoted by M°, is now homotopy-equivalent to $S^2 \vee \cdots \vee S^2$: Indeed, the map f can be easily arranged to avoid the missing 4–ball, and it then induces an isomorphism of the whole homologies of

---

[1] The next section starts on page 149.

[2] Remember that $A \vee B$ is obtained by identifying a random point of A with a random point of B. (One can realize $A \vee B$ as $A \times b \cup a \times B$ inside $A \times B$.) Thus, $S^2 \vee \cdots \vee S^2$ is a bunch of spheres with exactly one point in common; it is called a **bouquet** of spheres.

the two spaces. Invoking a celebrated result of Whitehead[3] implies that f is in fact a homotopy equivalence

$$M^\circ \sim \mathbb{S}^2 \vee \cdots \vee \mathbb{S}^2 .$$

Since M can be reconstructed by gluing the 4–ball back to M°, we deduce that the homotopy type of M can equivalently be obtained from $\bigvee m\, \mathbb{S}^2$ by gluing a 4-ball $\mathbb{D}^4$ to it:

$$M \sim \bigvee m\, \mathbb{S}^2 \cup_\varphi \mathbb{D}^4 .$$

The attachment of the ball is made through some suitable map

$$\varphi \colon \partial \mathbb{D}^4 \longrightarrow \bigvee m\, \mathbb{S}^2 .$$

In conclusion, the homotopy type of M is completely determined by the homotopy class of this φ; this class should be viewed as an element of $\pi_3(\bigvee m\, \mathbb{S}^2)$.

To prove Whitehead's theorem, we need only show that the homotopy class of φ is completely determined by the intersection form of M. This can be seen in two ways, an algebro-topologic argument and a more geometric (but longer) argument. We present both of them:

Homotopy-theoretic argument

For the following proof, the reader is assumed to have a friendly relationship with algebraic topology; if not, skip to the alternative argument.

At the outset, it is worth noticing that, through the homotopy equivalence $M \sim \bigvee m\, \mathbb{S}^2 \cup_\varphi \mathbb{D}^4$, the fundamental class $[M] \in H_4(M;\mathbb{Z})$ corresponds to the class of the attached 4–ball $\mathbb{D}^4$; indeed, since the latter has its boundary entirely contained in the 2–skeleton $\bigvee m\, \mathbb{S}^2$, it represents a 4–cycle.

Think of each $\mathbb{S}^2$ as a copy of $\mathbb{CP}^1$ inside $\mathbb{CP}^\infty$. Then embed

$$\mathbb{S}^2 \vee \cdots \vee \mathbb{S}^2 \subset \mathbb{CP}^\infty \times \cdots \times \mathbb{CP}^\infty ,$$

and consider the exact homotopy sequence

$$\pi_4(\times m\, \mathbb{CP}^\infty) \to \pi_4(\times m\, \mathbb{CP}^\infty,\ \bigvee m\, \mathbb{S}^2) \to \pi_3(\bigvee m\, \mathbb{S}^2) \to \pi_3(\times m\, \mathbb{CP}^\infty) .$$

Since $\mathbb{CP}^\infty$ is an Eilenberg–MacLane $K(\mathbb{Z},2)$–space, the only non-zero homotopy group of $\times m\, \mathbb{CP}^\infty$ is π_2, and thus the above sequence exhibits an isomorphism

$$\pi_4(\times m\, \mathbb{CP}^\infty,\ \bigvee m\, \mathbb{S}^2) \approx \pi_3(\bigvee m\, \mathbb{S}^2) .$$

---

[3] The statement is: *If between two simply-connected CW–complexes there exists a map that induces isomorphisms on all homology groups, then this map must be a homotopy equivalence.* Note that an abstract isomorphism of homologies is not sufficient.

The above π_4 is made of maps $\mathbb{D}^4 \to \times m\, \mathbb{CP}^\infty$ that take $\partial\, \mathbb{D}^4$ to $\vee m\, \mathbb{S}^2$. The isomorphism associates to $\varphi\colon \partial\, \mathbb{D}^4 \to \vee m\, \mathbb{S}^2$ in π_3 the class of any of its extensions

$$\widetilde{\varphi}\colon \mathbb{D}^4 \longrightarrow \times m\, \mathbb{CP}^\infty .$$

Further, since the inclusion $\vee m\, \mathbb{S}^2 \subset \times m\, \mathbb{CP}^\infty$ induces an isomorphism on π_2's, a different portion of the same homotopy exact sequence implies that both π_2 and π_3 of the pair $(\times m\, \mathbb{CP}^\infty, \vee m\, \mathbb{S}^2)$ must vanish. Therefore, Hurewicz's theorem shows that we have a natural identification

$$\pi_4(\times m\, \mathbb{CP}^\infty, \vee m\, \mathbb{S}^2) \approx H_4(\times m\, \mathbb{CP}^\infty, \vee m\, \mathbb{S}^2; \mathbb{Z}) .$$

Through this identification, the class of $\widetilde{\varphi}$ from π_4 is sent to the class

$$\widetilde{\varphi}_*[\mathbb{D}^4] \in H_4(\times m\, \mathbb{CP}^\infty, \vee m\, \mathbb{S}^2; \mathbb{Z}) ,$$

where $\widetilde{\varphi}_*$ is the morphism induced on homology by the map $\widetilde{\varphi}$.

Moreover, since both H_4 and H_3 of $\vee m\, \mathbb{S}^2$ vanish, the homology exact sequence makes appear the isomorphism

$$H_4(\times m\, \mathbb{CP}^\infty, \vee m\, \mathbb{S}^2; \mathbb{Z}) \approx H_4(\times m\, \mathbb{CP}^\infty; \mathbb{Z}) .$$

For example, since $\widetilde{\varphi}_*[\mathbb{D}^4]$ represents a 4–class and its boundary is included in the 2–skeleton of $\times m\, \mathbb{CP}^\infty$, it follows that $\widetilde{\varphi}_*[\mathbb{D}^4]$ can be viewed as a 4–cycle directly in $H_4(\times m\, \mathbb{CP}^\infty; \mathbb{Z})$.

Owing to the lack of torsion, we also have a natural duality

$$H^4(\times m\, \mathbb{CP}^\infty; \mathbb{Z}) = \operatorname{Hom}(H_4(\times m\, \mathbb{CP}^\infty; \mathbb{Z}), \mathbb{Z}) .$$

This shows that, in order to determine $\widetilde{\varphi}_*[\mathbb{D}^4]$ in H_4, it is enough to evaluate all classes from H^4 on it. In other words, the class $\varphi \in \pi_3(\vee m\, \mathbb{S}^2)$ (and thus the homotopy type of M) are completely determined by the set of values $\alpha_k(\widetilde{\varphi}_*[\mathbb{D}^4])$ for some basis $\{\alpha_k\}_k$ of $H^4(\times m\, \mathbb{CP}^\infty; \mathbb{Z})$.

Such a basis can be immediately obtained by cupping the classes dual to each $\mathbb{S}^2$, that is to say, we have

$$H^4(\times m\, \mathbb{CP}^\infty; \mathbb{Z}) = \mathbb{Z}\{\omega_i \cup \omega_j\}_{i,j} ,$$

where ω_k denotes the 2–class dual to $\mathbb{CP}^1$ inside the k^{th} copy of $\mathbb{CP}^\infty$. Furthermore, since

$$H^2(\times m\, \mathbb{CP}^\infty; \mathbb{Z}) \approx H^2(\vee m\, \mathbb{S}^2; \mathbb{Z}) \approx H^2(M^\circ; \mathbb{Z}) \approx H^2(M; \mathbb{Z}) ,$$

we see that each class ω_k of $\times m\, \mathbb{CP}^\infty$ can in fact be viewed as a 2–class w_k of M itself.

Specifically, the inclusion $\iota\colon \vee m\, \mathbb{S}^2 \subset \times m\, \mathbb{CP}^\infty$ extends by $\widetilde{\varphi}$ to the map

$$M \sim \vee m\, \mathbb{S}^2 \cup_\varphi \mathbb{D}^4 \xrightarrow{\iota + \widetilde{\varphi}} \times m\, \mathbb{CP}^\infty .$$

The w_k's appear as the pull-backs $w_k = (\iota + \widetilde{\varphi})^* \omega_k$ and make up a basis of $H^2(M; \mathbb{Z})$.

Evaluating $\omega_i \cup \omega_j$ on $\widetilde{\varphi}_*[\mathbb{D}^4]$ inside $\times m\, \mathbb{CP}^\infty$ yields the same result as pulling ω_i and ω_j back to M, cupping there, and then evaluating on $[\mathbb{D}^4]$:

$$\begin{aligned}(\omega_i \cup \omega_j)\big(\widetilde{\varphi}_*[\mathbb{D}^4]\big) &= \big((\iota + \widetilde{\varphi})^*(\omega_i \cup \omega_j)\big)[\mathbb{D}^4] \\ &= \big((\iota + \widetilde{\varphi})^*\omega_i\big) \cup \big((\iota + \widetilde{\varphi})^*\omega_j\big)[\mathbb{D}^4] \\ &= (w_i \cup w_k)[\mathbb{D}^4]\,.\end{aligned}$$

However, as we noticed at the outset, the class $[\mathbb{D}^4]$ coincides with the fundamental class $[M]$ of M, and hence

$$(w_i \cup w_k)[\mathbb{D}^4] = Q_M(w_i, w_k)\,.$$

Since $\{w_1, \ldots, w_m\}$ is a basis in $H^2(M; \mathbb{Z})$, we deduce that the set of values $Q_M(w_i, w_k)$ fills-up a complete matrix for the intersection form Q_M of M.

On the other hand, as we have argued, by staying in $\times m\, \mathbb{CP}^\infty$ and evaluating all the $\omega_i \cup \omega_j$'s on $\widetilde{\varphi}_*[\mathbb{D}^4]$ we fully determine the class of φ in $\pi_3\big(\vee m\, \mathbb{S}^2\big)$ and thus fix the homotopy type of M.

This concludes one proof of Whitehead's theorem. □

Pontryagin–Thom argument

We have seen that the homotopy type of M can be represented as the result of gluing a 4–ball $\mathbb{D}^4$ to a bouquet of spheres $\mathbb{S}^2 \vee \cdots \vee \mathbb{S}^2$ by using some map $\varphi\colon \partial\mathbb{D}^4 \to \vee m\, \mathbb{S}^2$. Thus, the homotopy type of M corresponds to the homotopy class of φ. We need to argue that φ is determined by the intersection form of M.

A geometric way of seeing how the intersection form Q_M determines the attaching map

$$\varphi\colon \mathbb{S}^3 \longrightarrow \vee m\, \mathbb{S}^2$$

comes from what is known as the *Pontryagin–Thom construction.* The latter technique will be detailed in more generality in the end-notes of this chapter (page 230).

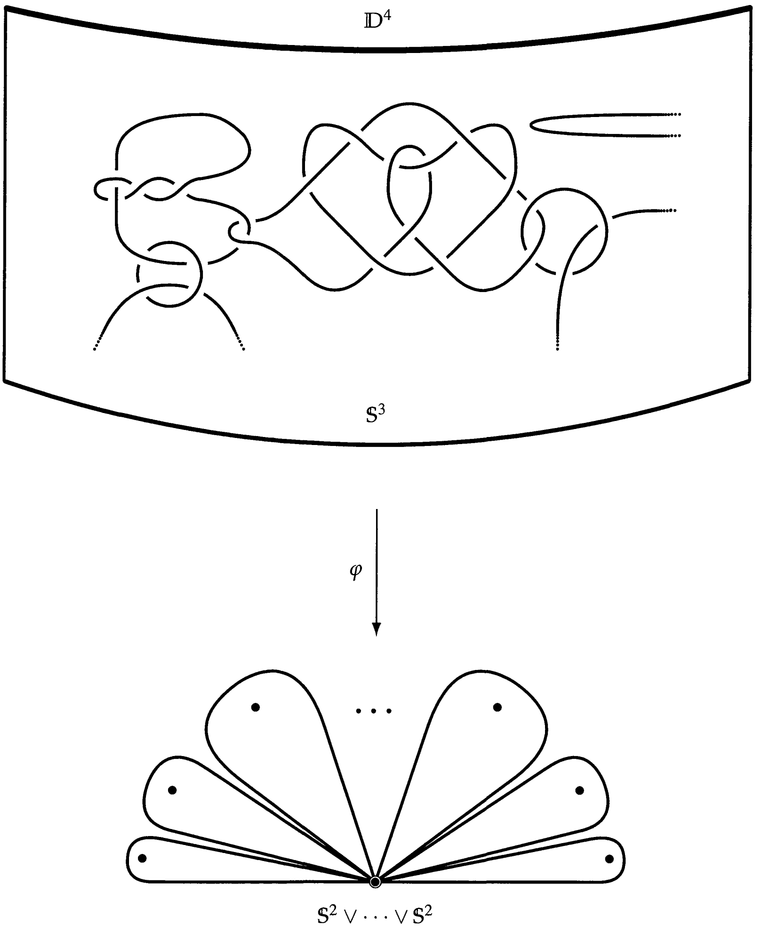
The framed link. Pick some points $p_1, \ldots, p_m$, one from each 2–sphere of $\vee m\, \mathbb{S}^2$. Arrange by a small homotopy that φ be transverse to these points. Also, wiggle φ until each pre-image $\varphi^{-1}[p_k]$ is connected.[4] Then each $L_k = \varphi^{-1}[p_k]$ is an embedded circle in $\mathbb{S}^3$ (a knot), and so the union

$$L = L_1 \cup \cdots \cup L_m$$

is a link in $\mathbb{S}^3$, as suggested in figure 4.1 on the following page.

---

[4] If "wiggle" is not convincing, read from the end-notes of this chapter (page 230).

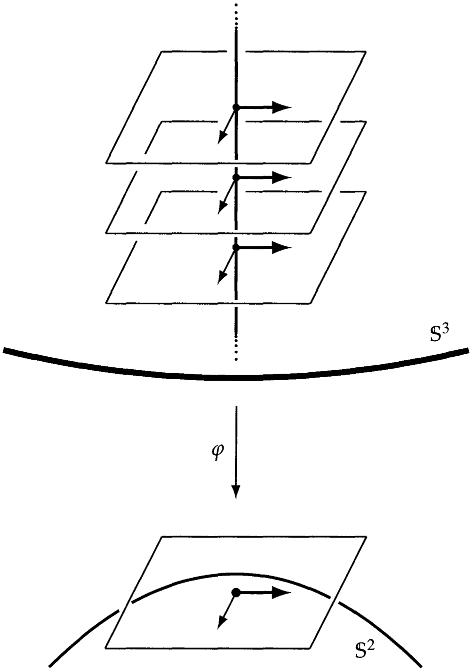


4.1. Framed link, from attaching a 4–ball to $\mathbb{S}^2 \vee \cdots \vee \mathbb{S}^2$

The way this link L appears out of the map φ endows it with an extra bit of structure, namely a *framing:* For each L_k, embed its normal bundle $N_{L_k/\mathbb{S}^3}$ as a subbundle of $T_{\mathbb{S}^3}$ over L_k. Since φ is transverse to p_k and can be assumed to be differentiable all around L_k, it follows that $d\varphi\colon T_{\mathbb{S}^3}|_{L_k} \longrightarrow T_{\mathbb{S}^2}|_{p_k}$ restricts to a map $N_{L_k/\mathbb{S}^3} \to T_{\mathbb{S}^2}|_{p_k}$ that is an isomorphism on fibers, see figure 4.2 on the next page. The effect is that the normal bundle $N_{L_K/\mathbb{S}^3}$ is thus trivialized. Such a trivialization of the normal bundle of L_k is called a **framing** of the knot L_k. Doing this for each p_k results in a **framed link** $L = L_1 \cup \cdots \cup L_m$. Also notice that each component of the link gains a natural orientation.[5]

---

[5] We have $T_{\mathbb{S}^3}|_{L_k} = T_{L_k} \oplus N_{L_k/\mathbb{S}^3}$; since $\mathbb{S}^3$ is oriented and $N_{L_k/\mathbb{S}^3}$ lifts an orientation from $\mathbb{S}^2$ (at the same time with the framing), this induces an orientation of T_{L_k}.



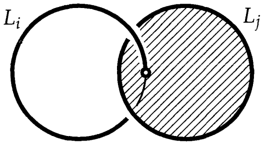
4.2. Pulling-back a framing

The linking matrix. We now focus on some simple numerical data that is expressed by our L. On one hand, for every two components L_i and L_j, we have the **linking numbers**

$$\mathrm{lk}(L_i, L_j) \ .$$

This integer measures how many times L_i twists around L_j.

More rigorously, one chooses in S^3 an oriented surface F_j bounded by[6] *L_j and counts the intersection number of F_j with L_i in S^3, as in figure 4.3. The linking number does not depend on the choice of F_j and is symmetric on link components:* $\mathrm{lk}(L_i, L_j) = \mathrm{lk}(L_j, L_i)$.



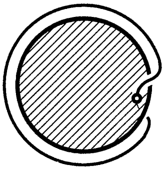
4.3. Linking number of two knots

We also have the self-linkings numbers $\mathrm{lk}(L_k, L_k)$, induced from the framing. These count the twists of the trivialization of L_k's normal bundle.

---

[6] Such a surface always exists and is called an (orientable) **Seifert surface** for L_j; we will say a bit more in a second.

The self-linking number can be defined by picking some section of $N_{L_k/\mathbb{S}^3}$ *that follows the trivialization of* $N_{L_k/\mathbb{S}^3}$ *given by the framing, then thinking of that section as drawing a parallel copy* L'_k *of* L_k *in* $\mathbb{S}^3$*, and finally setting* $\operatorname{lk}(L_k, L_k)$ *to equal the linking number* $\operatorname{lk}(L'_k, L_k)$ *of* L_k *with this parallel copy, as suggested in figure 4.4. In our context, this self-linking number can also be defined directly: since* $L_k = \varphi^{-1}[p_k]$*, pick a point* p'_k *close to* p_k*, and define* $\operatorname{lk}(L_k, L_k) = \operatorname{lk}(\varphi^{-1}[p_k],\ \varphi^{-1}[p'_k])$.



4.4. Self-linking number of a framed knot

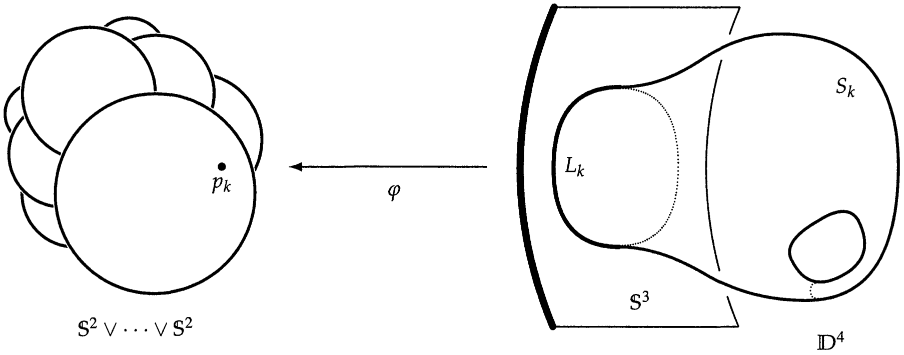
All these self/linking numbers can be fit together into a matrix

$$\left[\operatorname{lk}(L_i,\ L_j)\right]_{i,j}\ ,$$

which is called the **linking matrix** of the framed link L.

On one hand, it turns out that this linking matrix is exactly the matrix of the intersection form of M, as we will argue shortly. On the other hand, a Pontryagin–Thom framed-bordism argument[7] can be used to show that the homotopy class of φ is entirely determined by this linking matrix.

The intersection form. To see that the linking matrix of L indeed governs intersections in M, start by choosing for each L_k an oriented surface S_k inside $\mathbb{D}^4$ that is bounded by L_k, as in figure 4.5.

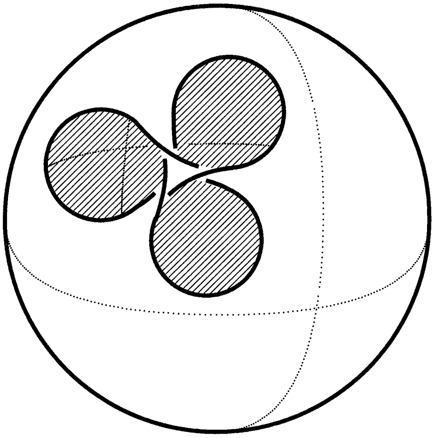


4.5. Building intersections out of a link.

---

[7] See the end-notes of this chapter (page 230).

Such S_k's exist because, as we mentioned before, every knot K in $\mathbb{R}^3$ bounds an orientable surface that is bounded by K, called a **Seifert surface** *for K. (If not convinced, draw a knot, then try to draw its Seifert surface.*[8] *Take a peek at figure 4.6 for inspiration. In any case, this is merely a particular case of the general fact that homologically-trivial codimension–2 submanifolds must bound codimension–1 submanifolds.) To get the S_k's above, one can start with Seifert surfaces in $\mathbb{S}^3$ for each L_k, then push their interiors into $\mathbb{D}^4$.*



4.6. A Seifert surface for the trefoil knot

The fundamental fact to notice is that $\operatorname{lk}(L_i, L_j)$ is in fact the intersection number $S_i \cdot S_j$ of the corresponding surfaces in $\mathbb{D}^4$:

$$\operatorname{lk}(L_i, L_j) = S_i \cdot S_j .$$

See figure 4.7 on the following page for an argument.

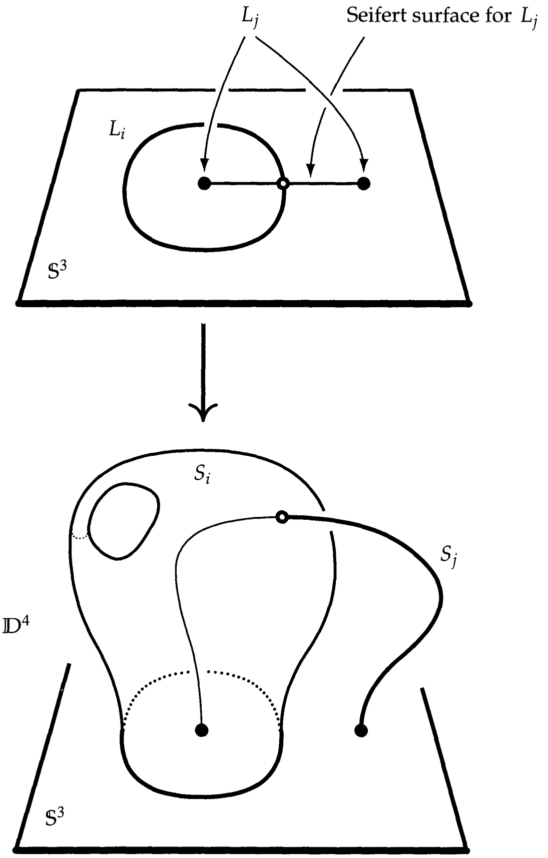
Therefore, when rebuilding the homotopy type of M through attaching $\mathbb{D}^4$ to $\bigvee m\, \mathbb{S}^2$ *via* the map φ, each S_k has its boundary L_k collapsed to the point p_k, and thus creates a closed surface S_k^*. Since the intersection numbers $S_i^* \cdot S_j^*$ in (the homotopy type of) M are exactly $\operatorname{lk}(L_i, L_j)$, we conclude that the linking matrix captures part of the intersection form of M.

To conclude the proof, all we need to do is argue that the intersections of the S_k^*'s in fact exhaust the whole intersection form of M. In other words, we need to argue that the S_k^*'s represent a basis for $H_2(M;\mathbb{Z})$. For this, recall that the homology $H_2(M;\mathbb{Z})$ was generated by the classes of the spheres of $\bigvee m\, \mathbb{S}^2$. The classes S_k^* intersect the classes of those spheres exactly once. Since the intersection form of M is unimodular, this implies that the S_k^*'s make up the dual basis[9] to the basis exhibited by the spheres of $\bigvee m\, \mathbb{S}^2$. This concludes the alternative proof of Whitehead's theorem. □

---

[8] Be careful to not draw a non-orientable surface.

[9] Two classes α and β were called dual to each other if $\alpha \cdot \beta = 1$; see back on page 117.



4.7. Linking numbers are intersection numbers of bounded surfaces

Example. Let us conclude the discussion of Whitehead's theorem with a very simple example. If we take $\varphi\colon \mathbb{S}^3 \to \mathbb{S}^2$ to be the Hopf map,[10] then its link is the unknot[11] with framing $+1$, and the homotopy type obtained by attaching $\mathbb{D}^4$ to $\mathbb{S}^2$ using this φ is none other than $\mathbb{CP}^2$'s.

Upside-down handle diagrams. *In a certain sense, the whole procedure from the above proof is an upside-down version of a handle decomposition: the framed link L is nothing but a Kirby diagram[12] for attaching 2-handles to $\mathbb{D}^4$. The closing of S_k into S_k^* by collapsing L_k to p_k is homotopy-equivalent to gluing along L_k a disk with center p_k: the core of a 2-handle. Then the framings can be used to thicken this disk to an actual 2-handle and eventually transform the whole procedure from gluing $\mathbb{D}^4$ to $\vee m\, \mathbb{S}^2$ into attaching 2-handles to $\mathbb{D}^4$ along the link L in $\partial\, \mathbb{D}^4$.*

---

[10] The Hopf map was recalled back in footnote 34 on page 129.

[11] A knot K is called the **unknot** if it is trivial, or not knotted. Specifically, this means that K bounds some embedded disk.

[12] Kirby diagrams were explained back in the end-notes of chapter 2 (page 91).

However, the framed link L is just one of many Kirby diagrams that can be obtained through homotopies of φ. The intersection form (i.e., the homotopy class of φ) is far from determining precisely the shape of this link. Most of these links will not even lead to constructions that close-up to a smooth closed 4–manifold. (They always close-up as topological *4–manifolds by using Freedman's fake 4–balls, since if one starts with a unimodular matrix, then the resulting boundary will be a homology 3–sphere.*[13]*) The framed link L is just one of many diagrams for a handle decomposition of a creature homotopy-equivalent to M, but rarely of M itself.*

4.2. Wall's theorems and *h*–cobordisms

We will now present a series of results due to C.T.C. Wall, which culminates with the statement that, if two *smooth* simply-connected 4–manifolds have isomorphic intersection forms, then they are not merely homotopy-equivalent, but in fact are h–cobordant. Combining this with Freedman's topological h–cobordism theorem will yield immediately that, if two smooth simply-connected 4–manifold have the same intersection form, then they must be homeomorphic.

Sum-stabilizations

Two smooth 4–manifolds M and N are often h–cobordant without being diffeomorphic. To obtain a diffeomorphism, we can first "stabilize" the manifolds. A **sum-stabilization**[14] of a 4–manifold means connect-summing with copies of $\mathbb{S}^2 \times \mathbb{S}^2$. The world of smooth 4–manifolds considered up to such stabilizations is considerably simplified:

Wall's Theorem on Stabilizations. *If M and N are smooth, simply-connected and h–cobordant, then there is an integer k such that we have a diffeomorphism*

$$M \# k\mathbb{S}^2 \times \mathbb{S}^2 \;\cong\; N \# k\mathbb{S}^2 \times \mathbb{S}^2 \,.$$

Proof. Adding $\mathbb{S}^2 \times \mathbb{S}^2$'s essentially allows us to go through with the h–cobordism theorem's program. This is owing to the fact that the new spheres can be used to undo unwanted intersections of surfaces, such as self-intersections of immersed Whitney disks.

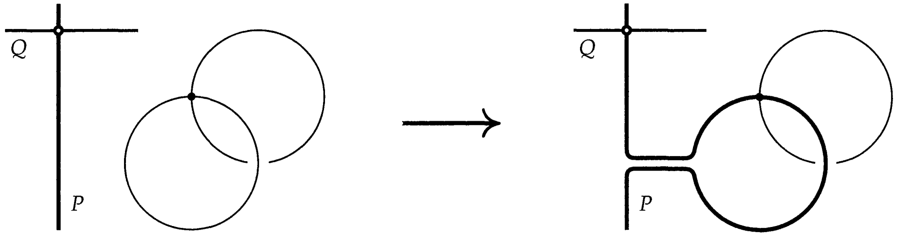
Imagine that two surfaces P and Q have an intersection point that we want to be rid of. First, since $\mathbb{S}^2 \times \mathbb{S}^2$ contains two spheres meeting in exactly one point, we can join P with one such sphere by using a thin tube, as in figure 4.8 on the next page; the result is that P is now

---

13. This last fact will be proved in the the end-notes of the next chapter (page 261).

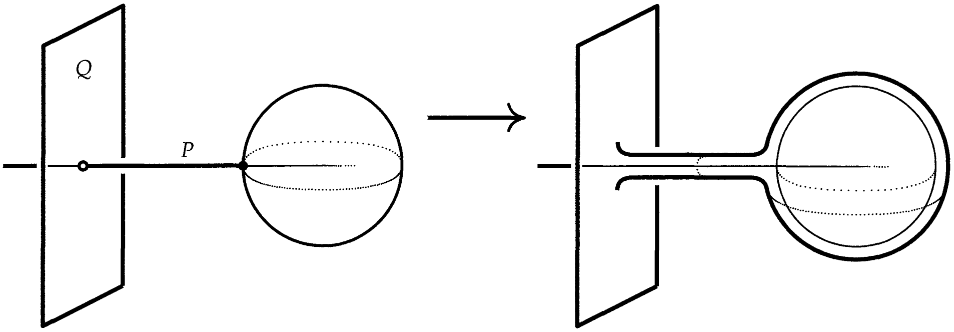
14. The name "stabilization" is in tune with, for example, stable properties of vector bundles—those preserved after adding trivial bundles; or stable homotopy groups—the part preserved after suspensions.

meeting the *other* sphere in exactly one point. (A sphere meeting a surface P in exactly one point is sometimes called a *transverse sphere* for P.)



4.8. Joining a sphere

Second, we pick a path in P from the intersection point with Q to the intersection point with the transverse sphere. Then, using a thin tube following this chosen path, we can connect Q to a parallel copy of the sphere, as in figure 4.9. The intersection point of P and Q has vanished.



4.9. Eliminating an intersection by sliding over a sphere

Notice that none of these maneuvers changed the genus of either P or Q. Thus, one can use this procedure to eliminate self-intersections of immersed Whitney disks and proceed with the h–cobordism program.

Finally, for dealing with the framing obstruction for the Whitney trick in dimension 4, which was observed back in the end-notes of chapter 1 (page 57), one can connect-sum the Whitney disk with the diagonal or anti-diagonal sphere[15] *of an extra* $S^2 \times S^2$*, which changes the framing of the disk by* ± 2*. Since having intersection points of opposite signs guarantees that the framing of a Whitney disk is even, summing with enough such diagonal spheres achieves the vanishing of the framing, and hence allows us to proceed with the Whitney trick.*

---

[15] The diagonal sphere in $S^2 \times S^2$ is the image of the embedding $S^2 \to S^2 \times S^2 \colon x \mapsto (x, x)$ and has self-intersection $+2$. The anti-diagonal sphere is the image of $x \mapsto (x, -x)$, with self-intersection -2.

With luck, a same $\mathbb{S}^2 \times \mathbb{S}^2$–term could be used for eliminating several (if not all) intersections.[16] If not, add more. □

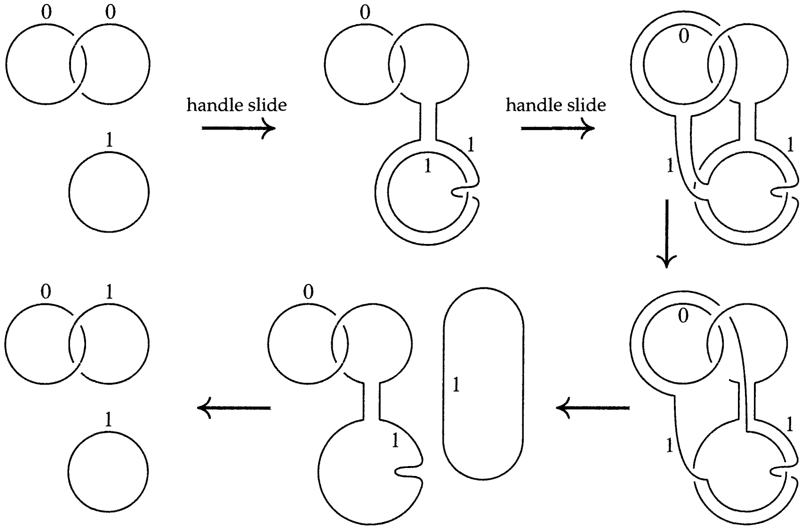
An alternative argument (more economical with $\mathbb{S}^2 \times \mathbb{S}^2$–terms) will be encountered on page 157, in the middle of the proof of Wall's theorem on h–cobordisms.

Of course $\mathbb{S}^2 \times \mathbb{S}^2$ is not the only summand that can be used with similar effects as above. One might imagine that, for example, the twisted product $\mathbb{S}^2 \mathbin{\tilde{\times}} \mathbb{S}^2$ would work just as well. However, on one hand, summing with $\mathbb{S}^2 \times \mathbb{S}^2$'s preserves the parity and signature of M, which is usually desirable; and, on the other hand, in many cases summing with $\mathbb{S}^2 \mathbin{\tilde{\times}} \mathbb{S}^2$ is nothing different, since one can prove directly that:

Lemma. *If M^4 has odd intersection form, then there is a diffeomorphism*

$$M \# \mathbb{S}^2 \times \mathbb{S}^2 \;\cong\; M \# \mathbb{S}^2 \mathbin{\tilde{\times}} \mathbb{S}^2 .$$

Idea of proof. Consider the simple case when M is $\mathbb{CP}^2$. For brevity, we use Kirby calculus, as outlined in the end-notes of chapter 2 (page 91). Then, after two handle slides and a bit of clean-up, it is done, as shown in figure 4.10. For the general case, one would slide over some odd-framed handle of M, then use similar tricks to untangle and separate $\mathbb{S}^2 \mathbin{\tilde{\times}} \mathbb{S}^2$ from M. □



4.10. Proof that $\mathbb{CP}^2 \# \mathbb{S}^2 \times \mathbb{S}^2 \;\cong\; \mathbb{CP}^2 \# \mathbb{S}^2 \mathbin{\tilde{\times}} \mathbb{S}^2$

---

[16] It is worth noting that in all *known* cases summing with just *one* copy of $\mathbb{S}^2 \times \mathbb{S}^2$ is enough. Currently, there are no devices able to detect cases when more than one copy of $\mathbb{S}^2 \times \mathbb{S}^2$ would be necessary.

Automorphisms of the intersection form

Wall also investigated algebraic automorphisms of intersection forms, and the question of their realizability by self-diffeomorphisms of an underlying 4–manifold.

Algebraic automorphisms. Let us consider for a moment the intersection form as an abstract algebraic creature, a symmetric bilinear unimodular form

$$Q\colon Z \times Z \longrightarrow \mathbb{Z}\,,$$

defined on some finitely-generated free $\mathbb{Z}$–module Z. An automorphism of Q is a self-isomorphism $\varphi\colon Z \approx Z$ that preserves the values of Q; that is to say, $Q(x,y) = Q(\varphi x,\, \varphi y)$.

The **divisibility** of an element x of Z is the greatest integer d such that x can be written as $x = dx_0$ for some non-zero $x_0 \in Z$. An element of divisibility 1 is called **indivisible.**

An element $\underline{w}$ of a $\mathbb{Z}$–module endowed with a symmetric bilinear unimodular form Q is called **characteristic** if it satisfies

$$Q(\underline{w}, x) = Q(x, x) \pmod 2$$

for all $x \in Z$. Notice that, if Q is even, then the divisibility of any characteristic element must be even; further, if Q is even, then $\underline{w} = 0$ is characteristic. An element is called **ordinary** if it is not characteristic. Whether some $x \in Z$ is characteristic or ordinary is called the **type** of x.

Wall's Theorem on Automorphisms. *If* $\operatorname{rank} Q - |\operatorname{sign} Q| \geq 4$, *then, given any two elements* x', $x'' \in Z$ *with the same divisibility, self-intersection and type, there must exist an automorphism* φ *of* Q *so that* $\varphi(x') = x''$. □

Since $\operatorname{rank} Q - \operatorname{sign} Q$ *is always even, the condition* $\operatorname{rank} Q - |\operatorname{sign} Q| \geq 4$ *only excludes definite forms (when* $\operatorname{sign} Q = \pm \operatorname{rank} Q$*) and forms with* $\operatorname{rank} Q - |\operatorname{sign} Q| = \pm 2$ *(which Wall calls* near-definite*). As we will see later,*[17] *the only excluded forms are* H *and* $[+1] \oplus m\,[-1]$ *and* $[-1] \oplus m\,[+1]$ *and all definite forms. Further, as far as smooth 4–dimensional topology is concerned, the only relevant definite forms are*[18] $\oplus\, m\,[+1]$ *and* $\oplus\, m\,[-1]$.

The characteristic elements of an intersection form will continue to play an important role and will be visited again in section 4.4 (page 168) ahead.

---

[17] From Serre's classification of indefinite forms; see section 5.1 (page 238).

[18] This follows from Donaldson's theorem; see section 5.3 (page 243) ahead.

Automorphisms and diffeomorphisms. It is obvious that any self-diffeomorphism of a 4–manifold induces an automorphism of its intersection form. The converse is true, but only after stabilizing once:

Wall's Theorem on Diffeomorphisms. *Let M be a smooth simply-connected 4–manifold with Q_M indefinite.*[19] *Then any automorphism of the intersection form of $M \# S^2 \times S^2$ can be realized by a self-diffeomorphism of $M \# S^2 \times S^2$.*

> ***Idea of the proof.*** One identifies a concrete set of generators for the group of automorphisms of $Q_M \oplus H$, then one shows directly that each of these generators corresponds to a self-diffeomorphism. □

Topological heaven. *It should no longer come as a surprise that, if we weaken to the realm of topological 4–manifolds, stabilization is no longer necessary:*

Theorem (*M. Freedman*). *Any automorphism of Q_M can be realized by a self-homeomorphism of M, unique up to isotopy.* □

Of course, the smooth version of such a result fails.[20]

Self-diffeomorphism from spheres. *For amusement, we briefly mention a couple of examples of self-diffeomorphisms of a 4–manifold. These are built around an embedded sphere S of self-intersection*[21] *± 1 or ± 2, and act on homology by $[S] \mapsto -[S]$ and by fixing the Q–complement of $[S]$; in other words, they act as reflections on the homology lattice. Of course, finding such spheres is an endeavor in itself and often they do not exist.*[22]

Reflection on a (± 1)–sphere. A neighborhood of a $(+1)$–sphere S in M is diffeomorphic to a neighborhood of $\mathbb{CP}^1$ in $\mathbb{CP}^2$, and furthermore $M = M' \# \mathbb{CP}^2$, with S appearing as $\mathbb{CP}^1$ in $\mathbb{CP}^2$. Our diffeomorphism acts on $\mathbb{CP}^2$ and fixes M'. We take coordinates $[z_0 : z_1 : z_2]$ on $\mathbb{CP}^2$ and consider the complex conjugation $\varphi_0 \colon \mathbb{CP}^2 \to \mathbb{CP}^2$, with $\varphi_0[z_0 : z_1 : z_2] = [\bar{z}_1 : \bar{z}_2 : \bar{z}_2]$. Away from the projective line $\mathbb{CP}^1 = \{z_0 = 0\}$, on $\mathbb{CP}^2 \setminus \mathbb{CP}^1 = \mathbb{C}^2$, this conjugation acts as $(z_1, z_2) \mapsto (\bar{z}_1, \bar{z}_2)$, or, in real coordinates, $(x_1, y_1, x_2, y_2) \mapsto (x_1, -y_1, x_2, -y_2)$. We pick a small 4–ball $\mathbb{D}^4$ around $0 \in \mathbb{C}^2$ and modify φ_0 as we approach $\mathbb{D}^4$ by increasingly rotating the (y_1, y_2)–plane by an angle growing from 0 to π, until φ_0 becomes the identity on all $\mathbb{D}^4$; see figure 4.11 on the following page. We have built a self-diffeomorphism φ of $\mathbb{CP}^2$ that flips $\mathbb{CP}^1$ but fixes a small 4–ball $\mathbb{D}^4$. If we think of $M = M' \# \mathbb{CP}^2$ as being built by cutting out $\mathbb{D}^4$ from $\mathbb{CP}^2$, then φ extends from $\mathbb{CP}^2$ to the whole M by the identity. (For a (-1)–sphere, reverse orientations.)

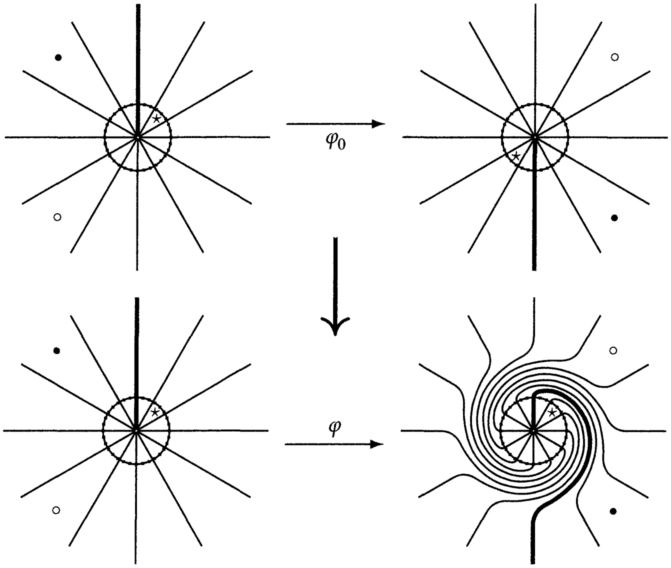
---

[19] Requiring that the intersection form of a *smooth* 4–manifold be indefinite is not a strong restriction, since in fact the only excluded forms are $\oplus\, m\, [\pm 1]$; see section 5.3 (page 243) ahead.

[20] For example, a simple obstruction is that any automorphism of Q_M that can be realized by diffeomorphisms must send Seiberg–Witten basic classes to basic classes (for these notions, see chapter 10, starting on page 375 ahead), but even that in general is not sufficient.

[21] For the extent of this inserted note, we will call such spheres (± 1)– and (± 2)–spheres.

[22] Nonetheless, recall that we did identify twenty (-2)–spheres inside the $K3$ surface, see page 133.

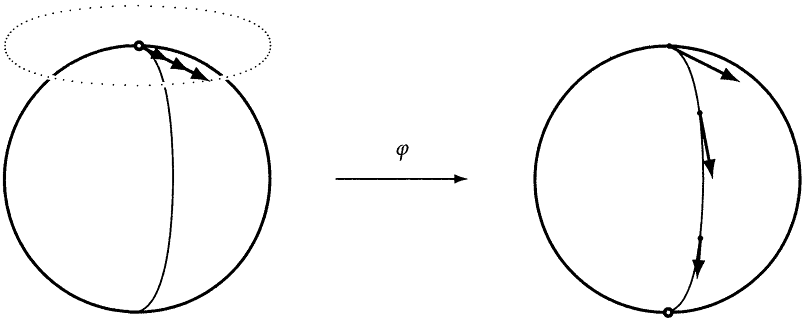


4.11. Modification toward reflection on a (-1)-sphere

Reflection on a (± 2)*-sphere. A neighborhood of a* $(+2)$*-sphere* S *in* M *is diffeomorphic to the unit-disk bundle* $\mathbb{D}T_{\mathbb{S}^2}$*. We think of* $\mathbb{D}T_{\mathbb{S}^2}$ *as* $\{(v,w) \in \mathbb{R}^3 \times \mathbb{R}^3 \mid |v| = 1,\ |w| \leq 1,\ v \perp w\}$ *and define a self-diffeomorphism* $\varphi\colon \mathbb{D}T_{\mathbb{S}^2} \to \mathbb{D}T_{\mathbb{S}^2}$ *by*

$$\varphi(v,w) = \begin{cases} \big(\cos\vartheta \cdot v \ + \ \sin\vartheta \cdot \frac{1}{|w|}w\,,\ \cos\vartheta \cdot w \ + \ \sin\vartheta \cdot |w|\,(-v)\big) & \text{if } w \neq 0 \\ (-v, 0) & \text{if } w = 0 \end{cases}$$

with $\vartheta = (1 - |w|)\pi$*. Specifically, each tangent vector* w *determines a great circle in* $\mathbb{S}^2$ *and we slide* w *along this circle by a distance depending on* $|w|$*: the shorter* w *is, the more we travel; see figure 4.12. The resulting* φ *restricts as the antipodal map on the sphere* $S = \{(v,0)\}$*, but as the identity on* $\partial\,\mathbb{D}T_{\mathbb{S}^2}$ *and thus can be extended by the identity to the rest of* M*, yielding a self-diffeomorphism* φ *of* M*. (For a* (-2)*-sphere, reverse orientations.)*



4.12. Reflection on a $(+2)$-sphere

Wall's theorem on diffeomorphisms plays an essential role in proving the fundamental result that we present next.

Intersection forms and h–cobordisms

Going quite further than Whitehead's theorem, C.T.C. Wall proved that two smooth manifolds with the same intersection form are more than merely homotopy-equivalent:

Wall's Theorem on h–Cobordisms. *If M and N are smooth, simply-connected, and have isomorphic intersection forms, then M and N must be h–cobordant.*

If we combine with the earlier theorem on stabilizations, this yields:

Corollary. *If M and N are smooth, simply-connected, and have the same intersection form, then there is an integer k such that we have a diffeomorphism*

$$M \# k\mathbb{S}^2 \times \mathbb{S}^2 \cong N \# k\mathbb{S}^2 \times \mathbb{S}^2 . \qquad \square$$

On the other hand, if we combine the above theorem on h–cobordisms with M. Freedman's topological h–cobordism theorem, then we deduce the following most remarkable result:

Corollary (*M. Freedman*). *If two smooth simply-connected 4–manifolds have isomorphic intersection forms, then they must be homeomorphic.* □

This came almost twenty years after Wall's results. Even today the attempt to strengthen the above to diffeomorphisms does not get farther than the preceding direct combination of Wall's old results.

Because of this striking consequence, in what follows we will present a fairly complete proof of Wall's theorem on h–cobordisms; it will take the rest of this section.[23]

Proof of Wall's theorem on h–cobordisms

Since M and N have the same signature, $\overline{M} \cup N$ has signature zero, and thus it must bound some 5–manifold; in other words, there is some oriented W^5 that establishes a cobordism between M and N.

The proof of the theorem consists in modifying this W (without changing its boundary) until it becomes simply-connected and homologically-trivial, in other words, until it becomes an h–cobordism from M to N.

---

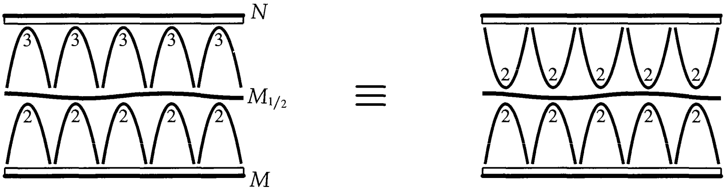
[23] The next section starts on page 160.

Kill the fundamental group. The first step is to modify W^5 to make it simply-connected. We choose a set of generating loops $\ell_1, \ldots, \ell_n$ for $\pi_1(W)$, realized as disjointly embedded circles. We will add disks to kill these homotopy classes. Specifically, for each ℓ_k we take a tubular neighborhood $\mathbb{S}^1 \times \mathbb{D}^4$ of ℓ_k and cut it out. This leaves a hole with boundary $\mathbb{S}^1 \times \mathbb{S}^3$, which we fill by gluing-in a copy of $\mathbb{D}^2 \times \mathbb{S}^3$. In the resulting 5–manifold, the class of ℓ_k is trivial. Repeating for all ℓ_k's yields a new cobordism between M and N, still denoted by W, that is simply-connected.

Divide and conquer. Choose now a handle decomposition of W^5. Since W is connected, we can cancel all 0– and 5–handles. Further, since W is simply-connected, all its 1–handles can be traded for 3–handles, and, upside-down, all 4–handles for 2–handles. We end up with a handle decomposition of W that only contains 2– and 3–handles, and view W as

$$W^5 = M^4 \times [0,\varepsilon] \cup \{2\text{–handles}\} \cup \{3\text{–handles}\} \cup N^4 \times [-\varepsilon, 0] ,$$

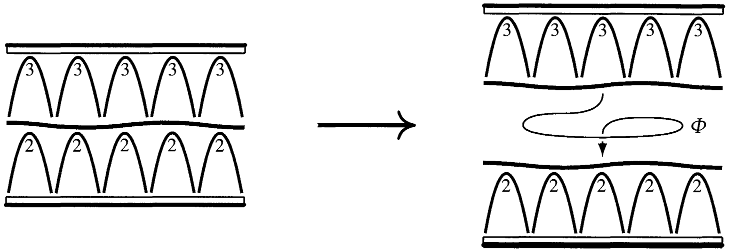
which we split into the two obvious halves: on one side, M and the 2–handles, on the other, N and the 3–handles, as on the left of figure 4.13. Looking upside-down at the upper half of W, instead of seeing the 3–handles as glued to the lower half, we can view them as 2–handles glued upwards to $N \times [-\varepsilon, 0]$.



4.13. The two halves of a simply-connected cobordism

Hence the middle level $M_{1/2}$, in between the 2– and the 3–handles, is a 4–manifold that can be obtained either from M by adding regular 2–handles attached downwards, or from N by adding upside-down 2–handles attached upwards.

The strategy for the remainder of the proof is the following: We will cut W into its two halves, then glue them back after twisting by a suitable self-diffeomorphism Φ of $M_{1/2}$, as in figure 4.14 on the next page. This cut-and-reglue procedure will create a new cobordism from M to N. If we choose the right diffeomorphism Φ, then the 3–handles from the upper half will cancel algebraically the 2–handles from the lower half. This means that the newly created cobordism between M and N will have no homology relative to its boundaries, and so will indeed be an h–cobordism from M to N.



4.14. Modifying a cobordism into an h–cobordism

On the frontier. Let us first clarify the shape of $M_{1/2}$. Think of it as obtained from M after adding the 2–handles of W.

A 5–dimensional 2–handle is a copy of $\mathbb{D}^2 \times \mathbb{D}^3$, to be attached by gluing $\mathbb{S}^1 \times \mathbb{D}^3$ to M^4. To attach such a 2–handle to M, we need to specify where the attaching circle $\mathbb{S}^1 \times 0$ is being sent, but a circle in a 4–manifold is isotopic to any other embedded circle. We also need to specify how the "thickening" of the attaching circle is to be glued to M. Since[24] $\pi_1 SO(3) = \mathbb{Z}_2$, there are only two ways of doing that, depending on whether the 3–disk $\mathbb{D}^3$ in M twists an even or an odd number of times around the attaching circle.[25] Therefore, to fully describe $M_{1/2}$ all we need is to specify how many "odd" and how many "even" 2–handles are to be attached.

Attaching a 2–handle $\mathbb{D}^2 \times \mathbb{D}^3$ deletes a copy of $\mathbb{S}^1 \times \mathbb{D}^3$ from M and, as a step toward $M_{1/2}$, replaces it with a copy of $\mathbb{D}^2 \times \mathbb{S}^2$. On one hand, if the 2–handle is even, then the disk $\mathbb{D}^2$ from $\mathbb{D}^2 \times \mathbb{S}^2$ can be closed to a 2–sphere of self-intersection 0: unite the disk with a small Seifert disk of the attaching circle in M; the self-intersection of such a Seifert disk in M is the same with the framing modulo 2 (compare with page 148 earlier). Hence, the result of adding this even 2–handle is the same as connect-summing with $\mathbb{S}^2 \times \mathbb{S}^2$. On the other hand, if the 2–handle is odd, then the disk closes to a sphere of self-intersection $+1$, and one can see that attaching it is the same as connect-summing with $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$. In conclusion, we have

$$M_{1/2} = M^4 \# m' \mathbb{S}^2 \times \mathbb{S}^2 \# m'' \mathbb{S}^2 \tilde{\times} \mathbb{S}^2 .$$

We will assume in the sequel that no $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$–terms are present.

No twists, and a proof of Wall's theorem on stabilizations. The assumption that there are no $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$–summands can be argued quite rigorously:

---

[24] Think of $SO(3)$ as the space of all oriented orthonormal frames in $\mathbb{R}^3$. Thus, $\pi_1 SO(3)$ will measure how many distinct trivializations of the 3–plane bundle $\mathbb{S}^1 \times \mathbb{R}^3$ exist. Some comments on $\pi_1 SO(m)$ will be made in the the end-notes of this chapter (page 177).

[25] Contrast this with what happens when, instead of building a 5–manifold as above, we build a 4–manifold. The framing for attaching a 2–handle is then determined by an element of $\pi_1 SO(2) = \mathbb{Z}$, an integer.

On one hand, if the intersection form of M is odd, then adding $\mathbb{S}^2 \widetilde{\times} \mathbb{S}^2$ or adding $\mathbb{S}^2 \times \mathbb{S}^2$ produces the same result, as we mentioned a bit earlier.[26]

On the other hand, if the intersection form of M is even, then a deeper result shows that $\overline{M} \cup N$ can be safely assumed to bound a 5–manifold that does not contain any odd handles.[27] This odd-less manifold should then be the one used as our W right back from the start of the argument.

By the way, if we accept that we can indeed avoid $\mathbb{S}^2 \widetilde{\times} \mathbb{S}^2$–summands, then we have stumbled upon another proof for Wall's theorem on stabilizations: from the lower half of W we have $M_{1/2} = M \# m\mathbb{S}^2 \times \mathbb{S}^2$, while from the upper half we have $M_{1/2} = N \# m\mathbb{S}^2 \times \mathbb{S}^2$, since $M_{1/2}$ can also be obtained by attaching even 2–handles upwards to N. Therefore

$$M \# m\mathbb{S}^2 \times \mathbb{S}^2 \cong N \# m\mathbb{S}^2 \times \mathbb{S}^2 .$$

This was, in fact, C.T.C. Wall's original argument for this result.

In any case, getting back to proving Wall's theorem on h–cobordisms, in what follows we assume that we have $M_{1/2} = M^4 \# m\mathbb{S}^2 \times \mathbb{S}^2$.

Negotiating the reunification. We are trying to find a self-diffeomorphism Φ of $M_{1/2}$ such that, after re-gluing W through it, the homology of W disappears. In other words, we wish to arrange Φ so that the 3–handles from the upper half cancel *algebraically* the 2–handles of the lower half.

Whether a certain Φ is good or not for this purpose is entirely determined by the self-isomorphism Φ_* that Φ induces on the 2–homology of $M_{1/2}$. Therefore, for finding a good diffeomorphism Φ, we will proceed by reverse-engineering: we will determine a good algebraic automorphism

$$\widetilde{\varphi}\colon H_2(M_{1/2}; \mathbb{Z}) \approx H_2(M_{1/2}; \mathbb{Z}) ,$$

preserving the intersection form of $M_{1/2}$, and then use Wall's earlier theorem on diffeomorphisms to claim that $\widetilde{\varphi}$ can be realized as Φ_* of some self-diffeomorphism Φ of $M_{1/2}$. Wall's theorem on diffeomorphisms might require that we add an extra copy of $\mathbb{S}^2 \times \mathbb{S}^2$, but that can be achieved immediately by the creation in W^5 of a (geometrically) canceling pair of a 2– and a 3–handle—the trace of such a pair in $M_{1/2}$ is exactly the required extra $\mathbb{S}^2 \times \mathbb{S}^2$–summand.

Each $\mathbb{S}^2 \times \mathbb{S}^2$–summand in $M_{1/2}$ appears from a 2–handle $\mathbb{D}^2 \times \mathbb{D}^3$, attached to M along $\mathbb{S}^1 \times \mathbb{D}^3$. The belt sphere of this 2–handle is $0 \times \mathbb{S}^2$. The homological hole created by the addition of the 2–handle is represented by the

---

[26] Back on page 151.

[27] This result is due to V. Rokhlin, and states: *Any spin 4–manifold with zero signature must bound a spin 5–manifold.* For the concept of spin manifold, look ahead at section 4.3 (page 162); the result itself will be restated on page 165.

first sphere-factor of $S^2 \times S^2$ in $M_{1/2}$, while the belt sphere of the handle survives as the *second* factor of $S^2 \times S^2$ and is filled by the handle itself.

Looking now at the upper half of W^5, a 3-handle is a copy of $\mathbb{D}^3 \times \mathbb{D}^2$, attached to the lower half through $S^2 \times \mathbb{D}^2$. The attaching sphere of the 3-handle is $S^2 \times 0$. Therefore, if the 3-handle is to algebraically cancel a 2-handle from the lower half, then the attaching sphere $S^2 \times 0$ of the 3-handle must intersect the belt sphere $0 \times S^2$ of the 2-handle algebraically exactly once.[28] Indeed, in "handle homology", we would then have $\partial(3\text{-}handle) = (2\text{-}handle)$. (Intuitively, view the 3-handle as algebraically filling the homological hole $S^2 \times 1$ created by the 2-handle.)

Algebraization. To translate everything into algebra, we proceed as follows: We view $M_{1/2}$ as

$$M_{1/2} = M \# m S^2 \times S^2 ,$$

and we denote by α_k the class of $S^2 \times 1$ and by $\overline{\alpha}_k$ the class of $1 \times S^2$ in the k^{th} $S^2 \times S^2$-summand. The classes $\overline{\alpha}_k$ are the classes of the belt spheres of the lower 2-handles, and they bound in the lower cobordism. We write

$$H_2(M_{1/2}; \mathbb{Z}) = H_2(M; \mathbb{Z}) \oplus \mathbb{Z}\{\alpha_1, \overline{\alpha}_1, \ldots, \alpha_m, \overline{\alpha}_m\} ,$$

with corresponding intersection form $Q_{M_{1/2}} = Q_M \oplus m H$.

Now we look at $M_{1/2}$ from upwards as

$$M_{1/2} = N \# m S^2 \times S^2 .$$

This decomposition is obtained by adding upside-down 2-handles to N in the upper half of W. For trivial algebraic reasons, the $S^2 \times S^2$-summands added to N are just as many as those added to M, but the respective summands in the two decompositions *do not* correspond by, say, a diffeomorphism (unless $M \cong N$).

Denote by β_k the class of $S^2 \times 0$ and by $\overline{\beta}_k$ the class of $0 \times S^2$ in the k^{th} $S^2 \times S^2$-summand of this latter splitting. The classes β_k are the classes of the attaching spheres of the upper 3-handles, and they bound in the upper cobordism. And we write

$$H_2(M_{1/2}; \mathbb{Z}) = H_2(N; \mathbb{Z}) \oplus \mathbb{Z}\{\beta_1, \overline{\beta}_1, \ldots, \beta_m, \overline{\beta}_m\} ,$$

with corresponding intersection form $Q_{M_{1/2}} = Q_N \oplus m H$.

A good self-diffeomorphism Φ of $M_{1/2}$ will be one that sends the class β_k onto α_k, thus guaranteeing that the attaching sphere β_k of each 3-handle has algebraic intersection $+1$ with the belt sphere $\overline{\alpha}_k$ of the corresponding 2-handle.

---

[28] Requiring more, such as only one geometric intersection, *i.e.*, that $S^2 \times 0$ from the 3-handle be sent to $S^2 \times 0$ from the 2-handle, implies that these 3- and 2-handles cancel. However, if we could do that for all handles, we would end with a diffeomorphism $M \cong N$, which cannot happen in general.

The final dance. The hypothesis of this theorem states that the intersection forms of M and N are isomorphic. Denote by

$$\varphi\colon H_2(N;\mathbb{Z}) \approx H_2(M;\mathbb{Z})$$

such an intersections-preserving isomorphism. Then we can extend φ to

$$\widetilde{\varphi}\colon H_2(N;\mathbb{Z}) \oplus \mathbb{Z}\{\beta_1,\dots,\overline{\beta}_m\} \approx H_2(M;\mathbb{Z}) \oplus \mathbb{Z}\{\alpha_1,\dots,\overline{\alpha}_m\}$$

by setting

$$\widetilde{\varphi}(\beta_k) = \alpha_k \qquad \text{and} \qquad \widetilde{\varphi}(\overline{\beta}_k) = \overline{\alpha}_k\,.$$

This extended $\widetilde{\varphi}$ is easily seen to still preserve intersections. Therefore, by Wall's theorem on diffeomorphisms, there must exist an actual self-diffeomorphism Φ of $M_{1/2}$ that realizes $\widetilde{\varphi}$ as $\Phi_* = \widetilde{\varphi}$.

Then, if we cut our W^5 into its two halves and glue them back using this Φ, then the resulting cobordism will be simply-connected and with no 2–homology. That is to say, an h–cobordism between M and N. □

4.3. Intersection forms and characteristic classes

Time has come to comment on the other classical invariants of a 4–manifold, specifically on the characteristic classes of its tangent bundle. Only $w_2(T_M)$, $e(T_M)$ and $p_1(T_M)$ are actually relevant in this realm. After first reviewing these, we will relate them to intersection forms.

We start with the **Stiefel–Whitney classes**

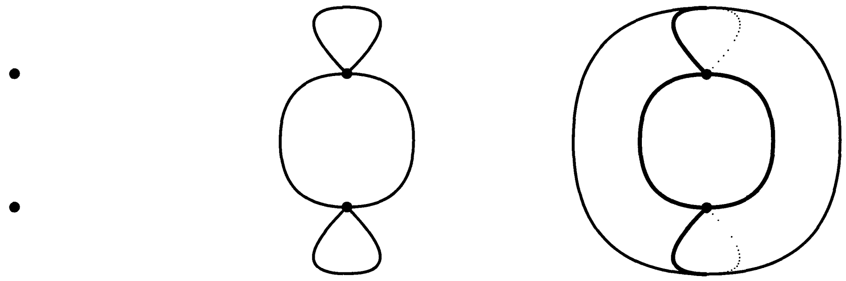
$$w_k(T_M) \in H^k(M^4;\mathbb{Z}_2)\,.$$

The class $w_k(T_M)$ measures the obstruction to finding a field of $4-k+1$ linearly-independent vectors over the k–skeleton of M.

> ***Skeleta.*** *Remember that, for a cellular complex, its k–**skeleton** is the union of all its cells of dimension $\leq k$, as in figure 4.15 on the facing page—similarly, for simplicial complexes (triangulations).*[29] *For a manifold M, one can also think in thickened terms and view the k–skeleton of M as the union of all the handles of order $\leq k$, in some handle decomposition of M; see figure 4.16 on the next page. Of course, the skeleta depend on the choice of cellular/handle decompositions.*

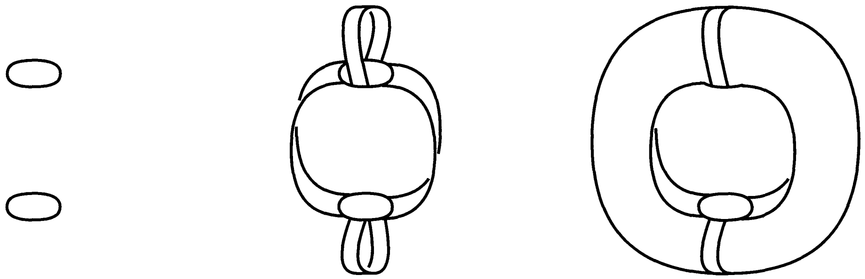
---

[29] Simplices and triangulations are briefly recalled in footnote 5 on page 182 ahead.



0–skeleton 1–skeleton 2–skeleton

4.15. Skeleta of a torus, I: the cells



0–skeleton 1–skeleton 2–skeleton

4.16. Skeleta of a torus, II: the handles

Orientations and the first Stiefel–Whitney class

The class $w_1(T_M)$ measures the obstruction to finding a trivialization T_M over the 1–skeleton of M. It can be defined directly[30] by its values on embedded circles C in M, namely by setting

$$\begin{aligned} w_1(T_M)\cdot C &= 0 && \text{if and only if } T_M|_C \text{ is trivial;} \\ w_1(T_M)\cdot C &= 1 && \text{if and only if } T_M|_C \text{ is not trivial.} \end{aligned}$$

Since a 4–plane bundle over a circle is either trivial or non-orientable, we observe that the first Stiefel–Whitney class merely detects orientation-reversing loops in M. Therefore w_1 is the obstruction to M being orientable.

Along these lines, it is not hard to see that an **orientation** of M is equivalent to a choice of trivialization of T_M over the 0–skeleton that can be extended over the 1–skeleton, considered up to homotopies.

Since we restricted our attention to oriented 4–manifolds, this class is not very interesting to us. Quite the opposite, though, can be said about the next Stiefel–Whitney class:

---

[30] Since $H^1(M;\mathbb{Z}_2) = \mathrm{Hom}_{\mathbb{Z}_2}(H_1(M;\mathbb{Z}_2),\ \mathbb{Z}_2)$, nothing is lost.

Spin structures and the second Stiefel–Whitney class

The second Stiefel–Whitney class

$$w_2(T_M) \in H^2(M; \mathbb{Z}_2)$$

measures the obstruction to finding a 3–frame over the 2–skeleton. If w_1 was trivial and we picked an orientation of M, then by using this orientation we can complete any 3–frame to a 4–frame. Therefore we can say that, for oriented manifolds, $w_2(T_M)$ is the obstruction to trivializing T_M over the 2–skeleton[31] of M.

The origin of the $\mathbb{Z}_2$–coefficients of w_2 is in[32] $\pi_1 SO(4) = \mathbb{Z}_2$. The generator of the latter is any path of rotations of angles increasing from 0 to 2π; if the angle keeps further increasing to 4π, then the resulting loop will be null-homotopic in $SO(4)$. For trivializations of T_M, it is best to think of $SO(4)$ as the space of orienting orthonormal frames in $\mathbb{R}^4$. The class $w_2(T_M)$ is obtained by patching together local obstructions over each 2–cell D of M: a trivialization of T_M over the 1–skeleton induces a map $\varphi\colon \partial D \to SO(4)$; the trivialization extends across D if and only if φ extends over D, in other words, if φ represent the trivial element of $\pi_1 SO(4)$.

> ***Displaying $w_2(T_M)$ as a cochain.*** *Given a random trivialization of T_M over the 1–skeleton of M, we can define a cellular cochain ϑ for $w_2(T_M)$ by assigning $1 \in \mathbb{Z}_2$ to any 2–cell D across which the chosen trivialization cannot be extended. This cochain will be trivial if and only if the trivialization extends over the 2–skeleton. Of course, one can try to go back and change the trivialization over the 1–skeleton, then check again. It turns out that all such changes modify our cellular cochain ϑ by the addition of a coboundary. Further, our cochain turns out to be a cocycle. Therefore, the existence of a trivialization that extends is equivalent to the cohomology class of ϑ being trivial.*[33] *(Observe that such a discussion can very well be carried out with 2–handles instead of 2–cells; the cocycle above assigns to each 2–handle the framing coefficient*[34] *modulo 2 of its attaching circle.)*

Look at surfaces. Since "2–skeleton" might not be your friendliest of notions, we can also rely upon

Lemma. *The second Stiefel–Whitney class $w_2(T_M) \in H^2(M; \mathbb{Z}_2)$ is the obstruction to trivializing T_M over the oriented surfaces embedded in M.*

---

[31] Keep in mind that, the manifold being oriented, T_M can already be trivialized over the 1–skeleton.

[32] The group $SO(4)$ is the group of orientation-preserving isometries of $\mathbb{R}^4$, *i.e.*, its group of rotations.

[33] This is obstruction theory and is better explained in the end-notes of this chapter (page 197).

[34] Compare also with Kirby calculus, in the end-notes of chapter 2 (page 91).

Proof. On one hand, we have $H^2(M;\mathbb{Z}_2) = \operatorname{Hom}_{\mathbb{Z}_2}\big(H_2(M;\mathbb{Z}_2),\,\mathbb{Z}_2\big)$, and thus w_2 is completely determined by its values $w_2 \cdot x$ on all modulo 2 classes $x \in H_2(M;\mathbb{Z}_2)$. On the other hand, when $H_1(M;\mathbb{Z})$ has no 2–torsion (for example when M is simply-connected), we further have that $H_2(M;\mathbb{Z}_2) = H_2(M;\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}_2$, or, in other words, classes in $H_2(M;\mathbb{Z}_2)$ are just modulo 2 reductions of integral classes from $H_2(M;\mathbb{Z})$. Therefore w_2 is completely determined by its values $w_2 \cdot S$ on the oriented surfaces S of M. Furthermore, $w_2(T_M) \cdot S = w_2(T_M|_S)$ is precisely the obstruction to trivializing T_M over S. □

Thus, when M is simply-connected, we can define $w_2(T_M)$ directly by

$$\begin{aligned} w_2(T_M) \cdot S &= 0 \qquad \text{if and only if } T_M|_S \text{ is trivial,} \\ w_2(T_M) \cdot S &= 1 \qquad \text{if and only if } T_M|_S \text{ is not trivial,} \end{aligned}$$

for each oriented surface S embedded in M.

Look at self-intersections. By using the obvious splitting of T_M over any surface S as $T_M|_S = T_S \oplus N_{S/M}$, we compute

$$\begin{aligned} w_2(T_M) \cdot S &= w_2(T_M|_S) \\ &= w_2(T_S \oplus N_{S/M}) \\ &= w_2(T_S) + w_2(N_{S/M}) + w_1(T_S) \cdot w_1(N_{S/M}) \,. \end{aligned}$$

Since both T_S and $N_{S/M}$ are orientable, the last term vanishes. More, since $w_2(T_S)$ is the modulo 2 reduction of the Euler class $\chi(S) = 2 - 2\,\mathrm{genus}(S)$, the first term on the right vanishes as well. We are left with $w_2(N_{S/M})$, which is the modulo 2 reduction of $e(N_{S/M})$. The latter measures the self-intersection of S in M. We have proved:

Wu's Formula. *For all oriented surfaces S embedded in M, we have:*

$$w_2(T_M) \cdot S = S \cdot S \pmod 2 \,. \qquad \square$$

This is the 4–dimensional case of the general Wu formula.[35] A verbose but more concrete alternative proof will appear on page 168 in the next section.

A nice consequence of Wu's formula is:

Corollary. *If $w_2(T_M) = 0$, then the intersection form of M is even.* □

The converse is true whenever $H_1(M;\mathbb{Z})$ has no 2–torsion.

---

[35] Wu's formula is a general statement about Stiefel–Whitney classes; see for example **J. Milnor** and **J. Stasheff**'s ***Characteristic classes*** [MS74].

Spin structures. Since $w_2(T_M)$ is the obstruction to trivializing T_M over the 2–skeleton of M, in the spirit of the earlier re-definition of orientations, we can define the concept of spin structure:

A **spin structure** on M is a choice of trivialization of T_M over the 1–skeleton that can be extended over the 2–skeleton, considered up to homotopies. Various alternative ways of defining spin structures and related matters are contained in the end-notes of this chapter.[36] A manifold endowed with a spin structure is called a **spin manifold.**[37]

Then we can state that $w_2(T_M) = 0$ if and only if M admits a spin structure. The simplest examples of spin 4–manifolds are $\mathbb{S}^4$, $\mathbb{S}^2 \times \mathbb{S}^2$, and the $K3$ surface. In general:

Corollary (*Spin structures and even forms*). *Any 4–manifold without 2–torsion, for example simply-connected, admits spin structures if and only if its intersection form is even.* □

Action of $H^1(M;\mathbb{Z}_2)$ on spin structures. *Let $\mathfrak{s}$ be a spin structure on M, described by a trivialization of T_M over the 1–skeleton of M (for some fixed triangulation of M). Choose a class $\alpha \in H^1(M;\mathbb{Z}_2)$ and represent it by its dual unoriented 3–submanifold Y_α in M. Arrange that Y_α does not touch any vertex of M's triangulation and is transverse to all its edges. Then one can define a new spin structure $\alpha \cdot \mathfrak{s}$ on M by twisting $\mathfrak{s}$'s trivialization over each edge ε that meets Y_α through the addition of a 2π–twist each time ε meets Y_α. For every loop ℓ in the 1–skeleton that bounds a 2–simplex D, the intersection of Y_α and D occurs along arcs linking the intersections points of ℓ and Y_α; therefore there must be an even number of such intersection points, and so the trivialization offered by $\alpha \cdot \mathfrak{s}$ along ℓ differs from $\mathfrak{s}$'s by an even number of 2π–twists; hence the trivialization of $\alpha \cdot \mathfrak{s}$ still extends across D—it is indeed a spin structure.*

The resulting action of $H^1(M;\mathbb{Z}_2)$ on the set of all spin structures of M is free and transitive.[38] Therefore, after fixing a spin structure on M, this action establishes a bijective correspondence between the elements of $H^1(M;\mathbb{Z}_2)$ and the set of all spin structures on M (the correspondence depends on the choice of "base" spin structure). In particular, if M is simply-connected and has $w_2(T_M) = 0$, then M admits a unique spin structure.

---

[36] For the more usual, differential-geometric definition, see the end-notes of this chapter (page 174); see also section 10.2 (page 383) ahead. A homotopy-theoretic definition is presented in the end-notes of this chapter (page 204).

[37] Often, one calls "spin manifold" any manifold that admits a spin structure, even if no specific structure has been chosen, instead of more honestly naming it, for example, "spinnable manifold".

[38] The action of a group G on a set S is called **transitive** if for every two elements s' and s'' of S there is some $g \in G$ so that $g \cdot s' = s''$. The action is called **free** if we can have $g \cdot s = s$ for some $s \in S$ only when $g = 1$.

Signatures and bounding spin-manifolds. In the context of spin structures, an important result is the spin version of the bounding theorem from section 3.2 (page 123). The latter stated that all zero-signature 4–manifolds must bound some oriented 5–manifold. For spin 4–manifolds, the following refinement is true:

Theorem (*V. Rokhlin*). *If a closed 4–manifold M is endowed with a spin structure and has*

$$\operatorname{sign} Q_M = 0\,,$$

then there exists a spin 5–manifold W^5 that is bounded by M so that the spin structure of W induces the spin structure of M. □

Spin structures on 5–manifolds are defined exactly as for manifolds of dimension 4: they are trivializations of T_W over the 1–skeleton that extend over the 2–skeleton.[39] A spin structure on W^5 induces a spin structure on ∂W by using an outward-pointing trivialization of the normal bundle $N_{\partial W/W}$ to obtain a trivialization of $T_{\partial W}$ over its 1–skeleton, *etc.*

In particular, it follows that:

Corollary (*Spin cobordism*). *If two spin 4–manifolds M and N have the same signature, then they can be linked by a cobordism W^5 that is a spin 5–manifold, and its spin structure induces on M and N their respective spin structures.* □

Notice that we have already relied on this result in the proof of Wall's theorem on h–cobordisms (page 157).

Third Stiefel–Whitney class

The third Stiefel–Whitney class $w_2(T_M) \in H^3(M; \mathbb{Z}_2)$ turns out to be rather uninteresting:

On one hand, if M is orientable and admits spin structures, equivalently if both $w_1(T_M)$ and $w_2(T_M)$ vanish, then $w_3(T_M)$ must vanish as well. Indeed, any spin structure offers a trivialization of T_M over the 2–skeleton, and since the group $\pi_2 SO(4)$ is trivial, this trivialization can always be extended across the whole 3–skeleton[40] of M.

---

[39] More geometrically, a 5–manifold W admits spin structures if and only if every surface embedded in W has trivial normal bundle. As we saw, a 4–manifold M admits spin structures if and only if every surface embedded in M has normal bundle of even Euler class.

[40] Indeed, think of $SO(4)$ as the space of orthonormal frames in $\mathbb{R}^4$. Take a 3–cell E with $T_M|_{\partial E}$ trivialized. The trivialization determines a map $\partial E \to SO(4)$, which, since $\pi_2 SO(4) = 0$, must be null-homotopic and thus extend to a map $E \to SO(4)$; but the latter is just a trivialization of $T_M|_E$. The relation between the $\pi_k SO(m)$'s and w_k's is probably best viewed under the light of the concepts presented in the end-notes of this chapter, on page 197 and page 204.

> *Since we can always choose handle decompositions of M with exactly one 4-handle and then shrink that 4-handle toward a point, we deduce that every spin 4-manifold M has T_M trivial over $M \setminus \{point\}$; such manifolds are called* **almost-parallelizable.**[41]

In general, the values of $w_3(T_M) \in H^3(M; \mathbb{Z}_2)$ do not matter—they are determined by the other characteristic classes of M, as will become clear a bit ahead, from the Dold-Whitney theorem.

The Euler class

The fourth and last Stiefel–Whitney class $w_4(T_M) \in H^4(M; \mathbb{Z}_2)$ is not the only remaining obstruction to trivializing T_M over the whole M. In fact, if M is oriented, then $w_4(T_M)$ can be refined to the integral **Euler class**

$$e(T_M) \in H^4(M; \mathbb{Z}) = \mathbb{Z}\,.$$

The Euler class counts the self-intersections of M, viewed as the zero-section inside the manifold T_M. Equivalently, it counts the zeros of a generic vector field on M, and we have $e(T_M) = \chi(M)$. If $e(T_M) = 0$, then T_M admits a nowhere-zero section. Clearly though, all simply-connected 4–manifolds have $e(T_M) = 2 + \operatorname{rank} Q_M$ and hence $e(T_M) > 0$.

Signatures and the Pontryagin class

Another relevant class is the **Pontryagin class**

$$p_1(T_M) \in H^4(M; \mathbb{Z}) = \mathbb{Z}\,.$$

It is defined in terms of Chern classes as $p_1(T_M) = -c_2(T_M \otimes \mathbb{C})$ and can be interpreted as the obstruction to finding three $\mathbb{C}$–linearly-independent global sections in $T_M \otimes \mathbb{C}$.

> *More obscurely, the Pontryagin number also coincides with -3 times the algebraic count of triple-points of a generic immersion*[42] $M^4 \to \mathbb{R}^6$.

On a 4–manifold the Pontryagin class is completely determined by its intersection form, owing to the 4–dimensional instance of F. Hirzebruch's celebrated signature theorem:

Hirzebruch's Signature Theorem. *For every closed 4–manifold M we have*

$$p_1(T_M) = 3 \operatorname{sign} Q_M\,. \qquad \square$$

---

[41] A manifold is called **parallelizable** if its tangent bundle is trivial over the whole manifold. An example of parallelizable 4–manifold is $S^1 \times S^3$; there are no simply-connected examples.

[42] See **R. Herbert**'s ***Multiple points of immersed manifolds*** **[Her81]**; also proved in **R. Kirby**'s ***The topology of 4–manifolds*** [Kir89, ch IV].

Signatures and bounding manifolds, revisited. *We quoted earlier*[43] *the fact that, if a 4-manifold has vanishing signature, then it must bound an oriented 5-manifold. A proof of that statement can be assembled by using the signature theorem, together with the above interpretation of p_1 in terms of triple-points of immersions.*

First, one builds an immersion of M into $\mathbb{R}^6$ (by using immersion theory, it is enough to build a candidate for the normal bundle of the immersed M inside $\mathbb{R}^6$, and thus the problem is reduced to a characteristic class computation). Such an immersion will have double-points, forming surfaces in M, and will have isolated triple-points. Since $3\,\mathrm{sign}\,Q_M = p_1(M)$, and the latter is an algebraic count of these triple-points, we conclude that the triple-points cancel algebraically. Furthermore, there is a modification of M inside $\mathbb{R}^6$ that geometrically eliminates all these triple-points[44] *and changes M merely by a cobordism inside $\mathbb{R}^6$. After that, the double points can be eliminated without obstruction (think of our method for eliminating double-points of surfaces in 4-space*[45] *and cross with $\mathbb{R}^2$), and this further changes M by a cobordism inside $\mathbb{R}^6$. We end up with a 4-manifold embedded in $\mathbb{R}^6$. Since the result is homologically-trivial and embedded, it must bound a 5-manifold W inside*[46] *$\mathbb{R}^6$. Putting together the cobordisms used to modify M with this last 5-manifold yields a filling 5-manifold for our initial 4-manifold.*[47]

That's it, the bundle is done

The above-mentioned characteristic classes completely determine T_M as a vector bundle. In fact, only w_2, e and p_1 are needed:

Dold–Whitney Theorem. *If two oriented 4-plane bundles over an oriented 4-manifold have the same second Stiefel–Whitney class w_2, Pontryagin class p_1 and Euler class e, then they must be isomorphic.* □

All these three characteristic classes can be related to intersection forms. In review, by using the partial Betti numbers $b_2^{\pm}$ we can write, for every simply-connected 4-manifold M,

$$e(T_M) = b_2^+(M) + b_2^-(M) + 2\,,$$
$$p_1(T_M) = b_2^+(M) - b_2^-(M)\,,$$

and recall that $w_2(T_M)$ vanishes exactly when Q_M is even.

---

[43] See back in section 3.2 (page 123).

[44] Somewhat in the spirit of figure 11.7 on page 486.

[45] Look back at figure 3.1 on page 113.

[46] Owing to a general result of R. Thom, stated back in footnote 3 on page 112.

[47] See **R. Kirby**'s ***The topology of 4-manifolds*** [**Kir89**, ch VIII] for the full argument.

4.4. Rokhlin's theorem and characteristic elements

We continue the story of the second Stiefel–Whitney class $w_2(T_M)$, but this time by focusing on the integral classes that reduce to it. Afterwards, we state a fundamental theorem for topology in general, namely Rokhlin's theorem: a smooth spin 4–manifold can only have a multiple of 16 as its signature.

Characteristic elements of the intersection form

We defined $w_2(T_M) \in H^2(M;\mathbb{Z}_2)$ as the obstruction to trivializing T_M over the 2–skeleton of M. We now look at representations of the class $w_2(T_M)$ by oriented surfaces and integral classes.

Make it a surface. Assume that $w_2(T_M)$ can be realized as an *oriented* surface Σ embedded in M. In other words, assume that $[\Sigma] \in H_2(M;\mathbb{Z})$ is (Poincaré-dual to) an integral lift $\underline{w}$ of the class w_2. Such a surface Σ with

$$\Sigma = w_2(T_M) \pmod 2$$

is called a **characteristic surface** of M, while its class $\underline{w} \in H_2(M;\mathbb{Z})$ is called a **characteristic element.**[48] Characteristic elements are certainly not unique: just add to such a $\underline{w}$ any even class 2γ to obtain another integral lift of w_2. Remember that we encountered characteristic elements before, in Wall's theorem on the automorphisms of an intersection form.[49]

Wu, again. Take now a random surface S in M. The obstruction to trivializing T_M over S is then given by $w_2(T_M) \cdot S \pmod 2$ or, in other words, by $\Sigma \cdot S \pmod 2$. We have already seen that this coincides modulo 2 with the self-intersection $S \cdot S$, but we prove it once again using a slightly different argument.

Wu's Formula. *Let M be a simply-connected 4–manifold. An oriented surface Σ is characteristic if and only if*

$$\Sigma \cdot S = S \cdot S \pmod 2$$

for all oriented surfaces S inside M.

Proof. Let $\tau \in \Gamma(T_S)$ be a vector field tangent to S, and let $\nu \in \Gamma(N_{S/M})$ be a field normal to S. If τ and ν are generic, then they are zero only at isolated points of S. Arrange that τ and ν are never zero at a same point of S. Pick a vector field τ^* complementary to τ in T_S, so

---

[48] Another customary name is *characteristic class*, but we will use "characteristic element" throughout, to avoid any chance of confusion with characteristic classes of the tangent bundle.

[49] See back in section 4.2 (page 152).

that τ^* is zero only at the zeros[50] of τ. Also pick a complement ν^* to ν in $N_{S/M}$ that is zero only at the zeros of ν. Then the vector field $\tau^* + \nu^*$ is nowhere-zero on S. The 3–frame $\{\tau,\ \nu,\ \tau^* + \nu^*\}$ can be completed to a full 4–frame of T_M, well-defined on S away from the zeros of τ and the zeros of ν.

Against extending this frame across the remaining points of S lies a $\mathbb{Z}_2$–obstruction: indeed, a neighborhood of a singularity is a copy of $\mathbb{D}^2 \setminus 0$, and the frame-field around 0 defines a map $f\colon \mathbb{S}^1 \to SO(4)$; the frame-field can be extended across 0 if and only if f is homotopically-trivial in $\pi_1 SO(4) = \mathbb{Z}_2$. It is not hard to argue that the obstructions at various singularities can be added together,[51] and thus yield a global $\mathbb{Z}_2$–obstruction to extending the frame-field over the whole surface S. Since $\tau^* + \nu^*$ is nowhere-zero, this obstruction comes entirely from the zeros of τ and ν.

Since τ and ν were chosen generic, their zeros are simple, and thus the obstruction can be computed as

$$\text{obstruction} = \#\{\text{zeros of } \tau\} + \#\{\text{zeros of } \nu\} \pmod 2 .$$

However, the number of zeros of a tangent vector field like τ is equivalent modulo 2 to $\chi(S)$, which is always even and thus disappears from the above formula. We are left with the number of zeros of the normal vector field ν, which is equivalent modulo 2 to $S \cdot S$. In conclusion,

$$\text{obstruction} = S \cdot S \pmod 2 .$$

However, the same obstruction can also be seen to be $w_2(T_M|_S) = w_2(T_M) \cdot S = \Sigma \cdot S \pmod 2$, and this concludes the proof. □

It might be amusing to look back at page 163 and compare the two proofs that relate w_2 to self-intersections—the version above is essentially just a more concrete version of the computations made there.

In any case, the property that $w_2 \cdot x = x \cdot x \pmod 2$ for all $x \in H^2(M; \mathbb{Z})$ completely determines the class $w_2(T_M)$ inside $H^2(M; \mathbb{Z}_2)$. In particular, if we find an integral class $\underline{w} \in H_2(M; \mathbb{Z})$ satisfying

$$\underline{w} \cdot x = x \cdot x \pmod 2 ,$$

then the modulo 2 reduction of $\underline{w}$ must be $w_2(T_M)$: we have found a *characteristic element* of the intersection form.

---

[50] For example, pick a complex structure on T_S and define $\tau^* = i\tau$.

[51] For example, by using an argument similar to the classic Poincaré–Hopf theorem on indices of vector fields: if the sum of indices is zero, then there is a nowhere-zero vector field. Here, since τ and ν are generic, the indices are ± 1; further, since we are dealing with a 4–plane bundle over a surface, the sum of indices only matters modulo 2.

They do exist. Characteristic elements (and hence characteristic surfaces) exist in all 4–manifolds:

Lemma. *On every 4–manifold M, there always exist integral classes $\underline{w}$ such that*

$$\underline{w} \cdot x = x \cdot x \pmod 2$$

for all $x \in H_2(M;\mathbb{Z})$.

Proof. This is a purely algebraic argument. Let $Q\colon Z \times Z \to \mathbb{Z}$ be a symmetric bilinear unimodular form, defined over a free $\mathbb{Z}$–module Z. We can build its modulo 2 reduction by taking $Z'' = Z/2Z$ and $Q'' = Q \pmod 2$. We obtain a symmetric $\mathbb{Z}_2$–bilinear unimodular form

$$Q''\colon Z'' \times Z'' \longrightarrow \mathbb{Z}_2\,.$$

The unimodularity of Q'' over $\mathbb{Z}_2$ translates as the following property: for every $\mathbb{Z}_2$–linear function $f\colon Z'' \to \mathbb{Z}_2$ there must be some element $x_f \in Z''$ so that $f(\,\cdot\,) = Q''(x_f,\,\cdot\,)$. However, since $(a+b)\cdot(a+b) = a\cdot a + b\cdot b + 2a\cdot b \equiv a\cdot a + b\cdot b \pmod 2$, we notice that the correspondence $x \longmapsto Q''(x,x)$ is additive, and thus is $\mathbb{Z}_2$–linear. Therefore there must exist an element $w'' \in Z''$ so that $Q''(x,x) = Q''(w'',x)$; in other words, we have

$$w'' \cdot x = x \cdot x \pmod 2 \qquad \text{for all } x \in Z''\,.$$

Since the element $w'' \in Z'' = Z/2Z$ represents a coset of Z, there must be integral elements $\underline{w} \in Z$ whose modulo 2 reduction is w''. In other words, there always exist characteristic elements for Q, *i.e.*, elements $\underline{w} \in Z$ with $\underline{w} \cdot x = x \cdot x \pmod 2$ for all $x \in Z$. □

The existence of integral lifts of $w_2(T_M)$ is important also because of spin$^{\mathbb{C}}$ structures (complexified spin structures). As we will see later[52] the existence of $\underline{w}$'s is equivalent to the existence of spin$^{\mathbb{C}}$ structures on M; the latter will play an essential role in Seiberg–Witten theory.

Rokhlin's theorem

First, an algebraic argument shows that:

Van der Blij's Lemma. *For every characteristic element $\underline{w}$ we must have*

$$\operatorname{sign} Q_M = \underline{w} \cdot \underline{w} \pmod 8\,. \qquad \square$$

We prove this statement in the end-notes of the next chapter (page 263).[53]

In particular, it follows that every spin manifold (for which we can always pick $\underline{w} = 0$) must have signature multiple of 8. Surprisingly, more is true:

---

[52] In section 10.2 (page 382).

[53] The reason for this postponement is not the difficulty of the argument, but merely its reliance on the classification of algebraic forms, which is discussed in the next chapter.

Rokhlin's Theorem. *If M^4 is smooth and has $w_2(T_M) = 0$, then its intersection form must have*

$$\operatorname{sign} Q_M = 0 \pmod{16} . \qquad \square$$

In part for reasons of space, proofs of this theorem are exiled to the endnotes of chapter 11 (one proof starting on page 507, another starting on page 521).

> ***Three's company.*** *Notice that we have already encountered several statements due to V. Rokhlin: one from page 123 (about zero-signature manifolds bounding), one from a few pages back (about zero-signature spin-manifolds spin-bounding), and the one right above.*[54] *In this volume, only the last result will be called "Rokhlin's theorem".*

Smooth exclusions. A first consequence of Rokhlin's theorem is that E_8 can never be the intersection form of a smooth simply-connected 4–manifold: indeed, E_8 is an even form with signature 8. In particular it follows that, as we claimed earlier, the E_8–manifold $\mathcal{M}_{E_8}$ does not admit any smooth structures at all.

> *Historically, we should note that, even though it was clear from Rokhlin's theorem that the E_8–form would never appear as the intersection form of a smooth 4–manifold, it was not known until Freedman's work that the E_8–form does nonetheless appear as the intersection form of a topological 4–manifold. Indeed, recall*[55] *that the definition of $\mathcal{M}_{E_8}$ involves Freedman's contractible Δ's, whose construction in turn needs Freedman's major result on Casson handles.*

More generally, since E_8 has signature 8 and H has signature 0, we deduce:

Corollary. *If M is smooth and has no 2–torsion, for example when M is simply-connected, and its intersection form is*

$$Q_M = \oplus \pm m \, E_8 \oplus n \, H ,$$

then m must be even. $\square$

As we will see shortly, *all* even indefinite intersection forms do in fact fall under the jurisdiction of this corollary.

We should note that the absence of 2–torsion is essential: the complex Enriques surface (doubly-covered by $K3$) has intersection form $-E_8 \oplus H$ but fundamental group $\pi_1 = \mathbb{Z}_2$; its 2–torsion allows the intersection form to be even without w_2 vanishing, and hence Rokhlin's theorem does not apply.

---

[54] Furthermore, all three results appeared in the same *four-pages*-long paper, ***New results in the theory of four-dimensional manifolds*** [**Rok52**].

[55] From section 2.3 (page 86).

It is also worth noting the fact that, for the thirty years between Rokhlin's and Donaldson's work, *no* new methods of excluding intersection forms from the smooth realm were discovered. Indeed, Rokhlin in the 1950s excluded E_8 from ever being the intersection form of a smooth 4–manifold, but the form $E_8 \oplus E_8$ was only excluded by Donaldson in the 1980s.

Other consequences. Rokhlin's theorem is a fundamental result in topology. Its consequences extend quite far, as we will comment in the various notes at the end of this chapter. For example, Rokhlin's theorem sends its tentacles into dimension 3 (the Rokhlin invariant, defined in the end-note on page 224), as well as into high dimensions (the Kirby–Siebenmann invariant, governing whether a topological manifold admits smooth structures, see the end-note on page 207); the theorem is essentially equivalent to the fact that for big n we have $\pi_{n+3} S^n = \mathbb{Z}_{24}$ instead of $\mathbb{Z}_{12}$.

Rokhlin's theorem also admits generalizations in dimension 4, such as:

Corollary (*M. Kervaire & J. Milnor*). *Let M be any smooth 4–manifold. If Σ is a characteristic sphere in M, then we must have:*

$$\operatorname{sign} M = \Sigma \cdot \Sigma \pmod{16} . \qquad \square$$

This last result was put to use for determining which characteristic elements cannot be represented by embedded spheres, and a fuller discussion will be carried through in section 11.1 (page 482).

An even further generalization of Rokhlin's theorem, due to M. Freedman and R. Kirby, is the formula

$$\operatorname{sign} M = \Sigma \cdot \Sigma + 8 \operatorname{Arf}(M, \Sigma) \pmod{16} ,$$

involving general characteristic surfaces Σ and needing a correction term $\operatorname{Arf}(M, \Sigma)$, with values in $\mathbb{Z}_2$ and depending only on the homology class of Σ. This last statement will be fully explained and proved in the end-notes[56] of chapter 11. Since the Freedman–Kirby formula will be proved from scratch, in particular it will offer a complete proof of Rokhlin's theorem. If one wishes so, one can skip ahead and read it right now.[57]

---

[56] Statement and heuristics starting on page 502 and detailed proof starting on page 507. An alternative spin-flavored proof starts on page 521.

[57] It is recommended, though, to first visit with the end-notes of chapter 10 (the characteristic cobordism group, page 427) and the end-notes of chapter 11 (the Arf invariant, page 501). This late placement of the proof of Rokhlin's theorem owes more to reasons of space organization of this volume, than to logical structure.

4.5. Notes

Introduction

Half of the following notes can be viewed as comments on the concept of spin structure. Part of this emphasis can be justified by the foundational role that their complex cousins—spin$^{\mathbb{C}}$ structures—play in the definition of the Seiberg–Witten invariants that we will encounter in chapter 10. Another (non-disjoint) half of the notes can be viewed as comments on Rokhlin's theorem.

In the main text we defined spin structures as extendable trivializations. The more usual definition is in terms of a reduction of the structure group of T_M to the group $Spin(4)$. The first note (page 174) is devoted to explaining this definition. For this purpose, the concept of cocycle defining a vector bundle is first introduced. The note ends with a comment on the non-spin case and with the definition of principal bundles and their relation to spin structures.

The second note (page 181) contains a hands-on proof that the two definitions of spin structures are indeed equivalent. It is a direct argument involving triangulations and cover spaces, and was included owing to its absence from the standard literature.

The third note (page 189) develops the concept of cocycle for a bundle in its natural context: Čech cohomology. We develop this notion just enough to encompass bundle cocycles, but not general sheaf-cohomology. This leads in particular to concrete representations of the Chern class of a complex line bundle and of the second Stiefel–Whitney class of an oriented vector bundle, together with its relation to spin structures.

The fourth note (page 197) is a quick presentation of obstruction theory for bundles; this is a method for encoding the obstacles to building a section of a fiber bundle into suitable cohomology classes. To this is added, in the fifth note (page 204), the concept of classifying spaces for G–bundles. Besides relating these to spin structures and $w_2(T_M)$, both obstruction theory and classifying spaces are needed in the subsequent note.

The sixth note (page 207) presents the theory of endowing topological manifolds with smooth structures, as developed among others by S. Cairns, J. Munkres, J. Milnor, M. Hirsch, B. Mazur, R. Kirby, and L. Siebenmann. For this, tangent bundles for topological manifolds are defined. In dimensions at least 5, a suitable reduction of their structure group (a smoothing of the bundle) can be integrated to a smooth structure on the manifold itself. The obstacles toward this group reduction are investigated using classifying spaces and obstruction theory, and lead to the Kirby–Siebenmann invariant as primary obstruction, as well as to higher obstructions. This theory is weak in dimension 4, but the Kirby–Siebenmann invariant is still defined, and we conclude the note (page 221) by commenting on its 4–dimensional behavior, its strong relation to Rokhlin's theorem, and with a nod toward exotic $\mathbb{R}^4$'s.

We should mention that this note on smoothing theory is a node in the parallel threads of this volume. Inwards, it is a far-reaching consequence of Rokhlin's theorem; a full understanding of it is helped by reading the earlier note on exotic spheres, at the end of chapter 2 (page 97), and the notes ahead on obstruction theory (page 197) and on classifying spaces (page 204). Outwards, it underlies Freedman's classification to be presented in the next chapter. It offers the right contrasting background for the results on smooth 4–manifolds that come from gauge theory, starting with Donaldson's theorem in section 5.3 (page 243) and passing through the exotic $\mathbb{R}^4$'s of section 5.4 (page 250); and it further motivates the Freedman–Kirby generalized Rokhlin theorem to be explained at the end of chapter 11 (page 502).

The seventh note (page 224) presents briefly the Rokhlin invariant of 3–manifolds that appears as a consequence of Rokhlin's theorem. Along the way, the Novikov additivity of signatures for 4–manifolds glued along their boundaries is stated.

The eighth note (page 227) presents the groups that appear by considering two manifolds equivalent if they are cobordant. The oriented cobordism group and the spin cobordism group are displayed.

The ninth note (page 230) explains the Pontryagin–Thom construction. This technique was already used during the geometric proof of Whitehead's theorem and is placed here in its proper place, as a framed cobordism theory. Relations with homotopy groups of spheres are outlined.

Finally, on page 234 are gathered the usual end-of-chapter bibliographical comments. The next chapter starts on page 237; for the sake of continuity the reader is strongly recommended to skip all these notes at a first reading and resume reading there.

Note: Spin structures, the structure group definition

The customary definition of a spin structure is in terms of the *Spin* group, namely as reduction of the structure group of T_M from $SO(4)$ to its simply-connected double-cover $Spin(4)$. In this note we discuss this definition. The equivalence with the definition presented in the main text will be detailed in the next note (page 181). The structure group approach will also be taken up in section 10.2 (page 382), where we will present spin$^{\mathbb{C}}$ structures in order to define the Seiberg–Witten invariants.

Describing vector bundles by using cocycles. A **vector bundle** E of rank k over X^m (also called a **k-plane bundle** over X) is an open $(m+k)$-manifold E together with a map $p\colon E \longrightarrow X$ such that its fibers $p^{-1}[x]$ are vector spaces isomorphic to $\mathbb{R}^k$, and p locally looks like projections $U \times \mathbb{R}^k \longrightarrow U$. In other words, there is an open covering $\{U_\alpha\}$ of X and an atlas of maps

$$\{\varphi_\alpha\colon p^{-1}[U_\alpha] \cong U_\alpha \times \mathbb{R}^k\} ,$$

with $\mathrm{pr}_1 \circ \varphi_\alpha = p$, and so that the overlaps $\varphi_\alpha \circ \varphi_\beta^{-1}$ are described by

$$(x,\, w) \longmapsto (x,\, g_{\alpha\beta}(x) \cdot w)$$

for some suitable change-of-coordinates functions[1]

$$g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow GL(k) ,$$

thus ensuring that the $\mathbb{R}^k$-factors are identified linearly.

The maps $g_{\alpha\beta}$ are in fact all that is needed to describe E: One can just glue-up E from trivial patches $U_\alpha \times \mathbb{R}^k$ by identifying (x, w_α) from $U_\alpha \times \mathbb{R}^k$ with (x, w_β) from $U_\beta \times \mathbb{R}^k$ whenever $w_\alpha = g_{\alpha\beta}(x) \cdot w_\beta$.

For an open covering $\{U_\alpha\}$ of X together with a random collection of maps

$$\{g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow GL(k)\}$$

to actually define a k-plane bundle, certain simple compatibility relations need to be satisfied. These are:

$$g_{\alpha\alpha}(x) = id , \qquad g_{\beta\alpha}(x) = g_{\alpha\beta}(x)^{-1} , \qquad g_{\alpha\gamma}(x) = g_{\alpha\beta}(x) \cdot g_{\beta\gamma}(x) .$$

These three can be contracted into just one condition:

$$g_{\alpha\beta}(x) \cdot g_{\beta\gamma}(x) \cdot g_{\gamma\alpha}(x) = id .$$

The latter is called the **cocycle condition.** Any collection $\{U_\alpha, g_{\alpha\beta}\}$ satisfying it will be called a **cocycle.** (The name of "cocycle" comes from Čech cohomology; this setting will be detailed in the note on page 189 ahead.)

As a simple example of cocycle defining a bundle, if $\{\Phi_\alpha\colon U_\alpha \simeq U'_\alpha \subset \mathbb{R}^m\}$ is an atlas of charts for the smooth manifold X^m, then the cocycle

$$g_{\alpha\beta}(x) = d(\Phi_\alpha \circ \Phi_\beta^{-1})|_x ,$$

made from the derivatives of the overlaps, defines the tangent bundle T_X of X.

Sections. Given a section $s\colon X \longrightarrow E$ of some bundle $E \longrightarrow X$, we can use the charts $\{\varphi_\alpha\colon E|_{U_\alpha} \approx U_\alpha \times \mathbb{R}^k\}$ to express s in coordinates. We obtain a collection of maps $\{s_\alpha\colon U_\alpha \longrightarrow \mathbb{R}^k\}$ given by $s_\alpha = \varphi_\alpha \circ s$. The various local maps s_α are compatible through the relations

$$s_\alpha(x) = g_{\alpha\beta}(x) \cdot s_\beta(x) .$$

Conversely, in terms of cocycles alone, given a set of maps $\{s_\alpha\colon U_\alpha \longrightarrow \mathbb{R}^k\}$, if they satisfy the above compatibility with some cocycle $\{g_{\alpha\beta}\}$, then they define a section in the vector bundle described by $\{g_{\alpha\beta}\}$.

Morphisms. Bundle morphisms can be described in terms of cocycles as well. Consider two bundles $E' \longrightarrow X$ and $E'' \longrightarrow X$ with fibers $\mathbb{R}^m$ and $\mathbb{R}^n$, both over a same base X endowed with

---

[1] In case one finds the notations $GL(m)$ and $SO(m)$ somewhat obscure, they are reviewed later, in section 9.2 (page 333).

a covering $\{U_\alpha\}$. Let E' be described by charts $\{\varphi'_\alpha\}$ and E'' by $\{\varphi''_\alpha\}$, inducing corresponding cocycles $\{g'_{\alpha\beta}\colon U_\alpha \cap U_\beta \to GL(m)\}$ and $\{g''_{\alpha\beta}\colon U_\alpha \cap U_\beta \to GL(n)\}$. Consider any linear bundle morphism $f\colon E' \to E''$, covering the identity $X \to X$. The morphism f can be expressed as a collection of maps $\{f_\alpha\colon U_\alpha \to \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)\}$ obtained by writing f in coordinates through the formulae $\varphi''_\alpha(f(w)) = f_\alpha(x) \cdot \varphi'_\alpha(w)$ for all $w \in E'$ and $x = p(w) \in X$. These f_α's satisfy the relations

$$f_\alpha(x) \cdot g'_{\alpha\beta}(x) = g''_{\alpha\beta}(x) \cdot f_\beta(x) \ .$$

Conversely, in terms of cocycles alone, given a set of maps $\{f_\alpha\colon U_\alpha \to \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)\}$, if they satisfy the above compatibility with some cocycles $\{g'_{\alpha\beta}\}$ and $\{g''_{\alpha\beta}\}$, then they must define a bundle morphism from the bundle defined by $\{g'_{\alpha\beta}\}$ to the one defined by $\{g''_{\alpha\beta}\}$.

Two $GL(k)$–valued cocycles $\{g'_{\alpha\beta}\}$ and $\{g''_{\alpha\beta}\}$, associated to a same covering $\{U_\alpha\}$, describe the same bundle (up to isomorphisms) if and only if there exists a collection of maps $\{f_\alpha\colon U_\alpha \to GL(k)\}$ such that

$$g''_{\alpha\beta}(x) = f_\alpha(x) \cdot g'_{\alpha\beta}(x) \cdot f_\beta(x)^{-1} \ .$$

Indeed, these f_α's are just a description in local coordinates of a vector-bundle isomorphism between the bundles defined by $\{g'_{\alpha\beta}\}'$ and $\{g''_{\alpha\beta}\}$.

By ignoring the underlying vector bundles, we will say directly that two cocycles $\{g'_{\alpha\beta}\}$ and $\{g''_{\alpha\beta}\}$ are **isomorphic** whenever they can be linked with f_α's as above.

For comparing two cocycles $\{g'_{\alpha'\beta'}\}$ and $\{g''_{\alpha''\beta''}\}$ associated to two different coverings $\{U'_{\alpha'}\}$ and $\{U''_{\alpha''}\}$ of M, we can first move to the common subdivision $\{U'_{\alpha'} \cap U''_{\beta''}\}$, then proceed as above.

Keep in mind that any bundle over a contractible set must be trivial, and thus, if one starts with a covering $\{U_\alpha\}$ of X by, say, disks, then such a covering can alone be used to describe *all* bundles over X.

Reductions of structure groups. Let E be a k–plane bundle, and let G be some subgroup of $GL(k)$. If we manage to describe E using a G–valued cocycle $g'_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$, then we say that we have **reduced the structure group** of E from $GL(k)$ to its subgroup G.

This notion can also be described in terms of cocycles alone: Given some cocycle $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to GL(k)$, we say that we reduced its structure group to G if we can find a G–valued cocycle $g'_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$ so that $\{g'_{\alpha\beta}\}$ is isomorphic to $\{g_{\alpha\beta}\}$.

For example, every vector bundle E can be endowed with a fiber-metric (*i.e.*, an inner product in each fiber, varying smoothly from fiber to fiber). Then, by restricting our choice of charts $\varphi_\alpha\colon E|_{U_\alpha} \approx U_\alpha \times \mathbb{R}^k$ to those φ_α's that establish *isometries* between the fibers of E and $\mathbb{R}^k$ (with its standard inner product), we are led to a description of E by an $O(k)$–valued cocycle

$$g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow O(k) \ .$$

We then say that a fiber-metric has reduced the structure group of E from $GL(k)$ to its subgroup $O(k)$.

If our bundle is orientable and we choose an orientation, then, by further restricting the φ_α's to those providing *orientation-preserving* isometries from the fibers of E to $\mathbb{R}^k$, we obtain a $SO(k)$–valued cocycle for E

$$g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow SO(k) \ .$$

We say that an orientation has further reduced the structure group of E from $O(k)$ to its subgroup $SO(k)$.

A spin structure on E can itself be described as a further "reduction" of the structure group of E from $SO(k)$ to the group $Spin(k)$. However, since $Spin(k)$ is not a subgroup of $GL(k)$, this "reduction" has to be developed abstractly, at the level of cocycles and not directly on the vector bundles.

Definition of a spin structure. While the notion of spin structure can be developed for general vector bundles E, for concreteness in what follows we will restrict to the case of the tangent bundle of a 4–manifold. The extension to the general case should be obvious enough.

Start with an oriented 4–manifold M and pick a random Riemannian metric on it. This reduces the structure group of T_M to $SO(4)$, and thus T_M can be described by an $SO(4)$–valued cocycle $\{U_\alpha, g_{\alpha\beta}\}$ with

$$g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow SO(4) .$$

The group $SO(4)$ is connected, but has fundamental group

$$\pi_1 SO(4) = \mathbb{Z}_2 .$$

This fundamental group is generated by a path of rotations of angles increasing from 0 to 2π. On the other hand, if one keeps rotating until reaching 4π, then the resulting loop in $SO(4)$ will be null-homotopic; this can be observed in figure 4.17 on the following page, if properly interpreted. In conclusion, a loop $\ell\colon S^1 \to SO(4)$ is homotopically-trivial if and only if it twists $\mathbb{R}^4$ by an even multiple of 2π, and nontrivial if it twists by an odd multiple.

The fundamental group is unfolded in $SO(4)$'s universal cover, specifically in the Lie group

$$Spin(4) ,$$

which double-covers[2] $SO(4)$.

> ***Ledger.*** *One can think of the Spin group as a method for bookkeeping 2π–rotations: Consider a random loop $\ell\colon [0,1] \to SO(4)$, with $\ell(0) = \ell(1)$. On one hand, if ℓ is homotopically-trivial, then it can be lifted to a loop $\tilde{\ell}$ in $Spin(4)$, with $\tilde{\ell}(0) = \tilde{\ell}(1)$. On the other hand, if ℓ describes a rotation of 2π, then it can only be lifted to an open path with $\tilde{\ell}(0) = -\tilde{\ell}(1)$.*

A **spin structure** on M is defined as a lift of the $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ of T_M to a $Spin(4)$–valued cocycle, considered up to isomorphisms. Specifically, given the $SO(4)$–cocycle

$$g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow SO(4)$$

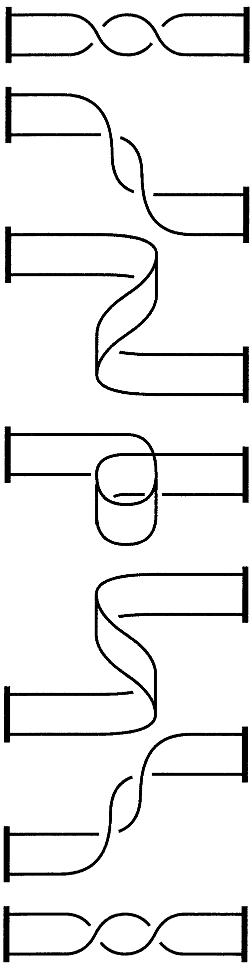
of T_M, we lift these maps against the projection $Spin(4) \to SO(4)$ to get maps[3]

$$\tilde{g}_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow Spin(4) .$$

---

[2] As a bit of help in visualizing $Spin(4) \to SO(4)$ with its $\pi_1 SO(4) = \mathbb{Z}_2$, one can invoke for a moment the thought of $S^2 \to \mathbb{RP}^2$. Or, even better, of $S^3 \to \mathbb{RP}^3$. "Better", because in fact $S^3 = Spin(3)$ and $\mathbb{RP}^3 = SO(3)$. In dimension 4, we have $Spin(4) = S^3 \times S^3$ and $SO(4) = S^3 \times S^3/\pm 1$.

[3] Such a lift is always possible: choose the covering $\{U_\alpha\}$ so that all $U_\alpha \cap U_\beta$'s are simply-connected.



4.17. $\pi_1 SO(n) = \mathbb{Z}_2$ (when $n \geq 3$)

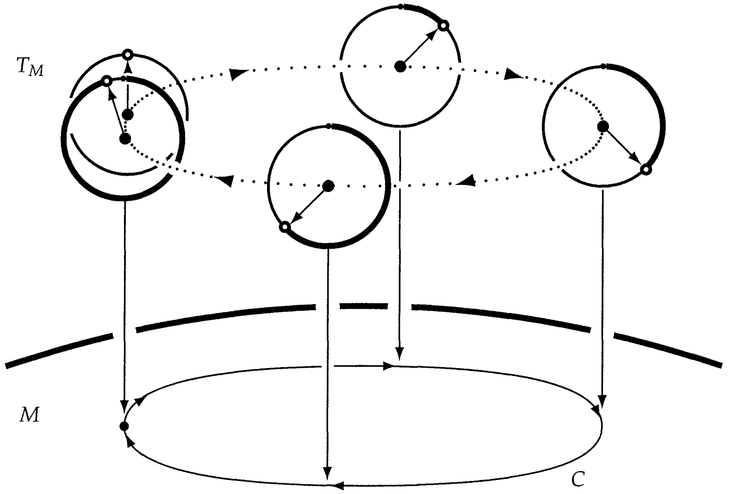
The problem is that, since $Spin(4) \to SO(4)$ is a double-cover, on triple-intersections $U_\alpha \cap U_\beta \cap U_\gamma$ such lifts *a priori* satisfy merely the conditions

$$\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha} = \pm\, id \,.$$

The appearance of an actual minus-sign makes $\{\tilde{g}_{\alpha\beta}\}$ fail from being a cocycle.

Hence, the manifold M is said to admit spin structures if and only if one can find a good $SO(4)$–cocycle $\{U_\alpha, g_{\alpha\beta}\}$ of T_M that can be lifted to $Spin(4)$–valued maps $\{U_\alpha, \tilde{g}_{\alpha\beta}\}$ for which *no* minus-signs appears in the equality above, and which thus make up a $Spin(4)$–cocycle.

No oddities. *Intuitively, a $Spin(4)$–valued cocycle $\{\tilde{g}_{\alpha\beta}\}$ for T_M exists if and only if odd multiples of 2π can be avoided when gluing up T_M. Explicitly, take a circle C bounding a disk in M and imagine that there are a few locally-trivialized patches $U_\alpha \times \mathbb{R}^4$ of T_M covering C that, when matched up, describe a rotation of 2π when travelling along C (see figure 4.18 on the next page). Then, since these patches describe the nontrivial class in $\pi_1 SO(4) = \mathbb{Z}_2$, they and their*



4.18. A non-extendable trivialization of T_M over the circle C

gluing maps $g_{\alpha\beta}$ cannot be used toward lifting to a $Spin(4)$-cocycle. This will be made more clear later.

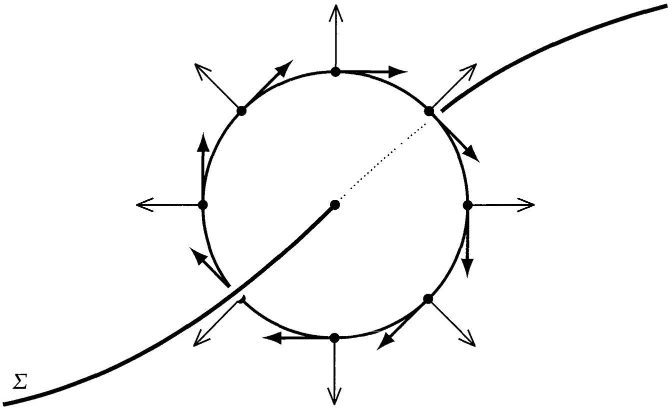
Homotopic simplifications. *Choosing an orientation on M reduces the structure group of T_M from the disconnected group $O(4)$ to the connected group $SO(4)$. Choosing a spin structure on M reduces the structure group of T_M to the simply-connected group $Spin(4)$. This process of homotopy-simplification of the structure group ends here. We already have $\pi_2 SO(4) = 0$ (and thus $\pi_2 Spin(4) = 0$). Further asking of a Lie group G to have $\pi_3 G = 0$ would force G to be contractible, and thus the bundle to be topologically trivial.*

In the remainder of this note, we will comment on what happens when M does not admit spin structures and explain the principal bundle point-of-view on spin structures. The latter will help us argue in the next note (page 181) that the two definitions of spin structures, the one with cocycles and the one with trivializations, are indeed equivalent. The third note (page 189) will develop bundle cocycles in their natural habitat, Čech cohomology. The fourth note (page 197) will present a smattering of obstruction theory and apply it to spin structures, while the fifth note (page 204) will present the homotopy-theoretic point-of-view on spin structures. Some consequences of the cocycle definition of spin structures (spinor bundles, Dirac operators) will be outlined in section 10.2 (page 382), as a quick prelude to the introduction of spin$^{\mathbb{C}}$ structures. The standard reference for spin structures is **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** [**LM89**].

When not spinnable. The existence of a spin structure is equivalent to the vanishing of $w_2(T_M)$. We wish to note what happens when no spin structures exist, that is, when $w_2(T_M) \neq 0$. In the cocycle point-of-view, this means that every $Spin(4)$-valued maps $\{\tilde{g}_{\alpha\beta}\}$, lifted from the $SO(4)$-cocycle of T_M, must have triples α, β, γ with $U_\alpha \cap U_\beta \cap U_\gamma$ non-empty and such that $\tilde{g}_{\alpha\beta}(x) \cdot \tilde{g}_{\beta\gamma}(x) \cdot \tilde{g}_{\gamma\alpha}(x) = -id$.

We pick an integral lift $\underline{w} \in H^2(M; \mathbb{Z})$ of $w_2(T_M)$ and represent $\underline{w}$ by an embedded oriented surface Σ in M. Since the characteristic surface Σ is the incarnation of the obstruction to the existence of a spin structure on M, there exist spin structures away from Σ, on the complement $M \setminus \Sigma$. None of these outside spin structures can be extended across Σ. (In terms of cocycles, we can arrange that the failing triples α, β, γ occur when and only when we go around Σ.)

In the trivializations point-of-view, such an outside spin structure on $M \setminus \Sigma$ offers a trivialization of T_M over the 1–skeleton, which restricts to a trivialization of T_M over small circles surrounding Σ (*e.g.*, fibers of the normal circle-bundle $\mathbb{S}N_{\Sigma/M}$ of Σ in M). Since the outside spin structure cannot extend across Σ, it follows that the trivialization of T_M over each such circle around Σ must describe a twist of 2π, as in figure 4.19. In the note ahead on Čech cohomology (page 196), this description will be made rigorous by using a concrete representation of $w_2(T_M)$.



4.19. Outside spin structure, not extending across a characteristic surface Σ

Principal bundle point-of-view. For any group G, a **principal G–bundle** is a locally-trivial fiber bundle with fiber G and structure group G. In other words, a principal G–bundle over X is a space $\mathcal{P}_G$ together with a projection map $p\colon \mathcal{P}_G \to X$ so that there is some covering $\{U_\alpha\}$ of X and maps $\varphi_\alpha\colon p^{-1}[U_\alpha] \cong U_\alpha \times G$, with $\mathrm{pr}_1 \circ \varphi_\alpha = p$ and so that the overlaps $\varphi_\beta \circ \varphi_\alpha^{-1}$ are described by formulae $(x, \gamma) \mapsto (x, \bar{g}_{\alpha\beta} \cdot \gamma)$ for suitable functions $\bar{g}_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$, acting on G by multiplication. Hence $\mathcal{P}_G \to X$ can be obtained by gluing trivial pieces $U_\alpha \times G \to U_\alpha$ using the G–cocycle $\{\bar{g}_{\alpha\beta}\}$, identifying $(x, \gamma_\alpha) \in U_\alpha \times G$ with $(x, \gamma_\beta) \in U_\beta \times G$ if and only if $\gamma_\alpha = \bar{g}_{\alpha\beta}(x) \cdot \gamma_\beta$.

Notice that, unlike a vector bundle, a principal G–bundle does not admit any global sections, unless it is trivial.[4]

Bundle of frames. For example, the $SO(4)$–valued cocycle $\{g_{\alpha\beta}\}$ of T_M acts directly on the group $SO(4)$ itself. Then, by gluing trivial pieces $U_\alpha \times SO(4)$, one obtains from $\{g_{\alpha\beta}\}$ a principal $SO(4)$–bundle

$$\mathcal{P}_{SO(4)} \to M\,.$$

---

[4] The fibers of $\mathcal{P}_G$ may look like G, and G itself acts on them, but they are merely "affine" copies of G, without, for example, a specified identity element. A global section in $\mathcal{P}_G$ can be viewed as offering a coherent choice of identity element, and thus yields an isomorphism $\mathcal{P}_G \approx X \times G$.

The bundle $\mathcal{P}_{SO(4)}$ depends only on T_M, not on the particular choice of $SO(4)$–cocycle $\{g_{\alpha\beta}\}$. Geometrically, one should think of $\mathcal{P}_{SO(4)} \to M$ as the bundle of orienting orthonormal frames of T_M.

A local section $\tau \colon U \to \mathcal{P}_{SO(4)}$ is a frame-field in T_M over U. It is thus equivalent to a trivialization of T_M over U. In particular, a trivialization of T_M over the 1–skeleton $M|_1$ of M is the same as a section $M|_1 \to \mathcal{P}_{SO(4)}$. The trivialization is extendable over the 2–skeleton $M|_2$ if and only if the corresponding section of $\mathcal{P}_{SO(4)}$ can be extended across $M|_2$.

Spin structures. Assume now that the $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ lifts to some $Spin(4)$–valued maps $\{\tilde{g}_{\alpha\beta}\}$ that satisfy the cocycle condition. Then we can use this lifted cocycle to glue a principal $Spin(4)$–bundle

$$\mathcal{P}_{Spin(4)} \to M$$

from trivial pieces $U_\alpha \times Spin(4)$. More, the double-cover $Spin(4) \to SO(4)$ defines fiber-to-fiber a natural map $\mathcal{P}_{Spin(4)} \to \mathcal{P}_{SO(4)}$, fitting in the diagram

$$\begin{tikzcd}
Spin(4) \arrow[r, phantom, "\subset"] \arrow[d, "2"'] & \mathcal{P}_{Spin(4)} \arrow[r] \arrow[d, "2"'] & M \arrow[d, equal] \\
SO(4) \arrow[r, phantom, "\subset"] & \mathcal{P}_{SO(4)} \arrow[r] & M\,.
\end{tikzcd}$$

The map $\mathcal{P}_{Spin(4)} \to \mathcal{P}_{SO(4)}$ is itself a double-cover of $\mathcal{P}_{SO(4)}$.

A spin structure can thus be redefined as a principal $Spin(4)$–bundle $\mathcal{P}_{Spin(4)}$ that double-covers the bundle $\mathcal{P}_{SO(4)}$ (and fits in the diagram above).

Note: Equivalence of the definitions of a spin structure

In what follows, we will prove hands-on the equivalence between defining spin structures as extendable trivializations of T_M and defining them as lifted $Spin(4)$–cocycles. Reading the preceding note is, obviously, a requisite.

Of course, more streamlined arguments exist. (Here is the best one: both the existence of an extendable trivialization and of a $Spin(4)$–cocycle are equivalent with the vanishing of $w_2(T_M)$; the end.) Nonetheless, in what follows we favor a concrete approach, which is rather expensive; we choose to present it here owing to its absence from the literature.

Our argument is rather long and involves some careful play with triangulations, principal bundles and double-covers, but the basic idea is pretty straightforward: Let $E \to \mathbb{D}^2$ be a vector bundle over a disk, with fiber $\mathbb{R}^4$. Since $\mathbb{D}^2$ is contractible, E must be trivial; for definiteness fix a reference trivialization $E \approx \mathbb{D}^2 \times \mathbb{R}^4$. Consider some other random trivialization $\varphi \colon E|_{\mathbb{S}^1} \approx \mathbb{S}^1 \times \mathbb{R}^4$ over the boundary of the base. Think of φ as a field of frames in E over $\partial\,\mathbb{D}^2$, that is to say, as a map $\varphi_f \colon \mathbb{S}^1 \to SO(4)$. The trivialization φ will extend across all $\mathbb{D}^2$ if and only if the frame-field φ_f can be extended over $\mathbb{D}^2$. That happens if and only if the loop φ_f in $SO(4)$ is homotopically-trivial, that is to say, if and only if

$\varphi_f\colon S^1 \to SO(4)$ can be lifted to a closed loop $\widetilde{\varphi}_f\colon S^1 \to Spin(4)$ (and not to an open path $\widetilde{\varphi}_f\colon [0,1] \to Spin(4)$, with $\widetilde{\varphi}_f(0) = -\widetilde{\varphi}_f(1)$).

Throughout this note, assume that M has been triangulated, in other words, exhibited as a simplicial complex.[5] Denote by $M|_1$ the 1–skeleton of M, by $M|_2$ the 2–skeleton, and so on. Further, for any bundle E over M, denote by $E|_k$ the restriction of E to the k–skeleton of M (and not the k–skeleton of the manifold E).

From cocycles to trivializations. Assume first that a $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ of T_M lifts to some maps $\{\widetilde{g}_{\alpha\beta}\}$ that actually satisfy the cocycle condition. Then a corresponding principal $Spin(4)$–bundle $\mathcal{P}_{Spin(4)}$ is well-defined. We will show that the existence of the bundle $\mathcal{P}_{Spin(4)}$ implies that T_M can be trivialized over the 2–skeleton $M|_2$. Specifically, we will show that the frame-bundle $\mathcal{P}_{SO(4)}$ admits a section over $M|_2$. For that, we define a section $\widetilde{\tau}$ of $\mathcal{P}_{Spin(4)}$ over $M|_2$ and project it to a section of $\mathcal{P}_{SO(4)}$. The section $\widetilde{\tau}$ is defined using a simplex-by-simplex construction.[6]

We start with the vertices of M and define each $\widetilde{\tau}(vertex)$ in some random manner as an element of $\mathcal{P}_{Spin(4)}$ in the fiber above it.

Any edge ε of M is contractible, and thus $\mathcal{P}_{Spin(4)}|_\varepsilon$ is trivial. Choose some trivialization $\mathcal{P}_{Spin(4)}|_\varepsilon \approx \varepsilon \times Spin(4)$. The section $\widetilde{\tau}$ is already defined at the endpoints (vertices) of ε. By looking through the trivialization, we see that the fact that $Spin(4)$ is connected implies that $\widetilde{\tau}$ can always be extended over ε, and thus eventually across the whole 1–skeleton $M|_1$.

There remain the 2–simplices. Any 2–simplex D is contractible and thus $\mathcal{P}_{Spin(4)}|_D$ can be trivialized as $D \times Spin(4)$. The section $\widetilde{\tau}$ is already defined over the edges that make up the boundary ∂D. Looking through the trivialization and using that $Spin(4)$ is simply-connected allows us to extend $\widetilde{\tau}$ over D, and eventually across the whole 2–skeleton $M|_2$.

The resulting section $\widetilde{\tau}\colon M|_2 \to \mathcal{P}_{Spin(4)}$ can be projected through the double-cover $\mathcal{P}_{Spin(4)} \to \mathcal{P}_{SO(4)}$ to a section $\tau\colon M|_2 \to \mathcal{P}_{SO(4)}$. The latter is a field of frames in T_M that trivializes T_M over $M|_2$.

Notice that, since we have $\pi_2 SO(4) = 0$ (and thus $\pi_2 Spin(4) = 0$), a bit more can be done: the section $\widetilde{\tau}$ of $\mathcal{P}_{Spin(4)}$ can be further extended across the 3–skeleton of M, yielding a trivialization of T_M over $M|_3$, which can be viewed as a trivialization over $M \setminus \{point\}$.

---

[5] A **triangulation** is a decomposition of M into simplices. A 0–simplex, or *vertex*, is a point. A 1–simplex, or *edge*, is a copy of $[0,1]$; its faces are its endpoint-vertices. A 2–simplex is a triangle (interior included); its faces are its three edges. A 3–simplex is a tetrahedron (interior included); its faces are the obvious four 2–simplices. A 4–simplex is whatever you want to call what follows; its faces are 3–simplices. If a simplex is part of a triangulation, then all its faces must be simplices of the triangulation. All simplices of a triangulation of M must be embedded in M and must either have exactly a whole sub-simplex (= face, or face-of-face, or...) in common with another simplex or be disjoint from it. In short, a triangulation of M means making M look like a polyhedron with simple "triangular" faces.

[6] This simplex-by-simplex method is just a most simple application of the method of *obstruction theory*, which will be explained in generality in the note on page 197 ahead. If you do not like the word "simplex", you can substitute "handle" or "cell" throughout.

Uniqueness. It is worth noting that the trivialization τ of $T_M|_1$ that we obtained above is uniquely determined, up to homotopies, by the spin structure $\mathcal{P}_{Spin(4)}$. Indeed, assume two random sections $\tilde{\tau}'$ and $\tilde{\tau}''$ of $\mathcal{P}_{Spin(4)}$ are given over $M|_1$. We will define a homotopy between them over the 1-skeleton of M. For that, we view a homotopy as a section of the bundle $\mathcal{P}_{Spin(4)} \times [0,1] \longrightarrow M \times [0,1]$ that limits to $\tilde{\tau}'$ over $M \times 0$ and to $\tilde{\tau}''$ over $M \times 1$. Since $\tilde{\tau}'$ and $\tilde{\tau}''$ are given, such a section is already defined over the vertices of $M \times [0,1]$. It can be extended across the edges connecting $M \times 0$ with $M \times 1$, using as above that $Spin(4)$ is connected. Then it can be extended over the 2-simplices of $M \times [0,1]$ by using that $\pi_1 Spin(4) = 0$. Thus, we have defined a homotopy between $\tilde{\tau}'$ and $\tilde{\tau}''$ over $M|_1$. This descends to a homotopy between the induced trivializations τ' and τ'' of T_M, proving uniqueness.

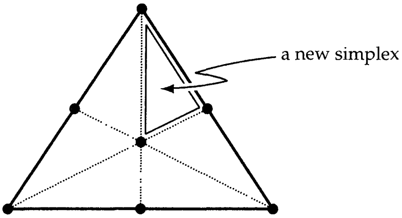
In conclusion, a spin structure defined *via* cocycles determines an extendable trivialization of $T_M|_1$, unique up to homotopies.

From trivializations to cocycles: Preparation. The converse argument involves a rather cumbersome setup that will allow us to link 1-skeletons and trivializations to cocycles and their lifts. It will take the rest of this note (through page 189).

Assume that M has been endowed with a fixed triangulation $\mathscr{T}$. For definiteness, fix a Riemannian metric on M. We will prove that any trivialization of $T_M|_1$ that extends across $M|_2$ defines a $Spin(4)$-cocycle for T_M.

First, remember that any triangulation $\mathscr{T}$ admits a *dual cellular decomposition* $\mathscr{T}^*$.

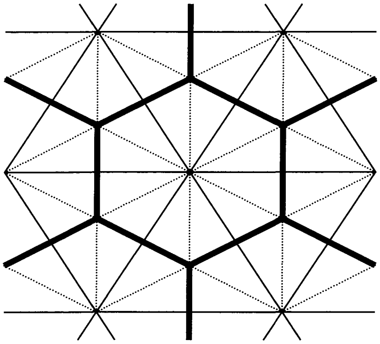
Given a triangulation $\mathscr{T}$ of M^4, its **dual cellular decomposition** *$\mathscr{T}^*$ is obtained by taking the barycentric subdivision[7] $\mathscr{T}'$ of $\mathscr{T}$, then, for each $(4-k)$-simplex Δ_α of $\mathscr{T}$, defining its dual k-cell Δ_α^* in $\mathscr{T}^*$ by taking the union of all k-simplices of $\mathscr{T}'$ that touch the barycenter of Δ_α. For example, the vertices of $\mathscr{T}^*$ are the barycenters of the 4-simplices of $\mathscr{T}$, the 1-cells of $\mathscr{T}^*$ are arcs normal to the 3-simplices of $\mathscr{T}$ (and link the vertices of $\mathscr{T}^*$), while the 4-cells of $\mathscr{T}^*$ are neighborhoods of the vertices of $\mathscr{T}$. See figure 4.21 on the following page. The dual cellular decomposition is an especially nice cellular decomposition, in that it fails from being a triangulation only by using more general "polygonal" cells rather than just "triangular" simplices; otherwise, all cells are embedded, etc. (On the side, note that dual cellular decompositions can be used to offer a nice visualization of Poincaré duality.)*



4.20. Barycentric subdivision of a 2-simplex

---

[7] The **barycenter** of a simplex Δ is simply a canonical center for it. The barycenter of a vertex is the vertex itself. The **barycentric subdivision** $\mathscr{T}'$ of $\mathscr{T}$ is obtained by taking as new k-simplices every *join* of the barycenter of an old k-simplex of $\mathscr{T}$ with the barycenter of a face and the barycenter of a face of that face and... For example, a 2-simplex in $\mathscr{T}'$ is the triangle that appears by joining the barycenter of a triangle of $\mathscr{T}$ with the center of one of its edges and with the vertex at one end of that edge. See figure 4.20. The **join** of two subsets A and B of $\mathbb{R}^n$ is the union of all segments that start in A and end in B.



4.21. Cellular decomposition dual to a triangulation

Since we have to deal with trivializations of T_M over the 1–skeleton $M|_1$ and their extendability over the 2–skeleton $M|_2$, we will only use cocycles $\{U_\alpha,\ g_{\alpha\beta}\}$ of T_M that are nicely compatible with the chosen triangulation $\mathscr{T}$ of M.

Namely, we will take the U_α's to be small neighborhoods of the 4–cells Δ^*_α of the dual decomposition $\mathscr{T}^*$ of M. The 4–cell Δ^*_α is a closed set surrounding a vertex v_α and touching the barycenters of all 4–simplices that contain v_α. In particular, each edge ε of $\mathscr{T}$ links the center of U_α with the center of U_β and passes through the overlap $U_\alpha \cap U_\beta$. The latter intersection is just a small neighborhood of the 3–cell (dual to ε) that Δ^*_α and Δ^*_β have in common.

Since each U_α is contractible, $T_M|_{U_\alpha}$ is trivial. Using the Riemannian metric of M, we choose trivializations

$$\varphi_\alpha\colon\ T_M|_{U_\alpha} \approx U_\alpha \times \mathbb{R}^4$$

that are isometries on the fibers. We compare these trivializations over $U_\alpha \cap U_\beta$ and obtain transition maps

$$g_{\alpha\beta}\colon\ U_\alpha \cap U_\beta \longrightarrow SO(4) \qquad \text{with} \qquad \varphi_\alpha = g_{\alpha\beta} \cdot \varphi_\beta\ .$$

These will be the cocycles $\{U_\alpha,\ g_{\alpha\beta}\}$ of T_M that we will consider. Notice that these cocycles depend essentially only on the choice of trivializations φ_α over the U_α's.

Trivializations and partial Spin–bundles. Given any trivialization

$$\Theta\colon\ T_M|_1 \approx M|_1 \times \mathbb{R}^4$$

of T_M over the 1–skeleton of M, we express Θ in coordinates with respect to the charts $\varphi_\alpha\colon\ T_M|_{U_\alpha} \approx U_\alpha \times \mathbb{R}^4$. Namely, we describe Θ by a collection of $SO(4)$–valued maps τ_α, defined on the part of the 1–skeleton of M that is included in U_α, which we denote by $U_\alpha|_1$ (see figure 4.22 on the next page).

Specifically, the maps

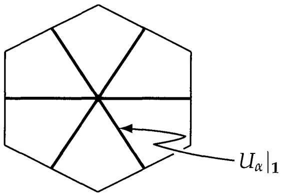
$$\tau_\alpha\colon\ U_\alpha|_1 \longrightarrow SO(4)$$

are defined by the equations $\tau_\alpha(x) \cdot w = \varphi_\alpha\big(\Theta^{-1}(x, w)\big)$ and will satisfy compatibility relations

$$\tau_\alpha = g_{\alpha\beta} \cdot \tau_\beta\ .$$

An alternative view of the τ_α's is as defining a section

$$\tau\colon\ M|_1 \longrightarrow \mathcal{P}_{SO(4)}|_1\ ,$$



4.22. Open set U_α, and the 1–skeleton of M

corresponding to the frame-field induced by the trivialization Θ.

Consider a random lift of the maps $\tau_\alpha \colon U_\alpha|_1 \to SO(4)$ to some maps

$$\tilde{\tau}_\alpha \colon U_\alpha|_1 \longrightarrow Spin(4) \ .$$

Given such a collection $\{\tilde{\tau}_\alpha\}$, we can correspondingly choose lifts

$$\tilde{g}_{\alpha\beta} \colon U_\alpha \cap U_\beta \longrightarrow Spin(4)$$

of the $g_{\alpha\beta}$'s in such manner as to fit the various $\tilde{\tau}_\alpha$'s, namely so that

$$\tilde{\tau}_\alpha = \tilde{g}_{\alpha\beta} \cdot \tilde{\tau}_\beta \ .$$

Since this fitting amounts merely to a choice of sign for each $\tilde{g}_{\alpha\beta}$ and owing to the special shape of our covering $\{U_\alpha\}$, such a lift can always be made.

Of course, $\{\tilde{g}_{\alpha\beta}\}$ is most likely not a cocycle. Whether it is or not depends only on the τ_α's, not on the random lifts $\tilde{\tau}_\alpha$. To see this, consider two random lifts $\{\tilde{\tau}'_\alpha\}$ and $\{\tilde{\tau}''_\alpha\}$. They can differ at most by a collection of signs $\varepsilon_\alpha \in \mathbb{Z}_2 = \{-1, +1\}$ with $\tilde{\tau}''_\alpha = \varepsilon_\alpha \tilde{\tau}'_\alpha$. The corresponding transition maps are then related by $\tilde{g}''_{\alpha\beta} = \varepsilon_\alpha \varepsilon_\beta \tilde{g}'_{\alpha\beta}$. Clearly, we have $\tilde{g}''_{\alpha\beta} \cdot \tilde{g}''_{\beta\gamma} \cdot \tilde{g}''_{\gamma\alpha} = +1$ if and only if $\tilde{g}'_{\alpha\beta} \cdot \tilde{g}'_{\beta\gamma} \cdot \tilde{g}'_{\gamma\alpha} = +1$. In particular, when one choice of $\tilde{\tau}_\alpha$'s leads to a cocycle, then so will any other choice, and the various choices lead to isomorphic cocycles, *i.e.*, a unique spin structure.

By definition, the maps $\tilde{g}_{\alpha\beta}$ satisfy $\tilde{g}_{\alpha\beta} = \tilde{g}_{\beta\alpha}^{-1}$. Therefore, if we avoid all triple intersections $U_\alpha \cap U_\beta \cap U_\gamma$, then the lifts $\tilde{g}_{\alpha\beta}$ can be used to define a principal $Spin(4)$–bundle away from the $U_\alpha \cap U_\beta \cap U_\gamma$'s. In particular, we get a bundle

$$\mathcal{P}_{Spin(4)}|_1$$

well-defined over the 1–skeleton of M.

Of course, $\mathcal{P}_{Spin(4)}|_1$ is a double-cover of $\mathcal{P}_{SO(4)}|_1$, built fiberwise from the projection $Spin(4) \to SO(4)$. Furthermore, the maps $\tilde{\tau}_\alpha$ can be viewed as defining a section $\tilde{\tau} \colon M|_1 \to \mathcal{P}_{Spin(4)}|_1$.

Trivial versus nontrivial covers. Since the bundle $\mathcal{P}_{Spin(4)}|_1$ defined above is a principal bundle, having a section $\tilde{\tau}$ implies that it is a trivial bundle over $M|_1$. Nonetheless, it can project in a nontrivial way onto $\mathcal{P}_{SO(4)}|_1$. In what follows we will investigate how this nontriviality can be detected. Since $\mathcal{P}_{Spin(4)}|_1 \to \mathcal{P}_{SO(4)}|_1$ is a cover projection, fundamental groups will play a prominent role in the argument.

Restrict to the boundary ∂D of a fixed 2–simplex D. Since both $\mathcal{P}_{SO(4)}|_{\partial D}$ and $\mathcal{P}_{Spin(4)}|_{\partial D}$ admit sections τ and $\tilde{\tau}$, they are trivial, and thus $\mathcal{P}_{SO(4)}|_{\partial D} \approx \partial D \times SO(4)$ and $\mathcal{P}_{Spin(4)}|_{\partial D} \approx \partial D \times Spin(4)$. Therefore

$$\pi_1\big(\mathcal{P}_{SO(4)}|_{\partial D}\big) = \mathbb{Z} \oplus \mathbb{Z}_2 \qquad \text{and} \qquad \pi_1\big(\mathcal{P}_{Spin(4)}|_{\partial D}\big) = \mathbb{Z}\,.$$

Denote by d the double-cover map

$$d\colon \mathcal{P}_{Spin(4)}|_{\partial D} \longrightarrow \mathcal{P}_{SO(4)}|_{\partial D}\,,$$

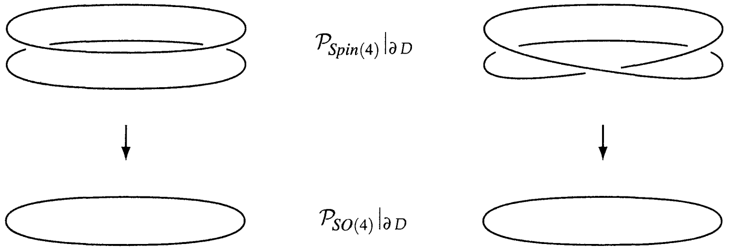
fitting in the diagram

$$\begin{tikzcd}
\mathcal{P}_{Spin(4)}|_{\partial D} \arrow[r, "d"'] \arrow[d] & \mathcal{P}_{SO(4)}|_{\partial D} \arrow[d] \\
\partial D \arrow[r, equal] & \partial D
\end{tikzcd}
\qquad \text{or, on } \pi_1\text{'s:} \qquad
\begin{tikzcd}
\mathbb{Z} \arrow[r, "d_*"'] \arrow[d] & \mathbb{Z} \oplus \mathbb{Z}_2 \arrow[d, "\mathrm{pr}_1"] \\
\mathbb{Z} \arrow[r, equal] & \mathbb{Z}
\end{tikzcd}\,.$$

Being a cover map, d's induced morphism d_* must be injective. We deduce that there are only two choices: either

$$d_*(1) = 1 \oplus 0 \qquad \text{or} \qquad d_*(1) = 1 \oplus 1 \quad \in \mathbb{Z} \oplus \mathbb{Z}_2\,.$$

The case $d_*(1) = 1 \oplus 0$ corresponds to the case when the cover $\mathcal{P}_{Spin(4)}|_{\partial D} \to \mathcal{P}_{SO(4)}|_{\partial D}$ is trivial, while $d_*(1) = 1 \oplus 1$ happens when the fiber of $\mathcal{P}_{Spin(4)}|_{\partial D}$ twists once as we go around ∂D, as suggested in figure[8] 4.23.



4.23. Trivial and nontrivial covers

To better visualize how this can happen, consider the trivial bundles $\mathbb{S}^1 \times \mathbb{S}^3$ *and* $\mathbb{S}^1 \times \mathbb{RP}^3$ *over* $\mathbb{S}^1$. *There are two possible double-cover projections* d *of* $\mathbb{S}^1 \times \mathbb{S}^3$ *onto* $\mathbb{S}^1 \times \mathbb{RP}^3$ *that both commute with the bundle projections and hence fit in a diagram*

$$\begin{tikzcd}
\mathbb{S}^1 \times \mathbb{S}^3 \arrow[r, "d"'] \arrow[d] & \mathbb{S}^1 \times \mathbb{RP}^3 \arrow[d] \\
\mathbb{S}^1 \arrow[r, equal] & \mathbb{S}^1
\end{tikzcd}\,.$$

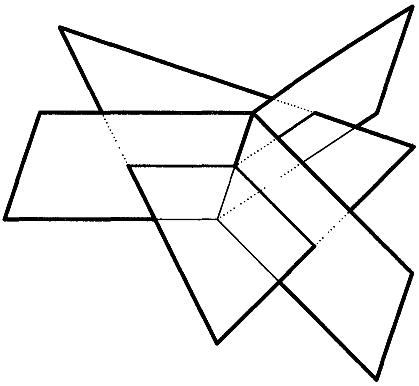
One possible double-cover is the obvious one, the product of the identity on $\mathbb{S}^1$ *with the double-cover* $\mathbb{S}^3 \to \mathbb{RP}^3$. *The other can be seen as follows: start with* $[0,1] \times \mathbb{S}^3$ *and glue the ends* $0 \times \mathbb{S}^3$ *and* $1 \times \mathbb{S}^3$ *using the antipodal map on* $\mathbb{S}^3$; *project each* $\mathbb{S}^3$ *to* $\mathbb{RP}^3$ *to get a double-cover of* $\mathbb{S}^1 \times \mathbb{RP}^3$. *However, since the antipodal of* $\mathbb{S}^3$ *is homotopic to the identity, what we glued is still* $\mathbb{S}^1 \times \mathbb{S}^3$. *The first map has* $d_*(1) = 1 \oplus 0$, *while the second has* $d_*(1) = 1 \oplus 1$. *In fact, this example is pretty close to our concerns, since* $\mathbb{S}^3 = Spin(3)$ *and* $\mathbb{RP}^3 = SO(3)$.

---

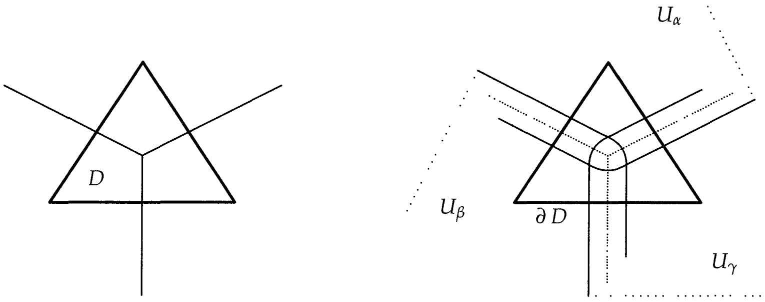
[8] Owing to dimension-reduction, figure 4.23 is misleading: on both sides, the space $\mathcal{P}_{Spin(4)}|_{\partial D}$ should be the same trivial bundle over ∂D.

Detecting nontriviality with cocycle candidates. The two cases $d_*(1) = 1 \oplus 0$ and $d_*(1) = 1 \oplus 1$ are detected both by the lifted $Spin(4)$–valued maps $\widetilde{g}_{\alpha\beta}$ and by the section $\widetilde{\tau}$ of $\mathcal{P}_{Spin(4)}|_1$. We start with the $\widetilde{g}_{\alpha\beta}$'s.

Since D is a 2–simplex of M, it is surrounded by three of the open sets from our covering, say U_α, U_β and U_γ, with the center of D right in the middle of $U_\alpha \cap U_\beta \cap U_\gamma$, as suggested in figures 4.24 and 4.25.



4.24. Set-up for equivalence argument, I



4.25. Set-up for equivalence argument, II

We claim that, for the indices α, β, γ around D, we have

$$\widetilde{g}_{\alpha\beta} \cdot \widetilde{g}_{\beta\gamma} \cdot \widetilde{g}_{\gamma\alpha} = +1$$

if and only if the cover $\mathcal{P}_{Spin(4)}|_{\partial D} \to \mathcal{P}_{SO(4)}|_{\partial D}$ is trivial, that is to say, if and only if $d_*(1) = 1 \oplus 0$.

Assume first that the product of the $\widetilde{g}_{\alpha\beta}$'s around D is $+1$. Then the $\widetilde{g}_{\alpha\beta}$'s can safely be used to extend $\mathcal{P}_{Spin(4)}|_{\partial D}$ over D as a bundle $\mathcal{P}_{Spin(4)}|_D$, fitting in

$$\begin{tikzcd}
\mathcal{P}_{Spin(4)}|_{\partial D} \arrow[r, phantom, "\subset"] \arrow[d, "d"'] & \mathcal{P}_{Spin(4)}|_D \arrow[d] \\
\mathcal{P}_{SO(4)}|_{\partial D} \arrow[r, phantom, "\subset"] & \mathcal{P}_{SO(4)}|_D
\end{tikzcd}$$

or, on π_1's:

$$\begin{tikzcd}
\mathbb{Z} \arrow[r] \arrow[d, "d_*"'] & 0 \arrow[d] \\
\mathbb{Z} \oplus \mathbb{Z}_2 \arrow[r, "\mathrm{pr}_2"] & \mathbb{Z}_2
\end{tikzcd} .$$

Thus the only possibility for d_* is

$$d_*(1) = 1 \oplus 0 .$$

Conversely, assume that $d_*(1) = 1 \oplus 0$. Then $\mathcal{P}_{Spin(4)}|_{\partial D} \to \mathcal{P}_{SO(4)}|_{\partial D}$ must be the trivial double-cover, with

$$\mathcal{P}_{Spin(4)}|_{\partial D} \approx \mathcal{P}_{SO(4)}|_{\partial D} \times \{-1, +1\} .$$

Therefore it can be extended to a double-cover $\mathcal{P}$ of $\mathcal{P}_{SO(4)}$ across the whole D, with $\mathcal{P}|_{\partial D} = \mathcal{P}_{Spin(4)}|_{\partial D}$. Such a double-cover, when projected down to D, can only have as fibers copies of $Spin(4)$. Moreover, since $\mathcal{P}$ projects to $\mathcal{P}_{SO(4)}|_D$, its cocycle must project to the cocycle $g_{\alpha\beta}$ of $\mathcal{P}_{SO(4)}$. Further, since $\mathcal{P} \to D$ is glued over ∂D by the $\widetilde{g}_{\alpha\beta}$'s, it must be that it is glued over the whole D by the $\widetilde{g}_{\alpha\beta}$'s. This in particular implies that the $\widetilde{g}_{\alpha\beta}$'s, since they glue an actual bundle over D, must be a genuine cocycle over D, and thus

$$\widetilde{g}_{\alpha\beta} \cdot \widetilde{g}_{\beta\gamma} \cdot \widetilde{g}_{\gamma\alpha} = +1 .$$

In conclusion, $\widetilde{g}_{\alpha\beta} \cdot \widetilde{g}_{\beta\gamma} \cdot \widetilde{g}_{\gamma\alpha} = +1$ if and only if $d_*(1) = 1 \oplus 0$.

Detecting nontriviality with trivializations. Now we will see how to distinguish between the two cases $d_*(1) = 1 \oplus 0$ and $d_*(1) = 1 \oplus 1$ by using the trivialization $\Theta \colon T_M|_1 \approx M|_1 \times \mathbb{R}^4$.

The trivialization Θ expresses itself through the section τ of $\mathcal{P}_{SO(4)}|_1$, with local coordinates $\tau_\alpha \colon U_\alpha \to SO(4)$. Recall that we chose random lifts $\widetilde{\tau}_\alpha \colon U_\alpha \to Spin(4)$ and then picked the maps $\widetilde{g}_{\alpha\beta}$ in such manner as to ensure that the $\widetilde{\tau}_\alpha$'s would define a section in the partial $Spin(4)$–bundle $\mathcal{P}_{Spin(4)}|_1$ that is glued by the $\widetilde{g}_{\alpha\beta}$'s.

Over the boundary ∂D, we have the diagram

$$\begin{tikzcd}
\mathcal{P}_{Spin(4)}|_{\partial D} \arrow[r, "d"'] & \mathcal{P}_{SO(4)}|_{\partial D} \\
\partial D \arrow[u, "\widetilde{\tau}"] \arrow[r, equal] & \partial D \arrow[u, "\tau"']
\end{tikzcd}
\quad \text{or, on } \pi_1\text{'s:} \quad
\begin{tikzcd}
\mathbb{Z} \arrow[r, "d_*"'] & \mathbb{Z} \oplus \mathbb{Z}_2 \\
\mathbb{Z} \arrow[u, "\widetilde{\tau}_*"] \arrow[r, equal] & \mathbb{Z} \arrow[u, "\tau_*"']
\end{tikzcd} .$$

Since from commuting we must have that $\tau_*(1) = d_*(1)$, it follows that either $\tau_*(1) = 1 \oplus 0$ or $\tau_*(1) = 1 \oplus 1$.

Trivialize $\mathcal{P}_{SO(4)}$ over D as $D \times SO(4)$ and use the inclusion

$$\mathcal{P}_{SO(4)}|_{\partial D} \subset \mathcal{P}_{SO(4)}|_D \approx D \times SO(4)$$

to obtain from $\tau \colon \partial D \to \mathcal{P}_{SO(4)}|_{\partial D}$ a map $\tau_0 \colon \partial D \to SO(4)$. Then the section τ of $\mathcal{P}_{SO(4)}|_{\partial D}$ can be extended to a section of $\mathcal{P}_{SO(4)}$ over all D if and only if the induced map $\tau_0 \colon \partial D \to SO(4)$ is homotopically-trivial. In other words, if and only if we have $\tau_*(1) = 1 \oplus 0$ and not $1 \oplus 1$.

In conclusion, the trivialization Θ of T_M over the 1–skeleton can be extended over the 2–simplex D if and only if $d_*(1) = 1 \oplus 0$.

Final twirl. Gathering our toys, we notice that we have proved the statement:

Given a trivialization Θ of $T_M|_1$, it can be extended over a 2–simplex D surrounded by the open sets U_α, U_β, U_γ if and only if $\widetilde{g}_{\alpha\beta} \cdot \widetilde{g}_{\beta\gamma} \cdot \widetilde{g}_{\gamma\alpha} = +1$.

In particular, if Θ is a trivialization of T_M over the 1–skeleton that extends across the whole 2–skeleton, then it can be used to define lifted maps $\{\widetilde{g}_{\alpha\beta}\}$ that will constitute a $Spin(4)$–cocycle.

The proof is concluded: an extendable trivialization defines a unique $Spin(4)$–cocycle, up to isomorphisms.

Note: Bundles, cocycles, and Čech cohomology

In this note we describe the Čech cohomology of a manifold, with constant coefficients in an Abelian group G. Then we extend this concept, on one hand, to non-Abelian groups and, on the other hand, to non-constant coefficients. (We will not take the next step of defining the general cohomology of a sheaf.)

This will enable us to present a cocycle defining a bundle as a Čech cocycle that defines a cohomology class in $\check{H}^1(M; C^\infty GL(k))$. Consequently, $\check{H}^1(M; C^\infty GL(k))$ can be viewed as the set of all k–plane bundles over M, up to isomorphisms. This approach will allow us to get concrete descriptions of a few characteristic classes and will be used to touch upon the obstruction and uniqueness of spin structures on M.

Čech cohomology. One should think of Čech cohomology as a cohomology theory that uses open coverings and the way their open sets assemble (intersect) patching-up the manifold M, in order to detect the topology of M.

Let $\{U_\alpha\}$ be a covering of M by open sets, and G an Abelian group. We consider collections of G–valued functions defined on intersections of the U_α's. Pick an integer n and choose a set of *locally-constant* functions

$$\varphi = \{\varphi_{\alpha_0 \ldots \alpha_n} : U_{\alpha_0} \cap \cdots \cap U_{\alpha_n} \longrightarrow G\} ,$$

each defined on the intersection of $n+1$ of the open sets U_α. This collection is called a **Čech n–cochain** with values in G. We denote by

$$\check{C}^n(\{U_\alpha\}; G)$$

the Abelian group of all such Čech n–cochains.

The **coboundary** operator $\delta \colon \check{C}^n \to \check{C}^{n+1}$ sends each φ to an $(n+1)$–cochain $\delta\varphi$, a set of functions defined on intersections of $n+2$ of the U_α's, each described as an alternating sum of restrictions of φ's. Namely, we set

$$(\delta\varphi)_{\alpha_0 \ldots \alpha_{n+1}} : U_{\alpha_0} \cap \cdots \cap U_{\alpha_{n+1}} \longrightarrow G$$

$$(\delta\varphi)_{\alpha_0 \ldots \alpha_{n+1}}(x) = \sum (-1)^k \varphi_{\alpha_0 \ldots \widehat{\alpha}_k \ldots \alpha_{n+1}}(x)$$

(where $\widehat{\alpha}_k$ signals the removal of α_k).

If an n–cochain φ has $\delta\varphi = 0$, then φ is called a **Čech cocycle.** If $\varphi = \delta\alpha$ for some $(n-1)$–chain α, then φ is called a **Čech coboundary.** The **Čech cohomology group**

$\check{H}^*(\{U_\alpha\};G)$ of the covering $\{U_\alpha\}$ of M is then defined in the usual fashion, as cocycles modulo coboundaries:

$$\check{H}^n(\{U_\alpha\};G) \;=\; \{\varphi \in \check{C}^n(\{U_\alpha\};G) \mid \delta\varphi = 0\} \Big/ \{\delta\alpha \mid \alpha \in \check{C}^{n-1}(\{U_\alpha\};G)\} \,.$$

A priori these groups depend on the chosen open covering $\{U_\alpha\}$. Eliminating this dependence, the **Čech cohomology group** of M is defined as the direct limit

$$\check{H}^*(M;G) = \varinjlim \check{H}^*(\{U_\alpha\},G) \,,$$

taken over refinements of the open cover.

If M is a manifold, then $\check{H}^*(M;G)$ coincides with the usual singular cohomology of M, as we will prove directly in an instant.

Taking the direct limit is rather unpleasant, and it is almost never done. Indeed, it is enough to consider a fine enough covering, for example a covering $\{U_\alpha\}$ of M by contractible open sets, with all intersections $U_{\alpha_0} \cap \cdots \cap U_{\alpha_n}$ contractible as well.[9] For such coverings we have $\check{H}^*(M;G) = \check{H}^*(\{U_\alpha\};G)$.

Simple examples. The group $\check{H}^0(\{U_\alpha\};G)$ comes from 0–cocycles, that is to say, from collections $\varphi = \{\varphi_\alpha : U_\alpha \to G\}$ of locally-constant functions defined on the U_α's and satisfying $\delta\varphi = 0$. In this case, the cocycle condition is

$$\delta\varphi = 0 \qquad \Longleftrightarrow \qquad \varphi_\alpha = \varphi_\beta \quad \text{on } U_\alpha \cap U_\beta \,,$$

and therefore immediately

$$\check{H}^0(\{U_\alpha\};G) \;=\; \{\text{locally-constant functions } M \to G\} \,.$$

Hence $\check{H}^0$ detects the components of M: if M is connected, then $\check{H}^0(M;G) = G$.

The first group $\check{H}^1(\{U_\alpha\};G)$ comes from 1–cocycles, that is to say, from families $\varphi = \{\varphi_{\alpha\beta} : U_\alpha \cap U_\beta \to G\}$ satisfying $\delta\varphi = 0$, where

$$\delta\varphi = 0 \qquad \Longleftrightarrow \qquad \varphi_{\alpha\gamma} = \varphi_{\alpha\beta} + \varphi_{\beta\gamma} \quad \text{on } U_\alpha \cap U_\beta \cap U_\gamma \,.$$

In particular, notice that a 1–cocycle must satisfy the skew-symmetry $\varphi_{\alpha\beta} = -\varphi_{\beta\alpha}$. These 1–cocycles yield cohomology classes in $\check{H}^0(\{U_\alpha\};G)$ by considering them up to the addition of a coboundary. That is, for any two cocycles φ' and φ'', we have:

$$[\varphi'] = [\varphi''] \quad \text{in } \check{H}^1 \qquad \Longleftrightarrow \qquad \varphi'_{\alpha\beta} = \varphi''_{\alpha\beta} + f_\alpha - f_\beta$$

for some 0–cochain $f = \{f_\alpha \colon U_\alpha \to G\}$.

And the usual suspects. We now prove directly that nothing new is obtained:

Lemma. *If X is a simplicial complex (e.g., a triangulated manifold), then*

$$\check{H}^*(X;G) \;=\; H^*(X;G) \,,$$

where on the right we have the simplicial cohomology of X.

---

[9] A typical geometric method for building such coverings is to pick a Riemannian metric on M and choose *geodesically convex* open sets for the U_α's. A more topological method would use a triangulation of M and take the U_α's to be the *stars* of the vertices of M.

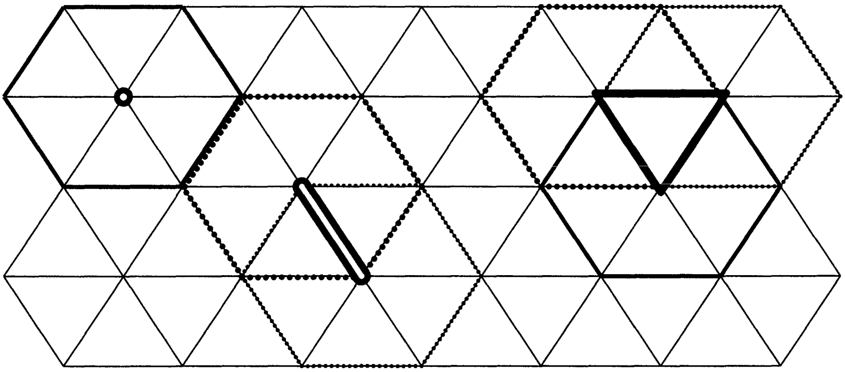
Proof. For every vertex v of X we define its **star**, denoted by $\operatorname{star}(v)$, as the union of all simplices of X that contain v. List the vertices of X as $\{v_\alpha\}$ and define the open sets U_α as

$$U_\alpha = \text{interior of } \operatorname{star}(v_\alpha) \ .$$

Then we have that

$$U_{\alpha_0} \cap \cdots \cap U_{\alpha_n} \neq \varnothing \quad \text{if and only if} \quad v_{\alpha_0}, \ldots, v_{\alpha_n} \text{ span an } n\text{-simplex.}$$

See also figure 4.26.



4.26. Linking Čech cochains with simplicial cochains

Each of these intersections is connected, and therefore every Čech n–cochain φ is constant on it. Thus, a Čech n–cochain φ simply assigns to every n–simplex $\langle v_{\alpha_0}, \ldots, v_{\alpha_n} \rangle$ of X an element $\varphi_{\alpha_0 \ldots \alpha_n}$ of G, and hence corresponds bijectively to a simplicial n–cochain.

Finally, it is not hard to check that the Čech and simplicial coboundary operators correspond perfectly, and thus

$$\check{H}^*(\{U_\alpha\}; G) = H^*(X; G) \ .$$

Going to the limit with the coverings is not a problem, *e.g.*, by using subdivisions of the simplicial complex. □

Even though nothing new appears at the outset, Čech theory admits a remarkable extension from coefficients in a group to coefficients in a presheaf and leads to the sheaf cohomology that is essential in complex geometry. We will not fully pursue that avenue, but the reader is encouraged to consult **P. Griffiths** and **J. Harris**'s ***Principles of algebraic geometry*** [**GH78, GH94**].

Another remarkable extension of the theory is to non-commutative groups:

Non-commutative Čech cohomology. One should notice that the whole cohomology apparatus depends on G being Abelian, and thus the extension to the non-Abelian case will have serious restrictions. Namely, $H^1(M; G)$ ceases to be a group and $H^2(M; G)$ ceases to be altogether. However, since vector bundles are glued using non-commutative groups such as $GL(k)$, $SO(k)$, $U(k)$, we do need to pursue this direction. Thus, let G be a *non-Abelian* group, written *multiplicatively*. We can define Čech cochains just as before. However, when it comes to defining the coboundary operator, we need to be careful.

We are only interested in $H^1(M;G)$, so let $\varphi = \{\varphi_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G\}$ be a G-valued 1-cochain. Switching from additive to multiplicative writing, we write

$$(\delta\varphi)_{\alpha\beta\gamma} = \varphi_{\alpha\beta} \cdot \varphi_{\beta\gamma} \cdot \varphi_{\gamma\alpha} \,.$$

A 1-cocycle must then be any φ with $(\delta\varphi)_{\alpha\beta\gamma} = 1$ for every α, β, γ. In particular, every cocycle has $\varphi_{\alpha\beta} = \varphi_{\beta\alpha}^{-1}$.

Now let $f = \{f_\alpha\colon U_\alpha \to G\}$ be a 0-cochain. Its coboundary is, naturally,

$$(\delta f)_{\alpha\beta} = f_\alpha \cdot f_\beta^{-1} \,.$$

Nonetheless, when it comes to defining when two 1-cochains φ' and φ'' are cohomologous, that is, when φ' and φ'' are considered to differ by δf, the noncommutativity of G makes essential a specific choice of order. The right one is:

$$[\varphi'] = [\varphi''] \qquad \Longleftrightarrow \qquad \varphi'_{\alpha\beta} = f_\alpha \cdot \varphi''_{\alpha\beta} \cdot f_\beta^{-1} \,.$$

Then we can define in the usual manner the Čech cohomology *set* $\check{H}^1(\{U_\alpha\};G)$ of the covering $\{U_\alpha\}$, and thereafter its limit $\check{H}^1(M;G) = \varinjlim \check{H}^1(\{U_\alpha\};G)$. Since the coboundaries cannot be expected to make up a normal subgroup of the cocycles, this $\check{H}^1$ is not a group, but merely a set with a distinguished element, the class of the trivial cocycle given by $\mathbf{1}_{\alpha\beta} = 1$.

The similarities with the cocycles that glue bundles should be obvious by now. Nonetheless, to fully engulf that case we need to extend the notion of cochain a bit to allow for non-locally-constant functions.

Non-constant cochains. We extend the notion of cochain. Namely, given a *topological* group G and a covering $\{U_\alpha\}$ of M, we define a *continuous* n-cochain $\varphi = \{\varphi_{\alpha_0\ldots\alpha_n}\}$ as a collection of *continuous* functions

$$\varphi_{\alpha_0\ldots\alpha_n}\colon U_{\alpha_0} \cap \cdots \cap U_{\alpha_n} \longrightarrow G \,.$$

The rest of the theory flows just as before and leads to what one should properly call the *Čech cohomology with coefficients in the sheaf of continuous G-valued functions,* and denote it by something like

$$\check{H}^*(M;\, C^0(G)) \,.$$

Notice that, if G is a *discrete* group (such as $\mathbb{Z}$), then the cochains will be forced to be locally-constant, and so in particular $\check{H}^*(M;\, C^0(\mathbb{Z})) = \check{H}^*(M;\mathbb{Z})$.

Assuming that M is a smooth manifold and G is a *Lie group,* we can further require the cochains to be made of *smooth* functions, thus leading to the *Čech cohomology with coefficients in the sheaf of smooth G-valued functions,*

$$\check{H}^*(M;\, C^\infty(G)) \,.$$

It is important to note that, if one merely chooses G to be the additive groups $\mathbb{R}$ or $\mathbb{C}$, then nothing much happens, since it is proved that $\check{H}^n(M;\, C^\infty(\mathbb{R})) = 0$ for every $n \geq 1$.

Finally, note in passing that, if M and G happen to be complex manifolds, then we can require the cocycles to be holomorphic. This leads to the Čech cohomology with coefficients in the sheaf of holomorphic G-valued bundles, denoted by $\check{H}^*(M;\, \mathcal{O}(G))$. If one then takes G to be the additive group $\mathbb{C}$, then $\check{H}^n(M;\, \mathcal{O}(\mathbb{C}))$ —usually denoted by $H^n(M; \mathcal{O})$—is very much nontrivial, and plays an essential role in complex geometry.

A further generalization of Čech cohomology allows, in a sense, for the coefficient-group G to vary from point to point, and that leads to **sheaf cohomology**, but not in this volume. For ramifications in complex geometry, see **P. Griffiths** and **J. Harris**'s ***Principles of algebraic geometry*** **[GH78, GH94]**. For algebraic topology applications, see **R. Godement**'s ***Topologie algébrique et théorie des faisceaux*** **[God58, God73]**. For topological use in combination with differential forms, see **R. Bott** and **L. Tu**'s ***Differential forms in algebraic topology*** **[BT82]**.

Finally, we reached the bundles. We now combine the two extensions above, allowing both non-commutative groups and non-constant cochains. Assume that G is a subgroup of $GL(k)$. Then

$$\check{H}^1(M;\, C^\infty(G))$$

is the set of all k–plane bundles with structure group G, up to isomorphisms. Its distinguished element $\left[\{\mathbf{1}_{\alpha\beta}\}\right]$ is the trivial bundle $M \times \mathbb{R}^k \to M$.

To convince ourselves, let us notice that a class in $\check{H}^1(M;\, C^\infty(G))$ is determined by a G–valued 1–cochain

$$\{g_{\alpha\beta} \colon U_\alpha \cap U_\beta \longrightarrow G\}$$

that is coclosed, meaning that we must have $g_{\alpha\beta} \cdot g_{\beta\gamma} \cdot g_{\gamma\alpha} = 1$. Two such cocycles g' and g'' define a same class if they differ by a coboundary, that is to say,

$$[g'] = [g''] \qquad \Longleftrightarrow \qquad g'_{\alpha\beta} = f_\alpha \cdot g''_{\alpha\beta} \cdot f_\beta^{-1}$$

for some collection $\{f_\alpha \colon U_\alpha \to G\}$. However, this defines nothing but a smooth vector bundle, unique up to isomorphisms and with structure group G, as was explained back on page 176.

More generally, for any group G the set $\check{H}^1(M;\, C^\infty(G))$ is the set of all principal G–bundles, with distinguished element $M \times G \to M$.

Let us now look at a few examples:

Complex line bundles. Since any complex-line bundle can be endowed with a Hermitian metric, which reduces its structure group from $GL_{\mathbb{C}}(1)$ to $U(1) = \mathbb{S}^1$, it becomes clear that

$$H^1(M;\, C^\infty(\mathbb{S}^1))$$

is the set of all *smooth*[10] complex-line bundles on M. Since $\mathbb{S}^1$ is Abelian, the set $H^1(M;\, C^\infty(\mathbb{S}^1))$ turns out to be a group; its operation corresponds to tensor products of line bundles.

Further, $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ fits into the exact sequence of groups

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{R} \xrightarrow{e^{2\pi\cdot}} \mathbb{S}^1 \longrightarrow 0$$

---

[10] For *holomorphic* line bundles on a complex manifold M, one would look at $\check{H}^1(M;\, \mathcal{O}(\mathbb{C}^*))$, usually denoted by $H^1(M, \mathcal{O}^*)$.

(with the groups $\mathbb{Z}$ and $\mathbb{R}$ written additively, but $\mathbb{S}^1$ written multiplicatively). This short exact sequence leads, as usual, to a long exact sequence in cohomology, part of which is:

$$\cdots \longrightarrow \check{H}^1(M;\, C^\infty(\mathbb{R})) \longrightarrow \check{H}^1(M;\, C^\infty(\mathbb{S}^1)) \longrightarrow$$
$$\longrightarrow \check{H}^2(M;\, C^\infty(\mathbb{Z})) \longrightarrow \check{H}^2(M;\, C^\infty(\mathbb{R})) \longrightarrow \cdots$$

Since $\check{H}^n(M; C^\infty(\mathbb{R})) = 0$ and $\check{H}^n(M; C^\infty(\mathbb{Z})) = H^n(M;\mathbb{Z})$, exactness provides an isomorphism

$$0 \longrightarrow \check{H}^1(M;\, C^\infty(\mathbb{S}^1)) \underset{\approx}{\xrightarrow{c_1}} H^2(M;\mathbb{Z}) \longrightarrow 0\,.$$

In terms of bundles, this isomorphism is established by sending a line bundle L to its first Chern class:

$$L \longmapsto c_1(L)\,.$$

In particular, this proves (again) that every 2–class of M can be represented by a smooth complex-line bundle on M, and thus (by taking the zero-locus of a generic section) by a surface embedded in M.

Čech cocycle for Chern. By explicitly following the isomorphism $\check{H}^1(M;\, C^\infty(\mathbb{S}^1)) \approx \check{H}^2(M;\mathbb{Z})$, we obtain a concrete description of a cocycle for $c_1(L)$: Let L be a complex-line bundle, defined by a cocycle $\{g_{\alpha\beta}\}$ with values in $\mathbb{S}^1$. Lift each map $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to \mathbb{S}^1$ to some map $\vartheta_{\alpha\beta}\colon U_\alpha \cap U_\beta \to \mathbb{R}$ so that $g_{\alpha\beta}(x) = e^{2\pi\vartheta_{\alpha\beta}(x)}$ and $\vartheta_{\alpha\beta} = -\vartheta_{\beta\alpha}$. The cocycle condition $g_{\alpha\beta}\cdot g_{\beta\gamma}\cdot g_{\gamma\alpha} = 1$ only lifts to $\vartheta_{\alpha\beta} + \vartheta_{\beta\gamma} + \vartheta_{\gamma\alpha} \in \mathbb{Z}$. Then define the Čech 2–cocycle $\{c_{\alpha\beta\gamma}\colon U_\alpha \cap U_\beta \cap U_\gamma \to \mathbb{Z}\}$ by setting

$$c_{\alpha\beta\gamma} = \vartheta_{\alpha\beta} + \vartheta_{\beta\gamma} + \vartheta_{\gamma\alpha}\,.$$

Its cohomology class is $c_1(L) \in H^2(M;\mathbb{Z})$. This exhibits $c_1(L)$ as essentially a cohomological bookkeeping of the 2π rotations used while building L. (For that matter, so is $w_2(L)$, but only modulo 2.)

Orientable vector bundles. Since every k–plane bundle can be endowed with a fiber metric, the set

$$\check{H}^1(M;\, C^\infty O(k))$$

is still the set of all k–plane bundles on M. A vector bundle is *orientable* if its structure group can be reduced to $SO(k)$. The exact sequence

$$0 \longrightarrow SO(k) \longrightarrow O(k) \xrightarrow{\det} \mathbb{Z}_2 \longrightarrow 0$$

(with $\mathbb{Z}_2 = \{-1,+1\}$ written multiplicatively) leads to an exact sequence of sets[11]

$$\cdots \longrightarrow H^0(M;\mathbb{Z}_2) \longrightarrow \check{H}^1(M;\, C^\infty SO(k)) \longrightarrow$$
$$\longrightarrow \check{H}^1(M;\, C^\infty O(k)) \xrightarrow{w_1} H^1(M;\mathbb{Z}_2)\,.$$

The map denoted w_1 is the assignment of the first Stiefel–Whitney class

$$E \longmapsto w_1(E)\,.$$

By exactness, if a bundle $E \in \check{H}^1(M;\, C^\infty O(n))$ has $w_1(E) = 0$, then E must come from $\check{H}^1(M;\, C^\infty SO(n))$, that is to say, E can be oriented. If a bundle is orientable, then its various orientations are all classified by the elements of $H^0(M;\mathbb{Z}_2)$.

---

[11] An exact sequence of sets (each with a distinguished element) means that the image of one map coincides with the preimage of the distinguished element through the next map.

Čech cocycle for $w_1(E)$. Specifically, if E is defined by the $O(k)$-cocycle $\{g_{\alpha\beta}\}$, then $w_1(E) \in \check{H}^1(M;\mathbb{Z}_2)$ is determined by the $\mathbb{Z}_2$-valued Čech 1-cocycle $\{\det g_{\alpha\beta}\}$.

Spin structures. An oriented k–plane bundle (with k at least 3) is said to admit a *spin structure* if its $SO(k)$–cocycle lifts through the double-cover $Spin(k) \to SO(k)$ to a $Spin(k)$–valued cocycle. The exact sequence

$$0 \longrightarrow \mathbb{Z}_2 \longrightarrow Spin(k) \longrightarrow SO(k) \longrightarrow 0$$

(with $\mathbb{Z}_2 = \{-1, +1\}$ written multiplicatively) leads to an exact sequence

$$\cdots \longrightarrow H^1(M;\mathbb{Z}_2) \longrightarrow \check{H}^1(M;\, C^\infty Spin(k)) \longrightarrow$$
$$\longrightarrow \check{H}^1(M;\, C^\infty SO(k)) \xrightarrow{w_2} H^2(M;\mathbb{Z}_2)\ .$$

The map w_2 above simply ascribes the second Stiefel–Whitney class

$$E \longmapsto w_2(E)\ .$$

By exactness, if a bundle $E \in \check{H}^1(M;\, C^\infty SO(k))$ has $w_2(E) = 0$, then E must come from a $Spin(k)$-cocycle from $\check{H}^1(M;\, C^\infty Spin(k))$. Further, the spin structures on a bundle E with $w_2(E) = 0$ are classified by $H^1(M;\mathbb{Z}_2)$.

Čech cocycle for $w_2(E)$. Let $\{g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to SO(k)\}$ be a cocycle for an oriented bundle E. Assuming that the $U_\alpha \cap U_\beta$'s are all simply-connected, we can always lift the maps $g_{\alpha\beta}$ to maps

$$\tilde{g}_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow Spin(k)$$

with $\tilde{g}_{\alpha\beta} = \tilde{g}_{\beta\alpha}^{-1}$. The product $\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha}$ will take values in $\mathbb{Z}_2 = \{-1, +1\}$. We can then define a $\mathbb{Z}_2$–valued Čech 2–cochain $\{w_{\alpha\beta\gamma}\colon U_\alpha \cap U_\beta \cap U_\gamma \to \mathbb{Z}_2\}$ by setting

$$w_{\alpha\beta\gamma} = \tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha}\ .$$

Clearly, the cochain $\{w_{\alpha\beta\gamma}\}$ measures the failure of the $\tilde{g}_{\alpha\beta}$'s to define a spin structure on E. Moreover, $\{w_{\alpha\beta\gamma}\}$ is a cocycle: indeed, it is not hard to check that

$$(\delta w)_{\alpha\beta\gamma\delta} = (\tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\delta} \cdot \tilde{g}_{\delta\beta}) \cdot (\tilde{g}_{\alpha\gamma} \cdot \tilde{g}_{\gamma\delta} \cdot \tilde{g}_{\delta\alpha}) \cdot (\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha})$$

is constantly $+1$. The cocycle $\{w_{\alpha\beta\gamma}\}$ represents in Čech cohomology the second Stiefel–Whitney class of E:

$$w_2(E) = [w_{\alpha\beta\gamma}]\ .$$

This can be argued indirectly by using the fact that the vanishing of both $w_2(E)$ and of $[w_{\alpha\beta\gamma}]$ are equivalent to the existence of a spin structure on $E \to X$. Indeed, if $[w_{\alpha\beta\gamma}] = 0$, that means that $w_{\alpha\beta\gamma}$ is a coboundary. In other words, there must be a $\mathbb{Z}_2$–valued 1–cochain $\{\varepsilon_{\alpha\beta}\colon U_\alpha \cap U_\beta \to \mathbb{Z}_2\}$ so that $w_{\alpha\beta\gamma} = \varepsilon_{\alpha\beta} \cdot \varepsilon_{\beta\gamma} \cdot \varepsilon_{\gamma\alpha}$. However, that implies that $(\varepsilon_{\alpha\beta} \cdot \tilde{g}_{\alpha\beta}) \cdot (\varepsilon_{\beta\gamma} \cdot \tilde{g}_{\beta\gamma}) \cdot (\varepsilon_{\gamma\alpha} \cdot \tilde{g}_{\gamma\alpha}) = +1$ or, in other words, that the $\varepsilon_{\alpha\beta}$'s represent the *corrections* needed to make the $\tilde{g}_{\alpha\beta}$'s into a genuine $Spin(4)$–cocycle. Thus, $[w_{\alpha\beta\gamma}] = 0$ if and only if E admits a spin structure.

Simplicial cocycle for $w_2(E)$. Passing the identity

$$w_2(E) = \left[w_{\alpha\beta\gamma}\right]$$

through the isomorphisms between Čech and simplicial cohomology exhibited earlier, leads to the uncovering of a simplicial cocycle ϑ for $w_2(E)$:

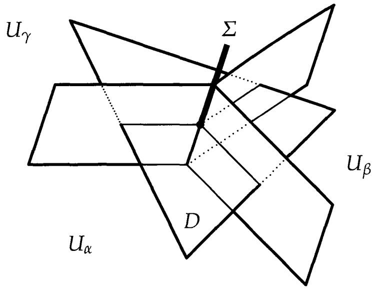
Triangulate the base X and use for all cocycles the covering $U_\alpha = \text{star}(v_\alpha)$ corresponding to the vertices v_α of X. Then a triple intersection $U_\alpha \cap U_\beta \cap U_\gamma$ is non-empty if and only if it corresponds to a 2–simplex $\langle v_\alpha, v_\beta, v_\gamma \rangle$ (and in that case the interior of $\langle v_\alpha, v_\beta, v_\gamma \rangle$ is included in $U_\alpha \cap U_\beta \cap U_\gamma$).

Choose a random lift of the $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ of E to some set of $Spin(4)$–valued maps $\{\tilde{g}_{\alpha\beta}\}$. Then the simplicial 2–cocycle ϑ for $w_2(E)$ is defined by assigning to every 2–simplex $\langle v_\alpha, v_\beta, v_\gamma \rangle$ the $\mathbb{Z}_2$–value of the product $\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha}$.

Switching from writing $\mathbb{Z}_2 = \{-1, +1\}$ multiplicatively to the more familiar additive writing $\mathbb{Z}_2 = \{0, 1\}$, we translate to having ϑ assign 0 to D if and only if $\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha} = +1$, and assign 1 if and only if $\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha} = -1$.

Around a characteristic surface. Let us focus on the case of 4–manifolds M and their tangent bundles T_M. Using the above description of a simplicial cocycle ϑ for $w_2(T_M)$, we can imagine a characteristic surface of M as a surface that manages to cross an odd number of times exactly those 2–simplices that ϑ assigns to 1.

An even better way to see this is probably in the slightly different setting used in the proof of equivalence of the spin structure definitions (preceding note, page 181), as is recalled in figure 4.27. Recall that in that case the U_α's were small neighborhoods of the 4–cells dual to the vertices of M.

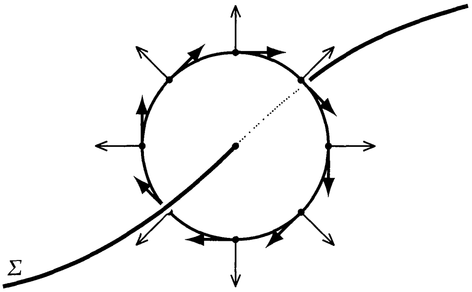


4.27. Drawing a characteristic surface

Assume now that D is a 2–simplex, surrounded by the open sets U_α, U_β, U_γ, with $\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha} = -1$. Then the 2–cell Σ dual to D is part of a simplicial chain that describes a (modulo 2) homology class Poincaré-dual to $w_2(T_M)$.

With a bit of luck in choosing the lifts $\tilde{g}_{\alpha\beta}$, the union of all these distinguished dual 2–cells will make up an actual (unoriented) embedded surface in M ("luck" is needed, because *a priori* there might be problems at the vertices). With a bit more luck, the surface Σ will be orientable, in which case it represents an integral homology class dual to $w_2(T_M)$, and thus is deserving of the name "characteristic surface".

This picture also has the advantage of exhibiting a characteristic surface Σ as surrounded by 2π–twists of T_M, as was mentioned earlier (page 179) and is recalled here through figure 4.28. Away from Σ, the maps $\widetilde{g}_{\alpha\beta}$ are a genuine cocycle and thus define a spin structure on the complement $M \setminus \Sigma$; clearly, this spin structure on $M \setminus \Sigma$ cannot be extended across Σ.



4.28. Outside spin structure, not extending across a characteristic surface Σ

Note: Obstruction theory

In this note, we give a short presentation of obstruction theory. On one hand, this will shed light on several constructions already seen in this chapter. On the other hand these techniques will be needed in the note on page 207 ahead, where the theory of smooth structures on topological manifolds is explained.

Obstruction theory deals with the problem of existence and uniqueness of sections of fiber bundles, encoding it into cohomology classes with coefficients in the homotopy groups of the fiber. At the outset, the case of a vector bundle E is uninteresting, since there always exist sections. However, obstruction theory can be applied to bundles associated to E, such as its sphere bundle SE (uncovering the obstruction to the existence of a nowhere-zero section in E), or the bundle $\mathcal{P}_{SO(E)}$ of frames in E (uncovering obstructions to the existence of a global frame-field in E, that is, obstructions to trivializing E), or bundles of partial frames—the resulting obstructions turn out to be the usual characteristic classes of E. In particular, from this note we will gain yet another point-of-view on the characteristic classes of a 4–manifold.

The argument to follow has two main components, each propelling the other: on one hand, defining things through cell-by-cell extensions and climbing from each k–skeleton to the $(k+1)$–skeleton; on the other hand, meshing the issue of extending sections with the issue of their uniqueness up to homotopy.

Set-up. A **fiber bundle** E with fiber F over a manifold X is a space E and a map $p\colon E \to X$ so that X is covered by open sets U over which the restriction of p to $p^{-1}[U]$ looks like the projection $U \times F \to U$.

Assume that that the fiber F is connected; furthermore, assume that F's *first nontrivial* homotopy group[12] is $\pi_m(F)$. (If $m = 1$, assume further that $\pi_1(F)$ is Abelian.)

Moreover, choose a random cellular decomposition[13] of X. We denote by $X|_k$ the k–skeleton of X, and by $E|_k$ the restriction of E to $X|_k$ (not the k–skeleton of the space E). Also, let $\sigma|_k$ denote the restriction of σ to $X|_k$.

Free ride, up to the m–skeleton. We try to define a section σ of E by defining it over the vertices of X, then try to extend σ over the 1–skeleton of X, then over the 2–skeleton, and so on, cell-by-cell. This plan proceeds without problems until we attempt to extend from the m–skeleton across the $(m+1)$–skeleton.

Indeed, to reach the m–skeleton, we start by defining $\sigma(vertex)$ in any random way. Then, assuming σ was already defined over the k–skeleton of X, we try to extend σ across the $(k+1)$–cells of X: For every $(k+1)$–cell C, we notice that the restriction $E|_C$ is trivial (since C is contractible)[14] and hence $E|_C \approx C \times F$. Our σ, already defined on the k–sphere ∂C, induces a map $\partial C \to F$. Then $\sigma|_{\partial C}$ can be extended across C if and only if the induced map $\partial C \to F$ is homotopically-trivial. However, as long as $k \leq m-1$, we have $\pi_k(F) = 0$ and thus every map $\partial C \to F$ can be extended over C, and hence so can σ. Therefore, we can always define sections σ over the m–skeleton of X.

Uniqueness so far. Let us investigate for a second the dependence (up to homotopy) of the resulting $\sigma|_m$ on the choices made along the way; again, we split the problem in stages between the k– and $(k+1)$–skeletons.

Take k and assume that σ is fixed over $X|_k$, then let σ' and σ'' be two extensions of σ from $X|_k$ across $X|_{k+1}$. A homotopy between $\sigma'|_{k+1}$ and $\sigma''|_{k+1}$ means a section Φ in the product-bundle $p \times id\colon E \times [0,1] \to X \times [0,1]$, defined over $(X|_{k+1}) \times [0,1]$ and limiting to σ' on $X \times 0$ and to σ'' on $X \times 1$.

We choose the obvious cellular decomposition of $X \times [0,1]$ induced from the chosen decomposition of X, with each j–cell C of X creating two j–cells $C \times 0$ and $C \times 1$ in $X \times [0,1]$, and a $(j+1)$–cell $C \times [0,1]$.

Certainly Φ must be defined to be $\sigma' \times 0$ on $(X|_{k+1}) \times 0$ and to be $\sigma'' \times 1$ on $(X|_{k+1}) \times 1$. Furthermore, since σ' and σ'' were taken to coincide over the k–skeleton of X, it follows that, for every j–cell C of X with $j \leq k$, we have $\sigma'|_C = \sigma''|_C$. We can then extend Φ across the $(j+1)$–cell $C \times [0,1]$ simply as $\sigma \times id$. Therefore, for fully extending Φ across $(X|_{k+1}) \times [0,1]$, we need only extend Φ across those

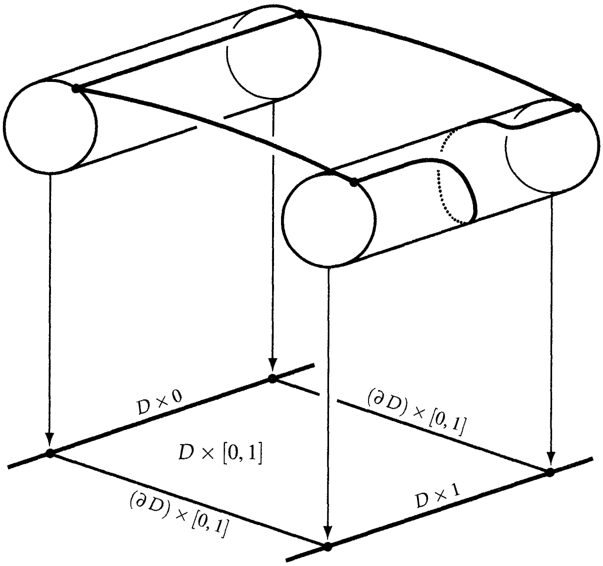
---

[12] Remember that the **homotopy group** $\pi_k(A)$ is the set of all homotopy-classes of maps $\mathbb{S}^k \to A$, with a suitable group operation. An element $f \in \pi_k(F)$ is trivial if and only if $f\colon \mathbb{S}^k \to A$ can be extended to a map $\tilde{f}\colon \mathbb{D}^{k+1} \to A$. Whenever k is at least 2, the group $\pi_k(A)$ is Abelian.

[13] Handle decompositions would work just as well. Just substitute the word "handle" for "cell" in all that follows. Or one could use a triangulation of X (as recalled back in footnote 5 on page 182) and substitute "simplex" for "cell" throughout.

[14] Technically, since the cell C is not necessarily embedded along ∂C, one should view $E|_C$ as the pull-back ι^*E, where $\iota\colon C \to X$ is the "inclusion" of the cell in X.

$(k+2)$–cells of $X \times [0,1]$ that are of shape $D \times [0,1]$ for some $(k+1)$–cell D of X. Compare with figure 4.29.



4.29. Toward a homotopy between two sections

Notice that Φ is already defined over the whole boundary $\partial(D \times [0,1])$. Thus, Φ restricted to the $(k+1)$–sphere $\partial(D \times [0,1])$ determines an element of $\pi_{k+1}(F)$. It follows that Φ extends across $D \times [0,1]$ if and only if the class of $\Phi|_{\partial(D\times[0,1])}$ in $\pi_{k+1}(F)$ is trivial.

Therefore, since all homotopy groups of F were assumed trivial up to dimension m, it follows that the extension of σ up to the $(m-1)$–skeleton of X must be unique up to homotopy. However, when we extend σ from the $(m-1)$–skeleton across the m–skeleton, the various options can differ over each m–cell by elements of $\pi_m(F)$. We will come back to this issue.

Across the $(m+1)$–cells: obstruction cocycles. In any case, our fibre bundle $E \to X$ admits a section σ defined over the m–skeleton of the base. When attempting to extend σ from the m–skeleton across the $(m+1)$–skeleton, obstructions appear. Indeed, if D is a $(m+1)$–cell, then $\sigma|_{\partial D}$ might describe a nontrivial element in $\pi_m(F)$, and then our σ will not extend across D. Compare with figure 4.30 on the following page.

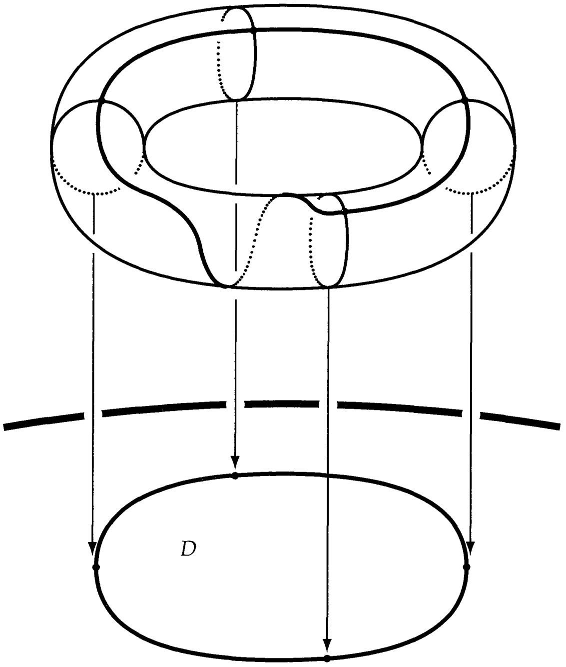
To measure this, we define the function

$$\vartheta_\sigma \colon \{(m+1)\text{–cells of } X\} \longrightarrow \pi_m(F) \qquad D \longmapsto [\sigma|_{\partial D}] ,$$

which takes an $(m+1)$–cell D to the element of $\pi_m(F)$ that is determined by $\sigma|_{\partial D}$ through some random trivialization $E|_D \approx D \times F$. We can then extend ϑ_σ by linearity, and think of it as a $\pi_m(F)$–valued[15] cellular $(m+1)$–cochain on X.

---

[15] In truth, the twists of our fiber bundle $E \to X$ might twist the way the π_m's of the various fibers can be assembled together. Thus, to get a well-defined map ϑ, one must in fact use *twisted coefficients*



4.30. Obstruction to extending a section

This cochain ϑ_σ is in fact *coclosed.* Indeed, on every $(m+2)$–cell B, we have

$$(\delta\,\vartheta_\sigma)(B) = \vartheta_\sigma(\partial\,B) = \left[\sigma|_{\partial\partial B}\right] = 0\,,$$

where ∂ denotes taking the homological boundary and we use that $\partial\,\partial = 0$. We call ϑ_σ the **obstruction cocycle** of σ. Our chosen section σ will extend over the $(m+1)$–skeleton if and only if $\vartheta_\sigma = 0$.

Even when the cocycle ϑ_σ happens to be nontrivial, we can still try to go back and change the way σ was defined over the m–skeleton of X, and maybe the new version σ' will have $\vartheta_{\sigma'} = 0$ and hence extend. We need to revisit the issue of uniqueness of sections of $E|_m$:

Uniqueness, revisited: difference cochains. Given any two sections σ' and σ'' of E defined over the m–skeleton, they cannot differ homotopically over the $(m-1)$–skeleton. Therefore there must exist a homotopy κ between $\sigma'|_{m-1}$ and $\sigma''|_{m-1}$.

We try to extend this κ to a homotopy Φ between $\sigma'|_m$ and $\sigma''|_m$. As before, we view Φ as a partial section of $E \times [0,1] \to X \times [0,1]$ and set Φ to be $\sigma'|_m \times 0$ on $(X|_m) \times 0$ and $\sigma''|_m \times 1$ on $(X|_m) \times 1$, and thereafter extend it across $(X|_{m-1}) \times [0,1]$ by spreading κ over it, thus linking $\sigma'|_{m-1} \times 0$ with $\sigma''|_{m-1} \times 1$.

To extend this to a full homotopy between $\sigma'|_m$ and $\sigma''|_m$, we need only extend Φ across every $(m+1)$–cell $C \times [0,1]$ that corresponds to some m–cell C of X. The

---

(better known as *local coefficients*) that twist $\pi_m(F)$ by the action of $\pi_1(X)$ on the fibers of F. Let us assume that X is *simply-connected* and move on as if nothing happened...

homotopic difference between σ' and σ'' can be encoded in the obstruction to this extension, namely in the function

$$d(\sigma'_{\kappa}\sigma'') \colon \{m\text{–cells of } X\} \longrightarrow \pi_m(F) \qquad C \longmapsto \left[\Phi|_{\partial(C\times[0,1])}\right] .$$

This $d(\sigma'_{\kappa}\sigma'')$ can be extended by linearity and defines a $\pi_m(F)$–valued[16] cellular m–cochain on X. It is called the **difference cochain** of σ' and σ''.

The homotopy κ between $\sigma'|_{m-1}$ and $\sigma''|_{m-1}$ can be extended to a full homotopy between σ' and σ'' if and only if $d(\sigma'_{\kappa}\sigma'')$ is identically-zero. However, a different choice of homotopy $\kappa \colon \sigma'|_{m-1} \sim \sigma''|_{m-1}$ might be a better choice toward obtaining a homotopy between $\sigma'|_m$ and $\sigma''|_m$. We will come back to this issue.

Back to obstruction cocycles: primary obstructions. We return to the extension problem, to see how different choices of sections over $X|_m$ influence our chances of extension across $X|_{m+1}$. Let σ' and σ'' be two sections of $E|_m$ and choose a random homotopy κ between $\sigma'|_{m-1}$ and $\sigma''|_{m-1}$. Consider the bundle $E \times [0,1] \to X \times [0,1]$ and denote by $\sigma'_{\kappa}\sigma''$ its section defined as $\sigma' \times 0$ over $X|_m \times 0$, as $\sigma'' \times 1$ over $X|_m \times 1$, and as κ over $(X|_{m-1}) \times [0,1]$.

Notice that this section $\sigma'_{\kappa}\sigma''$ is defined over the whole m–skeleton of the base $X \times [0,1]$. We can therefore define its obstruction cocycle $\vartheta_{\sigma'_{\kappa}\sigma''}$. We observe that this cocycle is made of three distinct parts: (1) the obstruction to extending $\sigma' \times 0$ across the $(m+1)$–cells $D \times 0$ of $X \times 0$, which can be identified with $\vartheta_{\sigma'}(D)$; (2) the obstruction to extending $\sigma'' \times 1$ across the $(m+1)$–cells $D \times 1$ of $X \times 1$, which can be identified with $\vartheta_{\sigma''}(D)$; finally, (3) the obstruction to extending κ across the $(m+1)$–cells of shape $C \times [0,1]$, which can be identified with $d(\sigma'_{\kappa}\sigma'')(C)$.

Take any $(m+1)$–cell D of X and consider the $(m+2)$–cell $D \times [0,1]$ of $X \times [0,1]$. Apply the above decomposition of $\vartheta_{\sigma'_{\kappa}\sigma''}$ to $\partial(D \times [0,1])$. On one hand, since $\vartheta_{\sigma'_{\kappa}\sigma''}$ is a cocycle, it must vanish on every boundary and in particular on $\partial(D \times [0,1])$. On the other hand, we have $\partial(D \times [0,1]) = D\times 1 \cup D\times 0 \cup (\partial D)\times[0,1]$, and we can evaluate the parts of $\vartheta_{\sigma'_{\kappa}\sigma''}$ on these pieces. We end up with $\vartheta_{\sigma'}(D)$, $\vartheta_{\sigma''}(D)$, and $d(\sigma'_{\kappa}\sigma'')(\partial D)$. Gathering up and keeping track of signs, we obtain the equality $\vartheta_{\sigma'}(D) - \vartheta_{\sigma''}(D) = d(\sigma'_{\kappa}\sigma'')(\partial D)$, which translates to

$$\vartheta_{\sigma'} - \vartheta_{\sigma''} = \delta\, d(\sigma'_{\kappa}\sigma'') .$$

The conclusion is that different choices of sections in $E|_m$ change the corresponding obstruction cochain by a *coboundary*. It follows that the obstruction cocycle determines a well-defined cohomology class

$$[\vartheta_{\sigma}] \in H^{m+1}(X; \pi_m(F)) .$$

This class depends only on the bundle $E \to X$ and *not* on the choice of section σ. Moreover, this class is trivial if and only if there is some m–cochain d such that $\vartheta_{\sigma} = \delta d$. In that case, we can change σ over the m–skeleton of X to a section σ' with $d(\sigma\sigma') = d$, and then the new σ' will have obstruction cocycle $\vartheta_{\sigma'} = 0$: it will extend across $X|_{m+1}$.

---

[16] Again, in general one needs twisted coefficients.

In conclusion, $E \to X$ admits sections over the $(m+1)$–skeleton of X if and only if the class $[\vartheta]$ vanishes. We call this class the **primary obstruction**[17] of $E \to X$.

Back to uniqueness: difference cocycles. If the primary obstruction $[\vartheta]$ vanishes, then conceivably there exist several distinct sections of $E|_m$ that extend across the $(m+1)$–skeleton of X.

Assume that $\sigma'|_m$ and $\sigma''|_m$ are two such extendable sections of $E|_m$ and take κ to be some homotopy between $\sigma'|_{m-1}$ and $\sigma''|_{m-1}$. We have $\vartheta_{\sigma'} - \vartheta_{\sigma''} = \delta\, d(\sigma'_\kappa\sigma'')$, but since both σ' and σ'' were assumed extendable, their obstruction cocycles vanish, and thus

$$\delta\, d(\sigma'_\kappa\sigma'') = 0 .$$

In other words, the difference cochain is now in fact a *cocycle.*

Further, the difference cochain $d(\sigma'_\kappa\sigma'')$ can in this case be understood as representing the whole obstruction cocycle $\vartheta_{\sigma'_\kappa\sigma''}$ of the section $\sigma'_\kappa\sigma''$ across the $(m+1)$–skeleton of $X \times [0,1]$. We can then apply the previous results about obstruction cocycles to this $d(\sigma'_\kappa\sigma'')$. It follows that changing the choice of homotopy $\kappa \colon \sigma'|_{m-1} \sim \sigma''|_{m-1}$ merely modifies $d(\sigma'_\kappa\sigma'')$ by the addition of a coboundary. Therefore, the difference cocycle itself determines a well-defined cohomology class

$$\left[d(\sigma'_\kappa\sigma'')\right] \in H^m(X; \pi_m(F)) .$$

This class depends only on the extendable sections σ' and σ'' and *not* on the choice of homotopy κ. Furthermore, if $[d(\sigma'\sigma'')] = 0$, then there exists a choice of κ that can be extended to a full homotopy Φ between $\sigma'|_m$ and $\sigma''|_m$.

Conclusion. For every fiber bundle $E \to X$ whose fiber F has its first nontrivial homotopy group in dimension m, the primary obstruction

$$[\vartheta] \in H^{m+1}(X; \pi_m(F))$$

vanishes if and only if there are sections of $E \to X$ defined over the m–skeleton of X that extend across the $(m+1)$–skeleton.

Moreover, if $[\vartheta] = 0$ and one chooses some extendable section σ, then all other sections σ' of $E|_m$ that extend across $X|_{m+1}$ are classified up to homotopy by the elements of

$$H^m(X; \pi_m(F))$$

via their corresponding difference classes $[d(\sigma\sigma')]$

Application: trivializing the tangent bundle. We will now apply the method of obstruction theory to the problem of trivializing the tangent bundle T_M of an oriented 4–manifold M. Since a trivialization of T_M over some subset U of M is equivalent to a field of frames over U, the problem becomes one of finding sections in the bundle of frames $\mathcal{P}_{SO(4)}$ of T_M.

The fiber of $\mathcal{P}_{SO(4)}$ is the Lie group $SO(4)$, which is connected and has

$$\pi_1 SO(4) = \mathbb{Z}_2 , \qquad \pi_2 SO(4) = 0 , \qquad \pi_3 SO(4) = \mathbb{Z} \oplus \mathbb{Z} .$$

---

[17] "Primary", because the project can conceivably be continued by attempting to further extend across higher skeleta, until we exhaust all X.

Therefore, when applying the obstruction theory method, we first encounter a primary obstruction in $H^2(M; \pi_1 SO(4))$. This obstruction class is none other than the Stiefel–Whitney class

$$w_2(T_M) \in H^2(M; \mathbb{Z}_2) .$$

Hence, if $w_2(T_M) = 0$, then $\mathcal{P}_{SO(4)}$ admits a section over the 2–skeleton of M, in other words, T_M can be trivialized over $M|_2$. Two such sections of $\mathcal{P}_{SO(4)}$ over $M|_2$ differ by difference cocycles from $H^1(M; \mathbb{Z}_2)$. Such trivializations of T_M over $M|_2$ are, of course, spin structures on M.

Assuming that $w_2(T_M)$ vanished and we did choose a section of $\mathcal{P}_{SO(4)}$ over $M|_2$, we can now try to further extend it over M. Since $\pi_2 SO(4) = 0$, extending across the 3–skeleton encounters no problems. The next significant obstruction appears in $H^4(M; \pi_3 SO(4))$, and it can be identified as the pair

$$\big(e(T_M),\, p_1(T_M)\big) \in H^4(M; \mathbb{Z} \oplus \mathbb{Z}) ,$$

made from the Euler class $e(T_M)$ and the Pontryagin class $p_1(T_M)$.

The Euler class appears as the obstruction to extending a nowhere-zero vector field over all M, that is to say, $e(T_M)$ is the primary obstruction to defining a section in the 3–sphere bundle ST_M of T_M; thus, it belongs to $H^4(M; \pi_3 S^3)$.

That the pair $(e,\, p_1)$ fully catches the secondary obstruction can be argued directly by computing characteristic classes of 4–plane bundles over S^4 that are built using equatorial gluing maps from $\pi_3 SO(4)$; an exposition can be found in **[Sco03]**.

If, besides $w_2(T_M)$ being trivial, we also have that both $e(T_M)$ and $p_1(T_M)$ vanish, then the tangent bundle T_M can be completely trivialized, $T_M \approx M \times \mathbb{R}^4$. This happens for example with $S^1 \times S^3$, but never for simply-connected 4–manifolds.

Similar results apply for general oriented 4–plane bundles over 4–manifolds. In particular, notice that moving along these lines one can quickly obtain a proof of the Dold–Whitney theorem (stated on page 167).

Application: characteristic classes. The obstruction-theoretic approach was in fact the one initially used by E. Stiefel and H. Whitney when they discovered characteristic classes.

Given a vector bundle $E \to X$ with fiber $\mathbb{R}^k$, consider the Stiefel bundle $\mathcal{V}_j(E) \to X$ of all j–frames in E. Then the corresponding primary obstruction $[\vartheta_j]$ of $\mathcal{V}_j(E)$ appears in H^{k-j+1} and determines the Stiefel–Whitney classes by[18]

$$w_{k-j+1}(E) = \begin{cases} [\vartheta_j] & \text{if } k-j+1 \text{ is even and} < k \\ [\vartheta_j] \pmod 2 & \text{if } k-j+1 \text{ is odd, or } j = 1 \end{cases} \in H^{k-j+1}(X; \mathbb{Z}_2) .$$

Thus, each class w_{k-j+1} reveals itself as an obstruction to defining a field of j–frames in E over the $(k-j+1)$–skeleton of X.

---

[18] The modulo 2 reduction in the following formula is done because in those cases ϑ appears at the outset with twisted $\mathbb{Z}$–coefficients. Still, if we know all the w_j's, no information is lost through this reduction.

Finally, if $E \to X$ is oriented, then for $j = 1$ the full obstruction of $\mathcal{V}_1(E) = \mathbb{S}E$ is caught by the Euler class

$$e(E) = [\vartheta_1] \in H^k(X, \mathbb{Z}),$$

which measures the obstruction to defining a nowhere-zero section of E over the k-skeleton of X.

A similar approach can be used for Chern classes.

References. Classic obstruction theory, including a description of Stiefel–Whitney and Chern classes, is presented in **N. Steenrod**'s ***The topology of fibre bundles*** **[Ste51, Ste99**, part III], and is still the best introduction. For a recent discussion of obstruction theory, see for example **J. Davis** and **P. Kirk**'s ***Lecture notes in algebraic topology*** **[DK01]**.

We will use obstruction theory again in the note on page 207 ahead, where we will explore the obstructions to the existence of smooth structures on topological manifolds.

Note: Classifying spaces and spin structures

In what follows, we will define spin structures in terms of maps to classifying spaces. We will start by saying a few words about the spaces $\mathscr{B}G$ that classify all fiber bundles with structure group G, then describe a spin structure on a bundle $E \to X$ as the lift of its classifying map $X \to \mathscr{B}SO(m)$ to a map $X \to \mathscr{B}Spin(m)$.

Part of this note will be better understood if one first reads the preceding note (starting on page 197) on obstruction theory.

This and the preceding note can be viewed both as a continuation of the survey of spin structures from earlier notes, and as preparing the ground for the theory of smoothing topological manifolds that will be explained in the next note (starting with page 207).

Fiber bundles and classifying spaces. A (locally-trivial) **fiber bundle** E with fiber F over a manifold X is a space E and a map $p\colon E \to X$ so that X is covered by open sets $\{U_\alpha\}$ and over each U_α the restriction of p to $p^{-1}[U_\alpha]$ looks like the projection $U_\alpha \times F \to U_\alpha$.

The fiber bundle E is said to have **structure group** G, or is called a G**–bundle**, if over every overlap $U_\alpha \cap U_\beta$ the two trivializations $p^{-1}[U_\alpha] \simeq U_\alpha \times F$ and $p^{-1}[U_\beta] \simeq U_\beta \times F$ are related by a self-homeomorphism of $(U_\alpha \cap U_\beta) \times F$ acting by $(x, f) \longmapsto (x, g_{\alpha\beta}(x) \cdot f)$, where $g_{\alpha\beta}$ is a map $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to G$ and G is a group acting on F by homeomorphisms.

For every topological group G there exists a space $\mathscr{B}G$, called the **classifying space** of G, and for every fiber F on which G acts there exists a G–bundle

$$\mathscr{E}_F G \longrightarrow \mathscr{B}G$$

with fiber F, called the **universal bundle** of fiber F and group G. The spaces $\mathscr{B}G$ and $\mathscr{E}_F G$ are unique up to homotopy-equivalence. The reason for the names "classifying" and "universal" is that that all G–bundles over any X are classified by the homotopy classes of maps $X \to \mathscr{B}G$.

This means that for every G–bundle $E \longrightarrow X$ with fiber F there must exist a map $\xi\colon X \longrightarrow \mathscr{B}G$ so that E is isomorphic to the pull-back through ξ of the universal bundle $\mathscr{E}_F G \longrightarrow \mathscr{B}G$; in other words, ξ can be covered by a bundle morphism $\overline{\xi}$, fitting in the diagram

$$\begin{tikzcd} E \arrow[r, "\overline{\xi}"'] \arrow[d, "p"'] & \mathscr{E}_F G \arrow[d] \\ X \arrow[r, "\xi"] & \mathscr{B}G\,, \end{tikzcd}$$

so that $\overline{\xi}$ is a G–homeomorphism on the fibers. Further, two bundles E' and E'' are isomorphic if and only if their corresponding maps $\xi', \xi''\colon X \longrightarrow \mathscr{B}G$ are homotopic. In brief, the set of all G–bundles can be identified with the set $[X, \mathscr{B}G]$ of homotopy classes of maps $X \longrightarrow \mathscr{B}G$.

> ***Construction.*** *The classifying space $\mathscr{B}G$ can be built as follows: Take G and start joining*[19] *it to itself, building $G * G$, then $G * G * G$, then $G * G * G * G$, then... In the limit, we obtain the space $\mathscr{E}G = G * G * G * \cdots$. The group G acts freely on $\mathscr{E}G$, and we can then build the quotient space $\mathscr{B}G = \mathscr{E}G/G$, which is the classifying space of G. The bundle $\mathscr{E}G \longrightarrow \mathscr{B}G$ is the universal bundle that classifies all principal G-bundles. To get the universal bundle for some other fiber F, pick some cocycle for $\mathscr{E}G \longrightarrow \mathscr{B}G$, let it act on F, and glue $\mathscr{E}_F G$ with it. More generally, if $\mathcal{E}$ is any* contractible *space on which G acts freely, then the map $\mathcal{E} \longrightarrow \mathcal{E}/G$ is a principal G-bundle, and in fact, up to homotopy equivalence, $\mathcal{E} \longrightarrow \mathcal{E}/G$ coincides with $\mathscr{E}G \longrightarrow \mathscr{B}G$. This construction is due to* **J. Milnor***'s* ***Construction of universal bundles. II*** **[Mil56a]**.

Vector bundles. A vector bundle of fiber $\mathbb{R}^k$ over X is a fiber bundle with group $GL(k)$. Then its classifying space can be determined to be

$$\mathscr{B}GL(k) = \mathscr{G}r_k(\mathbb{R}^\infty)\,,$$

i.e., the Graßmann space of all k–planes in $\mathbb{R}^\infty$, defined as $\varinjlim \mathscr{G}r_k(\mathbb{R}^m)$ when $m \longrightarrow \infty$. The universal bundle $\mathscr{E}_{\mathbb{R}^k} GL(k)$ is the vector bundle over $\mathscr{B}GL(k)$ that has as fiber over a point $P \in \mathscr{G}r_k(\mathbb{R}^\infty)$ the actual k–plane P. Intuitively, all twists and turns that a fiber of a vector bundle over X might have can be retrieved by positions of k–planes in $\mathbb{R}^\infty$, and a description of these positions yields the classifying map. More rigorously, one can show that for every bundle $E \longrightarrow X$ there exists a bundle $F \longrightarrow X$ so that $E \oplus F \approx X \times \mathbb{R}^N$, and thus the fibers of E can be identified with k–planes in $\mathbb{R}^N$. For example, if X has dimension m, then one can use $N = m + k + 1$ and $\mathscr{G}r_k(\mathbb{R}^{m+k+1})$ instead of the full $\mathscr{G}r_k(\mathbb{R}^\infty)$.

> *A similar approach works for complex bundles and shows that $\mathscr{B}GL_{\mathbb{C}}(k)$ is the complex Graßmannian $\mathscr{G}r_k(\mathbb{C}^\infty)$. In particular, complex-line bundles are classified by maps to $\mathscr{B}GL_{\mathbb{C}}(1) = \mathbb{CP}^\infty$. For line bundles on 4-manifolds, it is enough to consider maps to $\mathbb{CP}^2$.*

---

[19] The **join** $A * B$ of two spaces A and B is defined as follows: take $A \times B \times [0,1]$, then collapse $A \times 0$ to a point and $B \times 1$ to another point. The join is easiest to visualize if we imagine both A and B as embedded in general position in some high-dimensional $\mathbb{R}^N$; then $A * B$ is the union of all straight segments starting in A and ending in B. For example, the join of two 1–simplices (segments) will be a 3–simplex (a tetrahedron).

Metrics. The group $GL(k)$ is homotopy-equivalent to[20] $O(k)$. Since the whole theory is homotopy-flavored, it follows that $\mathscr{B}GL(k)$ and $\mathscr{B}O(k)$ are homotopy-equivalent, and thus a $GL(k)$-bundle is the same thing as an $O(k)$-bundle. In down-to-earth terms, this simply means that every vector bundle admits a fiber-metric.

Orientations. If the vector bundle is oriented, then its structure group can be further refined from $O(k)$ to $SO(k)$. In terms of classifying spaces, the inclusion $SO(k) \subset O(k)$ induces a map[21]

$$S\colon \mathscr{B}SO(k) \to \mathscr{B}O(k) .$$

Finding an orientation for a bundle E is the same as finding a lift of its classifying map $\xi\colon X \to \mathscr{B}O(k)$ to a map $\xi^s : X \to \mathscr{B}SO(k)$, fitting in

$$\begin{tikzcd} X \arrow[r, "\xi^s"'] \arrow[d, equal] & \mathscr{B}SO(k) \arrow[d, "S"] \\ X \arrow[r, "\xi"] & \mathscr{B}O(k) . \end{tikzcd}$$

The map $S\colon \mathscr{B}SO(k) \to \mathscr{B}O(k)$ is itself a fiber bundle with fiber $O(k)/SO(k) = \mathbb{Z}_2$. We can pull this bundle back over X by using the map $\xi\colon X \to \mathscr{B}O(k)$, and hence transform the problem of finding a lifted map ξ^s into the problem of finding a global section in the pulled-back bundle $\xi^*S \to X$ from

$$\begin{tikzcd} \xi^*S \arrow[r] \arrow[d] & \mathscr{B}SO(k) \arrow[d, "S"] \\ X \arrow[r, "\xi"] & \mathscr{B}O(k) . \end{tikzcd}$$

The fiber of $\xi^*S \to X$ is of course still $\mathbb{Z}_2$.

The obstruction to the existence of a section in ξ^*S can then be attacked by obstruction theory, similar to the outline from the preceding note.[22] This yields as unique obstruction the first Stiefel–Whitney class

$$w_1(E) \in H^1(X; \mathbb{Z}_2) .$$

If one such section (*i.e.*, an orientation of E) is chosen, then all other sections, up to homotopy, are classified by the elements of $H^0(X; \mathbb{Z}_2)$; in other words, you can change the orientation on each connected component of X.

Spin structures. The group $SO(k)$ is double-covered by the Lie group $Spin(k)$, and the double-cover projection $Spin(k) \to SO(k)$ induces a map of classifying spaces

$$Sp\colon \mathscr{B}Spin(k) \to \mathscr{B}SO(k) .$$

---

[20] Indeed, if we think of a matrix $A \in GL(k)$ as a frame in $\mathbb{R}^k$, then we can apply the Gram–Schmidt procedure to split A as a product $A = T \cdot R$ of an upper-triangular matrix T and an orthogonal matrix $R \in O(k)$; since all upper-triangular matrices make up a contractible space, the claim follows.

[21] Notice that $\mathscr{B}SO(k)$ can be represented as the Graßmannnian of all *oriented* k–planes inside $\mathbb{R}^\infty$.

[22] A rather special case of obstruction theory, since one plays with $\pi_0\,(\textit{fiber})$.

This map is a fiber bundle. Its fiber is denoted $SO(k)/Spin(k)$ and it is an Eilenberg–Maclane $K(\mathbb{Z}_2,1)$–space. This means that $\pi_1(SO(k)/Spin(k)) = \mathbb{Z}_2$ is its only non-zero homotopy group.[23]

A **spin structure** on an oriented bundle E is the same as a lift of its classifying map $\xi\colon X \to \mathscr{B}SO(k)$ to a map $\xi^{sp}\colon X \to \mathscr{B}Spin(k)$, made against the map $Sp\colon \mathscr{B}Spin(k) \to \mathscr{B}SO(k)$. Equivalently, by pulling back over X,

$$\begin{tikzcd}
\xi^* Sp \arrow[r] \arrow[d] & \mathscr{B}Spin(k) \arrow[d, "Sp"] \\
X \arrow[r, "\xi"] & \mathscr{B}SO(k)\,,
\end{tikzcd}$$

we see that a spin structure on E is the same as a global section of the bundle $\xi^* Sp \to X$, whose fiber is $SO(k)/Spin(k)$.

After applying obstruction theory to this setting, it turns out that the unique obstruction to the existence of such a section is the second Stiefel–Whitney class

$$w_2(E) \in H^2(X;\mathbb{Z}_2)\,.$$

Characteristic classes. *It is worth noting that, avoiding any obstruction theory, the characteristic classes of a vector bundle $E \to X$ can be viewed directly as pull-backs of cohomology classes of the appropriate classifying space. Indeed, we have isomorphisms between the cohomology rings of the $\mathscr{B}G$'s and polynomial rings generated by the various characteristic classes (endowed with suitable degrees). Specifically,*

$$\begin{aligned}
H^*(\mathscr{B}O(k);\mathbb{Z}_2) &= \mathbb{Z}_2[w_1,w_2,\dots,w_k]\\
H^*(\mathscr{B}SO(k);\mathbb{Z}_2) &= \mathbb{Z}_2[w_2,\dots,w_k]\\
H^*(\mathscr{B}SO(2i);\mathbb{Z}) &= \mathbb{Z}[p_1,p_2,\dots,p_{i-1},e]\\
H^*(\mathscr{B}SO(2i+1);\mathbb{Z}[1/2]) &= \mathbb{Z}[1/2][p_1,p_2,\dots,p_i]\\
H^*(\mathscr{B}U(k);\mathbb{Z}) &= \mathbb{Z}[c_1,c_2,\dots,c_k],
\end{aligned}$$

where $w_j \in H^j$ is the j^{th} Stiefel–Whitney class of the corresponding universal bundle $\mathscr{E}_{\mathbb{R}^k}$, while $p_j \in H^{4j}$ is its m^{th} Pontryagin class, $e \in H^{2i}$ is the Euler class, and $c_j \in H^{2j}$ is the j^{th} Chern class of the universal complex bundle $\mathscr{E}_{\mathbb{C}^k}$. The difference between the $SO(2i)$ and $SO(2i+1)$ cases is owing to the fact that in the first case $e \cup e = p_i$, while in the second $e = 0$; further, the ring $\mathbb{Z}[1/2]$ is needed to kill the 2–torsion (and $\mathbb{Q}$ or $\mathbb{R}$ could be used instead). Indeed, remember that $p_j(E) = (-1)^j c_{2j}(E\otimes\mathbb{C})$, but that the classes $c_{2j+1}(E\otimes\mathbb{C})$, which are all of order 2, escape. See also **D. Husemoller's *Fibre bundles*** **[Hus66, Hus94,** *ch 17]*

Note: Topological manifolds and smoothings

In what follows, we will outline the theory of topological manifolds and of their smooth structures. The theory works best in dimensions 5 or more, where it offers complete answers on the existence and classification of smooth structures on topological manifolds. The theory is quite weaker in dimension 4, but it is still relevant.

Requisites for understanding this note are the two previous notes, namely the one on page 197, where the rudiments of obstruction theory were presented, and the

---

[23] Since $Spin(k) \to SO(k)$ is a cover map, we have $\pi_m(Spin(k)) = \pi_m(SO(k))$ for all $m \geq 2$.

one on page 204, where general fiber bundles and their classifying spaces were explained. The groups of homotopy spheres Θ_m, described in the end-notes of chapter 2.5 (page 97), will also make an appearance. On the other hand, if one skips the paragraphs on smoothing bundles, then one merely needs the simple definition of a general fiber bundle, which can be read from the beginning of the note on page 204.

Historically, at first the realm of purely topological manifolds and pure homeomorphisms seemed unapproachable, so mathematicians attacked the gap between smooth and *piecewise-linear (PL) manifolds,* meaning manifolds structured by a nice triangulation (where "nice" means that the link[24] of every vertex is required to be simplicially-homeomorphic to a standard polyhedral sphere; such triangulated manifolds are also called *combinatorial manifolds*). Success with smoothing PL manifolds started with S. Cairns and continued with M. Hirsch and B. Mazur, which completely elucidated the gap between PL and smooth. The door on smoothing topological manifolds was opened by J. Milnor, who introduced the right concept of tangent bundle for a topological manifold. Finally R. Kirby breached the barrier toward the study of topological manifolds, and together with L. Siebenmann clarified the gap between topological and PL manifolds. See also the bibliographical comments on page 67 at the end of chapter 1, as well as the references ahead on page 219.

Since we are focused on 4–manifolds while the gap between PL and smooth manifolds only starts to make its presence felt in dimension 7, in our presentation below we will shortcut the PL level and discuss smoothing theory only in terms of the gap between topological and smooth manifolds.

Tangent bundles for topological manifolds. Remember that a topological manifold of dimension m is merely a separable metrizable topological space that locally looks like $\mathbb{R}^m$; in other words, X is covered by open sets U that are homeomorphic to $\mathbb{R}^m$.

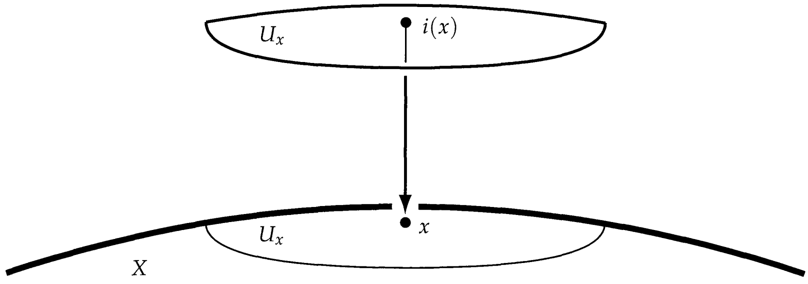
For smooth manifolds, one of the most useful objects used in their study is the tangent bundle, which gives the infinitesimal image of the manifold and thus approximates its structure by simpler spaces. A suitable analogue for topological manifolds can only prove useful.

A first idea would be to pick for each $x \in X$ a small open neighborhood U_x homeomorphic to $\mathbb{R}^m$ and consider it as the fiber of T_X at x, as in figure 4.31 on the next page. Parts of nearby such fibers would get identified just as the corresponding open sets in X: the fiber U_x over x and the fiber U_y over y have their common part $U_x \cap U_y$ identifiable, as suggested in figure 4.32 on the facing page.

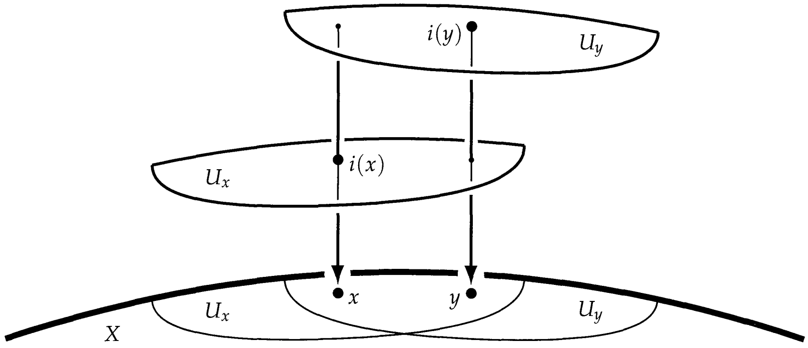
Such a tangent "bundle" has fiber $\mathbb{R}^m$ and has an obvious "zero section" i sending $x \in X$ to $x \in U_x = T_X|_x$. This creature is not a bundle: neighboring fibers cannot be identified with each other, since only parts of them overlap. However, it is conceivable that, by restricting to smaller neighborhoods of the zero-section

---

[24] Think of the **link** of a vertex v essentially as the (simplicial) boundary of a small simplicial neighborhood of v. Specifically, take all simplices σ that contain v and take the faces of σ that do not touch v; the union of all such faces makes up the link of v.



4.31. Building a tangent bundle, I



4.32. Building a tangent bundle, II

and deforming our structure by homotopies, one would end-up with a genuine fiber bundle, with fiber $\mathbb{R}^m$. If, for the resulting bundle, we take care to identify each fiber above x with $\mathbb{R}^m$ in such manner that x corresponds to 0, then the structure group of the bundle would be the group of all self-homeomorphisms $\varphi\colon \mathbb{R}^m \simeq \mathbb{R}^m$ that fix the origin, $\varphi(0) = 0$. Let us denote this group by

$$TOP(m)\ .$$

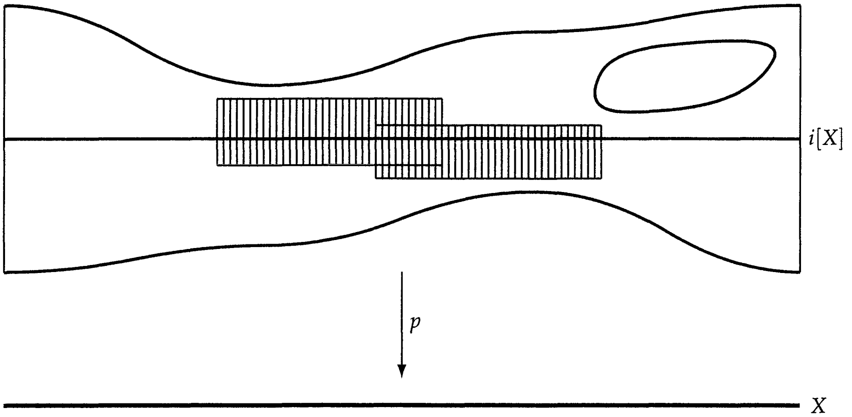
Thus, our proposed tangent structure appears to induce a $TOP(m)$–bundle.

The only real problem with such an approach is that the construction does not appear canonical, since the choice of neighborhoods/fibers U_x is random. It is important that each topological manifold have a *canonical* tangent bundle T_X. In order to achieve this, the main observation is that what really matters is what happens around x—whatever U_x has been chosen to be, the most important part of $T_X|_x$ is the immediate neighborhood of $x \in T_X|_x$ and how it relates to its neighboring fibers. Thus, one should consider, instead of the whole U_x's, just their *germs* at x. This idea was concretized in **J. Milnor**'s notion of a *microbundle,* which he introduced in ***Microbundles*** **[Mil64]**.

Microbundles and the topological tangent bundle. A k–**microbundle** ξ on X is a configuration

$$\xi\colon\ X \xrightarrow{i} E \xrightarrow{p} X\ ,$$

made of a topological space E (called the *total space*), together with two maps, $i\colon X \to E$ (called the *zero section*) and $p\colon E \to X$ (called the *projection*). These are required to satisfy two properties: (1) i must behave like a section, so we have $p \circ i = id$; and (2) $E \to X$ must be *locally trivial, i.e.,* for every $x \in X$, there is a neighborhood V_x of $i(x)$ in E such that $p|_{V_x}\colon V_x \to M$ looks like a projection $U \times \mathbb{R}^k \to U$. Notice that, as suggested in figure 4.33, nothing is required far from $i[X]$ or on the overlaps of the various local "charts": only parts of the fibers match.



4.33. A microbundle

You should think of a microbundle as a fiber bundle in which all that matters is what happens around the zero section, or as a vector bundle in which we are focused near the zero-section and all requirements of linearity have been dropped. Indeed, microbundles behave pretty much like vector bundles: they can be pulled-back, direct sums are defined, *etc.* We leave such amusements to the reader.

Two k-microbundles $\xi'\colon X \xrightarrow{i'} E' \xrightarrow{p'} X$ and $\xi''\colon X \xrightarrow{i''} E'' \xrightarrow{p''} X$ are called **isomorphic** if there are neighborhoods W' of $i'[X]$ in E' and W'' of $i''[X]$ in E'' and a homeomorphism $\varphi\colon W' \simeq W''$ fitting in the diagram

$$\begin{tikzcd}
X \arrow[r, "i'"'] \arrow[d, equal] & W' \arrow[r, "p'"'] \arrow[d, "\varphi"', "\simeq"] & X \arrow[d, equal] \\
X \arrow[r, "i''"] & W'' \arrow[r, "p''"] & X\,.
\end{tikzcd}$$

Of course, any actual bundle with fiber $\mathbb{R}^k$ is a k-microbundle, and two isomorphic fiber bundles are also isomorphic as microbundles.

Further, inside every microbundle actually lies a genuine bundle:

Kister–Mazur Theorem. *For every k-microbundle $X \xrightarrow{i} E \xrightarrow{p} X$ there is a neighborhood W of $i[X]$ in E such that $p|_W\colon W \to X$ is a locally-trivial fiber bundle with fiber $\mathbb{R}^k$ and zero-section i. The contained fiber bundle is unique up to isomorphism, and even up to isotopy.*

> ***Idea of proof.*** The crux of the argument is J. Kister's result that the space of topological *embeddings* $\mathbb{R}^k \to \mathbb{R}^k$ that fix the origin can be deformation-retracted to the space of *homeomorphisms* $\mathbb{R}^k \simeq \mathbb{R}^k$ that fix the origin. Thus, the partly-matching "charts" of a microbundle can be reduced and deformed

to get a small global genuine bundle. See **J. Kister**'s ***Microbundles are fibre bundles*** **[Kis64]**. □

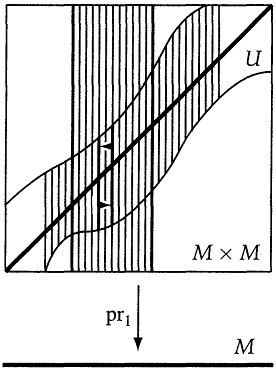
Thus, to every k–microbundle is associated a canonical fiber bundle with group $TOP(k)$ and fiber $\mathbb{R}^k$. Microbundles have the advantage that they are easy to describe. Thus, if we define a canonical *tangent microbundle* for a topological m–manifold, then we can pass it through Kister–Mazur to obtain a *canonical* tangent bundle, with structure group $TOP(m)$.

The **tangent microbundle** of a topological manifold X is defined simply as

$$X \xrightarrow{\Delta} X \times X \xrightarrow{\mathrm{pr}_1} X,$$

where Δ is the diagonal map $x \mapsto (x,x)$ and pr_1 is the projection $(x,y) \mapsto x$.

Close to the diagonal $\Delta[X]$, the fibers of pr_1 are just copies of neighborhoods of points in X. They are stacked next to each other according to their position in X: indeed, $z' \in \mathrm{pr}_1^{-1}[x']$ and $z'' \in \mathrm{pr}_1^{-1}[x'']$ are close to each other in $X \times X$ if and only if $\mathrm{pr}_2(z')$ and $\mathrm{pr}_2(z'')$ are close to each other in X. See also figure 4.34.



4.34. The tangent microbundle

We can then define the **topological tangent bundle**

$$T_X^{top}$$

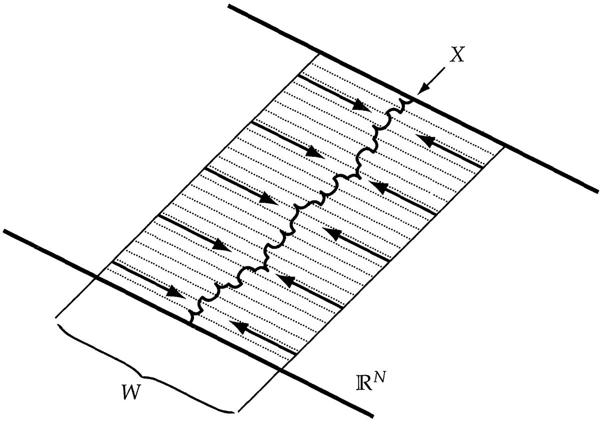
of the topological m–manifold X to be the $TOP(m)$–bundle contained inside the tangent microbundle of X. One can prove that, if X happens to be a smooth manifold and hence is endowed with a tangent vector bundle T_X, then T_X and T_X^{top} are isomorphic fiber bundles.

Using the topological tangent bundle for smoothing. Start with a topological m–manifold X. Embed X^m into some large[25] $\mathbb{R}^N$ and choose a neighborhood W of

---

[25] To build an embedding of a topological manifold in some $\mathbb{R}^N$, the easiest way is as follows: When X is compact, cover X^m by open sets $U_1, \ldots, U_n$, each homeomorphic to an open subset of $\mathbb{R}^m$ through embeddings $f_k \colon U_k \subset \mathbb{R}^m$; extend each f_k to a continuous maps $\widetilde{f}_k \colon M \to \mathbb{R}^m$, then gather all of them together to get an embedding $(\widetilde{f}_1, \ldots, \widetilde{f}_n) \colon M \to \mathbb{R}^{mn}$. In general, by dimension theory one can find an open covering $\{U_\alpha\}$ of X so that at every point of X no more than $m+1$ of the U_α's meet; eventually one gets an embedding in $\mathbb{R}^{m(m+1)}$.

X in $\mathbb{R}^N$ that *retracts* to X, *i.e.*, for which there is map $r\colon W \to X$ so that $r|_X = id$. See figure 4.35.



4.35. X embedded as a Euclidean neighborhood retract

Build the tangent bundle T_X^{top} of X and then use the retraction r to pull it back over the whole W; denote the total space of the result by $r^*T_X^{top}$.

$$\begin{tikzcd}
r^*T_X^{top} \arrow[r] \arrow[d] & T_X^{top} \arrow[d] \\
W \arrow[r, "r"] & X .
\end{tikzcd}$$

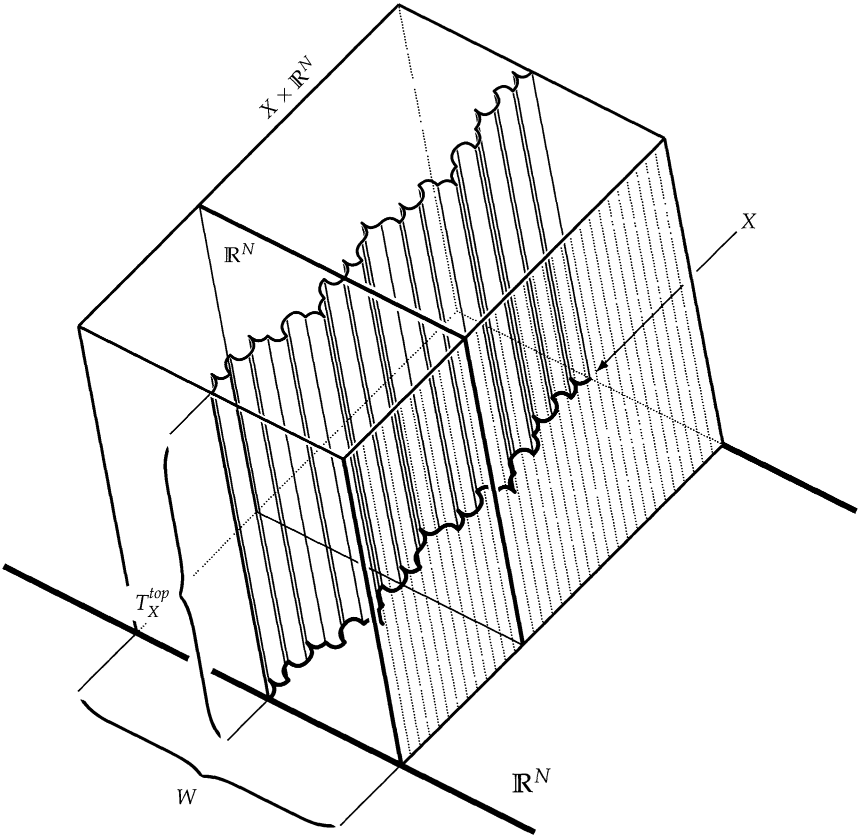
The total space of this pulled-back bundle can also be viewed as sitting on top of X, through the composition $r^*T_X^{top} \longrightarrow W \xrightarrow{r} X$; in reverse, X can be embedded in $r^*T_X^{top}$ through the composition of the inclusion $X \subset W$ with the zero section of the bundle $r^*T_X^{top} \to W$. We have the following remarkable property:

Lemma. *The space $r^*T_X^{top}$ is homeomorphic to $X \times \mathbb{R}^N$, with $X \subset r^*T_X^{top}$ corresponding to $X \times 0 \subset X \times \mathbb{R}^N$.*

Idea of proof. As a first approximation, think in terms of vector bundles: Assume that X was a smooth manifold, and T_X its tangent bundle. Smoothly embed X in $\mathbb{R}^N$, then choose a tubular neighborhood $W \approx N_{X/\mathbb{R}^N}$, which retracts to X through the bundle projection $r\colon N_{X/\mathbb{R}^N} \to X$. Then $r^*T_X \to X$ is isomorphic to the bundle $T_X \oplus N_{X/\mathbb{R}^N} = T_{\mathbb{R}^N}|_X = X \times \mathbb{R}^N$.

We can use a similar argument for our lemma if we start with a better W. Namely, we could start with an embedding of X into a large-enough $\mathbb{R}^N$, so that X admits a topological normal bundle $N_{X/\mathbb{R}^N}^{top}$ in $\mathbb{R}^N$, and take W to be the total space of $N_{X/\mathbb{R}^N}^{top}$ and r be its projection. □

Microbundle proof. *Without choosing a nice W and getting involved with topological normal bundles, one can also use a general argument, which is easiest to state in terms of microbundles: Consider T_X^{top} as the m-microbundle $X \xrightarrow{\Delta} X \times X \xrightarrow{\mathrm{pr}_1} X$. The pull-back over W has total space*



4.36. Smoothing $X \times \mathbb{R}^N$ by using the tangent bundle of X

$r^* T_X^{top} = \{(w, r(w), x) \in W \times X \times X\}$, projection $p\colon r^* T_X^{top} \to W$, $p(w, r(w), x) = w$, and zero-section $i\colon W \to r^* T_X^{top}$, $i(w) = (w, r(w), r(w))$. First, notice that the total space $r^* T_X^{top}$ is naturally homeomorphic to $X \times W$, by sending $(w, r(w), x)$ to (x, w). This homeomorphism $r^* T_X^{top} \simeq X \times W$ sends $i[X]$ to $\Delta[X]$. Then, by translating the inclusion $X \times W \subset X \times \mathbb{R}^N$ through the map $X \times \mathbb{R}^N \to X \times \mathbb{R}^N$: $(x, v) \mapsto (x, v - x)$, we obtain an embedding of $r^* T_X^{top}$ into $X \times \mathbb{R}^N$ that sends $i[X]$ to $X \times 0$. While this is a bit less than the statement of the lemma, all further developments could be slightly modified to be happy with this version.

Owing to this lemma, if we manage to make the total space $r^* T_X^{top}$ into a *smooth manifold*, then that means that we have endowed $X \times \mathbb{R}^N$ with a smooth structure. We would be a bit closer to smoothing X itself.

As mentioned, the tangent bundle T_X^{top} is a fiber bundle over X^m with fiber $\mathbb{R}^m$ and structure group $TOP(m)$. Denote now by

$$DIFF(m)$$

the group of diffeomorphisms $\varphi\colon \mathbb{R}^m \cong \mathbb{R}^m$ with $\varphi(0) = 0$. If we could reduce the structure group of T_X^{top} from $TOP(m)$ to $DIFF(m)$, then the pull-back $r^* T_X^{top}$ would be a bundle over W whose fibers are glued by smooth maps from $DIFF(m)$. Since W is open in $\mathbb{R}^N$, it is itself a smooth manifold. The base being smooth and the fibers being smoothly-matched, it follows that the total space of $r^*T_X^{top} \to W$ would itself be a smooth manifold. However, $r^*T_X^{top}$ is homeomorphic to $X \times \mathbb{R}^N$, and therefore the latter inherits a smooth structure.

Milnor's Smoothing Theorem. *Let X be a topological m–manifold. If its tangent bundle T_X^{top} admits a $DIFF(m)$–structure, then, for N big enough, $X \times \mathbb{R}^N$ must admit a smooth structure.* □

This was proved[26] by **J. Milnor**'s ***Microbundles*** **[Mil64]**, first announced in ***Topological manifolds and smooth manifolds*** **[Mil63c]**.

We postpone the investigation of the existence of $DIFF(m)$–structures on T_X^{top} for later. In the mean time, let us see how to get rid of the $\mathbb{R}^N$–factor, so that we may end up with a smooth structure on X itself.

Structures on products and products of structures. The following results are due to R. Kirby and L. Siebenmann. The first statement below is analogous to the Cairns–Hirsch theorem, which dealt with the PL case.

Product Structure Theorem. *Let X be a topological m–manifold, with m at least 5. If $X \times \mathbb{R}^N$ admits a smooth structure, then this structure must be isotopic to a product smooth structure on $X \times \mathbb{R}^N$, coming from a smoothing of X crossed with the standard smooth structure on $\mathbb{R}^N$.* □

Note that the isotopy conclusion above is stronger than a mere diffeomorphism between the two smooth structures on $X \times \mathbb{R}^N$.

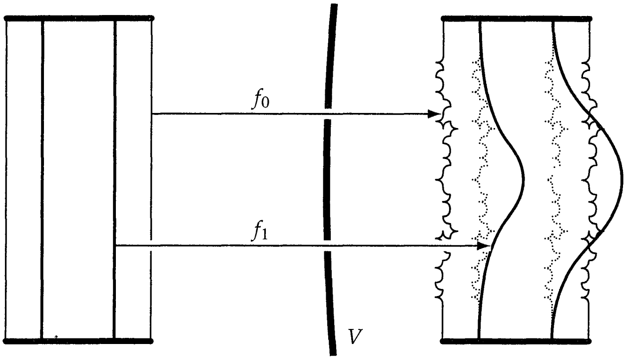
Isotopies of smoothings. *For convenience, call ξ the given smooth structure on $X \times \mathbb{R}^N$, by ζ the resulting smooth structure on X, and by $\zeta \times std$ the product structure on $X \times \mathbb{R}^N$. The existence of an isotopy between ξ and $\zeta \times std$ means two things: First, that ξ and $\zeta \times std$ are* concordant, *meaning that there exists a smooth structure on $(X \times \mathbb{R}^N) \times [0,1]$ that coincides with ξ near $(X \times \mathbb{R}^N) \times 0$ and with $\zeta \times std$ near $(X \times \mathbb{R}^N) \times 1$. Second, that there is a smooth map $h\colon (X \times \mathbb{R}^N) \times [0,1] \longrightarrow (X \times \mathbb{R}^N,\ \xi)$ so that each slice $h_t = h(\cdot, t)\colon X \times \mathbb{R}^N \times \{t\} \xrightarrow{\cong} X \times \mathbb{R}^N$ is a diffeomorphism onto $X \times \mathbb{R}^N$ smoothed by ξ. Thus, h_0 is the identity map from $(X \times [0,1],\ \xi)$ to itself, while h_1 is a diffeomorphism from $(X \times [0,1],\ \zeta \times std)$ to $(X \times [0,1],\ \xi)$, and h_t is the isotopy between them.*

Notice the dimensional restriction $m \geq 5$ that appears in the statement of the theorem. Its appearance is owing to the inevitable reliance of the proofs on the h–cobordism theorem (and its non-simply-connected cousin, the s–cobordism theorem). This is what prevents smoothing theory from fully applying to 4–dimensional manifolds.

Proving the product theorem. *The essential tool for proving the product structure theorem is the following handle-smoothing technique: Assume we have a smooth manifold V^m and a* smooth *embedding of a thickened sphere $\mathbb{S}^{k-1} \times [0,\varepsilon) \times \mathbb{R}^{m-k} \subset V^m$ (think of $\mathbb{S}^{k-1} \times [0,\varepsilon)$ as a collar on $\mathbb{S}^{k-1}$ in $\mathbb{D}^k$). Further assume that this smooth embedding can be extended as a* topological *embedding $f_0\colon \mathbb{D}^k \times \mathbb{R}^{m-k} \subset V^m$ of an open k–handle into V. We say that the handle f_0 can be* **smoothed** *in V if there is an isotopy f_t between f_0 and a map f_1 that restricts to a* smooth *embedding of the closed k-handle $f_1\colon \mathbb{D}^k \times \mathbb{D}^{m-k} \subset V^m$, and so that f_t fixes f_0 outside a compact neighborhood of $\mathbb{D}^k \times \mathbb{D}^{m-k}$. See figure 4.37 on the facing page.*

---

[26] Proved before the discovery of the Kister–Mazur theorem.



4.37. Smoothing a handle

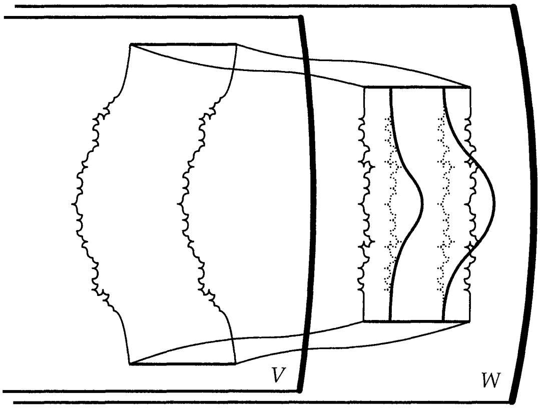
It turns out that the property of a handle f_0 to be smoothable is invariant under concordance:

Handle Smoothing Theorem. *Let f_0 be an open k-handle in V^m as above, and let W^m be a smooth manifold (of same dimension) containing V^m. Assume that there is an isotopy $F\colon \mathbb{D}^k \times \mathbb{R}^{m-k} \times [0,1] \longrightarrow W$ so that $F(\,\cdot\,,0) = f_0$, that F moves the attaching sphere smoothly, and that $F(\,\cdot\,,1)$ is a handle in W that can be smoothed in W. If $m \geq 5$, then f_0 itself can be smoothed inside V.* □

See also figure 4.38. This theorem is due to **R. Kirby** *and* **L. Siebenmann**, *see* ***Foundational essays on topological manifolds, smoothings, and triangulations*** [KS77]. *An essential ingredient for proving this handle smoothing theorem is, of course, the h-cobordism theorem. A consequence of it is the following stability property:*

Corollary. *Let $f_0\colon \mathbb{D}^k \times \mathbb{R}^{m-k} \subset V^m$ be some open k-handle as above and assume that $m \geq 5$. If the product-handle $f_0 \times id\colon \mathbb{D}^k \times \mathbb{R}^{m-k} \times \mathbb{R} \subset V \times \mathbb{R}$ can be smoothed inside $V \times \mathbb{R}$, then f_0 itself can be smoothed inside V.*

The proof of the product structure theorem then uses a chart-by-chart induction. Since each chart $\Phi\colon U \simeq U' \subset \mathbb{R}^m$ endows U with a smooth structure, this means that in each chart we can use handle decompositions, with handles that are then smoothed and made to fit on the overlaps of the charts.[27]



4.38. Handle smoothing theorem

---

[27] It is worth noticing how, even when investigating purely topological manifolds, it is the differential world that offers the local tools, which are then extended by careful patching and matching.

In conclusion, by combining Milnor's smoothing theorem with the Kirby–Siebenmann product structure theorem, we obtain:

Corollary. *Let X^m be a topological m–manifold, with m at least 5. If its topological tangent bundle T_X^{top} admits a $DIFF(m)$–structure, then X must itself admit a smooth structure.* □

In other words, we are able to "integrate" an infinitesimal differentiable structure on the tangent bundle to a differentiable structure on the manifold X itself.

It is now time to see what obstructions appear when trying to smooth the tangent bundle of a topological manifold X^m:

Smoothing bundles: the setting. The topological tangent bundle T_X^{top} is a bundle with fiber $\mathbb{R}^m$ and group $TOP(m)$; we wish to reduce its structure group to $DIFF(m)$. The method of choice will be obstruction theory, applied to classifying spaces. Thus, for a better understanding of the following, it is recommended to first read the earlier notes (on page 197 and on page 204).

At the outset, we should remark that the group $DIFF(m)$ of self-diffeomorphisms of $\mathbb{R}^m$ fixing the origin is homotopy-equivalent with the more familiar group $GL(m)$ of invertible matrices. Indeed, if $\varphi_1 \colon \mathbb{R}^m \cong \mathbb{R}^m$ is a diffeomorphism with $\varphi_1(0) = 0$, then the Alexander isotopy $\varphi_t(x) = \frac{1}{t}\varphi(tx)$ provides a deformation of φ_1 to $\varphi_0 = d\varphi_1|_0 \in GL(m)$, and thus contracts $DIFF(m)$ to $GL(m)$. This implies that a fiber bundle with structure group $DIFF(m)$ is nothing but a vector bundle. Therefore, to reduce the structure group of the tangent bundle T_X^{top} from $TOP(m)$ to $DIFF(m)$ means merely to organize T_X^{top} as a vector bundle.

The group $TOP(m)$ has a classifying space denoted by $\mathscr{B}TOP(m)$. As a consequence, the tangent bundle T_X^{top} is described by a classifying map

$$\tau \colon X \longrightarrow \mathscr{B}TOP(M)\ .$$

The group $DIFF(m)$ has a classifying space[28] $\mathscr{B}DIFF(m)$. The natural inclusion $DIFF(m) \subset TOP(m)$ induces a fibration

$$\mathscr{S} \colon \mathscr{B}DIFF(m) \longrightarrow \mathscr{B}TOP(m)$$

with fiber $TOP(m)/DIFF(m)$. Then endowing the tangent bundle T_X^{top} with a $DIFF(m)$–structure is the same as lifting the classifying map τ to a map $\tau^{sm} \colon X \to \mathscr{B}DIFF(m)$ that fits in

$$\begin{tikzcd}
X \arrow[r, "\tau^{sm}"'] \arrow[d, equal] & \mathscr{B}DIFF(m) \arrow[d, "\mathscr{S}"] \\
X \arrow[r, "\tau"] & \mathscr{B}TOP(m)\ .
\end{tikzcd}$$

We can pull the fibration $\mathscr{S} \colon \mathscr{B}DIFF(m) \longrightarrow \mathscr{B}TOP(m)$ back over X as

$$\begin{tikzcd}
\tau^*\mathscr{S} \arrow[r] \arrow[d] & \mathscr{B}DIFF(m) \arrow[d, "\mathscr{S}"] \\
X \arrow[r, "\tau"] & \mathscr{B}TOP(m)\ ,
\end{tikzcd}$$

and then smoothing T_X^{top} is equivalent to finding a section in this pulled-back fibration $\tau^*\mathscr{S}$. The fiber of $\tau^*\mathscr{S} \to X$ is $TOP(m)/DIFF(m)$.

---

[28] $\mathscr{B}DIFF(m)$ is the same (homotopy-equivalent) with $\mathscr{B}GL(m) = \mathscr{B}O(m) = \mathscr{G}r_m(\mathbb{R}^\infty)$.

Think of all this as a setting on which to use obstruction theory. We start with a random smoothing of the tangent bundle over the vertices of some cellular decomposition of X, viewed as a section of $\tau^*\mathscr{S}$ over the 0–skeleton of X. We then strive to extend this section cell-by-cell across all X. When extending from the k–skeleton of X across the $(k+1)$–skeleton of X, obstructions appear in

$$H^{k+1}\big(X;\ \pi_k(TOP(m)/DIFF(m))\big)\ .$$

Further, if a given section σ of $\tau^*\mathscr{S}$ over the k–skeleton is extendable across the $(k+1)$–skeleton, then the elements of

$$H^{k}\big(X;\ \pi_k(TOP(m)/DIFF(m))\big)$$

classify up to homotopy all other sections over the k–skeleton that are extendable across the $(k+1)$–skeleton and are homotopic to σ over the $(k-1)$–skeleton.

In terms of smooth structures on T_X^{top} or, equivalently when $m \geq 5$, in terms of the induced smooth structures on X^m, any homotopy of a section of $\tau^*\mathscr{S}$ corresponds to a *concordance* of smooth structures on X. Two smooth structures ζ' and ζ'' on X are called **concordant** if there is a smooth structure on $X \times [0,1]$ that is ζ' on $X \times 0$ and is ζ'' on $X \times 1$. Keep in mind that smooth structures can be diffeomorphic without being concordant; simple examples come from manifolds that do not admit orientation-reversing diffeomorphisms. (Furthermore, in high-dimensions concordance implies isotopy.)

Hence, obstruction theory can be used to clarify the existence and classification up to concordance of smooth structures on topological manifolds of dimension at least 5. Of course, in order to effectively put obstruction theory to work, we need to determine the homotopy groups of the fiber $TOP(m)/DIFF(m)$.

Smoothing bundles: computing the homotopy groups. This paragraph is rather dense and very sketchy. It can be safely skipped; the next paragraph starts on page 220.

High homotopy. Let us apply the above obstruction theory setting to the case of the sphere S^n. Since the topological manifold S^n admits smooth structures, no obstructions appear. Further, the only non-zero classifying cohomology group $H^k(X;\pi_k)$ appears when $k = n$, in which case we have

$$H^n\big(S^n;\ \pi_n(TOP(n)/DIFF(n))\big) \;=\; \pi_n\big(TOP(n)/DIFF(n)\big)\ .$$

Therefore, for $n \geq 5$ we have

$$\{\text{smooth structures on } S^n\} \;\approx\; \pi_n\big(TOP(n)/DIFF(n)\big)$$

(smooth structures considered up to concordance). That is to say:

Lemma. *When $n \geq 5$, we have $\pi_n\big(TOP(n)/DIFF(n)\big) = \Theta_n$, where Θ_n denotes the group of homotopy n–spheres.* □

The groups Θ_n have been presented in the end-notes of chapter 2 (page 97). They are defined as the set of all smooth homotopy n–spheres, considered up to h–cobordisms and with addition given by connected sums. We have seen that, when $n \geq 5$, the set Θ_n can be understood as the group of concordance classes of smooth structures on S^n; hence we could call Θ_n "the group of exotic n–spheres". These groups can be computed using surgery methods. All groups Θ_n are finite, and the first nontrivial one is $\Theta_7 = \mathbb{Z}_{28}$.

Further, after using stabilizations $TOP(k) \subset TOP(k+1)$, we are led to:

Theorem. *For all n and m with $5 \leq n \leq m+1$, we have*

$$\pi_n\big(TOP(m)/DIFF(m)\big) = \Theta_n . \qquad \square$$

This theorem follows from the delicate result that, for $n \leq m+1$ and $m \geq 4$, we have

$$\pi_n\big(TOP(m+1)/DIFF(m+1),\ \ TOP(m)/DIFF(m)\big) = 0 .$$

The cases when $m \geq 5$ were proved in **R. Kirby** *and* **L. Siebenmann***'s* ***Foundational essays on topological manifolds, smoothings, and triangulations*** [KS77]. *The cases when $m = 4$ were cleared in* **F. Quinn***'s* ***Ends of maps. III. Dimensions 4 and 5*** [Qui82] *for $n \leq 3$; in his* ***Isotopy of 4-manifolds*** [Qui86] *for $n = 5$; and in* **R. Lashof** *and* **L. Taylor***'s* ***Smoothing theory and Freedman's work on four-manifolds*** [LT84] *for $n = 3, 4$.*

Low homotopy. We now need to compute the low-dimensional homotopy groups of $TOP/DIFF$. For $n \geq 5$, we have used S^n to evaluate π_n. For $n \leq 4$, we can instead increase the dimension of S^n by thickening it to $S^n \times \mathbb{R}^k$ such that $n + k \geq 5$. Then, after using stabilizations, we have

$$\{\text{smooth structures on } S^n \times \mathbb{R}^k\} \approx \pi_n\big(TOP(m)/DIFF(m)\big)$$

for all $m \geq 4$. However, smooth structures on the open manifold $S^n \times \mathbb{R}^k$ are hard to approach directly. Instead, one considers smooth structures on $S^n \times \mathbb{T}^k$. On one hand, by climbing the universal cover $\mathbb{R}^k \to \mathbb{T}^k$, it is clear that each smooth structure on $S^n \times \mathbb{T}^k$ induces a smooth structure on $S^n \times \mathbb{R}^k$.

The fundamental fact is that, conversely, the smooth structures on $S^n \times \mathbb{R}^k$ correspond to smooth structures on $S^n \times \mathbb{T}^k$, more precisely to *homotopy smooth structures* on $S^n \times \mathbb{T}^k$. A **homotopy smooth structure** on a topological m-manifold X^m is a homotopy equivalence $X^m \sim V^m$ with some smooth m-manifold V^m (same dimension).

This converse is a consequence of the celebrated torus unfurling trick *of* **R. Kirby**, *which first appeared in* ***Stable homeomorphisms and the annulus conjecture*** [Kir69], *and was used in our context in* **R. Kirby** *and* **L. Siebenmann***'s* ***On the triangulation of manifolds and the Hauptvermutung*** [KS69] *(see also* ***Foundational essays…*** [KS77]*).*

When $n + k \leq 6$, the homotopy smooth structures on $S^n \times \mathbb{T}^k$ (thought of as smooth structures on $\mathbb{D}^n \times \mathbb{T}^k$ relative to the boundary) are known by surgery theory to be classified by the elements of $H^{3-n}(\mathbb{T}^k; \mathbb{Z}_2)$. Thus, for $n \geq 4$ there is only one homotopy smooth structure on $S^n \times \mathbb{T}^k$, the standard one. For $n \leq 2$, all structures are known to be finitely-covered by the standard one (and thus can be standardized after climbing a finite cover of $\mathbb{T}^k$). Finally, for $n = 3$ there is at most one structure that is not covered by the standard one. Therefore the conclusion is that, for all small n not 3, we have

$$\pi_n\big(TOP(m)/DIFF(m)\big) = 0$$

and, moreover, that $\pi_3\big(TOP(m)/DIFF(m)\big)$ has either one or two elements.

As was first noticed by L. Siebenmann, it turns out that π_3 cannot be trivial, and hence

$$\pi_3\big(TOP(m)/DIFF(m)\big) = \mathbb{Z}_2 .$$

If one accepts everything else that was claimed above, then, for proving this nontriviality of π_3, we need only exhibit *one* topological manifold of dimension less than 7 that does not admit any smooth structures.

In dimension 4, Freedman's E_8–manifold $\mathcal{M}_{E_8}$ is an example, as follows from Rokhlin's theorem. For dimensions higher than 4, we also have:

Lemma. *The topological manifold $\mathcal{M}_{E_8} \times S^k$ does not admit any smooth structures.*

> ***Proof.*** Assume that $\mathcal{M}_{E_8} \times S^k$ admits a smooth structure. Then, by writing $S^k = \mathbb{R}^k \cup \{\infty\}$, we obtain a smooth structure on $\mathcal{M}_{E_8} \times \mathbb{R}^k$. We apply the product structure theorem and deduce that $\mathcal{M}_{E_8} \times \mathbb{R}$ admits a smooth structure. Consider the projection map $\mathrm{pr}_2 \colon \mathcal{M}_{E_8} \times \mathbb{R} \to \mathbb{R}$. Then, since $\mathcal{M}_{E_8} \times \mathbb{R}$ is smooth, we can perturb pr_2 over $\mathcal{M}_{E_8} \times (0,\infty)$ so that it becomes smooth over $\mathcal{M}_{E_8} \times (\varepsilon,\infty)$ but remains unchanged on $\mathcal{M}_{E_8} \times (-\infty,0)$. Pick a positive regular value $c > \varepsilon$ of pr_2; then $\mathrm{pr}_2^{-1}[c]$ is a smooth 4–manifold. Since $\mathcal{M}_{E_8}$ has $w_2 = 0$, so must $\mathrm{pr}_2^{-1}[c]$. However, $\mathrm{pr}_2^{-1}[-1] = \mathcal{M}_{E_8}$, and hence the 5–manifold $\mathrm{pr}_2^{-1}[-1,c]$ is a cobordism between $\mathcal{M}_{E_8}$ and the smooth manifold $\mathrm{pr}_2^{-1}[c]$. Since signatures are cobordism-invariants, it follows that the smooth 4–manifold $\mathrm{pr}_2^{-1}[c]$ has signature 8, but $w_2 = 0$. This, of course, is forbidden by Rokhlin's theorem. □

The manifolds $\mathcal{M}_{E_8} \times S^k$ do not admit PL structures either. More important, notice the fundamental role that Rokhlin's theorem plays[29] in the nontriviality of $\pi_3\big(TOP(m)/DIFF(m)\big)$.

What was omitted. *A more detailed discussion would of course have taken into account the intermediate piecewise-linear level between smooth and topological, and infinite stabilizations.*

Stabilization means considering everything up to adding trivial bundles. This embeds $TOP(m)$ into $TOP(m+1)$ and in the limit yields the group $TOP = \varinjlim TOP(m)$, with its own classifying space $\mathscr{B}TOP$. Similarly, $DIFF(m)$ stabilizes to $DIFF = \varinjlim DIFF(m)$, with classifying space $\mathscr{B}DIFF$. The group of piecewise-linear self-homeomorphisms of $\mathbb{R}^m$ that fix 0 is denoted by $PL(m)$, stabilizing to PL and with classifying space $\mathscr{B}PL$. The inclusions $TOP \subset PL \subset DIFF$ lead to fibrations $\mathscr{B}PL \to \mathscr{B}TOP$ and $\mathscr{B}DIFF \to \mathscr{B}PL$, with corresponding fibers TOP/PL and $PL/DIFF$.

Between smooth and PL: The study of the smooth/PL gap was attacked by **S. Cairns** *in* ***The manifold smoothing problem*** **[Cai61]**. *Then* **R. Thom**'*s* ***Des variétés triangulées aux variétés différentiables*** **[Tho60]** *suggested that the smoothing problem should admit a setting in terms of obstruction theory. A natural simplex-by-simplex obstruction theory was developed by* **J. Munkres**' ***Obstructions to the smoothing of piecewise-differentiable homeomorphisms*** **[Mun59, Mun60b]** *(see also his* **[Mun64]** *and* **[Mun65]**). *A different obstruction theory was outlined in* **M. Hirsch**'*s* ***Obstruction theories for smoothing manifolds and maps*** **[Hir63]**, *and also proved a product structure theorem for the smooth/PL gap. Then appeared* **J. Milnor**'*s* ***Microbundles*** **[Mil64]**. *All this led to an obstruction theory based on the classifying spaces $\mathscr{B}DIFF$ and $\mathscr{B}PL$, developed by* **M. Hirsch** *and* **B. Mazur** *and eventually published in the volume* ***Smoothings of***

---

[29] Of course, R. Kirby and L. Siebenmann's result that $\pi_3(TOP/DIFF) = \mathbb{Z}_2$ was proved before M. Freedman built the fake 4–balls that are used in the construction of $\mathcal{M}_{E_8}$. Nonetheless, their examples also rest upon Rokhlin's theorem.

piecewise-linear manifolds **[HM74]**. *(A quick comparison of Munkres' and Hirsch–Mazur's approaches can be read from* **J. Munkres'** ***Concordance of differentiable structures—two approaches*** **[Mun67]**.*)*

The passing of the smooth/PL gap depends on the fiber DIFF/PL, which has homotopy groups

$$\pi_n(DIFF/PL) = 0 \quad \text{for all } n \leq 6, \qquad \text{and} \qquad \pi_n(DIFF/PL) = \Theta_n \quad \text{for all } n \geq 5 \ .$$

For proving the triviality of π_n in low-dimensions, the cases $n = 1, 2$ are boring, the case $n = 3$ follows from **J. Munkres**'s ***Differentiable isotopies on the 2–sphere*** **[Mun60a]** *and* **S. Smale**'s ***Diffeomorphisms of the 2–sphere*** **[Sma59]**. *The case $n = 4$ was proved by* **J. Cerf**'s *series of papers* ***La nullité de $\pi_0(\mathrm{Diff}\,\mathbb{S}^3)$*** **[Cer64]**, *later published in the volume* ***Sur les difféomorphismes de la sphère de dimension trois*** **($\Gamma_4 = 0$)** **[Cer68a]**. *The cases $n = 5, 6$ follow from the computations of Θ_n in* **M. Kervaire** *and* **J. Milnor**'s ***Groups of homotopy spheres*** **[KM63]**.

Thus, the first non-zero homotopy group of DIFF/PL is $\pi_7 = \mathbb{Z}_{28}$, coming from Milnor's exotic 7-spheres; geometrically, this first group corresponds to the existence of PL 8-manifolds that cannot be smoothed; an example is the 8-dimensional topological manifold $\mathcal{M}^8_{E_8}$ built by E_8-plumbing eight copies of $\mathbb{D}T_{\mathbb{S}^4}$ and capping with an 8-disk, see back on page 98. In general all $\mathcal{M}^{4k}_{E_8}$'s are PL and non-smoothable.

Between PL and topological: For the study of topological manifolds, some important steps along the way were **B. Mazur**'s ***On embeddings of spheres*** **[Maz59, Maz61]**, *followed by* **M. Brown**'s ***A proof of the generalized Schoenflies theorem*** **[Bro60]**, *then* **A. Černavskiĭ**'s ***Local contractibility of the group of homeomorphisms of a manifold*** **[Čer68b, Čer69]**. *Then came* **R. Kirby**'s *already mentioned torus unfurling trick, in* ***Stable homeomorphisms and the annulus conjecture*** **[Kir69]**, *which was then put to work together with L. Siebenmann.*

The passing of the PL/topological gap is governed by the fiber TOP/PL. The latter was shown to be an Eilenberg–Maclane $K(\mathbb{Z}_2; 3)$-space, that is to say,

$$\pi_3(TOP/PL) = \mathbb{Z}_2 \qquad \text{and} \qquad \pi_n(TOP/PL) = 0 \quad \text{for all } n \neq 3 \ .$$

This can be read from **R. Kirby** *and* **L. Siebenmann**'s ***Foundational essays on topological manifolds, smoothings, and triangulations*** **[KS77]**. *Examples of topological $(4 + k)$-manifolds that do not admit any PL structure are all $\mathcal{M}_{E_8} \times \mathbb{S}^k$ and $\mathcal{M}_{E_8} \times \mathbb{T}^k$. A recent exposition of the PL/topological gap can also be read from* **Y. Rudyak**'s ***Piecewise linear structures on topological manifolds*** **[Rud01]**.

The evaluation of the homotopy groups of TOP/PL rests upon the determination of all homotopy PL structures on $\mathbb{S}^n \times \mathbb{T}^k$ (viewed as structures on $\mathbb{D}^n \times \mathbb{T}^k$ relative to the boundary). These were cleared using surgery by **A. Casson**, *then by* **W-c. Hsiang** *and* **J. Shaneson**'s ***Fake tori, the annulus conjecture, and the conjectures of Kirby*** **[HS69]**, *based on the surgery techniques developed by* **C.T.C Wall**'s ***On homotopy tori and the annulus theorem*** **[Wal69b]** *(see also* ***Surgery on compact manifolds*** **[Wal70, Wal99,** ch 15**]***).*

Smoothing bundles: the Kirby–Siebenmann invariant. Reviewing the results outlined in the preceding paragraph, we can now state:

Theorem. *For every n and m with $5 \leq n \leq m + 1$, we have:*

$$\begin{aligned} \pi_n\big(TOP(m)/DIFF(m)\big) &= 0 \quad \text{for } 3 \neq n \leq 6 \\ \pi_3\big(TOP(m)/DIFF(m)\big) &= \mathbb{Z}_2 \\ \pi_n\big(TOP(m)/DIFF(m)\big) &= \Theta_n \quad \text{for } n \geq 5 \end{aligned}$$

where Θ_n is the group of homotopy n-spheres.[30] □

---

[30] For those who skipped the preceding paragraphs: The groups of homotopy spheres Θ_n have been presented in the end-notes of chapter 2 (page 97). They can be defined for $n \geq 5$ as the set of smooth

We can now apply obstruction theory to study smoothings of topological manifolds of dimension at least 5, *via* smoothings of their topological tangent bundles.

Since the first dimension with a nontrivial homotopy group is $n = 3$, it follows that the primary obstruction to endowing the topological tangent bundle of X with a $DIFF(m)$–structure appears as a class in $H^4(X;\mathbb{Z}_2)$. It is called the **Kirby–Siebenmann invariant** and is denoted by

$$\mathrm{ks}(X) \in H^4(M;\mathbb{Z}_2) .$$

The existence of this first obstruction rests upon Rokhlin's theorem. Further, the difference cocycles are elements of $H^3(X;\mathbb{Z}_2)$.

Past dimension 7, higher obstructions appear from $H^{n+1}(X;\Theta_n)$, the first one from $H^8(X;\mathbb{Z}_{28})$. Higher difference cocycles live in $H^n(X;\Theta_n)$, the first ones in $H^7(X;\mathbb{Z}_{28})$.

Bringing in the intermediate PL level, we should say: The Kirby–Siebenmann invariant $\mathrm{ks}(X) \in H^4(X;\mathbb{Z}_2)$ *is the complete obstruction to endowing a topological manifold* X^m *of dimension* $m \geq 5$ *with a PL structure. If such a structure exists, all other PL structures are classified up to concordance (and thus isotopy) by* $H^3(X;\mathbb{Z}_2)$. *The higher obstructions from* $H^{n+1}(X;\Theta_n)$ *govern the possibility of endowing a PL manifold* X^m *with a smooth structure and do not appear until* $m = 8$. *Notice also that every PL* 7*–manifold admits exactly* 28 *distinct smooth structures, up to concordance.*

Since $\mathbb{Z}_2$ and all the Θ_n's are finite, a consequence is that any topological manifold of dimension *not* 4 admits at most *finitely-many* distinct smooth structures.[31]

Another consequence of the theory is that, for all $m \geq 5$, any topological manifold homeomorphic to $\mathbb{R}^m$ admits a unique smooth structure. Since the cases $m \leq 3$ are similar, this leaves $\mathbb{R}^4$ as the only possible support of exotic structures.

Conclusion. If the Kirby-Siebenmann invariant $\mathrm{ks}(X)$ vanishes and $m \leq 7$, then the tangent bundle of X^m admits a $DIFF(m)$–structure. If moreover $m \geq 5$, then this bundle structure can be integrated to a smooth structure on X itself. For example, all simply-connected topological 5–manifolds admit smooth structures.[32]

Moreover, if X admits some smooth structure, then all other smooth structures on X are classified (up to concordance/isotopy, *via* difference cocycles) by the elements of $H^3(X;\mathbb{Z}_2)$. Starting with dimension 8, beside $\mathrm{ks}(X)$ appear higher obstructions to the existence of smooth structures, living in the groups $H^{n+1}(X;\Theta_n)$.

***The case of dimension* 4.** The Kirby–Siebenmann invariant can certainly still be defined in dimension 4. However, lacking the power of the (smooth) h–cobordism theorem behind it, it mainly has negation power.

For a topological 4–manifold M, the Kirby–Siebenmann invariant

$$\mathrm{ks}(M) \in H^4(M;\mathbb{Z}_2)$$

---

structures on S^n considered up to concordance, with addition defined by connected sums; all groups Θ_n are finite, and the first nontrivial one is $\Theta_7 = \mathbb{Z}_{28}$.

[31] The cases of dimension 2 and 3 being handled, of course, separately.

[32] Since $H^4(X^5;\mathbb{Z}_2) = H_1(X^5,\mathbb{Z}) = 0$.

is simply a $\mathbb{Z}_2$–valued invariant: it is either 0 or 1. Its value is strongly related to Rokhlin's theorem (and its generalizations). Specifically, $\mathrm{ks}(M)$ detects whether a smooth structure on M is prohibited by Rokhlin's or not.

Evaluating the Kirby–Siebenmann invariant. Let M be any topological 4–manifold with no 2–torsion in $H_1(M;\mathbb{Z})$ and with *even* intersection form Q_M (such a manifold can safely be called a "spin manifold"). We have:

$$\mathrm{ks}(M) = \tfrac{1}{8}\operatorname{sign} M \pmod 2 .$$

In particular, $\mathrm{ks}(\mathcal{M}_{E_8}) = 1$.

More generally, regardless of the parity of Q_M, if a characteristic element of M can be represented by a topologically embedded *sphere* Σ, then we have

$$\mathrm{ks}(M) = \tfrac{1}{8}\big(\operatorname{sign} M - \Sigma\cdot\Sigma\big) \pmod 2 .$$

This is related to the Kervaire–Milnor generalization of Rokhlin's theorem.

Finally, *via* the Freedman–Kirby generalization of Rokhlin's theorem, we have, for every topological 4–manifold M with an embedded characteristic surface Σ,

$$\mathrm{ks}(M) = \tfrac{1}{8}\big(\operatorname{sign} M - \Sigma\cdot\Sigma\big) + \mathrm{Arf}(M,\Sigma) \pmod 2 ,$$

where $\mathrm{Arf}(M,\Sigma)$ is a $\mathbb{Z}_2$–invariant that measures the obstruction to representing Σ by a sphere, and depends only on the homology class of Σ. The Freedman–Kirby theorem will be discussed and proved in the end-notes of chapter 11 (page 502); it is readable anytime.

When Kirby–Siebenmann vanishes. If M admits a smooth structure, then $\mathrm{ks}(M) = 0$. The converse is false: if $\mathrm{ks}(M) = 0$, then M might still not admit any smooth structures. Such examples were uncovered starting with Donaldson's work[33] and they are not rare. Nonetheless, if $\mathrm{ks}(M) = 0$, then the 5–manifolds $M \times \mathbb{R}$ or $M \times S^1$ do admit smooth structures. Further, without increasing dimension, if $\mathrm{ks}(M) = 0$, then for m big enough the stabilization $M \# m\, S^2 \times S^2$ must admit a smooth structure.

On the other hand, it was proved that all *open* 4–manifolds can be smoothed. In particular, any closed 4–manifold M can be endowed with a smooth structure off a point.

In the case when $\mathrm{ks}(M) = 0$, then, since $M \# m\, S^2 \times S^2$ can be smoothed, such a smoothing-off-points for M can be made in a controlled fashion:

Theorem (*F. Quinn*). *If M is a topological 4–manifold with $\mathrm{ks}(M) = 0$, then there is a finite set of points $p_1, \dots, p_m$ in M and a smooth structure on*

$$M \setminus \{p_1, \dots, p_m\} ,$$

such that, for each k, on one hand there is a neighborhood U_k of p_k in M, and on the other hand there is a self-homeomorphism $\varphi_k \colon S^2 \times S^2 \simeq S^2 \times S^2$ (isotopic to the identity), a neighborhood U_k' of $h_k[S^2 \vee S^2]$ in $S^2 \times S^2$; and we have a diffeomorphism

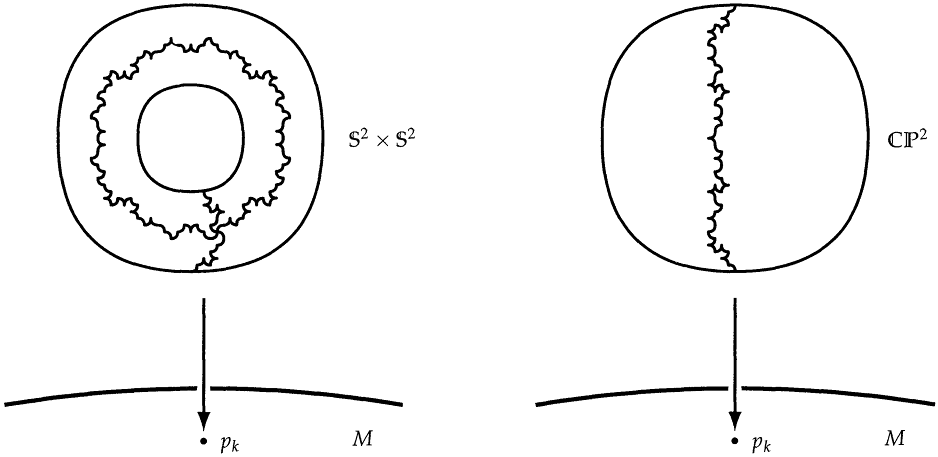
$$U_k \setminus p_k \cong U_k' \setminus \varphi_k[S^2 \vee S^2] .$$

---

[33] See ahead section 5.3 (page 243).

In other words, the complement of each p_k *is locally smoothed like the complement of a displacement of* $\mathbb{S}^2 \vee \mathbb{S}^2$ *in* $\mathbb{S}^2 \times \mathbb{S}^2$. □

See the left side of figure 4.39. This result was proved in **F. Quinn**'s ***Smooth structures on* 4*–manifolds*** [Qui84] and can also be read from **M. Freedman** and **F. Quinn**'s ***Topology of* 4*–manifolds*** [FQ90].



4.39. Almost-smoothing a 4–manifold with $\mathrm{ks}(M) = 0$

Since $\mathbb{S}^2 \times \mathbb{S}^2 \# \mathbb{CP}^2 = \# 2\, \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$, the theorem can immediately be restated by instead using displacements of $\mathbb{CP}^1$ in $\mathbb{CP}^2$ and diffeomorphisms

$$U_k \setminus p_k \cong U'_k \setminus \varphi_k[\mathbb{CP}^1]$$

(some of which which could reverse orientations). See the right side of figure 4.39, and also think in analogy with blow-ups of complex manifolds.[34]

A fundamental remark to be made in this context is that both $\mathbb{S}^2 \times \mathbb{S}^2 \setminus \varphi_k[\mathbb{S}^2 \vee \mathbb{S}^2]$ and $\mathbb{CP}^2 \setminus \varphi_k[\mathbb{CP}^1]$ are open smooth 4–manifolds that are homeomorphic to $\mathbb{R}^4$. This implies that, if M has $\mathrm{ks}(M) = 0$ but is not smoothable, then these open manifolds must exhibit non-standard smooth structures on $\mathbb{R}^4$. In other words, they must be **exotic** $\mathbb{R}^4$**'s.** This, in part, explains why the discovery of exotic $\mathbb{R}^4$'s had to wait for Donaldson's work.[35] Exotic $\mathbb{R}^4$'s will be discussed in section 5.4 (page 250) ahead.

When Kirby–Siebenmann does not vanish. If $\mathrm{ks}(M) = 1$, then M does not admit any smooth structures. If $\mathrm{ks}(M) = 1$, then stabilizations do not help: $\mathrm{ks}(M \# m\, \mathbb{S}^2 \times \mathbb{S}^2)$ will still be 1, and all the $M \# m\, \mathbb{S}^2 \times \mathbb{S}^2$'s will be non-smoothable. Indeed, the Kirby–Siebenmann invariant is nicely additive:

$$\mathrm{ks}(M \cup_\partial N) = \mathrm{ks}(M) + \mathrm{ks}(N) .$$

---

[34] Blow-ups are described in section 7.1 (page 286) ahead.

[35] Of course, it also had to wait for A. Casson's and M. Freedman's work. Nonetheless, one can still ask whether the existence of exotic $\mathbb{R}^4$'s can be obtained as a consequence of Rokhlin's theorem while avoiding Donaldson's theory or equivalents. No.

In particular $\operatorname{ks}(M \# N) = \operatorname{ks}(M) + \operatorname{ks}(N)$, and so, if $\operatorname{ks}(M) = 1$, then $\operatorname{ks}(M \# m S^2 \times S^2) = 1$. Another important property to note is that the Kirby–Siebenmann invariant is unchanged by cobordisms.[36]

The invariant ks misses most of the wildness of dimension 4: for example, the Kirby–Siebenmann invariant of $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8}$ vanishes; but the latter has intersection form $E_8 \oplus E_8$, which is excluded from the smooth realm by the results of Donaldson: Kirby–Siebenmann's does not see what Rokhlin's does not exclude.

Note: The Rokhlin invariant of 3–manifolds

The Rokhlin theorem has major consequences beyond dimension 4. As we have seen in the preceding note (starting on page 207), in high-dimensions it is fundamentally implied in the non-existence of smooth structures on topological manifolds. In dimension 3, the Rokhlin theorem permits the definition of invariants for 3–manifolds, which are the topic of this note. The invariants for 3–manifolds are a $\mathbb{Z}_2$–invariant

$$\rho(\Sigma) \in \mathbb{Z}_2$$

for homology 3–spheres Σ, and a $\mathbb{Z}_{16}$–invariant

$$\mu(N) \in \mathbb{Z}_{16}$$

for 3–manifolds N endowed with spin structures.

Preparation: additivity of signatures. We have already seen that, if we connect-sum two 4–manifolds M and N, then we have $Q_{M \# N} = Q_M \oplus Q_N$, and as a consequence

$$\operatorname{sign}(M \# N) = \operatorname{sign} M + \operatorname{sign} N .$$

Intersection forms can also be defined for 4–manifolds with non-empty boundary, but they will not be unimodular unless the boundary is a homology sphere.[37] Then the additivity properties above are easy to prove for two manifolds M and N whose boundaries are a same homology sphere with *opposite* orientations: if we glue M and N along their boundaries, then $Q_{M \cup_\partial N} = Q_M \oplus Q_N$ and hence $\operatorname{sign}(M \cup_\partial N) = \operatorname{sign} M + \operatorname{sign} N$.

Examples. *For example, the* 4*–manifold*[38] P_{E_8} *has intersection form* E_8 *and signature* 8*. The manifold* $P_{E_8} \cup_{\Sigma_P} \overline{P}_{E_8}$ *is a closed* 4*–manifold with intersection form* $E_8 \oplus -E_8 \approx \oplus 8H$ *and signature* 0*. Because of signature-vanishing, we expect* $P_{E_8} \cup_{\Sigma_P} \overline{P}_{E_8}$ *to bound a* 5*–manifold, and indeed, it is the boundary of* $P_{E_8} \times [0,1]$*, as in figure 4.40 on the facing page. It turns out that* $P_{E_8} \cup_{\Sigma_P} \overline{P}_{E_8}$ *is none other than* $\# 8 S^2 \times S^2$*. (Notice that, since* Σ_P *does not have an orientation-reversing self-diffeomorphism, a manifold like* $P_{E_8} \cup_{\Sigma_P} P_{E_8}$ *does not exist.*[39]*)*

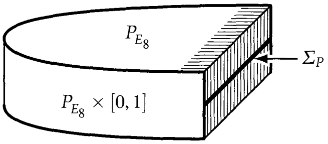
---

36. In fact, the topological cobordism group Ω_4^{top} of oriented topological 4–manifolds is $\Omega_4^{top} = \mathbb{Z} \oplus \mathbb{Z}_2$, with isomorphism given by $M \mapsto (\operatorname{sign} M, \operatorname{ks}(M))$. Cobordisms groups will be discussed in the note on page 227 ahead.

37. This will be fully proved in the end-notes of the next chapter (page 261).

38. Recall that P_{E_8} denotes the E_8–plumbing and is bounded by the Poincaré homology sphere Σ_P; see section 2.3 (page 86).

39. A roundabout argument: $P_{E_8} \cup_{\Sigma_P} P_{E_8}$ would be a *smooth* 4–manifold with definite intersection form $E_8 \oplus E_8$. However, that is excluded by Donaldson's theorem (see section 5.3, on page 243 ahead). Thus, this 4–manifold does not exist, and therefore Σ_P cannot admit an orientation-reversing self-diffeomorphism.



4.40. $P_{E_8} \cup_{\Sigma_P} \overline{P}_{E_8}$ is the boundary of $P_{E_8} \times [0,1]$

If two 4–manifolds have boundaries that are *not* homology spheres, then the additivity of the intersection forms ceases to hold. Nonetheless, signatures are still additive:

Novikov's Additivity Theorem. *Let M and N be two 4–manifolds with non-empty boundaries. Assume that their boundary 3–manifolds ∂M and ∂N admit an orientation-reversing diffeomorphism $\partial M \cong \overline{\partial N}$. Then the closed manifold $M \cup_\partial N$, built by identifying the boundaries ∂M and ∂N, has signature*

$$\operatorname{sign}(M \cup_\partial N) = \operatorname{sign} M + \operatorname{sign} N .$$

Outline of proof. Denote by Y^3 the (unoriented) boundaries of M and N as well as the resulting 3–submanifold in $M \cup_\partial N$. Take a random element $\alpha \in H_2(M \cup_\partial N)$, represented as surface transverse to Y. Then the intersection $\alpha \cap Y$ is a 1–cycle in Y.

On one hand, if $\alpha \cap Y$ is non-trivial in $H_1(Y^3;\mathbb{Z})$, then it admits a dual class $\beta \in H_2(Y;\mathbb{Q})$. (Notice that we must use rational coefficients, but that is no problem: signatures were defined by diagonalization over a field.) The class β can be included as a class in $M \cup_\partial N$. Since β in $M \cup_\partial N$ can be pushed off itself by using some nowhere-zero vector field normal to Y in $M \cup_\partial N$, it follows that $\beta \cdot \beta = 0$ in $M \cup_\partial N$. Therefore, the span of α and β in $H_2(M \cup_\partial N;\mathbb{Q})$ has intersection form

$$Q|_{\alpha\beta} = \begin{bmatrix} * & 1 \\ 1 & 0 \end{bmatrix} ,$$

whose signature is zero and thus does not contribute to $\operatorname{sign}(M \cup_\partial N)$.

On the other hand, if $\alpha \cap Y$ is homologically-trivial, then one shows, using a Mayer–Vietoris argument, that α must in fact be a sum $\alpha = \alpha_M + \alpha_N$ of classes from M and N. Therefore the contribution of α to the signature of $M \cup_\partial N$ is caught in $\operatorname{sign} M$ and $\operatorname{sign} N$. □

The complete proof can be found in **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89**, ch II**]**.

If two 4–manifolds are glued on only parts of their boundaries, then the additivity of the signature ceases to hold. Nonetheless, there is a well-determined correction term, see **C.T.C. Wall**'s ***Non-additivity of the signature*** **[Wal69a]**.

The Rokhlin invariant of homology 3–spheres. On 3–manifolds spin structures can be defined in the same way as on 4–manifolds. Since every 3–manifold N is parallelizable (*i.e.*, T_N is a trivial bundle), it admits spin structures. As in dimension 4, the group $H^1(N;\mathbb{Z}_2)$ acts transitively on the set of spin structures. In particular, if $H^1(N;\mathbb{Z}_2) = 0$, then N admits exactly *one* spin structure. Moreover, every spin 3–manifold N bounds a (smooth) spin 4–manifold M with the spin structure of M restricting to the chosen spin structure[40] of N.

Let Σ^3 be a homology 3–sphere. Let M be a smooth spin 4–manifold bounded by Σ. Being spin, the manifold M must have an *even* intersection form. Since Σ is a homology 3–sphere, the intersection form of M must be unimodular. Thus, using van der Blij's lemma, its signature must be a multiple of 8:

$$\operatorname{sign} M = 0 \pmod 8$$

(from the same algebraic argument[41] as for closed 4–manifolds). In other words, the residue of $\operatorname{sign} M$ modulo 16 is either 0 or 8.

We can then define the **Rokhlin invariant** of Σ by

$$\rho(\Sigma) = \tfrac{1}{8} \operatorname{sign} M \pmod 2 .$$

Due to Rokhlin's theorem, this is a well-defined invariant of Σ, which does not depend on the choice of the bounded 4–manifold M. Indeed, if Σ also bounds another spin 4–manifold M', then M and $\overline{M'}$ can be glued along Σ yielding a closed *spin* 4–manifold $M \cup_\Sigma \overline{M'}$, which must have

$$\operatorname{sign}\big(M \cup_\Sigma \overline{M'}\big) = 0 \pmod{16} ,$$

and thus $\operatorname{sign} M - \operatorname{sign} M' = 0 \pmod{16}$.

For example, since it bounds P_{E_8} whose signature is 8, the Poincaré homology 3–sphere Σ_P must have $\rho(\Sigma_P) = 1$.

The Rokhlin invariant of $\mathbb{Z}_2$–homology 3–spheres. Assume now that the 3–manifold N is a $\mathbb{Z}_2$–homology sphere, *i.e.*, a closed 3–manifold with

$$H^1(N; \mathbb{Z}_2) = 0 .$$

Then N admits a unique spin structure. Pick some smooth spin 4–manifold M that is bounded by N, with compatible spin structures. The intersection form of M is still even, but no longer unimodular, and so the best we can do is define the **Rokhlin invariant** (or **μ–invariant**) of N by

$$\mu(N) = \operatorname{sign} M \pmod{16} .$$

A similar reasoning as above shows that it is well-defined, independent of M.

The Rokhlin invariant of spin 3–manifolds. Finally, if N is just a random closed 3–manifold, then we can choose a spin structure $\mathfrak{s}$ on N, find a spin 4–manifold M that is spin-bounded by N, and define the invariant

$$\mu(N) = \operatorname{sign} M \pmod{16} .$$

This is an invariant that depends on the chosen spin structure $\mathfrak{s}$.

Two easy properties of the Rokhlin invariants, in any of the above versions, are:

$$\mu(\overline{N}) = -\mu(N) \qquad \text{and} \qquad \mu(N' \# N'') = \mu(N') + \mu(N'') .$$

---

[40] In the language of the next note (cobordism groups; page 227), we are saying that $\Omega_3^{Spin} = 0$.

[41] For the proof of van der Blij's lemma, see the end-notes of the next chapter (page 263).

Vice-versa: proving Rokhlin's theorem from μ-invariants. The reason why the μ-invariant of a spin 3-manifold is a well-defined invariant modulo 16, rather than modulo 8, is Rokhlin's theorem. Surprisingly, one can also go in reverse: If one proves by other means that the μ-invariant is well-defined modulo 16, then from this fact one can deduce Rokhlin's theorem for 4-manifolds.

This brief and elegant proof of Rokhlin's theorem can be discovered hidden as an appendix to **R. Kirby** and **P. Melvin**'s ***The 3-manifold invariants of Witten and Reshetikhin–Turaev for $\mathfrak{sl}(2,\mathbb{C})$*** **[KM91]**. Specifically, one starts with a presentation of the 3-manifold as a Kirby link diagram, then defines the μ-invariant in terms of that diagram and proves that it well-defined by using only Kirby calculus.[42]

References. The Rokhlin invariant first appeared, in a more general setting, in **J. Eells** and **N. Kuiper**'s ***An invariant for certain smooth manifolds*** **[EK62]**. Some early properties are explored in **F. Hirzebruch, W. Neumann** and **S. Koh**'s ***Differentiable manifolds and quadratic forms*** [HNK71].

The Rokhlin invariant can be refined into the much more powerful *Casson invariant* of homology 3-spheres, to the exposition of which is devoted **S. Akbulut** and **J. McCarthy**'s ***Casson's invariant for oriented homology 3-spheres*** **[AM90]**. This was extended by **K. Walker** to an invariant of rational homology 3-spheres in ***An extension of Casson's invariant*** [Wal92], and then finally to general 3-manifolds in **C. Lescop**'s ***Global surgery formula for the Casson–Walker invariant*** **[Les96]**. A recent survey of such invariants is **N. Saveliev**'s ***Invariants for homology 3-spheres*** **[Sav02]**. In a different direction, the Casson invariant admits a gauge-theoretic interpretation in terms of Donaldson's instantons, as was noticed by **C. Taubes**' ***Casson's invariant and gauge theory*** **[Tau90]**, and, even further, it is the Euler characteristic of an instanton-based homology theory built in **A. Floer**'s ***An instanton-invariant for 3-manifolds*** **[Flo88]**. However, all this is beyond the scope of the present volume.

Note: Cobordism groups

If we consider two m-manifolds as equivalent whenever there is a cobordism between them, then we separate manifolds into **cobordism classes**, and these can be organized as an Abelian group.

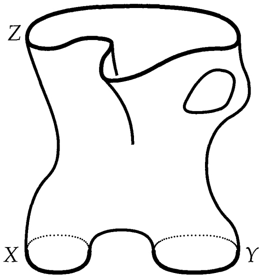
Oriented cobordism group. Consider the set of all oriented m-manifolds, together with the empty manifold $\varnothing$. Think of X^m and Y^m as equivalent if and only if they are cobordant, *i.e.*, if there is an oriented manifold W^{m+1} such that $\partial W = \overline{X} \cup Y$. The equivalence classes make up an Abelian group

$$\Omega_m^{SO}$$

called the **oriented cobordism group** in dimension m. Its addition comes from disjoint unions, $[X] + [Y] = [X \cup Y]$, as suggested in figure 4.41 on the next page.

---

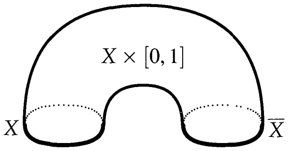
[42] A quick overview of Kirby calculus was made in the end-notes of chapter 2 (page 91).



4.41. Cobordisms: $[X] + [Y] = [Z]$ in Ω_m^{SO}

The identity element in Ω_m^{SO} is given by $0 = [\varnothing]$. Any bounding m-manifold represents 0, and thus in particular the identity can also be represented by the m-sphere $\mathbb{S}^m$—since $\mathbb{S}^m$ bounds $\mathbb{D}^{m+1}$, we have $[\mathbb{S}^m] = [\varnothing]$.

The inverse in Ω_m^{SO} is given by reversing orientations: we have $-[X] = [\overline{X}]$, as argued in figure 4.42.

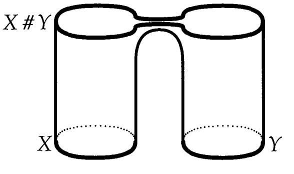


4.42. Cobordisms: $[X] + [\overline{X}] = 0$ in Ω_m^{SO}

It is worth noticing that $X \sqcup Y$ is always cobordant to $X \# Y$. This can be seen, for example, by using the boundary sum[43] $(X \times [0,1]) \natural (Y \times [0,1])$ as in figure 4.43. Thus, connected sum corresponds to addition in Ω_m^{SO}:

$$[X] + [Y] = [X \# Y] \; .$$

The diffeomorphisms $X \# \mathbb{S}^m = X$ reflect as $[X] + 0 = [X]$.



4.43. Cobordisms: $[X] + [Y] = [X \# Y]$ in Ω_m^{SO}

Cobordism ring. *Further, all the groups Ω_m^{SO} can in fact be put together to make up the* oriented cobordism ring Ω_*^{SO}, *with multiplication given by $[X] \cdot [Y] = [X \times Y]$, and unit the element $[+point] \in \Omega_0^{SO}$.*

As examples, it is easy to see that $\Omega_0^{SO} = \mathbb{Z}$, $\Omega_1^{SO} = 0$ and $\Omega_2^{SO} = 0$. It is a nontrivial result that $\Omega_3^{SO} = 0$. We have already mentioned that a 4-manifold is

---

[43] Boundary sums were recalled back on page 13.

the boundary of some oriented 5–manifold if and only if its signature is zero. It follows that

$$\Omega_4^{SO} = \mathbb{Z} ,$$

with isomorphism given by $[M] \mapsto \operatorname{sign} Q_M$. A generator of Ω_4^{SO} is $\mathbb{CP}^2$.

More cobordism groups are collected in table VI. The generator of Ω_5^{SO} *is the manifold* $\mathcal{Y}^5$ *described by the equation*[44] $(x_0 + x_1 + x_2)(y_0 + \cdots + y_4) = \varepsilon$ *in* $\mathbb{RP}^2 \times \mathbb{RP}^4$. *The generators of* Ω_8^{SO} *are* $\mathbb{CP}^2 \times \mathbb{CP}^2$ *and* $\mathbb{CP}^4$. *The generators of* Ω_9^{SO} *are* $\mathcal{Y}^5 \times \mathbb{CP}^2$ *and* $\mathcal{Y}^9$*, the latter being described by the equation* $(x_0 + x_1 + x_2)(y_0 + \cdots + y_8) = \varepsilon$ *in* $\mathbb{RP}^2 \times \mathbb{RP}^8$. *The generator of* Ω_{10}^{SO} *is* $\mathcal{Y}^5 \times \mathcal{Y}^5$. *The generator of* Ω_{11}^{SO} *is* $\mathcal{Y}^{11}$*, given by the equation* $(x_0 + \cdots + x_4)(y_0 + \cdots + y_8) = \varepsilon$ *in* $\mathbb{RP}^4 \times \mathbb{RP}^8$. *Keep in mind that Cartesian product organizes* $\oplus\, \Omega_k^{SO}$ *as a graded ring.*

VI. Oriented cobordism groups

m	0	1	2	3	4	5	6	7	8	9	10	11
Ω_m^{SO}	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	$\mathbb{Z}_2$	0	0	$\mathbb{Z} \oplus \mathbb{Z}$	$\mathbb{Z}_2 \oplus \mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$

Spin cobordism groups. The "*SO*" from the notation Ω_m^{SO} comes from the fact that an orientation of X^m is the same as a reduction of the structure group of T_X to $SO(m)$. The oriented cobordism group is not the only cobordism group—indeed, one can define a cobordism theory for most types of structure on manifolds.

In particular, the **spin cobordism group**

$$\Omega_m^{Spin}$$

is defined by starting with m–manifolds endowed with spin structures and considering X and Y as equivalent if and only if together they make up the boundary of a spin $(m+1)$–manifold W, with the spin structures on X and Y induced from the one on W.

In low-dimensions[45] we have $\Omega_1^{Spin} = \mathbb{Z}_2$, $\Omega_2^{Spin} = \mathbb{Z}_2$, and $\Omega_3^{Spin} = 0$. In dimension 4, we have

$$\Omega_4^{Spin} = \mathbb{Z} ,$$

with isomorphism given by $[M] \mapsto \frac{1}{16} \operatorname{sign} Q_M$ (always an integer, by Rokhlin's theorem). The generator is the $K3$ surface.

More groups are collected in table VII. The generator of Ω_4^{Spin} *is* $K3$. *The generators of* Ω_8^{Spin} *are* $\mathbb{HP}^2$ *and an* 8*-manifold* $\mathcal{K}^8$ *such that* $\#4\,\mathcal{K}$ *is spin cobordant to* $K3 \times K3$.

VII. Spin cobordism groups

m	0	1	2	3	4	5	6	7	8
Ω_m^{Spin}	$\mathbb{Z}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	0	$\mathbb{Z}$	0	0	0	$\mathbb{Z} \oplus \mathbb{Z}$

---

[44] The role of ε in the equation is merely to eliminate the singularities that would appear for $= 0$.

[45] Defining spin structures for 1– and 2–manifolds requires first stabilization (because $\pi_1 SO(n)$ begins to be $\mathbb{Z}_2$ only for $n \geq 3$). Thus, for 1–manifolds C we will look at trivializations of $T_C \oplus \underline{\mathbb{R}}^2$, while for surfaces S, we look at $T_S \oplus \underline{\mathbb{R}}$. These low-dimensional spin structures and their cobordisms will be discussed in more detail in the end-notes of chapter 11 (page 521).

Uses. The application of such cobordism results usually follows this pattern: In order to prove a general statement about manifolds, first prove that it is invariant under cobordisms, then prove that the statement holds on the generators.

For example, the signature sign Q_M is an oriented-cobordism invariant, and such an argument is used in **M. Freedman** and **R. Kirby**'s ***A geometric proof of Rochlin's theorem*** **[FK78]** to prove Rokhlin's theorem; we will present two versions of that argument in the end-notes of chapter 11 (page 502 and page 521).

The most famous results first proved *via* cobordism arguments are Hirzebruch's signature theorem and the Atiyah–Singer index theorem.

References. Cobordism groups were first studied by **R. Thom**'s ***Variétés différentiables cobordantes*** **[Tho53b]** and fully detailed in his ***Quelques propriétés globales des variétés différentiables*** **[Tho54]**. That Ω_3^{SO} is trivial was proved in **A. Wallace**'s ***Modifications and cobounding manifolds*** **[Wal60]** or **R. Lickorish**'s ***A representation of orientable combinatorial*** 3***–manifolds*** **[Lic62b]**. Both $\Omega_4^{SO} = 0$ and $\Omega_4^{Spin} = 0$ were first proved by **V. Rokhlin** in ***New results in the theory of four-dimensional manifolds*** **[Rok52]**.

R. Kirby's ***The topology of 4–manifolds*** **[Kir89]** contains geometric proofs of the low-dimensional cobordism statements mentioned above. A general study of cobordisms can start with chapter 7 of **M. Hirsch**'s ***Differential topology*** **[Hir76, Hir94]**, then continue with **R. Stong**'s monograph ***Notes on cobordism theory*** **[Sto68]**.

As far as we are concerned, we will also encounter the spin$^{\mathbb{C}}$ cobordism group and the characteristic cobordism group, both discussed in the end-notes of chapter 10 (page 427); the two are in fact isomorphic. Also, in the note that follows, we will explore the framed version of cobordisms.

Note: The Pontryagin–Thom construction

In what follows, we will present the Pontryagin–Thom construction, which relates homotopies of maps to framed bordisms of submanifolds. An instance of this method was encountered in the proof of Whitehead's theorem,[46] and the following should shed some extra light on that argument. It is also of independent interest, since it adds geometric content to homotopy groups of spheres. In particular, it was during the pursuit of this method that Rokhlin discovered his celebrated theorem.

The construction. Let

$$f\colon X^{m+k} \to S^m$$

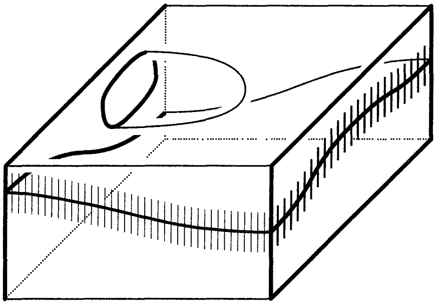
be any map, considered up to homotopy. Pick your favorite point p in S^m, then modify f slightly to make it smooth and transverse to p. The preimage $K = f^{-1}[p]$ is now a k–submanifold of X^{m+k}. Moreover, the differential $df\colon T_X \to T_{S^m}$ induces a map $df\colon N_{K/X} \to T_{S^m}|_p = \mathbb{R}^m$, which is an isomorphism on fibers and thus trivializes $N_{K/X}$. A submanifold together with a trivialization of its normal bundle is called a **framed submanifold.**

---

[46] Back in section 4.1 (page 143).

In the reversed direction, let K be any k–submanifold of X^{m+k} with trivial normal bundle. Assume that a trivialization of its normal bundle $N_{K/X}$ has been chosen. This means that there is a projection $f\colon N_{K/X} \to \mathbb{R}^m$ that is an isomorphism on fibers. Think of f as defined on a tubular neighborhood $N_{K/X}$ of K in X and compactify its codomain $\mathbb{R}^m$ to $\mathbb{S}^m$ by adding a point ∞. Then $f\colon N_{K/X} \to \mathbb{S}^m$ can be extended on $X \setminus N_{K/X}$ simply by setting $f|_{X\setminus N_{K/X}} = \infty$, thus yielding a map $f\colon X^{m+k} \to \mathbb{S}^m$.

The correspondence $K \rightleftarrows f$ becomes bijective if we consider f only up to homotopies, and K only up to framed bordisms. Specifically, two k–submanifold K' and K'' of X^{m+k}, both with trivialized normal bundles, are called **framed bordant** if there exist both a $(k+1)$–submanifold $\widetilde{K}$ of $X \times [0,1]$ such that $\partial\widetilde{K} = \overline{K'} \times 0 \cup K'' \times 1$, and a trivialization of the normal m–plane bundle $N_{\widetilde{K}\,/\,X\times[0,1]}$ of $\widetilde{K}$ such that it induces the chosen trivializations of $N_{K'/X}$ and $N_{K''/X}$ when restricted to $\widetilde{K}$'s boundary. See figure 4.44.



4.44. A framed bordism

Lemma *(Pontryagin–Thom Construction). We have the bijection*

$$[X^{m+k}, \mathbb{S}^m] \approx \Omega_k^{\mathit{framed}}(X^{m+k}),$$

where the former denotes the set of homotopy classes of maps $X \to \mathbb{S}^m$, while the latter denotes the set of framed bordism classes of k–submanifolds of X.

Sketch of proof. That $K \mapsto f \mapsto K$ is the identity is obvious. That $f_1 \mapsto K \mapsto f_0$ is the identity up to homotopy is shown by using the Alexander homotopy $f_t(x) = \frac{1}{t} f_1(tx)$ that links f_1 with $f_0 = df_1|_0$ (use coordinates on $\mathbb{S}^m = \mathbb{R}^m \cup \infty$ that set p at 0). Finally, apply the Pontryagin–Thom construction again to establish a correspondence between $(k+1)$–submanifolds of $X \times [0,1]$ and functions $X \times [0,1] \to \mathbb{S}^m$. Interpret the former as framed bordisms and the latter as homotopies. □

Lemma. *The bijection*

$$\pi_{m+k}\mathbb{S}^m \approx \Omega_k^{\mathit{framed}}(\mathbb{S}^{m+k})$$

is an isomorphism of groups. □

The group structure on the latter is the obvious bordism addition,

$$K' + K'' = K' \cup K'' \subset \mathbb{S}^m \# \mathbb{S}^m = \mathbb{S}^m .$$

Notice also that the bigger m becomes when compared to k, the less relevant the restriction to manifolds that embed in $\mathbb{S}^m$ becomes. In other words, the stable k–stem is given by abstract framed bordisms

$$\varinjlim_{m} \pi_{m+k}\mathbb{S}^m \approx \Omega_k^{framed} ,$$

where the latter is the cobordism group of k–manifolds endowed with a stable trivialization of their tangent bundle.[47]

Whitehead, revisited. Some claims made during the proof of Whitehead's theorem should now be clearer. First, going from $\pi_{m+k}(\mathbb{S}^m)$ to $\pi_{m+k}(\mathbb{S}^m \vee \cdots \vee \mathbb{S}^m)$ is trivial: just consider framed bordisms of several distinct (each maybe disconnected) submanifolds. After that, it is now obvious that any map $f\colon \mathbb{S}^{m+k} \to \mathbb{S}^m$ can be arranged to have $f^{-1}[p]$ connected (compare page 143), because it is easy to devise a framed bordism to a connected k–submanifold (connected sum inside $\mathbb{S}^m$ comes to mind). Similarly, the statement that the linking matrix of L determines the homotopy class of φ (page 146) can now be made rigorous, because the linking matrix is invariant under framed bordisms (allow the splitting of link components into disconnected pieces). It is in fact the only invariant, as will be suggested below.

References. The Pontryagin–Thom construction was created in the 1940s by **L. Pontryagin,** who used framed bordisms to compute homotopy groups of spheres, see his papers ***The homotopy group $\pi_{n+1}(K^n)$ $(n \geq 2)$ of dimension $n+1$ of a connected finite polyhedron K^n of arbitrary dimension, whose fundamental group and Betti groups of dimensions $2, \ldots, n-1$ are trivial*** **[Pon49a]**, and ***Homotopy classification of the mappings of an $(n+2)$–dimensional sphere on an n–dimensional one*** **[Pon50]**, or the book **[Pon55]** translated as ***Smooth manifolds and their applications in homotopy theory*** **[Pon59]**.

Then, after the development by J.P. Serre of more powerful methods for computing homotopy groups,[48] **R. Thom** in ***Quelques propriétés globales des variétés différentiables*** **[Tho54]** went backwards and used computations of homotopy groups in order to compute cobordism groups.[49] Framed bordisms are explained in a friendly manner in **J. Milnor**'s ***Topology from the differentiable viewpoint*** **[Mil65b, Mil97]**, but see also **A. Kosinski**'s ***Differential manifolds*** **[Kos93]**.

Application: homotopy groups of spheres. In what follows we will put to work the Pontryagin–Thom construction to offer geometric interpretations of certain simple homotopy groups of spheres. While this is how the homotopy groups below were first computed by L. Pontryagin and V. Rokhlin, the Pontryagin–Thom construction is a very weak method for evaluating homotopy groups when compared to Serre's later methods.

Lemma. $\pi_n \mathbb{S}^n = \mathbb{Z}$.

---

[47] A **stable bundle** is a bundle considered up to additions of trivial bundles. A stable trivialization of the tangent bundle T_K means an isomorphism $T_K \oplus \underline{\mathbb{R}}^m \approx \underline{\mathbb{R}}^{m+k}$, corresponding to a virtual embedding in $\mathbb{S}^{m+k}$ with $N_{K/\mathbb{S}^{m+k}}$ trivialized as $K \times \mathbb{R}^m$.

[48] See **J.P. Serre**'s ***Homologie singulière des espaces fibrés. III. Applications homotopiques*** **[Ser51]**.

[49] For a first taste of this approach, start with **M. Hirsch**'s ***Differential topology*** **[Hir76, Hir94, ch 7]**.

Sketch of proof. Not that this is not clear for all sorts of reasons, but it can also be argued in terms of framed bordisms: $\Omega_0^{\mathit{framed}}(\mathbb{S}^n)$ contains framed points; the framing of a point $x \in \mathbb{S}^n$ is a trivialization of $T_{\mathbb{S}^n}|_x$ considered up to homotopy, in other words, an orientation of $T_M|_x$. Comparing it with the fixed orientation of $\mathbb{S}^n$ exhibits the elements of $\Omega_0^{\mathit{framed}}(\mathbb{S}^n)$ as points with signs. The isomorphism

$$\Omega_0^{\mathit{framed}}(\mathbb{S}^n) \approx \mathbb{Z}$$

is given simply by counting those points with signs. (Of course, on one hand this is just a very roundabout way of getting to the degree of a map $\mathbb{S}^n \to \mathbb{S}^n$; on the other hand, though, this is just the easiest instance of a pattern that we will see developing below.) □

Lemma. $\pi_3 \mathbb{S}^2 = \mathbb{Z}$, *and* $\pi_{n+1} \mathbb{S}^n = \mathbb{Z}_2$ *when* $n \geq 3$.

Outline of proof. For $\pi_3\mathbb{S}^2$, we are looking at $\Omega_1^{\mathit{framed}}(\mathbb{S}^3)$, which contains framed links in $\mathbb{S}^3$. Each component of the link has a framing, determined by an integer, which can be added together to yield the isomorphism

$$\Omega_1^{\mathit{framed}}(\mathbb{S}^3) \approx \mathbb{Z} .$$

The framing is determined by an integer because we are talking about trivializations of 2–plane bundles over copies of $\mathbb{S}^1$, and $\pi_1 SO(2) = \mathbb{Z}$. As soon as the codimension increases, though, we have $\pi_1 SO(n) = \mathbb{Z}_2$ (detecting whether the bundle twists by an even or odd multiple of 2π), and thus

$$\Omega_1^{\mathit{framed}}(\mathbb{S}^{n+1}) \approx \mathbb{Z}_2 \quad \text{when } n \geq 3,$$

which concludes the argument. □

Lemma. $\pi_{n+2} \mathbb{S}^n = \mathbb{Z}_2$.

Outline of proof. Consider surfaces embedded in $\mathbb{S}^{n+2}$. Every surface S has a *skew*-symmetric bilinear unimodular intersection form on $H_1(S;\mathbb{Z})$, given by intersections of 1–cycles. It descends to an intersection form modulo 2 on $H_1(S;\mathbb{Z}_2)$.

Using the embedding of S in $\mathbb{S}^{n+2}$, we can define a **quadratic enhancement** q of the intersection forms, namely a map $q\colon H_1(S;\mathbb{Z}_2) \to \mathbb{Z}_2$ with

$$q(x+y) = q(x) + q(y) + x \cdot y \pmod 2 .$$

Such a q is defined as follows: represent $\ell \in H_1(S;\mathbb{Z}_2)$ by a circle embedded in S and consider the framing of $N_{S/\mathbb{S}^{n+2}}$ over ℓ: it is determined by a $\mathbb{Z}_2$–framing coefficient, and we define $q(\ell)$ to be that coefficient.

Any quadratic enhancement has an associated $\mathbb{Z}_2$–invariant, called its **Arf invariant**, which can be defined swiftly by setting

$$\mathrm{Arf}(q) = \sum q(e_k)\, q(\bar{e}_k)$$

for any choice of basis $\{e_1, \ldots, e_m, \bar{e}_1, \ldots, \bar{e}_m\}$ of $H_1(S;\mathbb{Z}_2)$ such that the only non-zero intersections are $e_k \cdot \bar{e}_k = 1$. A more thorough discussion of the algebra of the Arf invariant is made in the end-notes of chapter 11 (page 501).

In any case, the Arf invariant is the only framed bordism invariant and establishes the isomorphism

$$\Omega_2^{framed}(\mathbb{S}^n) \approx \mathbb{Z}_2 ,$$

and thus concludes the argument. □

All of the above computations are due to **L. Pontryagin** and can be read from his book **[Pon55]**, translated as ***Smooth manifolds and their applications in homotopy theory*** **[Pon59]**.

Finally, at the limits of Pontryagin–Thom's applicability, we have:

Theorem. $\pi_{n+3}\,\mathbb{S}^n = \mathbb{Z}_{24}$ *when* $n \geq 5$. □

This is already serious business and was first discovered by V. Rokhlin. While studying the problem of $\pi_{n+3}\,\mathbb{S}^n$ by using framed bordisms of 3–manifolds, V. Rokhlin first concluded that $\pi_{n+3}\,\mathbb{S}^n = \mathbb{Z}_{12}$. His mistake stemmed from thinking that a certain characteristic element in a 4–manifold could be represented by an embedded sphere. This was not the case, he corrected his mistake in ***New results in the theory of four-dimensional manifolds*** **[Rok52]**, and in the process discovered his theorem on the signature of almost-parallelizable 4–manifolds. The whole story can be followed in the volume ***À la recherche de la topologie perdue*** **[GM86a]**, edited by **L. Guillou** and **A. Marin,** with French translations of the relevant papers of Rokhlin, commentaries, *etc.*

For completeness, even though they were never obtained using the Pontryagin–Thom construction, we also state:

Theorem. $\pi_{n+4}\,\mathbb{S}^n = 0$, $\pi_{n+5}\,\mathbb{S}^n = 0$, $\pi_{n+6}\,\mathbb{S}^n = \mathbb{Z}_2$ *when n is big.* □

In particular it follows that $\Omega_4^{framed} = 0$. *This is not in contradiction with* $\Omega_4^{SO} = \mathbb{Z}$, *because not all* 4*–manifolds appear in* Ω_4^{framed}, *but only those that can be embedded in a sphere with* trivial *normal bundle, in other words, only those* 4*–manifolds* M *whose tangent bundle is stably-trivial, i.e.,* $T_M \oplus \underline{\mathbb{R}}^n = \underline{\mathbb{R}}^{n+4}$ *for some* n. *These* M*'s have vanishing Pontryagin class, and thus vanishing signature.*

Bibliography

Whitehead's theorem as stated was proved in **J. Milnor**'s ***On simply connected 4–manifolds*** **[Mil58b]**, using results from **J.H.C. Whitehead**'s ***On simply connected, 4–dimensional polyhedra*** **[Whi49b]**. Results similar to Whitehead's were proved independently by **L. Pontryagin** in ***On the classification of four-dimensional manifolds*** **[Pon49b]**. The algebraic-topology argument we presented is the proof from **J. Milnor** and **D. Husemoller**'s ***Symmetric bilinear forms*** **[MH73**, sec V.1**]**, while the long geometric argument involving the Pontryagin–Thom construction was taken from **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89**, ch II**]**. For background on knots and links, always look at **D. Rolfsen**'s ***Knots and links*** **[Rol76, Rol90, Rol03]**. Seifert surfaces were first introduced by **H. Seifert** in ***Über das Geschlecht von Knoten*** **[Sei35]**. The fundamental result that, if a map between simply-connected complexes induces isomorphisms on homology, then it is a homotopy equivalence (quoted in footnote 3 on page 141) was proved in such generality in **J.H.C.**

Whitehead's ***Combinatorial homotopy*** [**Whi49a**]; its proof can be found, for example, in **A. Hatcher**'s ***Algebraic topology*** [**Hat02**, ch 4] or **E. Spanier**'s ***Algebraic topology*** [**Spa66, Spa81**, ch 7].

C.T.C. Wall's theorem on algebraic automorphisms of intersection forms was published in ***On the orthogonal groups of unimodular quadratic forms*** [**Wal62**], and his work continued with the identification of generators in ***On the orthogonal groups of unimodular quadratic forms. II*** [**Wal64a**]. The theorem on diffeomorphisms is contained in ***Diffeomorphisms of 4–manifolds*** [**Wal64b**]. His theorems on stabilizations and h–cobordisms appeared in ***On simply-connected 4–manifolds*** [**Wal64c**]. The proof of the latter as outlined in this volume is from **R. Kirby**'s ***The topology of 4–manifolds*** [**Kir89**, ch X].

Characteristic classes of vector bundles are masterfully described in **J. Milnor** and **J. Stasheff**'s ***Characteristic classes*** [**MS74**]. Their chapter 12 presents the obstruction-theoretic view that we favored above. For the foundations of that view, one should look back at **N. Steenrod**'s wonderful ***The topology of fibre bundles*** [**Ste51, Ste99**, part III]. Another standard reference for bundle theory in general is **D. Husemoller**'s comprehensive ***Fibre bundles*** [**Hus66, Hus94**]. The Dold–Whitney theorem appeared in **A. Dold** and **H. Whitney**'s ***Classification of oriented sphere bundles over a 4–complex*** [**DW59**]. The definition of spin structures as extensible trivialization is due to **J. Milnor**'s ***Spin structures on manifolds*** [**Mil63b**].

Hirzebruch's signature theorem is a general statement about signatures and characteristic classes in all dimensions multiple of 4 and appeared in **F. Hirzebruch**'s ***On Steenrod's reduced powers, the index of inertia, and the Todd genus*** [**Hir53**], then in the book ***Neue topologische Methoden in der algebraischen Geometrie*** [**Hir56**], which eventually became the famous monograph ***Topological methods in algebraic geometry*** [**Hir66**], last printed as [**Hir95**]. It is worth noting that the 4–dimensional case of the signature theorem was also proved in **V. Rokhlin**'s ***New results in the theory of four-dimensional manifolds*** [**Rok52**]. The proof of the general case can be read in chapter 19 of **J. Milnor** and **J. Stasheff**'s ***Characteristic classes*** [**MS74**].

Van der Blij's lemma appeared in **F. van der Blij**'s ***An invariant of quadratic forms mod 8*** [**vdB59**].

A theorem of Rokhlin's. Rokhlin's theorem was published in a four-page paper, ***New results in the theory of four-dimensional manifolds*** [**Rok52**] (where it was also proved that if sign $= 0$, then the manifold bounds). It was translated in English in 1971. A French translation of this and three other remarkable papers of Rokhlin can be read as [**Rok86**] in the volume ***À la recherche de la topologie perdue*** [**GM86a**], edited by **L. Guillou** and **A. Marin**, where they are followed by a commentary [**GM86b**] that makes the dense style of Rokhlin easier to follow.

Rokhlin discovered his theorem by studying homotopy groups of spheres using the Pontryagin–Thom (framed-bordisms) approach, and by first mistakenly stating that $\pi_{n+3} S^n = \mathbb{Z}_{12}$; he then found his mistake, stated his theorem, and corrected to $\pi_{n+3} S^n = \mathbb{Z}_{24}$.

Most later proofs or textbook-treatments of Rokhlin's theorem actually deduce it from $\pi_{n+3} S^n = \mathbb{Z}_{24}$, with the latter fact obtained through the impressive machinery set up by **J.P. Serre**'s ***Homologie singulière des espaces fibrés. III. Applications homotopiques*** **[Ser51]** for computing homotopy groups, thus setting aside the direct geometric approach of Rokhlin's papers. For this homotopy-theoretic approach to proving Rokhlin's theorem, see **M. Kervaire** and **J. Milnor**'s ***Bernoulli numbers, homotopy groups, and a theorem of Rohlin*** **[KM60]**.

Rokhlin's theorem was generalized successively by **M. Kervaire** and **J. Milnor** in ***On 2–spheres in 4–manifolds*** **[KM61]** (see also section 11.1, page 482), and further, along an unpublished outline of **A. Casson** from around 1975, by **M. Freedman** and **R. Kirby** in ***A geometric proof of Rochlin's theorem*** **[FK78]**. The latter statement and its proof from scratch (thus in particular proving Rokhlin's theorem as well) will be discussed in the end-notes of chapter 11, with a warm-up starting on page 502 and a detailed proof on page 507.[50] Alternative proofs of a similar flavor can be read in **L. Guillou** and **A. Marin**'s ***Une extension d'un théorème de Rohlin sur la signature*** **[GM86c]** and **Y. Matsumoto**'s ***An elementary proof of Rochlin's signature theorem and its extension***[51] ***by Guillou and Marin*** **[Mat86]**, both inside the same wonderful volume ***À la recherche de la topologie perdue*** **[GM86a]**. It has been reported that **V. Rokhlin** was himself long aware (1964) of these generalizations, but only published them in ***Proof of a conjecture of Gudkov*** **[Rok72]**, when he found an application.

Another version of the proof is found in **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89**, ch XI**]**, where a nice streamlined argument with spin structures is used. This alternative proof is also explained in this volume, in the end-notes of chapter 11 (page 521).

A third and surprising proof of Rokhlin's theorem that starts with the μ–invariants of 3–manifolds can be read from the appendix of **R. Kirby** and **P. Melvin**'s ***The 3–manifold invariants of Witten and Reshetikhin–Turaev for*** $\mathfrak{sl}(2,\mathbb{C})$ **[KM91]**; it was briefly mentioned back on page 227.

---

[50] The reason for the exile of the proof of Rokhlin's theorem to chapter 11 is mainly one of space: even though logically that proof would better fit with the present chapter, the current group of end-notes is already quite extensive.

[51] The word "extension" from the last two titles refers to a refinement of the Kirby–Siebenmann formula from a $\mathbb{Z}_2$–equality to a $\mathbb{Z}_4$–equality, with the extra residues appearing only from non-orientable characteristic surfaces.

Chapter 5

Classifications and Counterclassifications

THE time has come to look at the algebraic classification of (abstract) symmetric integral unimodular forms. Afterwards we will classify topological 4–manifolds through their intersection forms (Freedman), but will notice that smooth manifolds are not so well-behaved (Donaldson). The chapter concludes with a presentation of exotic $\mathbb{R}^4$'s.

We start with the statement of J.P. Serre's classification of indefinite forms, followed by a frightening count of definite forms. The classification of indefinite odd forms is further argued in the end-notes on page 262. The end-notes also contain a proof of van der Blij's lemma, on page 263.

In section 5.2 (page 239) we present the striking result of M. Freedman that completely classifies simply-connected topological manifolds: for every even form there is exactly one topological manifold having it as intersection form, while for every odd form there are exactly two. The smooth realm is not that well-behaved: indeed, almost no definite forms can be represented by smooth 4–manifolds; this result of S.K. Donaldson and other exclusions from the smooth realm are presented in section 5.3 (page 243).

The collision between the smooth and topological realms spawns exotic $\mathbb{R}^4$. These open manifolds homeomorphic but not diffeomorphic to $\mathbb{R}^4$ appear whenever the smooth theory is in conflict with the topological theory. Indeed, there are plenty of them: there are in fact uncountably-many exotic $\mathbb{R}^4$'s, as we will see in section 5.4 (page 250).

5.1. Serre's algebraic classification of forms

Intersection forms in themselves are just a topological incarnation of an otherwise abstract algebraic entity: a symmetric bilinear unimodular form

$$Q\colon Z \times Z \longrightarrow \mathbb{Z}\,,$$

defined on some finitely-generated free $\mathbb{Z}$–module Z. As such, one can try to explore the various shapes of such creatures, without worrying about any geometric content.

Indefinite forms

The algebraic classification is quite successful in the case of indefinite symmetric bilinear unimodular forms:

Serre's Classification Theorem. *Let Q' and Q'' be two symmetric bilinear unimodular forms. If both Q' and Q'' are indefinite, then Q' and Q'' are isomorphic forms if and only if they have the same rank, signature, and parity.* □

In terms of concrete representatives, this becomes:

Corollary. *Let $Q\colon Z \times Z \to \mathbb{Z}$ be a symmetric bilinear unimodular form.*

— *If Q is indefinite and odd, then in a suitable basis it can be written*

$$Q = \oplus\, m\,[+1] \oplus n\,[-1]\ .$$

— *If Q is indefinite and even, then in a suitable basis it can be written*

$$Q = \oplus \pm m\, E_8 \oplus n\, H\ .$$ □

The reason for the relative simplicity of the classification of indefinite forms lies with the following nontrivial property:

Meyer's Lemma. *Let $Q\colon Z \times Z \to \mathbb{Z}$ be a symmetric bilinear unimodular form. If Q is indefinite, then there exists an element $x_0 \in Z$ so that*

$$x_0 \cdot x_0 = 0\,.$$ □

From this result can be quickly deduced the above classification of indefinite forms that are odd, as is explained in the end-notes of this chapter (page 262). Also, van der Blij's lemma can then be easily proved, see the end-notes of this chapter (page 263).

> ***Example.*** *A simple and rather random application of the algebraic classification of indefinite forms is the following: Start with a random smooth 4–manifold. Then $M \# \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$ will have indefinite odd intersection form, and thus must have $Q_{M\#\mathbb{CP}^2\#\overline{\mathbb{CP}}^2} = \oplus\, p\,[+1] \oplus q\,[-1]$, with $p + q = b_2(M) - 2$ and $p - q = \operatorname{sign} M$. Wall's theorem on stabilizations implies that, for an appropriate k, we have a diffeomorphism $M \# \mathbb{CP}^2 \# \, k S^2 \times S^2 \ \cong\ \# \, p\,\mathbb{CP}^2 \#$*

$q\overline{\mathbb{CP}}^2 \# k\,\mathbb{S}^2 \times \mathbb{S}^2$. However, we also have $\mathbb{S}^2 \times \mathbb{S}^2 \# \mathbb{CP}^2 \;\cong\; \# 2\,\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$, and so we end up with

$$M \# (k+1)\,\mathbb{CP}^2 \# (k+1)\,\overline{\mathbb{CP}}^2 \;\cong\; \#(p+k)\,\mathbb{CP}^2 \# (q+k)\,\overline{\mathbb{CP}}^2 \,.$$

Hence, no matter how complicated M might have been, we managed, by adding enough projective planes, to completely dissolve M into triviality.

In comparison to indefinite forms, definite forms are complexity itself:

Definite forms

There is no classification of definite forms. One has, of course, boring examples like $\oplus\, m\,[+1]$, or $E_8 \oplus [+1]$, or $E_8 \oplus E_8 \ldots$; but besides these,[1] the sheer number of definite forms is simply enormous.

Consider, for example, positive-definite even forms. From van der Blij's lemma, their rank is always a multiple of 8. Counting how many forms we have, for rank 24 we get 24 distinct forms; for the next available rank, 32, there are already more than eighty million of them; for rank 40, there are more than 10^{51}, and most likely that is a gross underestimate.[2] The situation for definite odd forms is similar. At least we can celebrate the fact that, for every fixed rank, it is known that there are only finitely-many forms representing it...

We gather our information about algebraic forms in table VIII.

VIII. Classification of integral symmetric unimodular forms

<table>
<tr><th></th><th><i>indefinite</i></th><th colspan="6"><i>(positive) definite</i></th></tr>
<tr><td><i>odd</i></td><td>$m[+1] \oplus n[-1]$</td><td colspan="6">too many</td></tr>
<tr><td rowspan="2"><i>even</i></td><td rowspan="2">$\oplus \pm m\,E_8 \oplus n\,H$</td><td><i>rank</i></td><td>8</td><td>16</td><td>24</td><td>32</td><td>40</td></tr>
<tr><td>#</td><td>1</td><td>2</td><td>24</td><td>$> 10^7$</td><td>$\gg 10^{51}$</td></tr>
</table>

5.2. Freedman's topological classification

The remarkable fact is that all the overwhelming wealth of algebraically-possible intersection forms can actually be incarnated into *topological* 4–manifolds:

---

[1] Two further examples, Γ_{4k} and the Leech lattice, are defined in the end-note on page 264.

[2] For further comments on this count, see the end-notes of this chapter (page 264).

Freedman's Classification Theorem. *For any integral symmetric unimodular form Q, there is a closed simply-connected topological 4-manifold that has Q as its intersection form.*

— *If Q is even, there is exactly one such manifold.*

— *If Q is odd, there are exactly two such manifolds, at least one of which does not admit any smooth structures.*

If we restrict to *smooth* 4–manifolds, this most remarkable theorem reduces to a statement that we have already presented,[3] specifically:

If M and N are smooth simply-connected 4–manifolds with isomorphic intersection forms, then M and N must be homeomorphic.

Indeed, Freedman's classification result above has no bearing on the diffeomorphic classification of 4–manifolds, and none either on the existence of smooth structures.

> ***Example.*** *Take the non-smoothable E_8-manifold, $\mathcal{M}_{E_8}$. On one hand, the connected sum $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8}$ has intersection form $E_8 \oplus E_8$ and is also non-smoothable, as we will see shortly. On the other hand, the manifold $\mathcal{M}_{E_8} \# \overline{\mathcal{M}}_{E_8}$ has intersection form $E_8 \oplus -E_8 \approx \oplus 8\, H$, and therefore must be homeomorphic to $\# 8\, S^2 \times S^2$:*
>
> $$\mathcal{M}_{E_8} \# \overline{\mathcal{M}}_{E_8} \simeq \# 8\, S^2 \times S^2 .$$
>
> *Thus, it can inherit a smooth structure from the latter.*

In the second case of Freedman's theorem, for odd forms, the two non-homeomorphic 4–manifolds M and $*M$ that represent a same form can be distinguished by their Kirby–Siebenmann invariants:[4] we have $\mathrm{ks}(M) = 0$ and $\mathrm{ks}(*M) = 1$. Then $*M$ is certainly non-smoothable, but M could also be non-smoothable.[5]

> ***Sketch of proof of Freedman's classification.***
>
> *Existence.* Let Q be an abstract symmetric unimodular bilinear form over a $\mathbb{Z}$–module Z. Then it can be realized as the intersection form of a simply-connected 4–manifold P_Q with non-empty boundary, obtained by suitably plumbing disk-bundles over spheres. Since Q is unimodular, the boundary $\Sigma = \partial P_Q$ must be a homology 3–sphere. Then so is its oppositely-oriented version $\overline{\Sigma}$, and the latter must bound some fake 4–ball Δ. We can then glue Δ to P_Q and obtain a closed oriented

---

[3] See back in section 4.2 (page 155).

[4] The Kirby–Siebenmann invariant was explained back in the end-notes of the preceding chapter (page 221); it is a $\mathbb{Z}_2$-valued invariant that must vanish for manifolds that admit smooth structures.

[5] We can distinguish M and $*M$ by saying that $M \times S^1$ admits a smooth structure, while $(*M) \times S^1$ does not. *Ditto* for $M \times \mathbb{R}$ and $(*M) \times \mathbb{R}$.

simply-connected topological 4–manifold $\mathcal{M}_Q = P_Q \cup_\partial \Delta$, whose intersection form is exactly Q. This argument will be presented in more detail in the end-notes of this chapter (page 260).

If plumbings are not to one's taste, one can always start with a framed link in $\mathbb{S}^3$ whose linking matrix is Q, then add a 2–handle to each component of the link while respecting the framing (as in Kirby calculus, see the note on page 91, at the end of chapter 2); one ends up again with a 4–manifold with intersection form Q and boundary a homology 3–sphere, which can be capped with a contractible Δ.

Uniqueness. If the manifolds admit smooth structures, then, as mentioned above, the result was already proved earlier, as a consequence of Wall's theorem on h–cobordisms and Freedman's topological h–cobordism theorem. For the general case, when the manifolds do not necessarily admit smooth structures, the result follows using surgery techniques of C.T.C. Wall and the Kirby–Siebenmann invariant. (An important role in Freedman's argument is played by the fake $\mathbb{CP}^2$ that we build next.) □

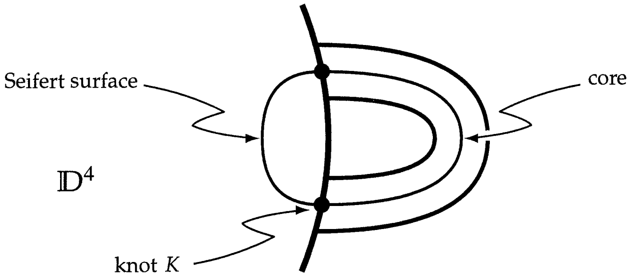
***An example:* $\mathbb{CP}^2$*, and fake* $\mathbb{CP}^2$.** We will be leaving the realm of topological 4–manifolds pretty soon. Before we do, we wish to build the two manifolds prescribed by Freedman's theorem to represent the form $[+1]$. One of them is $\mathbb{CP}^2$, the other is a non-smoothable 4–manifold homotopy-equivalent to $\mathbb{CP}^2$. The latter is called a **fake** $\mathbb{CP}^2$ and denoted by $*\mathbb{CP}^2$.

Take the 4-ball $\mathbb{D}^4$. To its boundary $\mathbb{S}^3$ we will attach a 2–handle $\mathbb{D}^2 \times \mathbb{D}^2$ by identifying the boundary of the core $\mathbb{D}^2 \times 0$ to a knot K in $\mathbb{S}^3$. Since we must attach the whole leg $\mathbb{S}^1 \times \mathbb{D}^2$ of the handle, this means that besides the destination K of the attaching circle, we also need to choose a precise way to attach around K its thickening. In other words, we need to pick a framing[6] for the knot. This can be specified by an integer describing how many twists occur along the knot, measured with respect to any oriented Seifert surface bounded by K. Notice that the core of the handle attaches to this Seifert surface and yields a closed surface, whose self-intersection is exactly the chosen framing, see figure 5.1 on the following page.

If we choose to attach a 2–handle to the unknot with framing $+1$, the result can be seen to be a disk bundle over a sphere (the base sphere is created by gluing the core of the handle to a disk bounded by the unknot). The Euler class of this disk bundle is $+1$ (like the framing). Therefore its boundary is the Hopf bundle; specifically, it is a 3–sphere. If we cap this 3–sphere with another 4–ball, then the resulting manifold will be $\mathbb{CP}^2$.

---

[6] Compare with Kirby calculus, described in the end-notes of chapter 2 (page 91).



5.1. Attaching a 2–handle creates a surface

On the other hand, if, instead of using the unknot, we attach our 2–handle to the (right) trefoil knot with framing $+1$ (see figure 5.2), then the boundary of the resulting manifold turns out to be Poincaré homology 3–sphere Σ_P. Its reverse $\overline{\Sigma}_P$ bounds a contractible topological 4–manifold, a fake 4–ball $\varDelta$. If we attach $\varDelta$ to our construction, what we obtain is a simply-connected 4–manifold with intersection form $[+1]$. By Whitehead's theorem, the result is homotopy-equivalent to $\mathbb{CP}^2$; but it is known to not be homeomorphic to it: we obtained a *fake* $\mathbb{CP}^2$, which will never admit any smooth structures. We denote it by $*\mathbb{CP}^2$.

5.2. $\mathbb{CP}^2$, and fake $\mathbb{CP}^2$

If we choose some other random knot K in $\mathbb{S}^3$ and attach a 2–handle to it with framing $+1$, then the boundary of the result will still be a homology 3–sphere, and thus there will exist a fake 4–ball that can be used to cap it off to a closed 4–manifold. The intersection form of this topological 4–manifold is still $[+1]$ and thus, by Freedman's classification, must be homeomorphic to either $\mathbb{CP}^2$ or $*\mathbb{CP}^2$—there are no other options.

More stars. *This manifold $*\mathbb{CP}^2$ can be used to create more pairs of non-homeomorphic manifolds with the same intersection form. For example, both $\mathbb{CP}^2 \# \mathcal{M}_{E_8}$ and $(*\mathbb{CP}^2) \# \mathcal{M}_{E_8}$ have the same intersection form $[+1] \oplus E_8$, but their Kirby–Siebenmann invariants differ: we have*[7] $\mathrm{ks}(\mathbb{CP}^2 \# \mathcal{M}_{E_8}) = 1$

---

[7] Here, we use the additivity of the Kirby–Siebenmann invariant and that $\mathrm{ks}(\mathcal{M}_{E_8}) = 1$. See the end-notes of the preceding chapter (page 221).

but $\mathrm{ks}\big((*\mathbb{CP}^2)\,\#\,\mathcal{M}_{E_8}\big) = 0$. *The form* $[+1] \oplus E_8$ *is definite, and from Donaldson's results (see ahead) neither of its manifolds can admit smooth structures. The sum-stabilization* $\mathbb{CP}^2\,\#\,\mathcal{M}_{E_8}\#\,\mathbb{S}^2 \times \mathbb{S}^2$ *is still not smooth, and no further stabilizing can help. However,* $(*\mathbb{CP}^2)\,\#\,\mathcal{M}_{E_8}\#\,\mathbb{S}^2 \times \mathbb{S}^2$ *must be homeomorphic to* $\#\,10\,\mathbb{CP}^2\,\#\,\overline{\mathbb{CP}}^2$ *and thus admits smooth structures. As another example, the 4–manifolds* $\mathbb{CP}^2\,\#\,\overline{\mathcal{M}}_{E_8}$ *and* $(*\mathbb{CP}^2)\,\#\,\overline{\mathcal{M}}_{E_8}$ *are non-homeomorphic, with intersection form* $[+1] \oplus -E_8 \approx [+1] \oplus 8\,[-1]$. *The manifold* $\mathbb{CP}^2\,\#\,\overline{\mathcal{M}}_{E_8}$ *is non-smoothable, but* $(*\mathbb{CP}^2)\,\#\,\overline{\mathcal{M}}_{E_8}$ *is homeomorphic to* $\mathbb{CP}^2\,\#\,8\,\overline{\mathbb{CP}}^2$ *and thus admits smooth structures.*

This $*$*–operation can be extended to all topological manifolds* M *with odd intersection forms, with* $*M$ *characterized by any homeomorphism* $(*M)\,\#\,\mathbb{CP}^2 \simeq M\,\#\,(*\mathbb{CP}^2)$ *that preserves the splitting* $Q_M \oplus [+1]$ *of the intersection forms. Of course,* $ks(*M) \neq ks(M)$.

5.3. Donaldson's smooth exclusions

As we have seen, all abstract candidates for intersection forms are actually realized by topological 4–manifolds, including the overwhelming crowds of definite forms. However, if one focuses on *smooth* manifolds, then most intersection forms are in fact excluded.

Besides a few old exclusions obtained from Rokhlin's theorem, the revelation came with S.K. Donaldson's remarkable application of differential geometry to smooth 4–dimensional topology, which ushered in a whole new stage in the development of smooth 4–dimensional topology.

Definite forms

For smooth manifolds the definite forms essentially disappear from the picture. Only one year after after M. Freedman's revolution, S.K. Donaldson's counter-revolution followed in 1982:

Donaldson's Theorem. *The bilinear symmetric unimodular forms*

$$\oplus\, m\,[+1] \qquad \text{and} \qquad \oplus\, m\,[-1]$$

are the only definite forms that can be realized as intersection forms of a smooth 4–manifold.

Thus, if a simply-connected smooth manifold has a definite intersection form, then it must be homeomorphic to either $\#\,m\,\mathbb{CP}^2$ or $\#\,m\,\overline{\mathbb{CP}}^2$.

It might be striking that, between Rokhlin's theorem (1952) and Donaldson's (1982), no new intersection form exclusions from the smooth realm appeared. Thus, while it had been clear for a while that E_8 *is not representable by a*

> *smooth 4–manifold, the possibility survived that, say, $E_8 \oplus E_8$ could be. After Donaldson, we know that it cannot (and in particular, $\mathcal{M}_{E_8} \# \mathcal{M}_{E_8}$ is non-smoothable). However, in a certain sense Rokhlin's theorem, through its far-reaching consequences, belongs to the topology of manifolds of all dimensions, while Donaldson's is confined to dimension 4.*

Combining Donaldson's theorem with Freedman's classification of topological manifolds and Serre's algebraic classification of forms yields the succinct corollary:

Corollary. *Two smooth simply-connected 4–manifolds are homeomorphic if and only if their intersection forms have the same rank, signature, and parity.* □

A more concrete version is:

Every smooth simply-connected 4–manifold is homeomorphic to either $\# m\, \mathbb{CP}^2 \# n\, \overline{\mathbb{CP}}^2$ *or*[8] $\# \pm m\, \mathcal{M}_{E_8} \# n\, S^2 \times S^2$.

However, since many of the $\# \pm m\, \mathcal{M}_{E_8} \# n\, S^2 \times S^2$'s are non-smoothable, this last statement is somewhat unsatisfactory. Fleshing out precisely which are and which are not smoothable is what the 11/8– and 3/2–conjectures try to do, as we will see shortly.

The proof of Donaldson's theorem uses differential geometry, specifically gauge theory, and it was the first of many results that came from that area. As the reader will certainly notice, the flavor of the argument is from a completely different planet than those encountered so far.

> ***Sketch of proof of Donaldson's theorem.*** Assume M is a smooth 4–manifold whose intersection form is *negative* definite.[9] Assume that M is simply-connected (Donaldson's methods work best in the simply-connected setting; this restriction disappears when using Seiberg–Witten theory, and a proof of the theorem using the latter techniques will be presented later.[10])
>
> Endow M with some Riemannian metric and build on M a complex-plane bundle
>
> $$E \to M$$
>
> with structure group $SU(2)$ and Euler class $+1$. We consider the collection of all $SU(2)$–connections[11] A on E that satisfy the curvature equation
>
> $$F_A^+ = 0\,.$$

---

[8] Of course, a negative coefficient of $\mathcal{M}_{E_8}$ should be understood as reversing the orientation.

[9] The positive-definite case follows from the negative-definite one after reversing orientation.

[10] In the end-notes of chapter 10 (page 454).

[11] The differential-geometric concept of "connection" is reviewed in chapter 9 (page 331). The whole part 4 of this volume is devoted to these differential-geometric approaches to smooth topology. Thus, the argument sketched here might be easier to follow after passing once through that part.

Here, F_A is the End(E)–valued curvature 2–form of A, and F_A^+ is the self-dual part of F_A. Such a connection is called an **instanton** or an **anti-self-dual connection.**

In fact, we consider the solutions A of the equation only up to **gauge equivalence**, *i.e.*, up to automorphisms of the bundle E. It is proved that, for a generic metric and slight perturbation of the equation, the moduli space

$$\mathfrak{M} = \{[A] \mid F_A^+ = 0\}$$

of (gauge classes) of solutions is in fact an *open* smooth 5–manifold, away from a few isolated singular points $p_1, \ldots, p_m$ where $\mathfrak{M}$ fails to be a manifold.

The singular points have neighborhoods which are cones on $\mathbb{CP}^2$. The number m of singular points is exactly half the number of homology classes of M that have self-intersection -1:

$$m = \tfrac{1}{2}\#\{\alpha \in H_2(M;\mathbb{Z}) \mid \alpha \cdot \alpha = -1\}$$

(half, because both $+\alpha$ and $-\alpha$ are in there).[12]

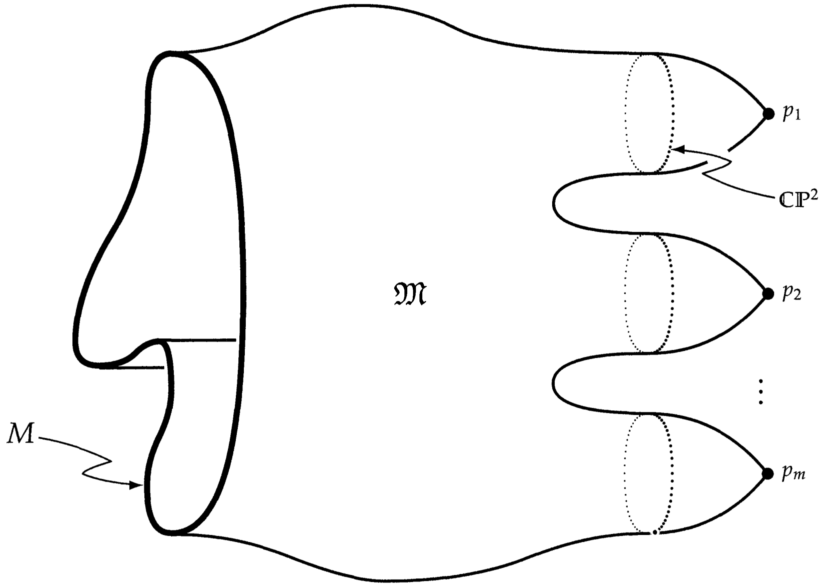
The fundamental fact is that $\mathfrak{M}$ can be compactified by adding our 4–manifold M as boundary to $\mathfrak{M}$. This happens because we can find, for each $x \in M$, a sequence of anti-self-dual connections whose curvatures are concentrating closer-and-closer to x, and so one can think of that sequence in $\mathfrak{M}$ as converging to the actual point x of M. In fact, for every divergent sequence of anti-self-dual connections, there is a subsequence whose curvatures concentrate at some $x \in M$, and so one can indeed think of M as the boundary[13] of $\mathfrak{M}$.

Therefore, by adding M as boundary to $\mathfrak{M}$, and then by cutting neighborhoods of the singular points out of it, we end up with a cobordism between, on one side, our 4–manifold M, and, on the other side, m_+ copies of $\mathbb{CP}^2$ and m_- copies of $\overline{\mathbb{CP}}^2$ (with the number m of singular points split as $m = m_+ + m_-$, depending on orientations), as in figure 5.3 on the following page.

---

[12] The singularities appear from **reducible solutions** (compare with the general discussion in chapter 9, page 331 ahead), in this case anti-self-dual connections on E that come from connections on a complex-line bundle L in a decomposition $E = L \oplus L^*$. Such a decomposition is equivalent to a choice of Chern class $\alpha = c_1(L)$ such that $\alpha \cdot \alpha = -1$. (Of course, choosing α or $-\alpha$ leads to the same decomposition.) It turns out that to every such decomposition corresponds (up to gauge) exactly one reducible anti-self-dual connection. Therefore, the count of singular points is the same as a count of α's as above.

[13] Of course, one actually needs to prove that a complement of a big compact set in $\mathfrak{M}$ is diffeomorphic to $M \times (0,1)$, so that we can add $M \times 0$ to $\mathfrak{M}$ as the smooth boundary of $\mathfrak{M}$, with $M \times [0,1)$ becoming a collar of $M = \partial\,\mathfrak{M}$ in $\mathfrak{M}$.



5.3. The 1–instanton moduli space for negative-definite Q_M

Since M together with the $\mathbb{CP}^2$'s make up the boundary of a 5–manifold, their total signature must be zero. Since the signature-contribution of a copy of $\mathbb{CP}^2$ is $+1$, while a copy of $\overline{\mathbb{CP}}^2$ contributes -1, we must have

$$\operatorname{sign} M = m_+ - m_- .$$

Since Q_M is negative-definite, we get $\operatorname{sign} M = -b_2(M)$, and thus

$$b_2(M) = -\operatorname{sign} M = -m_+ + m_- \leq m_+ + m_- = m \leq b_2(M) ,$$

which forces $b_2(M) = m$.

However, m was half the number of classes α with $\alpha \cdot \alpha = -1$. The unimodularity of Q_M implies that, if $\alpha \cdot \alpha = -1$, then Q_M must split as

$$Q_M = [-1] \oplus Q' .$$

Using up all such α's available, we end up with $Q_M = m[-1] \oplus Q^{(m)}$. Since $m = b_2(M)$, the α's must in fact exhaust all of Q_M, and hence $Q_M = \oplus\, m[-1]$, which concludes the proof. □

Please do not be misled by the breezy outline above, but keep in mind that we swept under the rug a whole elephant of hard analysis on which the above argument is founded.

A complete alternative proof using Seiberg–Witten theory will be detailed in the end-notes of chapter 10 (page 454).

Indefinite even forms

As far as representing indefinite forms by smooth manifolds, the odd ones are settled as each of them are realized by some $\#m\,\mathbb{CP}^2\,\#n\,\overline{\mathbb{CP}}^2$. On the other hand, even forms are subject to exclusions.

Algebraically, the even forms must all be of shape $Q = \oplus \pm m\,E_8 \oplus n\,H$. Rokhlin's theorem restricts the even forms corresponding to smooth 4–manifold to those with an even number of E_8's, that is, to

$$Q = \oplus \pm 2m\,E_8 \oplus n\,H\,.$$

Further, by Donaldson's result, they must contain at least one H to avoid definiteness.

Notice that increasing the number of H's is not a problem: just connect sum with a copy of $S^2 \times S^2$. Thus the question becomes: What is the *minimum* number of H's needed for a form as above to become representable by a smooth 4–manifold? It is conjectured that:

The $^{11}\!/_{8}$***–Conjecture*** *(open). Every smooth 4–manifold M with even intersection form must have*

$$b_2(M) \geq \tfrac{11}{8}\,|\mathrm{sign}\,M|$$

or, equivalently, we must have $n \geq 3\,|m|$, i.e., at least three H's for every couple of E_8's in Q_M.

If this conjecture were true, an immediate consequence would be:[14]

Every smooth simply-connected 4–manifold is homeomorphic to either of

$$\#m\,\mathbb{CP}^2\,\#n\,\overline{\mathbb{CP}}^2 \qquad \textit{or} \qquad \#\pm m\,K3\,\#n\,S^2 \times S^2\,.$$

However, without the $^{11}\!/_{8}$–conjecture proved, we do not know whether a statement of this type might not need to involve some mysterious smooth 4–manifolds in order to cover the topological ground between, say, $\#m\,K3$ and $\#(m+1)\,K3$.

An even stronger open conjecture applies to irreducible *4–manifolds:*

The $^{3}\!/_{2}$***–Conjecture*** *(open). Let M be a smooth 4–manifold M with even intersection form. Assume that M is irreducible, i.e., that it does not split into a connected sum of simpler 4–manifolds (not homeomorphic to S^4). Then we must have*

$$\chi(M) \geq \tfrac{3}{2}\,|\mathrm{sign}\,M|$$

or, equivalently, we must have $n \geq 4\,|m| - 1$, i.e., about four H's for every couple of E_8's in Q_M.

---

[14] Remember that the intersection form of $K3$ is $\oplus(-2)E_8 \oplus 3H$.

> *Keep in mind that our understanding of exactly who the collection of all irreducible 4–manifolds might be has receded in recent years.*[15]

A first step toward the 11/8–conjecture was:

Theorem (*S.K. Donaldson*). *The forms*

$$H \qquad and \qquad H \oplus H$$

are the only even forms that appear from smooth 4–manifold with $b_2^+ = 1$ or $b_2^+ = 2$. In other words, if $|m| \geq 1$, then $n \geq 3$, i.e., if there is an E_8, there must be at least three H's. □

Therefore, after S^4 and $\#n\, S^2 \times S^2$, the $K3$ surface has the simplest even intersection form.

Donaldson and others continued to use instantons and develop a whole machinery of invariants for smooth 4–manifolds. Then, in 1994, N. Seiberg and E. Witten introduced their **monopoles** in 4–dimensional topology. The Seiberg–Witten monopoles are much easier to manipulate than instantons, but contain essentially equivalent information. Indeed, all the results obtained using instantons were re-proved, usually more easily, by using Seiberg–Witten theory. We will discuss the latter theory in some detail starting with chapter 10 (page 375).

By studying the shape of the moduli space of solutions to the equations of Seiberg and Witten, M. Furuta proved:[16]

Furuta's 10/8–Theorem. *Every smooth 4–manifold M with even intersection form must have*

$$b_2(M) \geq \tfrac{10}{8}\,|\mathrm{sign}\, M| + 2$$

or, equivalently, we must have $n \geq 2\,|m| + 1$, i.e., at least two H's for every couple of E_8's in Q_M. □

Finally, a recent refinement:

Theorem (*M. Furuta, Y. Kametani and H. Matsue*). *The form $\oplus \pm 4\, E_8 \oplus 5\, H$ is not the intersection form of any smooth 4–manifold.* □

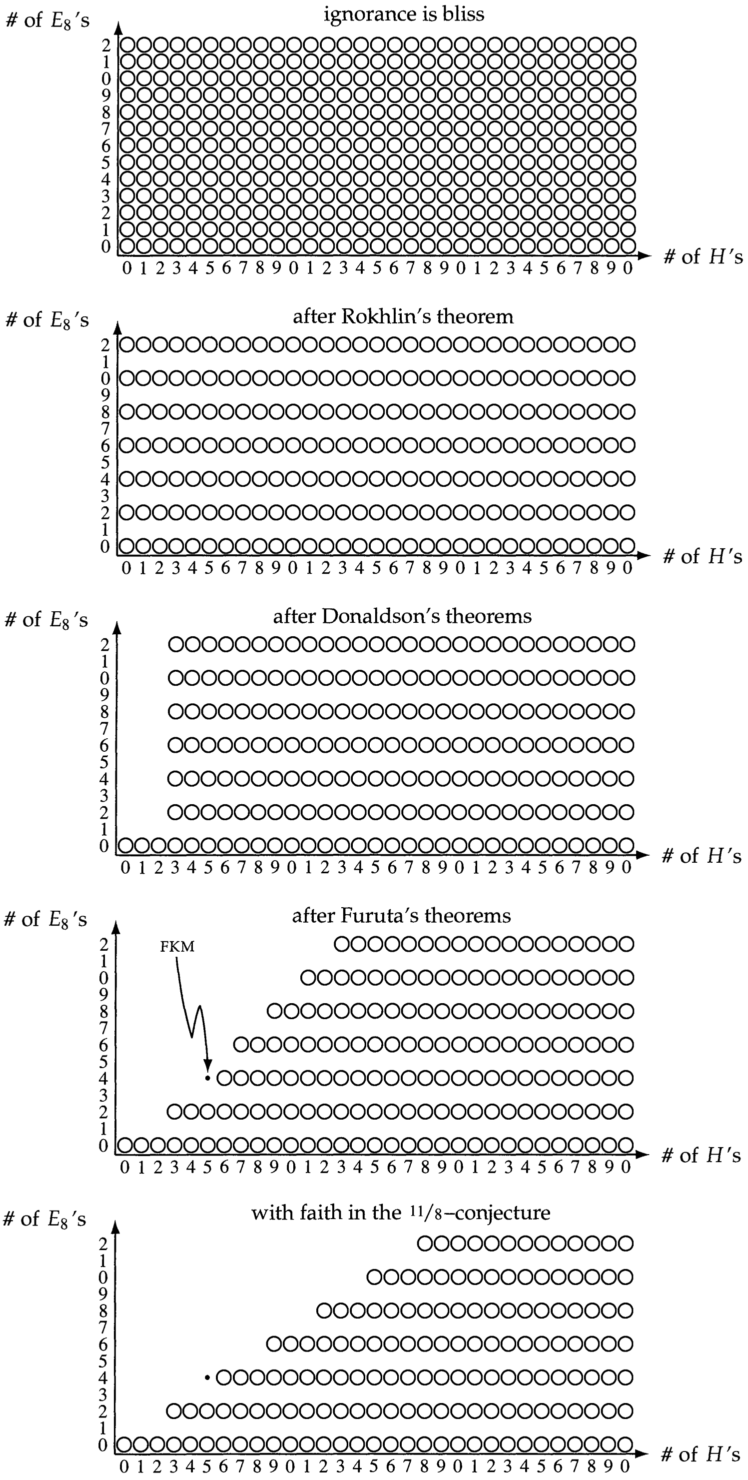
The overall current situation for even forms is summarized in figure[17] 5.4 on the next page.

---

[15] For a while, one entertained conjectures such as: all irreducible 4–manifolds are complex surfaces (with either orientation), then that they all are symplectic manifolds (with either orientations). Examples of irreducible 4–manifolds that are neither complex nor symplectic have left us with no current conjecture. Compare with the comments on page 553.

[16] The Seiberg–Witten equations are built on top of a choice of spin$^{\mathbb{C}}$ structure on M. If Q_M is even, then M admits a spin structure, which can be thought of as a very special type of spin$^{\mathbb{C}}$ structure, and this confers extra symmetry to the Seiberg–Witten moduli space, leading to Furuta's result.

[17] In figure 5.4, FKM denotes the Furuta–Kametani–Matsue theorem we just stated


ignorance is bliss

after Rokhlin's theorem

after Donaldson's theorems

after Furuta's theorems

with faith in the 11/8-conjecture

5.4. Smooth exclusions of indefinite even forms $\oplus \pm m\, E_8 \oplus n\, H$

5.4. Byproducts: exotic $\mathbb{R}^4$'s

Once Donaldson's exclusions appeared, it became clear that there must exist exotic $\mathbb{R}^4$'s. An **exotic** $\mathbb{R}^4$ is a smooth open 4–manifold that is homeomorphic to $\mathbb{R}^4$ but not diffeomorphic to it. There are two sources of exotic $\mathbb{R}^4$'s:

On one hand, they appear from the clash between the topological success of the h–cobordism theorem and its smooth failure: as we will see, every Akbulut cork[18] is surrounded by an exotic $\mathbb{R}^4$. These exotic $\mathbb{R}^4$'s can be embedded inside standard $\mathbb{R}^4$, and thus are called **small exotic $\mathbb{R}^4$'s.**

On the other hand, another suite of exotic $\mathbb{R}^4$'s appears from the collision between the topological success of connected-sum-splitting and its smooth failure. The resulting exotics do not embed in standard $\mathbb{R}^4$, and in fact contain a compact set that cannot be surrounded by any smoothly embedded 3–sphere. They are called **large exotic $\mathbb{R}^4$'s.**

To split or not to split

An exotic $\mathbb{R}^4$**.** Consider the smooth 4–manifold[19]

$$E = \mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2 \,.$$

Its intersection form is $Q_E = [+1] \oplus 9\,[-1]$. For a suitable choice of basis,[20] Q_E can also be written as

$$Q_E = -E_8 \oplus [-1] \oplus [+1] \ .$$

Topologically, this corresponds to a split $E = \overline{\mathcal{M}_{E_8}} \# \overline{\mathbb{CP}}^2 \# \mathbb{CP}^2$. Nonetheless, *smoothly* separating a copy of $\overline{\mathbb{CP}}^2 \# \mathbb{CP}^2$ from E is not possible, since $\overline{\mathcal{M}_{E_8}}$ is not smoothable, owing to Rokhlin's theorem.

We might then try to split off at least the $[+1]$–term, seeking a smooth connected-sum decomposition $E = N \# \mathbb{CP}^2$, for some smooth 4–manifold N. However, this must fail as well. Indeed, the needed smooth manifold N would have intersection form $-E_8 \oplus [-1]$, which is definite and thus excluded by Donaldson's theorem. This failure crystallizes into the existence of an exotic $\mathbb{R}^4$:

Denote by α an element of $H_2(E;\mathbb{Z})$ that spans the $[+1]$–term of

$$Q_E = -E_8 \oplus [-1] \oplus [+1] \ .$$

---

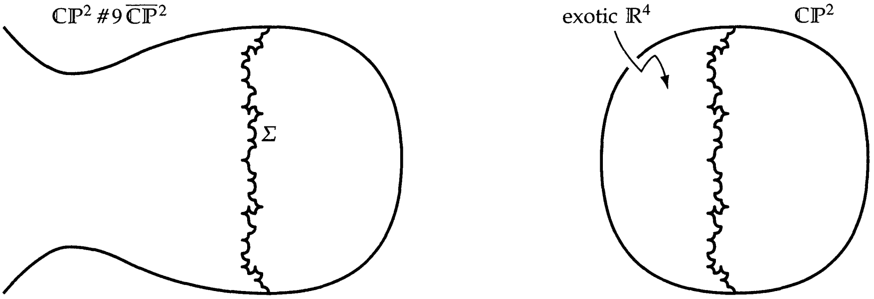
[18] Akbulut corks were encountered earlier in section 2.4 (page 89).

[19] The manifold E happens to be a complex surface, known as the *rational elliptic surface.*

[20] This follows since the two forms have the same rank, signature and parity. For a concrete basis, if e_k denotes the class of $\mathbb{CP}^1$ in the k^{th} copy of $\overline{\mathbb{CP}}^2$ inside $E = \mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2$, then the element $\alpha = 3e_0 - e_1 - \cdots - e_9$ spans the $[+1]$ in $Q_E = -E_8 \oplus [-1] \oplus [+1]$.

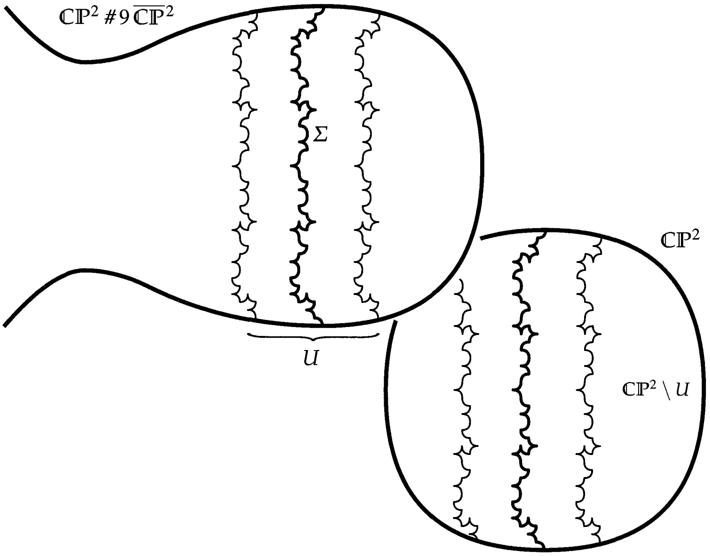
Assume first that α can be represented by a *smoothly*-embedded 2–sphere S in E. Then a neighborhood of S would look like a disk-bundle on S^2 with Euler number $+1$. In other words, a neighborhood of S in E would be diffeomorphic to a smooth tubular neighborhood of $\mathbb{CP}^1$ inside $\mathbb{CP}^2$, with boundary a copy of S^3 and complement a standard 4–ball. Then we could cut S and its neighborhood out of E, and glue-in that 4–ball instead, thus obtaining a smooth 4–manifold N with intersection form $-E_8 \oplus \left[-1\right]$, whose existence is prohibited.

Therefore α *cannot* be represented by a smoothly-embedded sphere in E. Nonetheless, by using Casson handles we can represent α by a *topologically-*embedded sphere Σ in E. More, the Casson handle itself provides a nice neighborhood U of Σ in E. Topologically, U is just a $+1$ disk-bundle over S^2, and therefore it is homeomorphic to a subset of $\mathbb{CP}^2$. Hence, by using an embedding of U into $\mathbb{CP}^2$, we can transport the topological sphere Σ from E to $\mathbb{CP}^2$. From Freedman's work (applied to open manifolds) it follows that the complement of Σ in $\mathbb{CP}^2$ must be homeomorphic to an open 4–ball. The ball $\mathbb{CP}^2 \setminus \Sigma$ inherits a smooth structure from its embedding in $\mathbb{CP}^2$, but this ball cannot be smoothly-standard: we have stumbled upon an exotic $\mathbb{R}^4$. See figure 5.5, then figure 5.6 on the next page.



5.5. Finding an exotic $\mathbb{R}^4$ inside $\mathbb{CP}^2$, I

Indeed, the open set $\mathbb{CP}^2 \setminus \Sigma$ cannot be diffeomorphic to $\mathbb{R}^4$: In standard $\mathbb{R}^4$, every compact subset can be surrounded by a smooth 3–sphere. If we had $\mathbb{CP}^2 \setminus \Sigma \approx \mathbb{R}^4$, then its compact subset $\mathbb{CP}^2 \setminus U$ could be surrounded inside $\mathbb{CP}^2 \setminus \Sigma$ by some smooth 3–sphere Z. This Z would live inside U, and in there Z would separate Σ from the end of U. Since U can also be viewed as a part of E, we could then transport Z to E. That would mean that Σ is surrounded in E by a neighborhood bounded by the smooth 3–sphere Z. This neighborhood could then be cut out of E and replaced by a standard 4–ball. However, this creates a forbidden manifold N with $Q_N = -E_8 \oplus \left[-1\right]$. Therefore, there cannot be such a smooth 3–sphere Z, and hence $\mathbb{CP}^2 \setminus \Sigma$ is not diffeomorphic to $\mathbb{R}^4$.



5.6. Finding an exotic $\mathbb{R}^4$ inside $\mathbb{CP}^2$, II

Notice that this exotic $\mathbb{R}^4$ embeds in $\mathbb{CP}^2$. Peculiarly enough, it does not embed in $\overline{\mathbb{CP}}^2$. Indeed, the exotic $\mathbb{R}^4$ identified here is not diffeomorphic to an oppositely-oriented copy of itself. For convenience, denote an exotic $\mathbb{R}^4$ obtained as above by $\mathbb{ER}^4_{\mathbb{CP}^2}$.

Also observe the deep relationship between the minimum genus needed for smoothly representing a fixed homology class and the peculiarities of smooth 4-dimensional topology. This minimum genus problem will be explored further in chapter 11 (starting on page 481).

Other large exotic $\mathbb{R}^4$***'s.*** Similarly to the above, we can alternatively start with the manifold

$$\mathbb{CP}^2 \# 10\overline{\mathbb{CP}}^2$$

and write its intersection form as

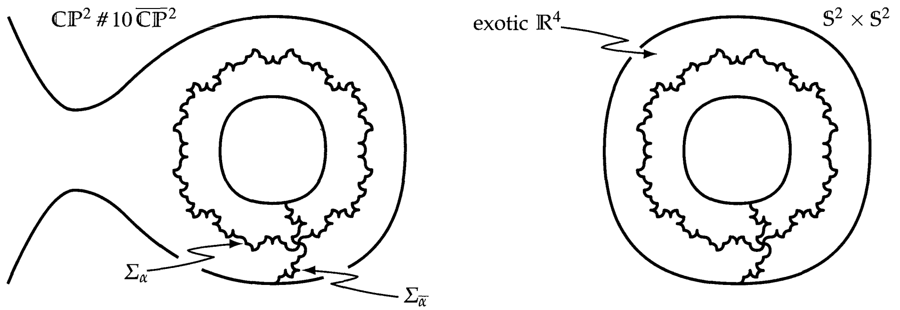
$$-E_8 \oplus [-1] \oplus H .$$

We then represent two classes α and $\bar{\alpha}$ that span H by two topological spheres Σ_α and $\Sigma_{\bar{\alpha}}$, which can then be transported (together with their surrounding Casson handles) into $S^2 \times S^2$, where they cut out an exotic $\mathbb{R}^4$ as their complement, as in figure 5.7 on the facing page.

Another example can be obtained if we start with the $K3$ surface, whose intersection form is

$$Q_{K3} = \oplus 2(-E_8) \oplus 3H ,$$

and we represent the classes generating $\oplus 3H$ by topologically-embedded spheres, transport them to $\# 3\, S^2 \times S^2$, and look at their complement—we have another sighting of an exotic $\mathbb{R}^4$.



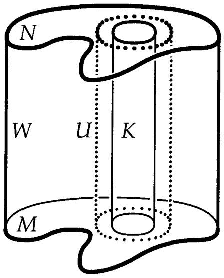
5.7. Finding an exotic $\mathbb{R}^4$ inside $\mathbb{S}^2 \times \mathbb{S}^2$

Many other such constructions can be imagined, but it is not clear if these variations can be made so as to yield diffeomorphic exotic $\mathbb{R}^4$'s or not. It is clear though that none of them embeds in standard $\mathbb{R}^4$, since each contains some compact set K that cannot be surrounded by a smooth 3–sphere, and thus cannot be smoothly embedded inside a standard 4–ball.

Exotic padding of corks

While the large exotic $\mathbb{R}^4$'s above are relatively easy to exhibit, they are hard to describe explicitly (say, in terms of a—necessarily infinite—handle decomposition). By contrast, small exotic $\mathbb{R}^4$ can be described explicitly.[21]

These small exotic $\mathbb{R}^4$'s appear in any nontrivial h–cobordism, where they surround the corresponding Akbulut cork, see figure 5.8. More precisely, we can enrich our earlier statement on corks (from section 2.4, page 89) as follows:



5.8. Exotic $\mathbb{R}^4$'s surrounding Akbulut corks

Theorem. *Let W^5 be any smooth h–cobordism between M^4 and N^4 so that W is not diffeomorphic to $M \times [0,1]$. Then there is an open sub–h–cobordism U^5 that is homeomorphic to $\mathbb{R}^4 \times [0,1]$ and contains a compact contractible sub–h–cobordism K^5, such that both W and U are trivial cobordisms outside K. In other*

---

[21] See for example **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** [**GS99**, ch 9].

words, we have diffeomorphisms

$$W \setminus K \;\cong\; (M \setminus K) \times [0,1] \qquad \textit{and} \qquad U \setminus K \;\cong\; (U \cap M \setminus K) \times [0,1]$$

(the latter can be chosen to be the restriction of the former). Furthermore, the open sets $U \cap M$ and $U \cap N$ at U's ends are homeomorphic to $\mathbb{R}^4$. If the h–cobordism W is not trivial, then the open sets $U \cap M$ and $U \cap N$ must be exotic $\mathbb{R}^4$'s. □

In conclusion, exotic $\mathbb{R}^4$'s can be found underlying all the peculiarities of smooth 4–dimensional topology. Every time the topological and smooth worlds collide, an exotic $\mathbb{R}^4$ is spawned, and there is quite a lot of them:

Let them multiply

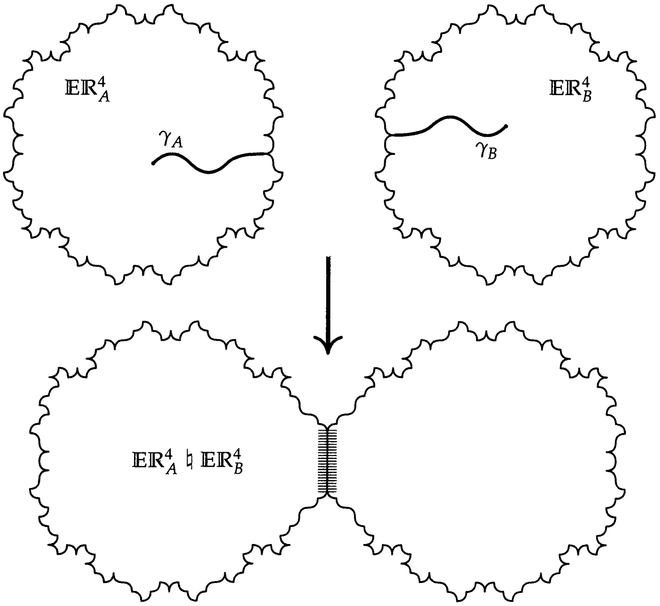
Countably-many exotic $\mathbb{R}^4$'s. Consider two random exotic $\mathbb{R}^4$'s, denoted by $\mathbb{ER}^4_A$ and $\mathbb{ER}^4_B$. In each, smoothly embed a path

$$\gamma_A \colon [0,\infty) \longrightarrow \mathbb{ER}^4_A \qquad \text{and} \qquad \gamma_B \colon [0,\infty) \longrightarrow \mathbb{ER}^4_B \,,$$

both going toward infinity, as in figure 5.9. Discard the start-points $\gamma_A(0)$ and $\gamma_B(0)$; the remainder of the paths have open tubular neighborhoods U_A and U_B, both diffeomorphic to $(0,\infty) \times \mathbb{R}^3$; the boundaries of such neighborhoods are copies of $\mathbb{R}^3$ (think of a test-tube shape).

We can then build the **end sum**

$$\mathbb{ER}^4_A \natural \mathbb{ER}^4_B$$



5.9. End summing two exotic $\mathbb{R}^4$'s by removing U_A from $\mathbb{ER}^4_A$ and U_B from $\mathbb{ER}^4_B$ and gluing together the newly-appeared boundary–$\mathbb{R}^3$'s. This creates a new 4–manifold, homeomorphic to $\mathbb{R}^4$ but not diffeomorphic to it.

For example, since $\mathbb{ER}^4_{\mathbb{CP}^2}$ does not embed in $\overline{\mathbb{CP}}^2$ while $\overline{\mathbb{ER}}^4_{\mathbb{CP}^2}$ does not embed in $\mathbb{CP}^2$, this implies that their end sum $\mathbb{ER}^4_{\mathbb{CP}^2} \natural\, \overline{\mathbb{ER}}^4_{\mathbb{CP}^2}$ will embed in neither $\mathbb{CP}^2$ nor $\overline{\mathbb{CP}}^2$. We have created a new exotic $\mathbb{R}^4$.

One can also build repeated end sums of n copies of a same exotic $\mathbb{ER}^4$, denoted by $\natural\, n\, \mathbb{ER}^4$, or even of countably-many copies, denoted by $\natural\, \infty\, \mathbb{ER}^4$. One can then show that:

Theorem (*R. Gompf*). *No two of* $\natural\, n\, \mathbb{ER}^4_{\mathbb{CP}^2}$ *are diffeomorphic.* □

We have thus exhibited countably-many distinct exotic $\mathbb{R}^4$'s. However, this is just a meek beginning:

Uncountably-many exotic $\mathbb{R}^4$***'s.*** We will use standard 4–balls of various radii to cut uncountably many new exotic $\mathbb{R}^4$'s inside a given exotic $\mathbb{R}^4$:

Pick some *large* exotic $\mathbb{R}^4$, for example $\mathbb{ER}^4_{\mathbb{CP}^2}$, and denote it by $\mathbb{ER}^4$. Then choose some homeomorphism

$$h\colon \mathbb{R}^4 \xrightarrow{\simeq} \mathbb{ER}^4$$

between standard $\mathbb{R}^4$ and our exotic $\mathbb{ER}^4$. Denote by

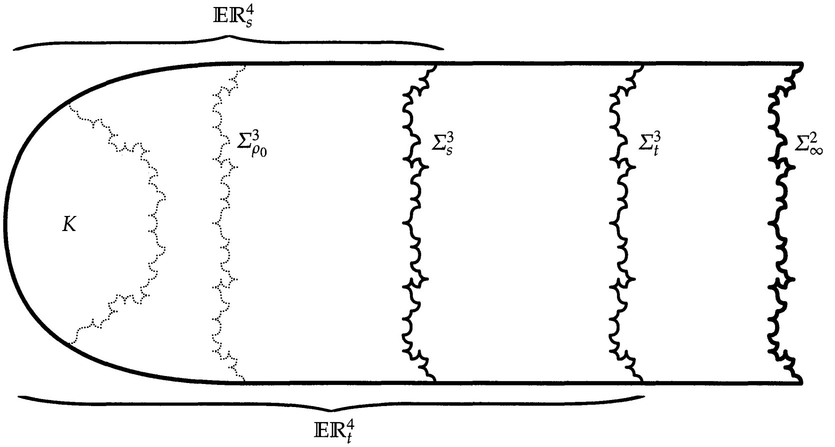
$$\mathbb{ER}^4_\rho = h\big[\mathbb{D}^4(\rho)\big]$$

the image through h of the standard open 4–ball $\mathbb{D}^4(\rho)$ of radius ρ in $\mathbb{R}^4$. See figure 5.10 on the next page. Since $\mathbb{ER}^4$ is large, there is a compact set K in $\mathbb{ER}^4$ that cannot be surrounded by any smooth 3–sphere. Nonetheless, there must be some ρ_0 big enough so that $K \subset \mathbb{ER}^4_{\rho_0}$. Then all bigger $\mathbb{ER}^4_\rho$'s are smoothly distinct:

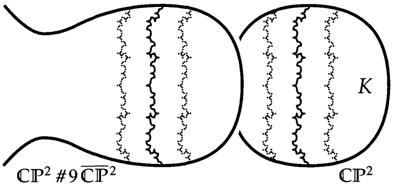
Theorem (*C. Taubes*). *If* $\mathbb{ER}^4$ *denotes the exotic* $\mathbb{R}^4$ *that we built earlier inside* $\mathbb{CP}^2$*, then, for any two distinct* $s, t > \rho_0$*, the slices* $\mathbb{ER}^4_s$ *and* $\mathbb{ER}^4_t$ *are never diffeomorphic.*

Therefore, there are uncountably-many distinct exotic $\mathbb{R}^4$'s.

> ***Proof.*** Recall how $\mathbb{ER}^4 = \mathbb{ER}^4_{\mathbb{CP}^2}$ was built: we first identified a topological 2–sphere Σ^2_∞ in $E = \mathbb{CP}^2 \# 9\, \overline{\mathbb{CP}}^2$ that spanned the orthogonal complement of $-E_8 \oplus \big[-1\big]$ in Q_E, then we transported Σ^2_∞ to $\mathbb{CP}^2$, where its complement $\mathbb{ER}^4 = \mathbb{CP}^2 \setminus \Sigma^2_\infty$ was our large exotic $\mathbb{R}^4$. The sphere Σ^2_∞ came together with a tubular neighborhood U that could be embedded around Σ^2_∞ both in E and in $\mathbb{CP}^2$. The compact set $K = \mathbb{CP}^2 \setminus U$ cannot be surrounded by any smooth 3–sphere in $\mathbb{ER}^4$.



5.10. Slicing new exotic $\mathbb{R}^4$'s



5.11. Remembering $\mathbb{E}\mathbb{R}^4_{\mathbb{CP}^2}$

Choose a homeomorphism $h \colon \mathbb{R}^4 \simeq \mathbb{E}\mathbb{R}^4$, then build the cuts $\mathbb{E}\mathbb{R}^4_\rho = h\big[\mathbb{D}^4(\rho)\big]$ and pick ρ_0 so that K be contained in $\mathbb{E}\mathbb{R}^4_{\rho_0}$. Denote by

$$\Sigma^3_\rho = h\big[\mathbb{S}^3(\rho)\big]$$

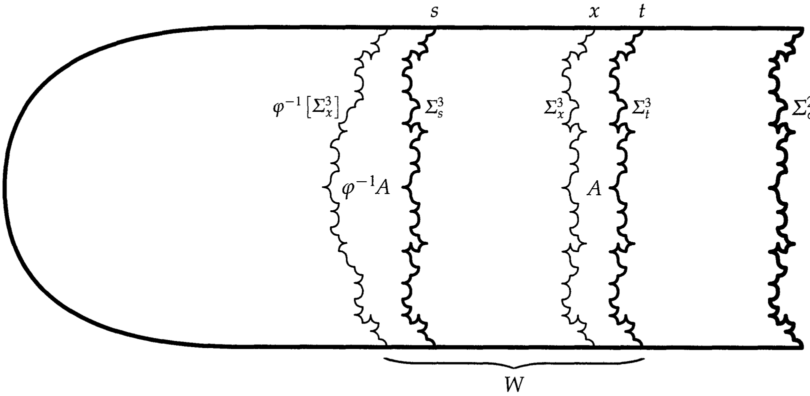
the image through h of the standard 3–sphere of radius ρ. Thus each Σ^3_ρ bounds $\mathbb{E}\mathbb{R}^4_\rho$ inside $\mathbb{E}\mathbb{R}^4$. See again figure 5.10. In particular, K is surrounded by the topological 3–sphere $\Sigma^3_{\rho_0}$.

Choose two random s and t with $\rho_0 < s < t$, and assume that $\mathbb{E}\mathbb{R}^4_s$ and $\mathbb{E}\mathbb{R}^4_t$ are diffeomorphic. We will argue that this cannot happen.

Building a brick. Denote by φ a diffeomorphism

$$\varphi \colon \mathbb{E}\mathbb{R}^4_s \xrightarrow{\cong} \mathbb{E}\mathbb{R}^4_t \,.$$

Choose some random x between s and t. Between Σ^3_x and Σ^3_t sits the open annulus $A = \mathbb{E}\mathbb{R}^4_t \setminus \mathbb{E}\mathbb{R}^4_x$, homeomorphic to $\mathbb{S}^3 \times (0,1)$ and contained in $\mathbb{E}\mathbb{R}^4_t$. This annulus can be pulled back through φ to an annulus $\varphi^{-1}A = \varphi^{-1}\big[\mathbb{E}\mathbb{R}^4_t \setminus \mathbb{E}\mathbb{R}^4_x\big]$, contained in $\mathbb{E}\mathbb{R}^4_s$ and bounded on one side by Σ^3_s and on the other by the 3–sphere $\varphi^{-1}[\Sigma^3_x]$. See figure 5.12 on the next page.

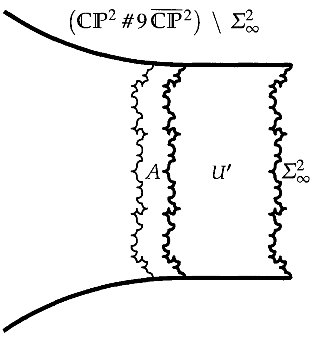


5.12. Slicing a double-edged segment

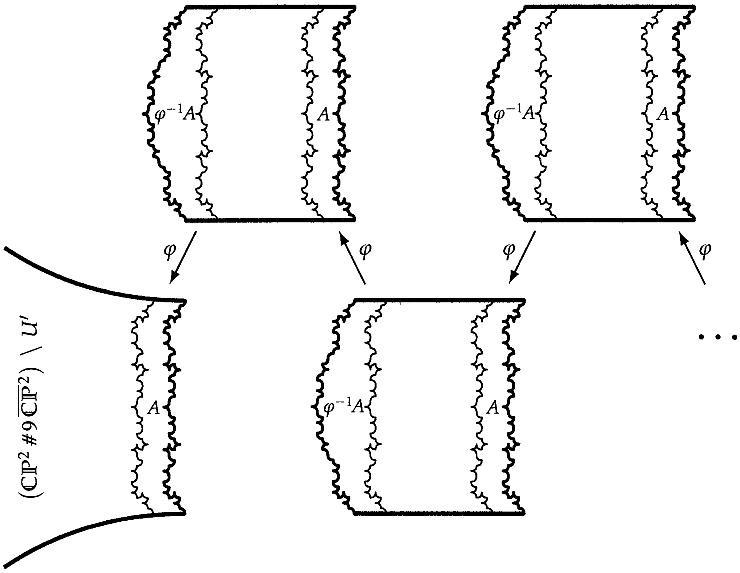
Consider the open 4–manifold W caught between $\varphi^{-1}[\Sigma^3_x]$ and Σ^3_t: its two ends A and $\varphi^{-1}A$ are diffeomorphic. Notice that we can use the diffeomorphism $\varphi\colon \varphi^{-1}A \cong A$ to glue end-to-end countably many copies of this W.

Let us stack. Remember that $\mathbb{E}\mathbb{R}^4 \setminus \mathbb{E}\mathbb{R}^4_s$ is a neighborhood of the 2–sphere Σ^2_∞ and that it sits inside U. Therefore it can be transported back to $E = \mathbb{CP}^2 \# 9\overline{\mathbb{CP}}^2$, together with all its contents. Also notice that $U \setminus \Sigma^2_\infty = \mathbb{E}\mathbb{R}^4 \setminus K$.

Inside E, we see Σ^2_∞ separated from the rest of E by the annulus A, as in figure 5.13. If we cut out from E the smaller neighborhood $U' = \mathbb{E}\mathbb{R}^4 \setminus \mathbb{E}\mathbb{R}^4_t$ of Σ^2_∞, we are left with an *open* 4–manifold, whose intersection form is $-E_8 \oplus \left[-1\right]$ and that has the annulus A as its end. To this end we can now attach a stack of countably-many copies of W glued end-to-end, as in figure 5.14 on the next page. The result is a smooth open 4–manifold, with *periodic end* and with intersection form $-E_8 \oplus \left[-1\right]$.



5.13. Preparing the stump



5.14. Building an open manifold with periodic end

It turns out that the end-periodicity offers just enough regularity at infinity to permit a version of the argument of Donaldson's theorem to be applied to this open manifold. The analysis of anti-self-dual connections on manifolds with periodic ends proceeds similarly, and thus it leads to the same exclusions of intersection forms as for closed 4–manifolds. Specifically:

Theorem *(C. Taubes). Let M° be a smooth open simply-connected 4–manifold with only one end. Assume that the end of M° is periodic. If the intersection form Q_{M° of M° is definite, then it has to be isomorphic to either $\oplus\, m\,[-1]$ or $\oplus\, m\,[+1]$.* □

In particular, this shows that our open manifold with intersection form $-E_8 \oplus [-1]$ cannot exist. The contradiction stems from our supposed ability to build a periodic end for it, which was a consequence of our assumption that $\mathbb{E}\mathbb{R}^4_s \cong \mathbb{E}\mathbb{R}^4_t$. Therefore each of the $\mathbb{E}\mathbb{R}^4_t$'s must be smoothly distinct. □

A corollary of these results is that every open 4–manifold with ends homeomorphic to $S^3 \times \mathbb{R}$ will admit uncountably many distinct smooth structures.[22] In particular, any closed 4–manifold M admits uncountably many

---

[22] Existence of at least one smooth structure is a consequence of work of **M. Freedman** and **F. Quinn**; see ***Topology of 4–manifolds*** **[FQ90]**. For uncountably-many, see **R. Gompf's** ***An exotic menagerie*** **[Gom93]**.

smooth structures on $M \setminus \{point\}$. It is unknown whether all *open* 4–manifolds admit uncountably-many smooth structures.[23]

An argument similar to the above can be applied to *small* exotic $\mathbb{R}^4$'s and yields series of uncountably many exotic $\mathbb{R}^4$'s, all smoothly-embedding in standard $\mathbb{R}^4$.

Even more, there is a maximal (or universal) exotic $\mathbb{R}^4$ in which all possible exotic $\mathbb{R}^4$'s embed smoothly.

It is also worth noting that every exotic $\mathbb{ER}^4$ becomes standard after crossing with a line: we have $\mathbb{ER}^4 \times \mathbb{R} \cong \mathbb{R}^5$. This diffeomorphism induces a smooth nowhere-zero vector field on $\mathbb{R}^5$, but not much is known concretely about such descriptions.

Finally, one could ask the question: *Does building exotic $\mathbb{R}^4$'s need Donaldson, or could exotic $\mathbb{R}^4$'s be exhibited as a consequence of Rokhlin's exclusions?* The answer is negative, and one reason was touched upon in the end-notes of the preceding chapter (smoothing topological manifolds, page 221). Specifically, a theorem of F. Quinn allows us to find, for every *non-smoothable* topological 4–manifold with vanishing Kirby–Siebenmann invariant, an exotic $\mathbb{R}^4$ inside $\mathbb{CP}^2$ or $\mathbb{S}^2 \times \mathbb{S}^2$; since in dimension 4 the Kirby–Siebenmann invariant only excludes what Rokhlin's theorem (or its generalizations) prohibits, it follows that other methods are needed to prove these $\mathbb{R}^4$'s to be exotic.

---

[23] Keep in mind that a *closed* 4–manifold can admit at most countably-many smooth structures.

5.5. Notes

Note: Plumbing topological 4–manifolds into existence

Let $Q\colon Z \times Z \to \mathbb{Z}$ be any symmetric bilinear form, not necessarily unimodular. By plumbing, we will build a simply-connected 4–manifold P_Q with non-empty boundary and intersection form Q. Then we will prove that ∂P_Q is a homology 3–sphere if and only if Q is unimodular. When that happens, we know there exists a fake 4–ball Δ that can be glued to P_Q and thus yield a closed topological 4–manifold with intersection form Q.

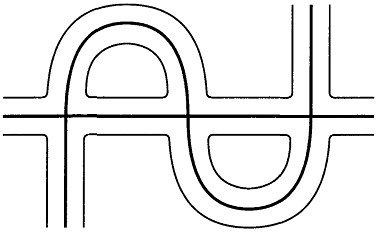
Plumbing manifolds with arbitrary intersection forms. Let Q be a random symmetric bilinear integral form and express it as an $m \times m$ matrix $[a_{ij}]$. We will use the a_{ij}–data as a plumbing recipe for building P_Q.

Pick m copies of $\mathbb{S}^2$ and denote them by $S_1, \dots, S_m$. On each S_k, build the 2–disk bundle E_k of Euler class a_{kk}; then $S_k \cdot S_k = a_{kk}$ as submanifold of E_k.

Now, plumb the bundles E_i and E_j a_{ij}–times; for example, if $a_{ij} = 2$, plumb E_i and E_j twice, arranging for their cores S_i and S_j to meet positively; if $a_{ij} = -2$, also plumb twice, but in such a manner that S_i and S_j meet negatively. Call E_Q the resulting smooth 4–manifold (with non-empty boundary).

Clearly, $H_2(E_Q;\mathbb{Z})$ is generated by the m classes $S_1, \dots, S_m$, and their intersections are given by the a_{ij}'s. Therefore, the intersection form of E_Q is Q.

The problem with E_Q is that, if any of the a_{ij}'s (with $i \neq j$) is different from -1, 0, or 1, then E_Q is not simply-connected. Indeed, whenever two bundles E_i and E_j are plumbed more than once, a loop is created, as hinted in figure 5.15.



5.15. Plumbing may create loops

The fundamental group of E_Q is free, with n generators for each $(n+1)$–plumbing of a pair of the E_k's. (The fundamental group of ∂E_Q is not free, but at least $\pi_1(\partial E_Q) \to \pi_1(E_Q)$ is surjective.) Represent the generators of $\pi_1(E_Q)$ as circles disjointly-embedded in ∂E_Q. Attach to each such circle a 2–handle $\mathbb{D}^2 \times \mathbb{D}^2$. This will kill $\pi_1(E_Q)$ without creating new 2–homology. (Notice that the result depends on the way we glue these handles, *i.e.*, on their framing.)

Call the result P_Q. It is a simply-connected smooth 4–manifold whose intersection form is Q. Its boundary ∂P_Q is in general not simply-connected. Nonetheless, as we will argue next, if Q is unimodular, then at least ∂P_Q is a homology 3–sphere.

Unimodular makes homology spheres.

Lemma. *Let M^4 be an oriented 4–manifold with boundary and let Q be its intersection form. Assume $H_1(M;\mathbb{Z}) = 0$. Then ∂M is a homology 3–sphere if and only if Q is unimodular.*

Proof. On one hand, Poincaré duality offers an isomorphism

$$H_2(M, \partial M;\ \mathbb{Z}) \approx H^2(M;\mathbb{Z})\ .$$

Since $H_1(M;\mathbb{Z}) = 0$, the universal coefficient theorem (page 15) implies that $H^2(M;\mathbb{Z}) = \mathrm{Hom}\big(H_2(M;\mathbb{Z}),\ \mathbb{Z}\big)$, and hence we further have an isomorphism

$$H_2(M, \partial M;\ \mathbb{Z}) \approx \mathrm{Hom}\big(H_2(M;\mathbb{Z}),\ \mathbb{Z}\big)\ .$$

On the other hand, the intersection form $Q\colon H_2(M;\mathbb{Z}) \times H_2(M;\mathbb{Z}) \to \mathbb{Z}$ can be viewed as a morphism

$$\widehat{Q}\colon H_2(M;\mathbb{Z}) \longrightarrow \mathrm{Hom}\big(H_2(M;\mathbb{Z}),\ \mathbb{Z}\big)\ .$$

The pairing Q is unimodular if and only if the morphism $\widehat{Q}$ is an isomorphism.

Let $j\colon M \subset (M, \partial M)$ be the inclusion and consider the diagram:

$$\begin{tikzcd}
H_2(M;\mathbb{Z}) \arrow[r, "j_*"'] \arrow[d, "\widehat{Q}"] & H_2(M, \partial M;\ \mathbb{Z}) \arrow[d, "PD", "\approx"'] \\
\mathrm{Hom}\big(H_2(M;\mathbb{Z}),\ \mathbb{Z}\big) \arrow[r, equal] & \mathrm{Hom}\big(H_2(M;\mathbb{Z}),\ \mathbb{Z}\big)\ .
\end{tikzcd}$$

An easy argument using the commuting of this diagram shows that $\widehat{Q}$ is injective or surjective if and only if j_* is as well. Thus, the intersection form Q is unimodular if and only if j_* is an isomorphism.

However, j_* fits into the homology exact sequence

$$H_3(M, \partial M;\ \mathbb{Z}) \xrightarrow{\partial_*} H_2(\partial M;\mathbb{Z}) \xrightarrow{i_*} H_2(M;\mathbb{Z})$$
$$\xrightarrow{j_*} H_2(M, \partial M;\ \mathbb{Z}) \xrightarrow{\partial_*} H_1(\partial M;\mathbb{Z}) \xrightarrow{i_*} H_1(M;\mathbb{Z})\ ,$$

where $i\colon \partial M \subset M$ is the inclusion. We assumed $H_1(M;\mathbb{Z}) = 0$; further, by Poincaré duality $H_3(M, \partial M;\ \mathbb{Z}) \approx H^1(M;\mathbb{Z}) = \mathrm{Hom}\big(H_1(M;\mathbb{Z});\ \mathbb{Z}\big) = 0$. Thus, the above sequence reduces to

$$0 \longrightarrow H_2(\partial M;\mathbb{Z}) \xrightarrow{i_*} H_2(M;\mathbb{Z}) \xrightarrow{j_*} H_2(M, \partial M;\ \mathbb{Z}) \xrightarrow{\partial_*} H_1(\partial M;\mathbb{Z}) \longrightarrow 0\ .$$

Exactness implies that $\operatorname{Ker} j_* = \operatorname{Im} i_*$, but the morphism i_* is injective. Hence, j_* is injective if and only if $H_2(\partial M; \mathbb{Z}) = 0$. Exactness also implies that $\operatorname{Im} j_* = \operatorname{Ker} \partial_*$, but ∂_* is surjective. Hence, j_* is is surjective if and only if $H_1(\partial M; \mathbb{Z}) = 0$.

In conclusion, Q is unimodular if and only if ∂M is a homology sphere. □

Higher plumbings. *The construction of P_Q can, of course, be replicated in all higher $4k$–dimensions (just as we did for P_{E_8} in the end-notes of chapter 2, back on page 97), by starting with $\mathbb{D}^{2k}$–bundles over $2k$–spheres. When $k \geq 2$, the result of plumbing has additionally the property that*[1] *$\pi_1(\partial E_Q^{4k}) = \pi_1(E_Q^{4k})$, and thus, after adding 2–handles, the resulting $4k$–manifold P_Q^{4k} will in fact have simply-connected boundary. If moreover Q is unimodular, then ∂P_Q^{4k} is not only a homology sphere, but in fact a homotopy sphere, and thus must be homeomorphic to $\mathbb{S}^{4k-1}$. For a more rigorous discussion of this case, see* **W. Browder***'s* ***Surgery on simply-connected manifolds*** [**Bro72**, ch V]. *Compare also with the end-notes of chapter 2 (plumbing exotic spheres, page 97).*

Closed 4–manifolds. Given a symmetric form Q, we build the 4–manifold P_Q. If Q is unimodular, then ∂P_Q is a homology 3–sphere, and by Freedman's theorem on fake 4–balls (page 83) its reverse $\overline{\partial P_Q}$ must bound a fake 4–ball Δ. Then we can build the closed topological 4–manifold

$$\mathcal{M}_Q = P_Q \cup_\partial \Delta \,,$$

whose intersection form is, of course, Q.

This concludes a complete argument for the existence part of Freedman's classification theorem.

Note: Classification of indefinite odd forms

If we accept the hard result of Meyer's lemma, then the classification of all indefinite odd forms follows pretty easily. In this note we will present the proof of this part of Serre's classification theorem.

Recall that Meyer's lemma stated that, for every indefinite bilinear symmetric unimodular form $Q \colon Z \times Z \to \mathbb{Z}$, there must exist an element $x \in Z$ so that

$$Q(x, x) = 0 \,.$$

We can, of course, assume that this x is indivisible, *i.e.*, not a multiple of some other element. Then, since Q is unimodular, there also exists some y so that

$$x \cdot y = 1 \,.$$

Denote by $Z^\perp$ the Q–orthogonal complement of the span of x and y. Clearly the restriction of Q to this $Z^\perp$ is still unimodular.

If $y \cdot y$ is even, then since Q was odd $Q|_{Z^\perp}$ must be odd. Therefore there is some element $z \in Z^\perp$ so that $z \cdot z$ is odd, but then $(y + z) \cdot (y + z) = y \cdot y + z \cdot z$ will also be odd, and we still have $x \cdot (y + z) = 1$. Therefore, maybe after replacing y with $y + z$, we can always assume that $y \cdot y$ is odd.

---

[1] In dimension 4, we merely have a surjection $\pi_1 \, \partial E_Q^4 \to \pi_1 \, E_Q^4$.

It follows that the restriction of Q to $A = \mathbb{Z}\{x, y\}$ is described by

$$Q|_A = \begin{bmatrix} 0 & 1 \\ 1 & odd \end{bmatrix} .$$

Since for $k \in \mathbb{Z}$ we have $(y + kx) \cdot (y + kx) = y \cdot y - 2k$, that means that there is a suitable k so that, after changing basis from $\{x, y\}$ to $\{x,\ y + kx\}$, we write

$$Q|_A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} .$$

Further, we have

$$\begin{aligned} (x - (y + kx)) \cdot (x - (y + kx)) &= -2x \cdot y + (y + kx) \cdot (y + kx) = -1 , \\ (x - (y + kx)) \cdot (y + kx) &= x \cdot y - (y + kx) \cdot (y + kx) = 0 . \end{aligned}$$

Therefore, if we change to the basis $\{x - (y + kx),\ \ y + kx\}$, then we get

$$Q|_A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} .$$

Recall that $Z = A \oplus Z^{\perp}$. If $Q|_{Z^{\perp}}$ is indefinite, then we can discard A and repeat the whole procedure above anew on $Z^{\perp}$. If $Q|_{Z^{\perp}}$ happens to be positive-definite, then the Q–orthogonal complement of $y + kx$ is $\mathbb{Z}\{x - (y + kx)\} \oplus Z^{\perp}$, over which Q must be indefinite. We can then discard $y + kx$ and restart the procedure above on $\mathbb{Z}\{x - (y + kx)\} \oplus Z^{\perp}$. If $Q|_{Z^{\perp}}$ is negative-definite, then Q is indefinite on the complement of $x - (y + kx)$.

Repeating this algorithm, until we eventually exhaust all of Z, proves that Q can always be written as

$$Q = \oplus m \, [-1] \oplus n \, [+1] \ .$$

Note: Proof of van der Blij's lemma

Let $Q\colon Z \times Z \to \mathbb{Z}$ be a symmetric bilinear unimodular form, defined over a free $\mathbb{Z}$–module Z. A characteristic element for Q is any $\underline{w} \in Z$ such that for all $x \in Z$ we have $\underline{w} \cdot x = x \cdot x \pmod 2$. Van der Blij's lemma states that for every characteristic element we have the congruence $\underline{w} \cdot \underline{w} = \operatorname{sign} Q \pmod 8$.

Any two characteristic elements $\underline{w}'$ and $\underline{w}''$ must differ by an even element, $\underline{w}'' = \underline{w}' + 2y$. Then we have

$$\begin{aligned} \underline{w}'' \cdot \underline{w}'' &= (\underline{w}' + 2y) \cdot (\underline{w}' + 2y) \\ &= \underline{w}' \cdot \underline{w}' \ + 4\,(\underline{w}' \cdot y + y \cdot y) \\ &\equiv \underline{w}' \cdot \underline{w}' \pmod 8 , \end{aligned}$$

since $\underline{w}' \cdot y + y \cdot y = 0 \pmod 2$ and thus must always be even.

Therefore, every symmetric bilinear unimodular form $Q\colon Z \times Z \to \mathbb{Z}$ has a mod 8 invariant determined by the self-intersection $\underline{w} \cdot \underline{w} \pmod 8$ of any characteristic element $\underline{w}$. In what follows, we prove that this modulo 8 invariant coincides with the signature.

(1) Assume first that our form Q is odd and indefinite. Then, by using the classification proved in the preceding note, we must have

$$Q = \oplus m \, [+1] \oplus n \, [-1] \ .$$

If $\{e_1^+, \ldots, e_m^+, e_1^-, \ldots, e_n^-\}$ is the corresponding basis of Z, then it is clear that

$$\underline{w} = e_1^+ + \cdots + e_m^+ + e_1^- + \cdots + e_n^-$$

is in fact a characteristic element of Q. Its self-intersection is exactly $\operatorname{sign} Q$. Thus for every characteristic element $\underline{w}'$ of Q we must have $\underline{w}' \cdot \underline{w}' = \operatorname{sign} Q \pmod 8$.

(2) If Q is *not* odd indefinite, then by adding $[-1]$ and $[+1]$ we obtain a new form

$$Q \oplus [+1] \oplus [-1] \ .$$

This new form is odd indefinite, and thus the above argument applies to it. Therefore, for every characteristic element $\underline{w}^\circ$ of $Q \oplus [-1] \oplus [+1]$, we must have that $\underline{w}^\circ \cdot \underline{w}^\circ = \operatorname{sign}(Q \oplus [-1] \oplus [+1]) \pmod 8 = \operatorname{sign} Q \pmod 8$. Now, if we choose some characteristic element $\underline{w}$ of Q and we add e^+ and e^- to it (where $e^\pm$ span the new $[\pm 1]$-terms), then we do obtain a characteristic element

$$\underline{w}^\circ = \underline{w} + e^+ + e^-$$

of $Q \oplus [-1] \oplus [+1]$. Therefore $\operatorname{sign} Q = (\underline{w} + e^+ + e^-) \cdot (\underline{w} + e^+ + e^-) \pmod 8$ but $(\underline{w} + e^+ + e^-) \cdot (\underline{w} + e^+ + e^-) = \underline{w} \cdot \underline{w}$, which concludes the argument. □

Note: Counting definite forms

In this note, we add a few details about the count of all symmetric unimodular forms that are definite. Along the way, two more examples of definite forms (the Leech lattice and the Γ_{4k}'s) are defined. For counting even definite forms, one uses the Minkowski–Siegel mass formula. Similar results exist for odd definite forms, but we will not discuss them here.

A consequence of van der Blij's lemma is that an even positive-definite form always has rank divisible by 8. Denote by $\mathcal{Q}_{8k}$ the set of isomorphism classes of even definite forms of rank $8k$. For any $Q \in \mathcal{Q}_{8k}$, we denote by g_Q the order (= cardinality) of the automorphism group $\operatorname{Aut}(Q)$ of Q. Often enough, this group is huge.

Minkowski–Siegel Mass Formula. *If $\mathcal{Q}_{8k}$ is the set of isomorphism types of even positive-definite forms of rank $8k$, then we have*

$$\sum_{Q \in \mathcal{Q}_{8k}} \frac{1}{g_Q} = 2^{1-8k} \frac{1}{(4k)!} B_{2k} \prod_{j=1}^{4k-1} B_j \, ,$$

where the B_n's are the Bernoulli numbers. □

A first thing to notice is this result does not offer a direct count of forms, but a count where each form is weighed by its "mass" $1/g_Q$, which can be very small.

The Bernoulli numbers B_n are a sequence of rational numbers that at first are less than 1, but soon enough begin to grow tremendously.[2] Therefore, the numbers provided by the Minkowski–Siegel formula quickly start getting very big. A few numerical examples are presented in table IX on the next page.

---

[2] The Bernoulli numbers are defined by the identity $\frac{x}{e^x - 1} = 1 - \frac{x}{2} - \sum_{n=0}^{\infty} (-1)^n \frac{B_n x^{2n}}{(2n)!}$.
Here are a few values: $B_1 = 1/6$, $B_2 = 1/30$, $B_3 = 1/42$, $B_4 = 1/30$, $B_5 = 5/66$, $B_6 = 691/2730$, $B_7 = 7/6$, $B_8 = 3617/510$, $B_9 = 43867/798$, $B_{10} = 174611/330$, $B_{11} = 854513/138$. Various explicit general formulae exist. The Bernoulli numbers are intimately connected with ζ-functions.

IX. Numbers from the Minkowski–Siegel formula

$k = 1$, rank $= 8$	$10^{-9} \cdot 1.4352\ldots$
$k = 2$, rank $= 16$	$10^{-18} \cdot 2.4885\ldots$
$k = 3$, rank $= 24$	$10^{-15} \cdot 7.9369\ldots$
$k = 4$, rank $= 32$	$10^{7} \cdot 4.0309\ldots$
$k = 5$, rank $= 40$	$10^{51} \cdot 4.3930\ldots$

The typical use of the formula is as follows: Choose a list of forms $Q_1, \ldots, Q_n$ from $\mathscr{Q}_{8k}$, determine each of their automorphism groups, compute the sum $\sum 1/g_{Q_k}$; if you achieved the same total as that from the Minkowski–Siegel formula, then your list must in fact exhaust all of $\mathscr{Q}_{8k}$.

Let us inspect a few low ranks:

(1) For rank 8, the E_8–form has $|\mathrm{Aut}(E_8)| = 2^{14} \cdot 3^5 \cdot 5^2 \cdot 7$ which numerically is

$$|\mathrm{Aut}(E_8)| = 696,729,600 .$$

After comparison with the Minkowski–Siegel number, it turns out that E_8 is the only form of rank 8. A most remarkable fact is that the E_8–lattice is highly symmetric: the subgroup of $\mathrm{Aut}(E_8)$ made from orientation-preserving automorphisms is a *simple* group, of order $174,182,400 = 2^{12} \cdot 3^5 \cdot 5^2 \cdot 7$.

(2) For rank 16, there are two isomorphism types, $E_8 \oplus E_8$ and Γ_{16}.

Definition of Γ_{16}. *In every rank $4k$, there is a positive-definite form Γ_{4k} defined as follows: consider the standard basis $e_1, \ldots, e_{4k}$ in Euclidean $\mathbb{R}^{4k}$, endowed with the standard inner product. Then Γ_{4k} is the lattice spanned by all $e_i + e_j$ together with $\frac{1}{2}(e_1 + \cdots e_{4k})$. In rank 8, Γ_8 is just another description of E_8. The Γ_{8k}'s are even, wile the Γ_{4k+4}'s are odd.*

(3) The list of twenty-four forms of rank 24 was determined by **H-V. Niemeier** in 1968, published as ***Definite quadratische Formen der Dimension* 24 *und Diskriminate* 1** **[Nie73]**. They include among them the Leech lattice Q_{Leech} whose automorphism group has order $|\mathrm{Aut}(Q_{Leech})| = 2^{22} \cdot 3^9 \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13 \cdot 23$. Numerically, this is merely

$$|\mathrm{Aut}(Q_{Leech})| = 8,315,553,613,086,720,000 .$$

Therefore the Leech form has a very small contribution to the Minkowski–Siegel number.

Definition of the Leech lattice. *Consider $e_1, \ldots, e_{25}$ to be the standard basis of $\mathbb{R}^{25}$. Endow $\mathbb{R}^{25}$ with the Lorentzian inner product $(x_1, \ldots, x_{25}) \cdot (y_1, \ldots, y_{25}) = x_1 y_1 + \cdots + x_{24} y_{24} - x_{25} y_{25}$. Consider also the vector of odd coordinates $w = (3,5,7,\ldots,49,51)$. Since $w \cdot w = -1$, its orthogonal complement $w^{\perp}$ is a positive-definite subspace for our inner product. The* **Leech lattice** *is made of the integral points of $w^{\perp}$. Another remarkable fact is that $\mathrm{Aut}(Q_{Leech})/Center$ is a* simple *group, of order $4,157,776,806,543,360,000 = 2^{21} \cdot 3^9 \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13 \cdot 23$.*

(4) For rank 32, the Minkowski–Siegel number is bigger that $4 \cdot 10^7$. Since every form has at least two automorphisms, we conclude that we have *at least* 80 million distinct forms. However, as the example of the Leech lattice seems to suggest, this is likely to be a gross underestimate!

A starting reference for this type of material can be **J.P. Serre**'s ***Cours d'arithmétique*** [**Ser70, Ser77**], translated as ***A course in arithmetic*** [**Ser73**]; or your favorite monograph on sphere-packings. As a general reference on forms, see **J. Milnor** and **D. Husemoller**'s ***Symmetric bilinear forms*** [**MH73**].

Bibliography

The algebraic classification of abstract integral symmetric bilinear forms is due to **J.P. Serre**'s ***Formes bilinéaires symétriques entières à discriminant*** $\pm\mathbf{1}$ [**Ser62**]; see also his ***Cours d'arithmétique*** [**Ser70, Ser77**], translated as ***A course in arithmetic*** [**Ser73**]. A proof also appeared in **J. Milnor**'s ***On simply connected*** $\mathbf{4}$***–manifolds*** [**Mil58b**]. A proof and exposition can be read in **J. Milnor** and **D. Husemoller**'s monograph ***Symmetric bilinear forms*** [**MH73**, ch II].

Meyer's lemma was proved (for rank ≥ 5) in **A. Meyer**'s ***Über die Auflösung der Gleichung*** $ax^2+by^2+cz^2+du^2+ev^2=0$ ***in ganzen Zahlen*** [**Mey84**] (that is, 1884). The result follows from hard-core algebraic number theory (*e.g.*, the Hasse–Minkowski theorem), and an outline of the argument can be found in [**MH73**, app 3]. Van der Blij's lemma appeared as **F. van der Blij**'s ***An invariant of quadratic forms mod*** $\mathbf{8}$ [**vdB59**].

The classification of simply-connected topological 4–manifolds was published in **M. Freedman**'s ***The topology of four-dimensional manifolds*** [**Fre82**] and can be found in book form in **M. Freedman** and **F. Quinn**'s ***Topology of*** $\mathbf{4}$***–manifolds*** [**FQ90**]. The latter also discusses in detail the non-simply connected case.

S.K. Donaldson's first theorem appeared in ***An application of gauge theory to four-dimensional topology*** [**Don83**]. His results were based on hard technical results obtained in **K. Uhlenbeck**'s ***Removable singularities in Yang–Mills fields*** [**Uhl82b**] and ***Connections with*** L^p ***bounds on curvature*** [**Uhl82a**] (a removable singularities theorem, leading to the compactification of the moduli space), as well as in **C. Taubes**' ***Self-dual Yang–Mills connections on non-self-dual*** $\mathbf{4}$***–manifolds*** [**Tau82**] (the moduli space is non-empty). The study of the moduli spaces of anti-self-dual connections started even earlier with **M. Atiyah, N. Hitchin** and **I. Singer**'s groundbreaking ***Self-duality in four-dimensional Riemannian geometry*** [**AHS78**] (where they prove that the moduli space is a manifold), and the topic itself has its origins in physics (Yang–Mills theory).

An exposition of this result of Donaldson can be found described briefly in **B. Lawson**'s ***The theory of gauge fields in four dimensions*** [**Law85**], extensively in **D. Freed** and **K. Uhlenbeck**'s monograph ***Instantons and four-manifolds*** [**FU84**], or included in the definitive account of Donaldson gauge theory, **S.K. Donaldson** and **P. Kronheiemer**'s ***The geometry of four-manifolds*** [**DK90**, sec 8.3]. We will come back to gauge theory later, in part 4.

The $^{11}/_{8}$–conjecture was first written down in **Y. Matsumoto** in ***On the bounding genus of homology*** $\mathbf{3}$***–spheres*** [**Mat82**], after having floated around in the topological community for some time. The first step toward it was made by **S.K. Donaldson** in ***Connections, cohomology and the intersection forms of*** $\mathbf{4}$***–manifolds*** [**Don86**], through the exclusion of the forms $\oplus \pm 2E_8 \oplus H$ and $\oplus \pm 2E_8 \oplus 2H$, in the case when $H_1(M;\mathbb{Z})$ has no torsion; the restriction on torsion was later removed in ***Irrationality and the h–cobordism conjecture*** **[Don87]**.

After the birth of Seiberg–Witten theory, in a lecture in 1994, **P. Kronheiemer** explained how to use the Seiberg–Witten equations to obtain Donaldson's exclusions; that argument is displayed in the end-notes of chapter 10 (page 454). **M. Furuta**'s $10/8$–theorem appeared in his ***Monopole equation and the*** $11/8$***–conjecture*** **[Fur01]**, and a brief outline can also be read from **[Akb96]**. The exclusion of the form $\oplus \pm 2 E_8 \oplus 5 H$ appeared in **M. Furuta, Y. Kametani** and **H. Matsue**'s ***Spin 4–manifolds with signature*** $= -32$ **[FKM01]**.

Flat exotica. The origins of exotic $\mathbb{R}^4$'s can be traced back to **A. Casson.** Roughly speaking, in a lecture in 1973 he presented (along with Casson handles) the idea later used for building large exotic $\mathbb{R}^4$'s **[Cas86,** lecture I**]**, while in a lecture in 1976 he discussed the idea for building small exotic $\mathbb{R}^4$'s from nontrivial h–cobordisms (noticing that h–cobordisms fail from being trivial owing to a set homotopy-equivalent to $\mathbb{R}^4$) **[Cas86,** lecture III**]**. Both lectures can be read from ***À la recherche de la topologie perdue*** **[GM86a]**. The only ingredients missing were Freedman's work, which would show that the proper homotopy $\mathbb{R}^4$'s of Casson must be homeomorphic to $\mathbb{R}^4$, and Donaldson's work, which would show that those cannot be diffeomorphic to $\mathbb{R}^4$.

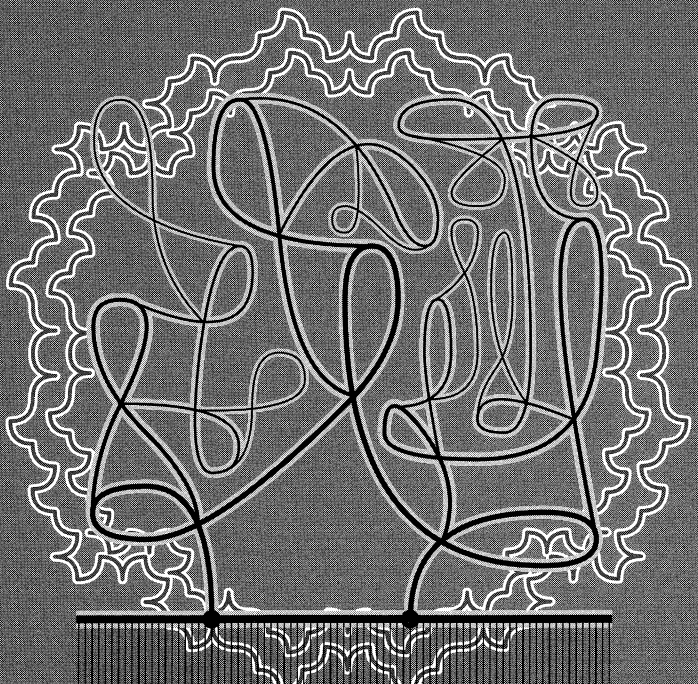
As soon as the latter ingredients appeared, many people put the pieces together and realized that exotic $\mathbb{R}^4$'s must exist. M. Freedman built an example of a small exotic $\mathbb{R}^4$ appearing from a nontrivial compact h–cobordism, but the problem with Freedman's argument at the time was that such a nontrivial h–cobordism had not yet been found (it was to be provided a bit later by **S.K. Donaldson**'s ***Irrationality and the h–cobordism conjecture*** **[Don87]**). Freedman's example can be read in **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89,** ch XIV**]**. Following these ideas, R. Kirby built a large exotic $\mathbb{R}^4$ through the failure of smooth connected sum splitting.[3] His example, together with three more, appeared in **R. Gompf**'s ***Three exotic*** $\mathbb{R}^4$***'s and other anomalies*** **[Gom83]**. Then **R. Gompf** used end sums to exhibit countably-many distinct exotic $\mathbb{R}^4$'s in ***An infinite set of exotic*** $\mathbb{R}^4$***'s*** **[Gom85]**. Soon after, **C. Taubes** found uncountably many in ***Gauge theory on asymptotically periodic 4–manifolds*** **[Tau87]**. Uncountably many small exotic $\mathbb{R}^4$'s were later exhibited in **S. DeMichelis** and **M. Freedman**'s ***Uncountably many exotic*** $\mathbb{R}^4$***'s in standard 4–space*** **[DMF92]**. The universal exotic $\mathbb{R}^4$ which contains all others is due to **M. Freedman** and **L. Taylor**'s ***A universal smoothing of four-space*** **[FT86]**. A two-parameter family of exotic $\mathbb{R}^4$'s, including both small and large exotics, was built in **R. Gompf**'s ***An exotic menagerie*** **[Gom93]**, together with handle decompositions of some exotic $\mathbb{R}^4$'s. Other explicit descriptions are contained in **Ž. Bižaka** and **R. Gompf**'s ***Elliptic surfaces and some simple exotic*** $\mathbb{R}^4$***'s*** **[BG96]**.

---

[3] The two thought at first that they had built the same example, until they started arguing whether the exotic $\mathbb{R}^4$ embeds in standard aR^4: Freedman's did, Kirby's did not...

The theorem on Akbulut corks and small exotic $\mathbb{R}^4$'s that we stated started with a preprint of **C. Curtis** and **W. Hsiang,** then was improved together with **M. Freedman** and **R. Stong** in ***A decomposition theorem for h–cobordant smooth simply-connected compact 4–manifolds*** [CFHS96], with further contributions by **R. Matveyev** in ***A decomposition of smooth simply-connected h–cobordant 4–manifolds*** [Mat96], by **R. Kirby** in ***Akbulut's corks and h–cobordisms of smooth, simply connected 4–manifolds*** [Kir96], and by **Ž. Bižaca** in ***A handle decomposition of an exotic*** $\mathbb{R}^4$ [Biž94].

A more comprehensive discussion of exotic $\mathbb{R}^4$'s can be found in **R. Gompf** and **A. Stipsicz**'s ***4–manifolds and Kirby calculus*** [GS99, ch 9].



Part III

A Survey of Complex Surfaces

ALL closed oriented 2–dimensional manifolds split as connected sums of tori. In dimension 3, W. Thurston's geometrization conjecture claims that all closed oriented 3–manifolds split into pieces that admit certain few Riemannian geometric structures, and it might have been proved in 2003.

For a while in 4–dimensional topology one entertained the conjecture that all closed oriented 4–manifolds might be decomposable as connected sums of complex algebraic surfaces (with complex or opposite orientations). The discovery of an indecomposable symplectic non-complex 4–manifold shattered that claim. The next conjecture proposed was that all closed oriented 4–manifolds might split as connected sums of symplectic manifolds (with induced or opposite orientations), but then an indecomposable 4–manifold that does not admit any symplectic structures was found... Alas, the current state of affairs is that there is *no* conjecture whatsoever.

Even though complex surfaces might have fallen from their conjectured high-status in 4–dimensional topology, they still offer a wide collection of essential examples. This part is devoted to their survey.

After an introductory chapter 6 that attempts to convey a smattering of complex geometry, in chapter 7 (starting on page 285) we outline the Enriques–Kodaira classification of complex surfaces, and then comment on the difference between the complex and the smooth points of view.

Then, in chapter 8 (starting on page 301), we describe in some detail elliptic surfaces (of which $K3$ is an example) and finish by culling from them infinite families of homeomorphic but non-diffeomorphic simply-connected 4–manifolds. This latter chapter is the more thorough of this rather cursory third part.

Contents of Part III

Chapter 6

Running through Complex Geometry

We now try to briefly cover some background on complex geometry, to help us make sense of the following chapters, but to also maybe get some vague intuition about these objects. Thus, not all concepts and statements presented in this very short chapter will actually be needed in the subsequent chapters. We attempted to make this exposition as comfortable as possible to read (as an example, we obsessively avoided using the word "divisor", and several concepts are only partly defined, if at all).

Be aware of the sometimes confusing terminology: owing to the superposition of the real and complex perspectives, a complex 2–manifold is called a **curve***, a complex 4–manifold is called a* **surface***, and an oriented real plane becomes a complex* **line***. A bit of attention and the context should be enough to clear any confusion.*

6.1. Surfaces

Zoology. Complex surfaces come in several species. A **complex analytic surface** is a 4–manifold admitting complex-holomorphic changes of coordinates. A **projective analytic surface** is an analytic surface that can be embedded in a projective space $\mathbb{CP}^m$. A **complex algebraic surface** is a complex surface defined by homogeneous polynomial equations in some complex projective space[1] $\mathbb{CP}^m$. Every algebraic surface is also an analytic surface. Conversely, by Chow's theorem, all analytic projective surfaces are

---

[1] All non-singular *abstract* algebraic surfaces are projective.

algebraic. (There are analytic surfaces that are not algebraic, such as those supported by $\mathbb{S}^3 \times \mathbb{S}^1$, but we will not discuss these in the sequel.) The projective algebraic and projective analytic worlds are essentially equivalent (bundles, sheaves, and all other *global* creatures) as a consequence of J.P. Serre's famous GAGA paper.[2]

A complex analytic surface M is **Kähler** if it admits a Riemannian metric highly-compatible with the complex structure. This compatibility can be expressed in several equivalent ways: One can say that a complex surface M is Kähler if it admits a Riemannian metric that makes M locally look like Euclidean $\mathbb{C}^2$ up to second order.[3] One can also say: a complex surface M is Kähler if it admits a Riemannian metric whose Levi–Cività connection ∇ is $\mathbb{C}$–linear, *i.e.*, it satisfies $\nabla_v(iw) = i\,\nabla_v w$. One can even say: a complex surface M is Kähler if it admits a compatible symplectic structure.

> *A* **symplectic structure** *on a 4–manifold M is any differential 2–form ω that is nowhere-degenerate, i.e., $\omega \wedge \omega$ is nowhere-zero and orients M, and is closed, i.e., $d\omega = 0$. A symplectic structure ω is said to be* **compatible** *with a complex structure if it is invariant under complex multiplication, i.e., $\omega(iv, iw) = \omega(v, w)$, and is positive on complex directions, i.e., $\omega(v, iv) > 0$. In this case ω defines a Riemannian metric on M by $\langle v, w\rangle = \omega(v, iw)$. The form ω is called a* **symplectic form.** *On a Kähler surface, it is called a* **Kähler form.** *Note that there exist 4–manifolds that admit both complex structures and symplectic structures, but the two are never compatible and thus these manifolds are not Kähler.*[4]

A complex surface M is Kähler if and only if it has $b_1(M)$ even. Thus, all simply-connected complex surfaces are in fact Kähler. Further, all Kähler surfaces admit complex-deformations[5] to ones for which the class $[\omega]$ is integral, in which case they embed in some projective space and thus are algebraic. (Notice that complex-deformations induce diffeomorphisms of the underlying smooth manifolds.) Conversely, all projective surfaces inherit a Kähler structure from their ambient $\mathbb{CP}^m$.

We choose to be blind. In what follows, we will view complex surfaces as essentially smooth 4–manifolds with some extra rigidity structure added. "Projective analytic", "algebraic" or "Kähler ", it will all be the same to us, and when we say **complex surface** we will mean "complex analytic algebraic projective Kähler ", with little distinction. Oh, the horror...

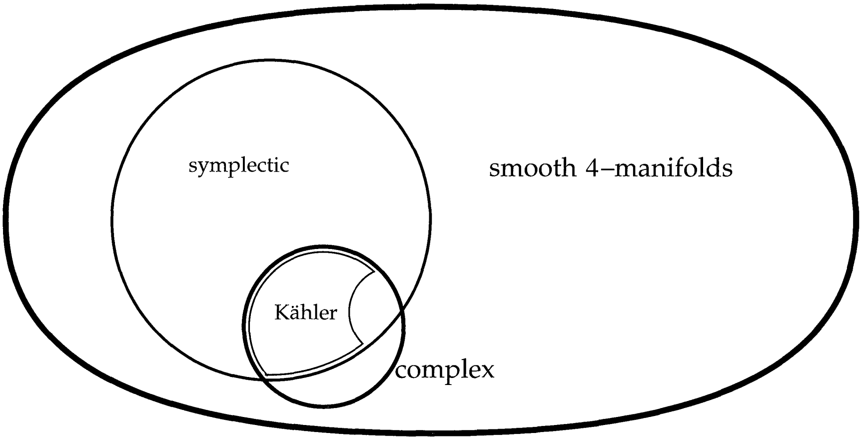
---

[2] ***Géometrie algébrique et géometrie analytique*** **[Ser56]**.

[3] More than second order kills the curvature, which is extremely restrictive.

[4] For example, the Thurston–Kodaira manifold (a torus-bundle over a torus), see **W. Thurston**'s ***Some simple examples of symplectic manifolds*** [**Thu76**].

[5] Roughly, two complex surfaces are **deformations** of each other if they appear as fibers of some holomorphic surjective map $E \twoheadrightarrow \mathbb{D}^2$, usually required to be smoothly-submersive.



6.1. Cutting the 4–manifold cake

Further, all the complex surfaces that we will consider will be projective, closed and *simply-connected.* All will be *non-singular* (*i.e.,* genuine 4–manifolds), but we will consider singular curves inside them. (We will take the term "curve" to always mean "irreducible curve".)

6.2. Curves on surfaces

Complex curves. An *(irreducible)* **curve** inside a complex surface is a closed complex submanifold (possibly singular) of real dimension 2, which cannot be split as the union of two simpler closed complex submanifolds. Note that two distinct curves always intersect positively, but a curve can have negative self-intersection.

To underline the rigidity of the complex world, it is worth mentioning right from the start that in a complex surface each homology class can be represented, up to smooth isotopy, by at most finitely-many distinct non-singular complex curves. If further the surface is simply-connected, then each class can be represented, up to isotopy, by at most *one* non-singular complex curve.

Only part of the cohomology of M can be represented by linear combinations of curves in M. For making this part visible, the usual method is to look at $H^2(M;\mathbb{C})$, think of it as de Rham cohomology, and split it according to the **type**[6] of the differential 2–forms representing the various classes.

---

[6] This split was mentioned in the end-notes of chapter 3 (page 136) and will be better explained in the end-notes of chapter 9 (page 365). We review: Pick local local real coordinates (x_1, y_1, x_2, y_2) on M so that, for $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, we have that (z_1, z_2) are local complex coordinates on M. Then define $dz_k = dx_k + i\,dy_k$ and $d\bar{z}_k = dx_k - i\,dy_k$. The span of all combinations of p of the dz_k's and q of the $d\bar{z}_k$'s inside $\Lambda^{p+q}(T^*_M) \otimes \mathbb{C}$ makes up the space $\Lambda^{p,q}$ of complex-valued forms of type (p, q). These lead to the cohomology groups $H^{p,q}(M)$.

Specifically,
$$\Lambda^2 \otimes \mathbb{C} = \Lambda^{2,0} \oplus \Lambda^{1,1} \oplus \Lambda^{0,2} .$$

Relevant for us here is type $(1,1)$: a complex-valued 2–form is a $(1,1)$–form if it can be written locally as $\alpha = f_{11}\, dz_1 \wedge d\bar{z}_1 + f_{12}\, dz_1 \wedge d\bar{z}_2 + f_{21}\, dz_2 \wedge d\bar{z}_1 + f_{22}\, dz_2 \wedge d\bar{z}_2$, for any local complex coordinates (z_1, z_2). As an example, any Kähler form can be written $\omega = \frac{i}{2}\, dz_1 \wedge d\bar{z}_1 + \frac{i}{2}\, dz_2 \wedge d\bar{z}_2$, and thus is always of type $(1,1)$. Corresponding to the above split of forms, we have, for cohomology:
$$H^2(M;\mathbb{C}) = H^{2,0}(M) \oplus H^{1,1}(M) \oplus H^{0,2}(M) ,$$
where the $H^{1,1}(M)$–summand comes from $(1,1)$–forms. While thinking of $H^2(M;\mathbb{Z})$ as a lattice inside $H^2(M;\mathbb{C})$, we have:

Lefschetz Theorem on* $(1,1)$*–Classes. *The integral points of $H^{1,1}(M)$ are exactly those classes of $H^2(M;\mathbb{Z})$ that can be represented by $\mathbb{Z}$–linear combinations of complex curves.*[7] □

Intersections of curves. The intersection form of a complex surface M has a special behavior on curves:

Hodge Signature Theorem. *If M is a Kähler surface with Kähler form ω, then the intersection form Q_M restricted to $H^{1,1}(M) \cap H^2(M;\mathbb{R})$ is negative-definite on the orthogonal complement of $[\omega]$.* □

Geometric Hodge Signature Theorem. *If L and E are such that $L \cdot L > 0$ and $L \cdot E = 0$, then either $E \cdot E < 0$, or E is homologically-trivial (or torsion).* □

Here, L and E are $\mathbb{Z}$–linear combinations of complex curves, but can also be thought of as Chern classes of holomorphic line bundles, as we will see shortly.

6.3. Line bundles

Holomorphic line bundles. While every class from $H^2(M;\mathbb{Z})$ can be represented by smooth complex-line bundles, *not* all classes from $H^2(M;\mathbb{Z})$ can be represented by *holomorphic* line bundles,[8] but only those that also belong to $H^{1,1}(M)$.

Further, in general there are non-isomorphic holomorphic line bundles with the same Chern class, but if one assumes that $H^1(M;\mathbb{Z}) = 0$ (*e.g.*, if M is

---

[7] A linear combination of complex curves of M is called a **divisor** on M. We have avoided introducing too many new terms, including "divisor", which has made certain statements somewhat cumbersome, but hopefully more comfortable to read for non-specialists.

[8] A **holomorphic bundle** is a complex vector bundle glued-up by using a cocycle $\{g_{\alpha\beta}\}$ in which each map $g_{\alpha\beta} \colon U_\alpha \cap U_\beta \to GL_{\mathbb{C}}(n)$ is holomorphic. In other words, the changes of coordinates are holomorphic. Cocycles were already encountered in the end-notes of chapter 4 (definition of spin structures, page 174).

simply-connected, as we agreed), then a line bundle is completely determined by its Chern class. For convenience, we will therefore denote by the same letter a holomorphic complex-line bundle and its first Chern class. Thus, we will write $L' + L''$ to denote $L' \otimes L''$, as well as mL for $L^{\otimes m}$ and $-L$ for $L^* = \overline{L}$.

Holomorphic sections. On one hand, a holomorphic line bundle L always admits *meromorphic* sections (such a section is not everywhere-defined, having points called *poles* where it explodes toward infinity). For any such section ℓ, the linear combination of curves $Zeros(\ell) - Poles(\ell)$ (zeros and poles counted with multiplicities) represents exactly the Chern class $c_1(L)$.

> ***Of curves and bundles.*** *In fact, complex geometry is so rigid that the pairs* (L, ℓ) *(up to scalar multiples of* ℓ*) are in bijection with such linear combinations of curves. In other words, for every* $D = \sum a_k C_k$ *with* C_k *complex curves in* M *and* $a_k \in \mathbb{Z}$*, there exists a* unique *holomorphic line bundle* L *with* $c_1(L) = [D]$ *and a meromorphic*[9] *section* ℓ *of* L*, unique up to scalar-multiplication, with* $\ell^{-1}[0] - \ell^{-1}[\infty] = D$.

On the other hand, a holomorphic line-bundle L might not admit any *holomorphic* sections. A bundle L has a holomorphic section if and only if its Chern class $c_1(L)$ can be represented as a linear combination of curves in M, all with *positive* coefficients.[10] This ensures that there is a meromorphic section of L that has no poles, and thus is a genuine holomorphic section. When dealing with holomorphic bundles, we will denote by $\Gamma(L)$ the space of all holomorphic sections of L.

Ample line bundles. In case the line-bundle L has at least two linearly-independent holomorphic sections, we can pick a basis $\{f_0, \dots, f_N\}$ of $\Gamma(L)$ over $\mathbb{C}$, and define the map

$$\Phi_L : M \dashrightarrow \mathbb{CP}^N \qquad \Phi_L(x) = \big[f_0(x) : \cdots : f_N(x)\big] .$$

In general, the map Φ_L is not defined over the whole M (and thus the dashed arrow above); it is undefined at the common zero-locus of the sections $f_0, \dots, f_N$. (A different choice of basis in $\Gamma(L)$ merely changes Φ_L by an automorphism of $\mathbb{CP}^N$, and therefore the map Φ_L depends essentially only on L.)

> *Notice that the higher the tensorial power, the more holomorphic sections a bundle might have: if* $\{f_i\}_i$ *are linearly independent holomorphic sections of* L*, then* $\{f_i \otimes f_j\}_{i,j}$ *are linearly-independent sections of* $2L = L \otimes L$*. Of course, if there are none to start with...*

---

[9] If all $a_k \geq 0$, then ℓ will be a holomorphic section of L.

[10] A divisor with all coefficients positive is called an **effective divisor.**

A line bundle L is called **ample** if there is some m such that Φ_{mL} is well-defined over the whole M and further is an *embedding* of M. A method for detecting ample bundles is offered by:

Grauert's Criterion. *A line bundle L on M is ample if and only if for every curve C in M there is some n such that $nL|_C$ has a holomorphic section that has at least one zero, but is not everywhere-zero on C.* □

The canonical bundle. Of fundamental importance throughout algebraic geometry is the **canonical bundle** of a complex manifold M, denoted by K_M. It is the line-bundle

$$K_M = \det_{\mathbb{C}} T^*_M ,$$

with[11] $c_1(K_M) = -c_1(T_M)$. In terms of complex-valued forms, we also have

$$K_M = \Lambda^{2,0} .$$

As usual, we will denote the Chern class $c_1(K_M)$ (called the **canonical class** of M) by the same symbol K_M as the line bundle itself. As an elementary example, for the projective plane $\mathbb{CP}^2$ we have $K_{\mathbb{CP}^2} = -3\,[\mathbb{CP}^1]$.

The tangent bundle T_M is a complex bundle and hence satisfies

$$p_1(T_M) = c_1(T_M)^2 - 2c_2(T_M) .$$

Therefore, by using the Hirzebruch signature theorem and the fact that K_M is $-c_1(T_M)$, we deduce:

Lemma. *For every complex surface M we have:*

$$K_M \cdot K_M = 3 \operatorname{sign} M + 2\chi(M) . \qquad \square$$

Thus, the self-intersection number of the canonical class is determined by the topology of M.

> ***Complex, or almost.*** *The above equality is the only obstruction to the existence of an almost-complex structure on M, i.e., of a complex structure on the fibers of T_M, not necessarily corresponding to any holomorphic charts on M. Namely, for every class $\kappa \in H^2(M;\mathbb{Z})$ such that $\kappa \cdot \kappa = 3 \operatorname{sign} M + 2\chi(M)$, there exists an almost-complex structure J such that $c_1(T_M, J) = -\kappa$. We will discuss almost-complex structures in more detail later.*[12]

---

[11] The reason for the minus sign in $c_1(T^*_M) = -c_1(T_M)$ was explained in the end-notes of chapter 3 (complex duals, page 134).

[12] See section 10.1 (page 376), and the end-notes of chapter 10 (page 420).

Pluri-genera. For the canonical bundle, the maps $\Phi_{mK_M} : M \dashrightarrow \mathbb{CP}^N$ are called the **pluri-canonical maps.** The numbers

$$P_m(M) = \dim_{\mathbb{C}} \Gamma(mK_M)$$

are called the **pluri-genera** of M. (In particular, they give the complex dimension $N = P_m(M) - 1$ of the codomain of Φ_{mK_M}.)

The first of these, $P_1(M) = \dim \Gamma(K_M)$, is at times denoted by p_g and called the **geometric genus**[13] of M. For Kähler surfaces, we have

$$b_2^+(M) = 2P_1(M) + 1 \,,$$

where the "+1" comes from the Kähler class $[\omega]$. (In particular, all Kähler surfaces have b_2^+ odd.) Thus $P_1(M) = \frac{1}{2}\big(b_2^+(M) - 1\big)$ and is determined by the intersection form of M; by contrast, the higher pluri-genera have *no* known smooth/topological interpretation.[14]

Genus of curves. The canonical class K_M is also important because it determines the genus of all curves in M:

Adjunction Formula. *If C is a non-singular curve in M, then we must have:*

$$\chi(C) + C \cdot C = -K_M \cdot C \,.$$

Proof. Since both T_C and $N_{C/M}$ are complex bundles, we compute:

$$\begin{aligned} -K_M \cdot C &= c_1(T_M) \cdot C = c_1(T_M|_C) \\ &= c_1\big(T_C \oplus N_{C/M}\big) = c_1(T_C) + c_1(N_{C/M}) \\ &= \chi(C) + C \cdot C \,. \end{aligned}$$ □

As a simple example of applying the adjunction formula, since $K_{\mathbb{CP}^2} = -3\,[\mathbb{CP}^1]$ and since every non-singular curve in $\mathbb{CP}^2$ that is defined by a polynomial equation of degree d must represent the class $d\,[\mathbb{CP}^1]$, it must be a real surface of genus

$$g = \tfrac{1}{2}(d-1)(d-2) \,.$$

In particular, all non-singular quadrics in $\mathbb{CP}^2$ are spheres (rational curves), while all non-singular cubics are tori (elliptic curves).

---

[13] Historically, the geometric genus p_g was the first one to arise and is the one that gave the name of "genera" to all P_n.

[14] They are nonetheless diffeomorphism-invariants, compare ahead with section 7.5 (page 296).

The Riemann–Roch theorem. For estimating the number of holomorphic sections a line-bundle might have, an invaluable resource is the celebrated Riemann–Roch theorem:

Riemann–Roch Theorem. *Let L be a holomorphic line-bundle on a complex surface M. We have:*

$$\dim_{\mathbb{C}} \Gamma(L) + \dim_{\mathbb{C}} \Gamma(K_M - L) \geq \tfrac{1}{2}\big(L \cdot L - K_M \cdot L\big) + \chi(\mathcal{O}),$$

where $\chi(\mathcal{O}) = \frac{1}{12}\big(K_M \cdot K_M + \chi(M)\big)$ *and is always an integer.*[15] □

From Riemann–Roch it follows that, if $L \cdot L > 0$, then for some $m \in \mathbb{Z}$ the bundle mL must have holomorphic sections. More precisely:

Corollary. *If $L \cdot L > 0$, then either nL or $-nL$ has a holomorphic section when n is big enough. The two cases are separated by* any *ample H:*

— *If $L \cdot L > 0$ and $L \cdot H > 0$, then nL must have a section for large n.*

— *If $L \cdot L > 0$ and $L \cdot H < 0$, then $-nL$ must have a section for large n.* □

And finally:

Nakai's Criterion. *A line-bundle L is ample if and only if*

$$L \cdot L > 0 \qquad \text{and} \qquad L \cdot C > 0 \quad \text{for every curve } C. \qquad \square$$

To be or not to be nef. Since to ask a bundle to be ample is to ask a lot, we can, inspired by the Nakai criterion above, define a weaker concept: a line bundle L is called **nef**[16] if it intersects all curves non-negatively:

$$L \cdot C \geq 0 \quad \text{for every curve } C.$$

A first property (easy but not obvious) is that

Lemma. *If L is nef, then $L \cdot L \geq 0$.* □

One should think of a nef line bundle as a bundle that is "almost ample". This concept of "nefness" will play a crucial role in the classification of complex surfaces that we will present next.

---

[15] The actual Riemann–Roch theorem is the equality $\chi(\Gamma(L)) - \chi(\Gamma(\underline{\mathbb{C}})) = \frac{1}{2}(L \cdot L - K_M \cdot L)$, where $\chi(\Gamma(L))$ is the Euler characteristic of the cohomology of the sheaf of holomorphic sections of L, while $\chi(\Gamma(\underline{\mathbb{C}})) = \chi(\mathcal{O})$ denotes the Euler characteristic of the cohomology $H^*(M; \mathcal{O})$ of the sheaf of holomorphic functions on M (or, if you prefer, the Euler characteristic of the Čech cohomology $\check{H}^*(M;\, \mathcal{O}(\mathbb{C}))$, using the notation from the end-notes of chapter 4, page 193). The version presented above is obtained after combining with Serre duality, then dropping the mysterious term $\dim_{\mathbb{C}} H^1(M;\, \Gamma(L))$ by weakening to an inequality, and finally quoting M. Noether's formula for $\chi(\mathcal{O})$. For a proof that indeed $K_M \cdot K_M + \chi(M)$ is a multiple of 12, see Noether's lemma in section 10.7 (page 422).

[16] The word "nef" is the abbreviation coined by M. Reid for "numerically eventually free", which is related to free linear systems. *Post factum,* we could also read "nef" as "numerically effective" (an effective divisor is a linear combination of curves with all coefficients positive).

6.4. Notes

Note: Nef as limit of ample

It can be shown that, in the appropriate context, "nef" is the boundary of "ample".

For that, extend to real coefficients and consider all classes $\alpha \in H_2(M;\mathbb{R})$ that can be written as $\alpha = \sum a_k C_k$ for some curves C_k and coefficients $a_k \in \mathbb{R}$. Such a class α is called **nef** if $\alpha \cdot C \geq 0$ for every curve C. The class α is called **ample** if it has $\alpha \cdot \alpha > 0$ and $\alpha \cdot C > 0$ for every curve C. Then we have:

Lemma. *If α is nef, then, for every rational $\varepsilon > 0$ and every ample class h, we have that $\alpha + \varepsilon h$ is an ample class.* □

Therefore, nef classes are limits of ample classes. Even more, nef classes make up a convex cone (called the **nef cone**, and usually denoted by Nef) inside $H_2(M;\mathbb{R})$, whose interior consists exactly of all ample classes.

Indeed, consider the **effective** classes, *i.e.*, those that can be represented as $\alpha = \sum a_k C_k$ with all $a_k > 0$. Their collection makes up the **effective cone** (usually denoted by Ef) inside $H_2(M;\mathbb{R})$. Then one can view the nef cone as the *dual* cone of the effective cone, in the sense that $\nu \in$ Nef if and only if $\nu \cdot \alpha \geq 0$ for all $\alpha \in$ Ef. The interior of Nef is defined by the condition $\nu \cdot \alpha > 0$ for all $\alpha \in$ Ef, which describes precisely the ample classes.

Bibliography

One could hardly have been more fleeting and superficial than we have. For digging the foundations of the subject, one can start by looking at a textbook for complex algebraic geometry, such as **P. Griffiths** and **J. Harris**'s comprehensive ***Principles of algebraic geometry*** [**GH78, GH94**], then for surfaces proceed to **W. Barth, C. Peters** and **A. Van de Ven**'s ***Compact complex surfaces*** [**BPVdV84**] or its second edition (with **K. Hulek**) [**BHPVdV04**]. For example, a proof that b_1 even implies Kähler can be read from section IV.3 of the latter (second edition only).

As general introductions, one can always try **D. Mumford**'s ***Algebraic geometry. I. Complex projective varieties*** [**Mum76, Mum95**] or **I. Shafarevich**'s ***Basic algebraic geometry*** [**Sha74, Sha94**].

Our presentation in this and the following chapter is inspired by **M. Reid**'s ***Chapters on algebraic surfaces*** [**Rei97**].

Chapter 7

The Enriques–Kodaira Classification

We outline in this short chapter the classification of complex surfaces, due to F. Enriques and K. Kodaira. As usual, we restrict to the simply-connected case. Even so, the classification is incomplete, leaving a vast realm of complex surfaces (surfaces of "general type") unexplored. Nonetheless, besides those, the $K3$ surfaces will find their proper place alongside rational, ruled, and elliptic surfaces.

We organize the classification around the nef/ampleness of the canonical bundle K_M. Thus, we start by explaining the blow-down construction and use it to reduce the classification to the case when K_M is nef. In section 7.2 (page 292), by measuring how far K_M is from being ample (using the so-called numerical dimension), we split complex surfaces into $K3$'s, proper elliptic surfaces, and surfaces of general type. A minuscule discussion of surfaces of general type is made on page 293. In section 7.3 (page 294) is presented the more customary statement of the classification in terms of Kodaira dimension, while in section 7.4 (page 295) the classification is restated, in the case of Kähler surfaces, in terms of the intersection of K_M with the Kähler class $[\omega]$.

Finally, looking back toward smooth topology, in section 7.5 (page 296) we compare the complex and the smooth points-of-view; surprisingly, many complex devices are in fact smoothly visible: numerical dimension, Kodaira dimension, pluri-genera, blow-downs and canonical bundles are all invariant under diffeomorphisms of complex surfaces.

7.1. Blow-down till nef

If M is a complex surface and K_M is ample, then for m big enough the pluricanonical maps Φ_{mK_M} offer embeddings of M in projective spaces. While in general we cannot expect K_M to be ample, we can still try to modify M to arrange that K_M become nef. This is the start of the classification scheme.

Remember that K_M would be nef if $K_M \cdot C \geq 0$ for all curves C. Thus, when trying to make K_M nef, we will wish to eliminate all negative intersections of K_M. We can do this by blowing-down everything we can. This blow-down process is founded on

Blow-Up / Blow-Down Lemma. *Let M' be a complex surface, and E a complex curve in M'. If E is a sphere with $E \cdot E = -1$, then there exists another complex surface M and a map*

$$\sigma\colon M' \to M$$

that contracts E to a point p and is biholomorphic[1] on the complement of E.

Conversely, if M is a complex surface and p is any of its points, then there exists another surface M', containing a complex curve E of genus zero and self-intersection $E \cdot E = -1$, and a map $\sigma\colon M' \to M$ as above, with $\sigma[E] = p$ and $\sigma\colon M' \setminus E \approx M \setminus \{p\}$. □

We call M' the **blow-up** of M at the point p, and we call M the **blow-down** of E from M'. Any complex curve E that is a *sphere* and has self-intersection $E \cdot E = -1$ is called a **(-1)–curve** (or **exceptional curve**) of M'.

Construction. Given a complex surface M and a point $p \in M$, in order to build the blow-up M' we start by choosing a small neighborhood U of p, which we holomorphically parametrize as a neighborhood of the origin in $\mathbb{C}^2$, with $p \equiv (0,0)$. Then we take a copy of $\mathbb{CP}^1$ (a sphere) and think of it as the space of all complex lines ℓ in $\mathbb{C}^2$ that pass through the origin p. We build

$$U' = \{(x,\ell) \in U \times \mathbb{CP}^1 \mid x \in \ell\} .$$

Observe that $U \setminus p$ and $U' \setminus (p \times \mathbb{CP}^1)$ are biholomorphic. Therefore we can cut U out of M and replace it by U'. The result is the blow-up M', and its (-1)–curve is just the sphere $E = p \times \mathbb{CP}^1$ from U'.

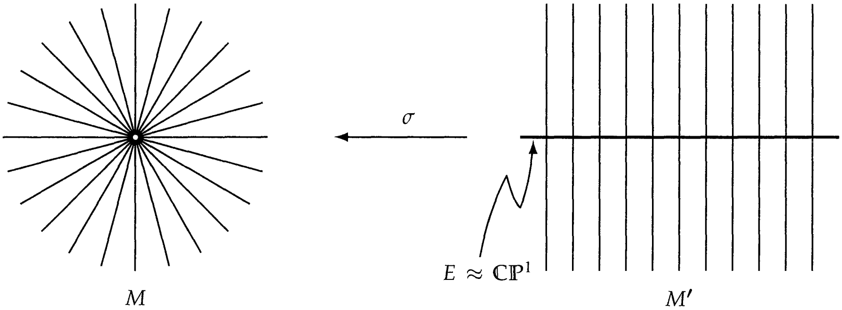
The difficult part of the above lemma is certainly its first part, namely, proving that as soon as you find a complex sphere of self-intersection -1, you can view your complex surface as the blow-up of a simpler surface.

The intuitive picture of blowing-up at a point p is that we replace p by representatives of all the complex lines passing trough p, in other words,

---

[1] A map is called **biholomorphic** if it is holomorphic and has an inverse that is holomorphic as well. It is the appropriate isomorphism in the complex category, just as diffeomorphisms were for smooth manifolds, *etc.*

that we replace p by a copy of $\mathbb{CP}^1 \approx S^2$. All the lines that used to cross through p are now passing through their corresponding direction in the added $\mathbb{CP}^1$, and thus are now disjoint. See figures[2] 7.1 and 7.2 on the following page.



7.1. Blow-up at a point, I

Of course, a blow-up/down will modify the canonical class.

Lemma. *If* $\sigma\colon M' \to M$ *is the blow-down of a* (-1)*–curve* E*, then we have*

$$K_{M'} = \sigma^* K_M + E\,. \qquad \square$$

Topological interpretation. Choose some local complex coordinates centered at p and take a small round neighborhood U of p. The complex lines ℓ of $\mathbb{C}^2$ intersect U as disks D_ℓ, centered at p. Thus, we can view the neighborhood U of p as built from the disks D_ℓ after identifying their centers with p. The disks D_ℓ are disjoint away from p, and their boundary-circles cover the 3–sphere ∂U and spell out the Hopf fibration[3] $S^3 \to \mathbb{CP}^1$. Thinking of $\mathbb{CP}^1$ as the space $\{\ell\}$ of complex directions in $\mathbb{C}^2$, we can view the map $\partial U \to \mathbb{CP}^1$, given by $x \mapsto \ell$ whenever $x \in \partial D_\ell$, as a circle-bundle projection.

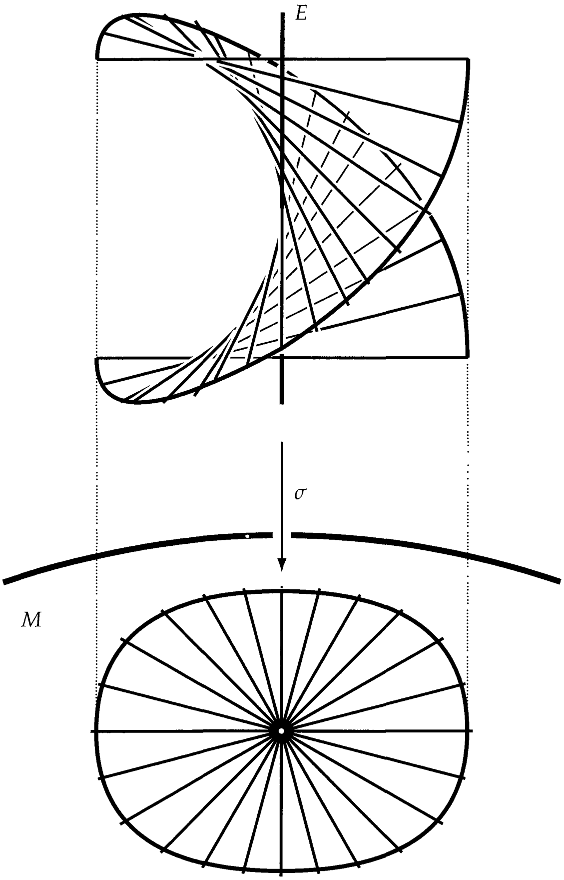
To obtain from the disks D_ℓ the blown-up set U' rather than U, we can simply *not* identify the centers of the D_ℓ's, but keep them distinct. The centers of the D_ℓ's, since they are parametrized by the ℓ's, will draw a copy of $\mathbb{CP}^1$ inside U'. In other words, we can define U' to be the total space of the disk-bundle over $\mathbb{CP}^1$ with fibers $D_\ell \mapsto \ell$. Its boundary is still the Hopf bundle with total space S^3, and thus we can cut U out of M, leaving a hole with boundary S^3 that we fill-in by gluing U' instead. The new manifold is the blow-up M'.

Since the disk bundle $D_\ell \mapsto \ell$ can be viewed as the normal bundle of $\mathbb{CP}^1$ in $\mathbb{CP}^2$, this leads to:

---

[2] Of course, keep in mind that dimensions have been reduced: a complex line is a real plane.

[3] The Hopf map was recalled in footnote 34 on page 129.



7.2. Blow-up at a point, II

Lemma. *Topologically, the blow-up of M at a point p is the connected sum*

$$M' = M\#\overline{\mathbb{CP}}^2,$$

with the (-1)*-curve appearing as* $\overline{\mathbb{CP}}^1 \subset \overline{\mathbb{CP}}^2$.

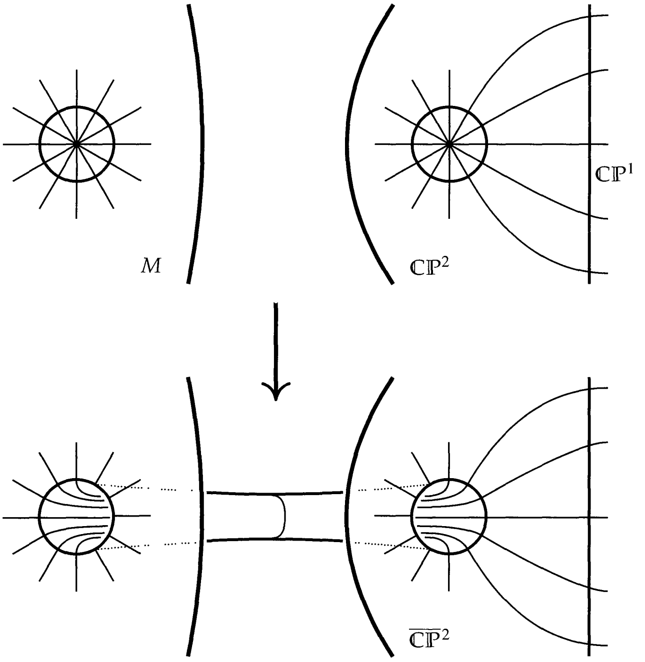
Sketch of proof. The preceding discussion falls short of being a proof of the lemma mainly by not arguing why we need $\overline{\mathbb{CP}}^2$ rather than $\mathbb{CP}^2$. Instead of a direct argument, we prefer a slightly different approach, essentially an inside-out version of the previous description. (More straightforward arguments can be found in the literature.[4])

As above, start by taking a point p in M and surround it by a 4–ball U, sliced into disks D_ℓ by the complex lines ℓ crossing through p. Its boundary is sliced into circles, and so on.

---

[4] For example, see **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** [**GS99**, sec 2.2].

On the other hand, pick your favorite point q of $\mathbb{CP}^2$ and a random projective line $\mathbb{CP}^1$ inside $\mathbb{CP}^2$, so that q does not belong to $\mathbb{CP}^1$. Consider now all projective lines ℓ in $\mathbb{CP}^2$ that pass through q: they are all disjoint away from q, and each intersects $\mathbb{CP}^1$ in exactly one point. In other words, aside from q every point x of $\mathbb{CP}^2$ belongs to a unique projective line ℓ_x that contains both x and q, and this line intersects $\mathbb{CP}^1$ in a unique point $y = \ell_x \cap \mathbb{CP}^1$. Hence the map $\mathbb{CP}^2 \setminus \{q\} \to \mathbb{CP}^1$ given by $x \mapsto y$ is a well-defined bundle projection, whose fibers are the punctured spheres $\ell_y \setminus \{q\}$. If we surround q by a small round neighborhood U_q (disjoint from $\mathbb{CP}^1$), then we obtain a bundle $\mathbb{CP}^2 \setminus U_q \to \mathbb{CP}^1$ whose fibers are the 2–disks $\ell_y \setminus U_q$. See also figure 7.3.



7.3. Topological blow-up

Choose the 4–ball U_q in $\mathbb{CP}^2$ in such manner that it is sliced by the projective lines passing through q in the same way as the neighborhood U of p in M is sliced by the complex lines there. In other words, choose U_q so that it is identifiable with U in a manner that respects their crossing lines. Since both q in $\mathbb{CP}^2$ and p in M sit like 0 in $\mathbb{C}^2$, lines and all, this is not hard to achieve. (Since we are trying to describe a blow-up, these lines are of course essential.)

The resulting identification of U with U_q induces an identification of their boundary 3–spheres ∂U and ∂U_q, again in a manner that preserves the crossing lines (as circle-fibers, oriented by the complex orientation of their respective lines).

In particular, $M \setminus U$ and $\mathbb{CP}^2 \setminus U_q$ have the same fibered boundary, and we can try to glue one to the other. However, for gluing we need to use an *orientation-reversing* diffeomorphism of the boundaries. Since we do not wish to alter the identification of ∂U with ∂U_q, we choose to reverse orientation by flipping the orientation of the whole $\mathbb{CP}^2 \setminus U_q$ (think of this as "turning $\mathbb{CP}^2 \setminus U_q$ inside-out"). Indeed, we have

$$\partial(M \setminus U) = \overline{\partial U} = \overline{\partial U_q} = \partial(\mathbb{CP}^2 \setminus U_q) = \overline{\partial(\overline{\mathbb{CP}^2 \setminus U_q})} ,$$

and hence $M \setminus U$ and $\overline{\mathbb{CP}^2 \setminus U_q}$ can be glued to each other in a manner that respects the fibered identification of ∂U with ∂U_q. The resulting manifold

$$M' = M \setminus U \cup_\partial \overline{\mathbb{CP}^2 \setminus U_q}$$

simply describes $M \# \overline{\mathbb{CP}}^2$.

We need to argue that it also describes the blow-up of M. Review what happened: we removed a neighborhood U of p and replaced it by $U' = \overline{\mathbb{CP}}^2 \setminus U_q$. For each complex line ℓ that crossed through p, this has the effect of replacing a small disk D_ℓ (centered at p) with a corresponding disk $\ell \setminus U_q$ from $\overline{\mathbb{CP}}^2 \setminus U_q$ (which can be thought of as centered at its intersection point with $\mathbb{CP}^1$). The main difference, as far as these complex lines are concerned, is that now they do not intersect each other at p anymore, but instead travel through $\overline{\mathbb{CP}}^2$ and pass through distinct points of $\mathbb{CP}^1$. This sounds like a blow-up. □

When M is a complex surface, the manifold $M \# \overline{\mathbb{CP}}^2$ is diffeomorphic to the complex blow-up M' of M. Thus, we can extend the nomenclature and, for every smooth 4–manifold M, call the connected sum $M \# \overline{\mathbb{CP}}^2$ the **topological blow-up** of M, whether or not M admits complex structures. (By way of contrast, notice that connect-summing with $\mathbb{CP}^2$ has no complex-geometric analogue.[5])

Conversely, whenever one finds a *sphere* S embedded in some smooth 4–manifold M and having self-intersection $S \cdot S = -1$, then M must split off a copy of $\overline{\mathbb{CP}}^2$ and can be written as

$$M = N \# \overline{\mathbb{CP}}^2$$

for some smooth 4–manifold N (which we can happily call the **topological blow-down** of S from M). Indeed, a tubular neighborhood of S in M is

---

[5] Connect-summing with $\mathbb{CP}^2$ is sometimes called an **anti-complex blow-up**.

diffeomorphic to a neighborhood of $\overline{\mathbb{CP}}^1$ in $\overline{\mathbb{CP}}^2$, and its boundary must be a (Hopf) 3–sphere splitting M into N and $\overline{\mathbb{CP}}^2$. (Similarly, if one finds a sphere with self-intersection $+1$, then one can split off a copy of $\mathbb{CP}^2$ and write $M = N \# \mathbb{CP}^2$.)

We now get back to the classification of complex surfaces:

Make it nef. If E is a (-1)–curve in M, then the adjunction formula dictates that we must have $K_M \cdot E = -1$. Therefore, blowing E down from M will eliminate a curve with negative intersection with K_M: a small step toward making K_M nef.

Even more, (-1)–curves are easily detected through their intersections:

Lemma. *If $C \cdot C < 0$ and $K_M \cdot C < 0$, then C is a (-1)–curve in M.* □

Thus, if a curve C has negative intersection with K_M, all we still need before being able to blow it down is that it have negative self-intersection. However, the Hodge signature theorem[6] shows that Q_M is negative-definite over most curves, and so it seems likely that in most cases $C \cdot C < 0$.

The miracle is that, indeed, we can always get rid of *all* negative intersections of K_M just by blowing-down everything we can:

Minimal Model Theorem. *Let M be a simply-connected complex surface. Then there is a chain of blow-downs*

$$M \xrightarrow{\sigma_1} M_1 \xrightarrow{\sigma_2} \cdots \xrightarrow{\sigma_m} M_m$$

such that either:

— K_{M_m} *is nef;*

— M_m *is a* $\mathbb{CP}^1$*–bundle over* $\mathbb{CP}^1$*;*

— $M_m = \mathbb{CP}^2$. □

Surfaces that blow-down to $\mathbb{CP}^2$ are called **rational surfaces**, while $\mathbb{CP}^1$–bundles over $\mathbb{CP}^1$ are called **ruled surfaces.**[7] The latter are diffeomorphic to either[8] $S^2 \times S^2$ or $S^2 \tilde{\times} S^2 = \mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$.

Notice that K_{M_m} being nef implies that M_m is minimal with respect to blow-downs. Any surface obtained from blowing M down till no further blows

---

[6] The Hodge signature theorem was stated back on page 278.

[7] Remember that we restricted the discussion to simply-connected complex surfaces. In general, a ruled surface is a $\mathbb{CP}^1$–bundle over a complex curve of any genus.

[8] On one hand, we have $\pi_1 SO(3) = \mathbb{Z}_2$ and hence only two S^2–bundles over S^2. On the other hand, since $\pi_1 U(1) = \mathbb{Z}$ (think: twists along the equator of the sphere-fiber), there are $\mathbb{Z}$–many distinct simply-connected ruled surfaces. The usual way to describe them is as the projectivization of $L_n \oplus \underline{\mathbb{C}} \to \mathbb{CP}^1$, where L_n is the complex-line bundle of Chern class n; these are called **Hirzebruch surfaces** and are often denoted by $\mathbb{F}_n$. All of them are deformation-equivalent to either $\mathbb{F}_0 = \mathbb{CP}^1 \times \mathbb{CP}^1$ or to $\mathbb{F}_1 = (\mathbb{CP}^2)'$ (which is rational).

are possible is called a **minimal model** of M. If M is neither rational nor ruled, then such a minimal model is guaranteed to be unique and can be called *the* minimal model of M.

We further investigate the case when K_M is nef:

7.2. How nef: numerical dimension

Recall that K_M would be ample if and only if we had both $K_M \cdot K_M > 0$ and $K_M \cdot C > 0$ for every curve C. After blow-downs, our K_M is now nef, meaning that we have $K_M \cdot K_M \geq 0$ and $K_M \cdot C \geq 0$.

The second stage of the classification is to determine how far K_M is from being ample. To measure this, we introduce the **numerical dimension** of M, denoted by $\mathrm{num}(M)$ and defined by the following three cases:

— $\mathrm{num}(M) = 0$: for every curve C, we have $K_M \cdot C = 0$;

— $\mathrm{num}(M) = 1$: there is a curve C with $K_M \cdot C > 0$, but $K_M \cdot K_M = 0$;

— $\mathrm{num}(M) = 2$: there is a curve C with $K_M \cdot C > 0$, and $K_M \cdot K_M > 0$.

What eventually emerges is the following:

Classification Theorem for K_M Nef. *Let M be a simply-connected complex surface, whose canonical bundle K_M is nef. We have:*

— *If* $\mathrm{num}(M) = 0$, *then M is a $K3$ surface.*

— *If* $\mathrm{num}(M) = 1$, *then M admits a (singular) fibration $M \to \mathbb{CP}^1$, with generic fiber a torus.*

— *If* $\mathrm{num}(M) = 2$, *then M is called of* **general type.** □

A $K3$ surface is a complex surface whose underlying smooth 4–manifold is the one we built in section 3.3 (page 127). From the perspective of algebraic geometry, there are many $K3$ surfaces. Their proper definition is: a complex surface is a **$K3$ surface** if it is simply-connected and has $K_M = 0$. They are all diffeomorphic, but span a moduli space of dimension 20.

Elliptic surfaces. A holomorphic fibration of a complex surface M with tori as generic fibers is called an **elliptic fibration** of M. (The name owes to the fact that a complex curve diffeomorphic to a torus is called an *elliptic curve.*) A concrete example of an elliptic fibration was explored earlier,[9] on the Kummer $K3$.

A complex surface that admits an elliptic fibration is called an **elliptic surface.** Keep in mind that, besides the generic torus-fibers, a random elliptic fibration will contain singular, as well as multiple, fibers.

---

[9] See back in section 3.3 (page 131).

Not all elliptic surfaces have $\operatorname{num}(M) = 1$: many of the $K3$'s are elliptic, as is the rational surface $\mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$. In consequence, complex surfaces with $\operatorname{num}(M) = 1$ are deserving of the name **proper elliptic surfaces.**

Elliptic surfaces are pretty well-understood, and the whole next chapter[10] is devoted to them.

Surfaces of general type. Surfaces of general type are the great unknown territory. Many things are known about them, but nothing amounting to any classification scheme.

> ***Fibrate it.*** *For example, similar to organizing an elliptic surface as a fibration by tori, if we start with a random surface M, then, maybe after blowing-up a few times, we can always organize M as a singular fibration over $\mathbb{CP}^1$ (if simply-connected), with the singularities of the curve-fibers merely simple double-points. Such a fibration is called a* Lefschetz fibration. *For surfaces of general type, it is known that the genus of the generic fiber must be at least 2. Besides this fact, at the outset not much information can be gleaned from such an approach.*[11]

The pluri-genera $P_m(M)$ cannot be used to distinguish surfaces of general type anymore than topological invariants, since we have

Lemma. *If M is a complex surface of general type, we have*

$$P_m(M) = \tfrac{6m^2-6m-1}{4} \operatorname{sign} M + \tfrac{4m^2-4m-1}{4} \chi(M) . \qquad \square$$

We are thus mostly left with numerical questions. For example:

Theorem. *Let M be a surface of general type, minimal with respect to blow-downs. Then we must have:* $3 \operatorname{sign} M + 2\chi(M) > 0$, $\chi(M) > 0$, *and*[12]

$$2 \operatorname{sign} M - 6\chi(M) - 12 \leq 12 \operatorname{sign} M \leq 9 \operatorname{sign} M + \chi(M) . \qquad \square$$

Wonderful.

Boring examples of surfaces of general type are $C' \times C''$, with C' and C'' complex curves of genus at least 2. These, of course, are not simply-connected.

An isolated example of a simply-connected surface of general type is the **Barlow surface**, which is homeomorphic to $\mathbb{CP}^2 \# 8 \overline{\mathbb{CP}}^2$ but not diffeomorphic to it. There are plenty more simply-connected surfaces of general type, as shown by the following result:

---

[10] The next chapter starts on page 301.

[11] S.K. Donaldson showed that every *symplectic* 4–manifold can be organized (maybe after a few blow-ups) as a (smooth) Lefschetz fibration. See the the end-notes of chapter 10 (page 416).

[12] The first inequality is called the *Noether inequality,* while the second is the *Bogomolov–Miyaoka–Yau inequality.*

Theorem *(U. Persson). For every pair m, n of integers satisfying $3m+2n > 0$, $m+n > 0$ and*

$$2m - 6n - 12 \;\leq\; 12m \;\leq\; 8m - \tfrac{40}{\sqrt[3]{3}}(m+n)^{2/3} \,,$$

there exists a minimal simply-connected complex surface M of general type, with $\operatorname{sign} M = m$ *and* $\chi(M) = n$. □

These surfaces are obtained as branched double covers of ruled surfaces, by using a careful control of the singularities of the branch locus.

The complex surfaces provided by this result are unfortunately quite big: for example, they must have $b_2^+(M) \geq 587$. The situation for small simply-connected surfaces of general type is unknown. Thus, we better leave this unfriendly neighborhood.

7.3. Alternative: Kodaira dimension

Equivalent to the numerical dimension used in the classification theorem above, one can use the better-known Kodaira dimension.

Recall the pluri-canonical maps

$$\Phi_{mK_M} : M \dashrightarrow \mathbb{CP}^N \,,$$

which were defined using holomorphic sections of mK_M. We define the **Kodaira dimension** of M as:

$$\operatorname{kod}(M) \;=\; \max\bigl\{\dim_{\mathbb{C}}\bigl(\operatorname{image}\Phi_{mK_M}\bigr) \bigm| m \gg 0\bigr\} \,.$$

An equivalent definition is in terms of the asymptotics of the pluri-genera $P_m(M)$, namely

$$\operatorname{kod}(M) = k \qquad \text{if and only if} \qquad P_m(M) \sim m^k \quad \text{for } m \gg 0 \,.$$

This translates into the following four cases:

- $\operatorname{kod}(M) = -\infty$: if for all m we have $P_m = 0$;
- $\operatorname{kod}(M) = 0$: if there is an m with $P_m \neq 0$, and $\{P_m\}$ is bounded;[13]
- $\operatorname{kod}(M) = 1$: if $\{P_m\}$ is unbounded, but $\{P_m/m\}$ is bounded;
- $\operatorname{kod}(M) = 2$: if $\{P_m/m\}$ is unbounded.

The Kodaira dimension effects the same partition of complex surfaces as the numerical dimension:

Theorem. *If K_M is nef, then* $\operatorname{kod}(M) = \operatorname{num}(M)$. □

Therefore one can rewrite the classification theorem by substituting Kodaira dimension instead of numerical dimension:

---

[13] In this case, each P_m is either 0 or 1.

Classification Theorem. *Let M be a simply-connected complex surface. Then:*

— *If* $\mathrm{kod}(M) = -\infty$, *then M is rational or ruled.*

— *If* $\mathrm{kod}(M) = 0$, *then M is a K3 surface.*

— *If* $\mathrm{kod}(M) = 1$, *then M is a (proper) elliptic surface.*

— *If* $\mathrm{kod}(M) = 2$, *then M is of general type.* □

Notice that, since the Kodaira dimension is defined for all surfaces, not necessarily minimal with respect to blow-downs, the above theorem does not require preparatory blow-downs. Indeed, $\mathrm{kod}(M)$ is invariant with respect to blow-ups/downs, and so are all the pluri-genera $P_m(M)$. Also, the Kodaira dimension can be $-\infty$, and, as stated above, that happens exactly for rational and ruled surfaces.

More than this, the Kodaira dimension can be defined for complex manifolds of any dimension. For example, it splits complex curves into three classes: *rational* curves (spheres) if $\mathrm{kod}(C) = -\infty$, *elliptic* curves (tori) if $\mathrm{kod}(C) = 0$, and curves of higher genus if $\mathrm{kod}(C) = 1$. This can be used in combination with:

Iitaka's Conjecture[14]. *If M is a complex surface that fibrates over a curve C with generic fiber F, then* $\mathrm{kod}(F) + \mathrm{kod}(C) \leq \mathrm{kod}(M)$. □

A simple consequence is that, as claimed, all products $C' \times C''$ of two curves of genera at least 2 must be surfaces of general type.

> *Finally, it is also worth mentioning that the pluri-genera are very powerful devices by themselves, and the simplest example is:*
>
> ***Castelnuevo's Criterion.*** *A surface M is rational if and only if $P_2(M) = 0$.* □

7.4. The Kähler case

In case we have a Kähler structure on M and we understand the class of the Kähler form, the classification can be simplified somewhat. First, some general comments about Kähler surfaces:

Assume that M is a projective surface, *i.e.*, M is embedded in some $\mathbb{CP}^N$. The projective space $\mathbb{CP}^N$ is endowed with a canonical Kähler metric (the *Fubini–Study metric*). Its Kähler form $\omega_{\mathbb{CP}^N}$ represents the Poincaré-dual of a hyperplane H in $\mathbb{CP}^N$, for example $H = \mathbb{CP}^{N-1} \subset \mathbb{CP}^N$. By restricting this standard Kähler structure of $\mathbb{CP}^N$ to M, we obtain an induced Kähler structure on M. The class $[\omega]$ of M's Kähler form is dual to any **hyperplane**

---

[14] The full Iitaka conjecture is that the inequality above is true for *all* complex dimensions and is of course still open. The particular statement above is proved.

section, *i.e.*, it is dual to the curve in M that is cut out by H. In other words, $[\omega] = [M \cap H]$. For a general Kähler surface, the class $[\omega]$ needs not be integral, but a complex deformation can take care of that.

Also, since we can integrate ω over every curve C and get $\int_C \omega > 0$, it follows that we must have

$$[\omega] \cdot C > 0$$

for every curve C in M, even for singular ones. In particular, if the Kähler class $[\omega]$ is integral, then it is *ample.*

A priori, the Kähler class $[\omega]$ and the canonical class K_M have nothing in common. Nonetheless, for *minimal* Kähler surfaces, their relation dictates the position of M in the classification scheme:

Kähler Classification. *Let M be a Kähler surface, and ω its Kähler form. If M is minimal with respect to blow-downs, then:*

— $K_M \cdot [\omega] < 0$ *if and only if M is rational or ruled.*

— $K_M \cdot [\omega] = 0$ *if and only if M is a K3 surface.*

— $K_M \cdot [\omega] > 0$ *and* $K_M \cdot K_M = 0$ *if and only if M is proper elliptic.*

— $K_M \cdot [\omega] > 0$ *and* $K_M \cdot K_M > 0$ *if and only if M is of general type.* □

7.5. Complex *versus* diffeomorphic

How does the other-worldly classification presented so far compare against the smooth topology of 4–manifolds?

First, by using Freedman's theorem, we can quickly identify the topological types that can occur through complex surfaces:

Theorem (*Topological types*). *If M is a simply-connected complex surface, then M must be homeomorphic either to*

$$\#m\,\mathbb{CP}^2\,\#n\,\overline{\mathbb{CP}}^2 \qquad \text{or} \qquad \#m\,K3\,\#\,n\,S^2 \times S^2\,. \qquad \square$$

> *Of course, not all the latter manifolds admit complex structures. For example, $\#3\,\mathbb{CP}^2$ admits almost-complex structures, but none of them can even be symplectic,*[15] *let alone integrable. Others, like $\mathbb{CP}^2\,\#\mathbb{CP}^2$, do not even admit almost-complex structures.*

Second, a count of complex structures:

Theorem (*Number of complex structures*). *Any simply-connected 4–manifold supports at most* finitely-many *distinct complex structures (up to complex deformations).* □

---

[15] An argument that $\#3\,\mathbb{CP}^2$ cannot be symplectic is outlined in footnote 11 on page 491.

Next, we quote results that show that most of the objects encountered in the Enriques–Kodaira classification are in fact diffeomorphism invariants, and therefore should be visible in the smooth topology of surfaces. All the remarkable results that follow were proved using gauge theory, either in the Donaldson or Seiberg–Witten flavor. (The condition $\operatorname{kod}(M) \geq 0$ that appears in some of the statements mainly ensures that there is a unique minimal model M_m of M.)

Theorem (*Smooth* (-1)*-curves*). *If M is a complex surface with* $\operatorname{kod}(M) \geq 0$, *then every smoothly-embedded sphere S with $S \cdot S = -1$ is homologous to a complex (-1)-curve (maybe with the opposite orientation).* □

Therefore a complex surface M with $\operatorname{kod}(M) \geq 0$ splits smoothly as $N \# \overline{\mathbb{CP}}^2$ if and only if M can be blown-down to a surface diffeomorphic to N.

> *As a consequence, it can be shown that, if M has* $\operatorname{kod}(M) \geq 0$, *then every smoothly-embedded sphere S with $S \cdot S \geq 0$ must in fact be homologically trivial. On the other hand, a complex surface is rational or ruled if and only if it contains a homologically-nontrivial smoothly- embedded sphere S with $S \cdot S \geq 0$.*

Theorem (*Kodaira dimension*). *If M and N are diffeomorphic complex surfaces, then they must have*

$$\operatorname{kod}(M) = \operatorname{kod}(N) \,.$$ □

This last statement is also known as the *Van de Ven conjecture.*

Corollary (*Numerical dimension*). *If M and N are diffeomorphic complex surfaces that are minimal with respect to blow-downs, then they must have*

$$\operatorname{num}(M) = \operatorname{num}(N) \,.$$ □

Theorem (*Pluri-genera*). *If M and N are diffeomorphic complex surfaces, then they must have*

$$P_m(M) = P_m(N) \qquad \textit{for all } m.$$ □

Theorem (*Canonical class and* (-1)*-curves*). *Let M and N be complex surfaces, with* $\operatorname{kod}(M) \geq 0$. *Let*

$$f\colon M \xrightarrow{\cong} N$$

be an orientation-preserving diffeomorphism. Let M_m be the minimal model of M, and N_m the minimal model of N. Let $E_1^M, \ldots, E_p^M$ be the (-1)-curves of M that were blown-down to get M_m; similarly let $E_1^N, \ldots, E_q^N$ be the curves of N blown-down to get N_m. Pull the canonical class of M_m back to M and denote it by K_{M_m}; similarly for K_{N_m}. Then we have:

$$p = q \qquad \textit{and} \qquad f^* E_i^N = \pm E_j^M \qquad \textit{and} \qquad f^* K_{N_m} = \pm K_{M_m} \,.$$ □

Notice how one must use the pull-back K_{M_m} of the canonical class of a minimal model of M and not K_M itself. Indeed, K_M is *not* a diffeomorphism invariant.

Finally, we have:

Conjecture *(open).* *Two simply-connected complex surfaces are diffeomorphic if and only if they are complex deformations of each other.*

The open part of this conjecture is the same with the unfinished part of the classification: the realm of surfaces of general type. Obviously, all known creatures abide by this conjecture. The conjecture fails for non-simply-connected surfaces.

7.6. Notes

Bibliography

The classification of complex surfaces was outlined by **F. Enriques**, see for example his book ***Le superficie algebriche*** [Enr49]. This was set on a solid basis and further refined by **K. Kodaira** in the series of papers ***On compact complex analytic surfaces. I–III*** [Kod63] and ***On the structure of compact complex analytic surfaces. I–IV*** [Kod68].

The classification scheme that we presented, in terms of intersections of K_M, follows the outline of **M. Reid**'s ***Chapters on algebraic surfaces*** [Rei97]. Presenting the classification in this way is in tune with recent developments in higher-dimensional complex geometry, specifically the Mori theory of complex 3–folds.

The standard reference for complex surfaces is **W. Barth, C. Peters** and **A. Van de Ven**'s ***Compact complex surfaces*** [BPVdV84]. This classic recently got a second edition [BHPVdV04], enlarged and including, besides a new coauthor (**K. Hulek**), discussions in terms of nef classes, as well as a glance at gauge theory. A short classical proof of the classification can be found in **A. Beauville**'s ***Surfaces algébriques complexes*** [Bea78], translated as ***Complex algebraic surfaces*** [Bea83, Bea96]. Of course, one cannot skip citing **P. Griffiths** and **J. Harris**'s bible ***Principles of algebraic geometry*** [GH78, GH94] for a general introduction to complex algebraic geometry, including surface theory.

A more detailed topologically-minded survey of surfaces of general type, including constructions of (non-simply-connected) examples, can be read from **R. Gompf** and **A. Stipsicz**'s ***4–Manifolds and Kirby calculus*** [GS99, sec 7.4].

The Barlow surface was constructed by **R. Barlow**'s ***A simply connected surface of general type with*** $p_g = 0$ [Bar85] and was proved to be an exotic $\mathbb{CP}^2 \# 8\,\overline{\mathbb{CP}}^2$ in **D. Kotschick**'s ***On manifolds homeomorphic to*** $\mathbb{CP}^2 \# 8\,\overline{\mathbb{CP}}^2$ [Kot89].

The simply-connected examples of surfaces of general type mentioned were constructed in **U. Persson**'s ***Chern invariants of surfaces of general type*** [Per81]. See also **U. Persson, C. Peters** and **G. Xiao**'s ***Geography of spin surfaces*** [PPX96].

For the comparison between the complex and smooth worlds, many of the results quoted have been obtained through the use of Donaldson theory and some are gathered in **R. Friedman** and **J. Morgan**'s book ***Smooth four-manifolds and complex surfaces*** [FM94a]. Finiteness of the number of complex structures on a smooth manifold is due to **R. Friedman** and **J. Morgan**'s ***Complex versus differentiable classification of algebraic surfaces*** [FM89] and their ***Smooth four-manifolds and complex surfaces*** [FM94b].

Hard work on the smooth classification of elliptic surfaces was essential to the results quoted. This was painfully achieved through **R. Friedman** and **J. Morgan**'s ***On the diffeomorphism types of certain algebraic surfaces. I & II*** [FM88], their ***Smooth four-manifolds and complex surfaces*** [FM94a], then **J. Morgan** and **K. O'Grady**'s ***Differential topology of complex surfaces*** [MO93], **S. Bauer**'s ***Some nonreduced moduli of bundles and Donaldson invariants for Dolgachev surfaces*** [Bau92] and his ***Diffeomorphism types of elliptic surfaces with*** $p_g = 1$,

R. Friedman's *Vector bundles and $SO(3)$–invariants for elliptic surfaces* [**Fri95**], and **J. Morgan** and **T. Mrowka**'s *On the diffeomorphism classification of regular elliptic surfaces* [**MM93**].

The invariance of the Kodaira dimension (Van de Ven conjecture) was proved in **R. Friedman** and **Z. Qin**'s *On complex surfaces diffeomorphic to rational surfaces* [**FQ95**], in **V. Pidstrigach** and **A. Tyurin**'s *The smooth structure invariants of an algebraic surface defined by the Dirac operator* [**PT92**], and in **V. Pidstrigach**'s *Some glueing formulas for spin polynomials and a proof of the Van de Ven conjecture* [**Pid94**].

The impact of Seiberg–Witten theory becomes quite apparent when one sees how easy it is to prove all the above results using the Seiberg–Witten invariants instead of Donaldson's. For example, the smooth invariance of the canonical class K_M, pluri-genera, and (-1)–curves can be read from **R. Brussee**'s *The canonical class and the C^∞ properties of Kähler surfaces* [**Bru96**] and from **R. Friedman** and **J. Morgan**'s *Algebraic surfaces and Seiberg–Witten invariants* [**FM97**], both using Seiberg–Witten theory. See also **C. Okonek** and **A. Teleman**'s *Les invariants de Seiberg–Witten et la conjecture de van de Ven* [**OT95b**] and *The coupled Seiberg–Witten equations, vortices, and moduli spaces of stable pairs* [**OT95a**].

A nice survey is **R. Friedman**'s *Donaldson and Seiberg–Witten invariants of algebraic surfaces* [**Fri97**].

Chapter 8

Elliptic Surfaces

FROM the point-of-view embraced in this volume, simply-connected complex surfaces are either trivial (ruled/rational surfaces), have already been discussed (the $K3$ surface), seem too big and are not understood (surfaces of general type), or are discussed in this chapter (elliptic surfaces).[1] Indeed, in the sequel we will build all simply-connected smooth 4–manifolds that support complex structures admitting elliptic fibrations. Then we will notice that they cluster into infinite families of homeomorphic but non-diffeomorphic 4–manifolds.

We start by presenting the simplest elliptic surface, the so-called rational elliptic surface, supported by $\mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$. In section 8.2 (page 306) we present the fiber-sum construction, which is used to build new elliptic surfaces. After that, section 8.3 (page 310) explains the logarithmic transformation, which essentially cuts out a fiber's neighborhood, twists it m times, then glues it back in. By applying these two techniques, all manifolds supporting simply-connected elliptic surfaces are obtained.

Then, in section 8.4 (page 314), we explain the classification of elliptic surfaces up to homeomorphisms and up to diffeomorphisms, which leads to examples of 4–manifolds supporting infinitely-many smooth structures (these are obtained from each other through logarithmic transformations).

---

[1] *[I]t is written that animals are divided into (a) those that belong to the emperor; (b) embalmed ones; (c) those that are trained; (d) suckling pigs; (e) mermaids; (f) fabulous ones; (g) stray dogs; (h) those that are included in this classification;* [...] — J.L. Borges **[Bor99]**.

8.1. The rational elliptic surface

Finding tori in $\mathbb{CP}^2$. A random polynomial of degree d, homogeneous in the complex coordinates z_0, z_1, z_2 of $\mathbb{CP}^2$, defines a complex curve C (maybe singular) in $\mathbb{CP}^2$, whose homology class is $[C] = d\,[\mathbb{CP}^1]$.

Since $K_{\mathbb{CP}^2} = -3\,[\mathbb{CP}^1]$, the adjunction formula[2]

$$\chi(C) + C \cdot C = -K_{\mathbb{CP}^2} \cdot C$$

implies that, if C is non-singular, then $\chi(C) = 3d - d^2$, and thus the genus of C is $g(C) = \frac{1}{2}(d-1)(d-2)$.

In particular, any non-singular curve of degree 3 must be a torus.

Building a fibration. Consider two *generic* polynomials P and Q, homogeneous of degree 3. The two curves defined by P and Q in $\mathbb{CP}^2$ are two tori, which must have $9 = 3 \times 3$ points $p_1, \ldots, p_9$ in common.

The family of curves described by

$$\{sP + tQ \mid [s:t] \in \mathbb{CP}^1\}$$

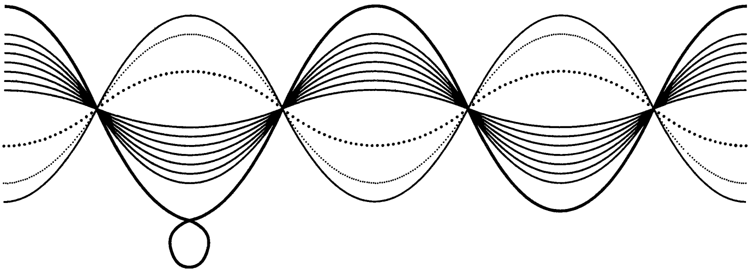
is a family of elliptic curves (some of them singular), all passing through $p_1, \ldots, p_9$, as suggested in figure 8.1. Aside from $p_1, \ldots, p_9$, each point of $\mathbb{CP}^2$ belongs to exactly one of these curves, and thus we can define a map

$$\mathbb{CP}^2 \setminus \{p_1, \ldots, p_9\} \longrightarrow \mathbb{CP}^1$$

that sends x to $[s:t]$ whenever x belongs to the curve $sP + tQ = 0$. See figure 8.2 on the next page. To extend this map over $p_1, \ldots, p_9$, we blow-up at these nine points (thus separating the tori) and end up with a well-defined map

$$\mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2 \longrightarrow \mathbb{CP}^1\,,$$

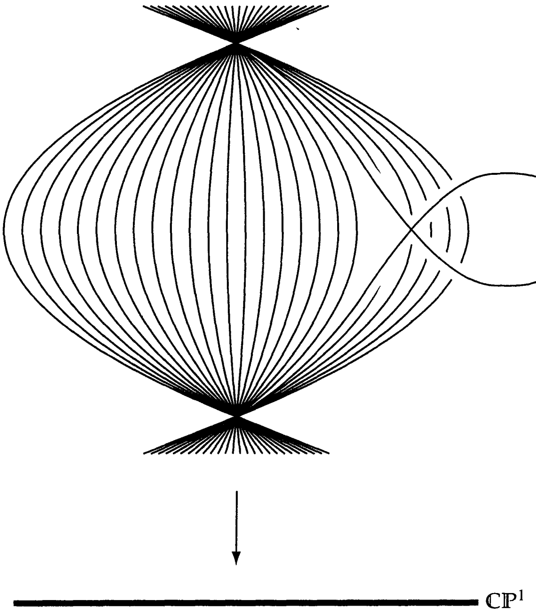
as sketched in figure 8.3 on the facing page.



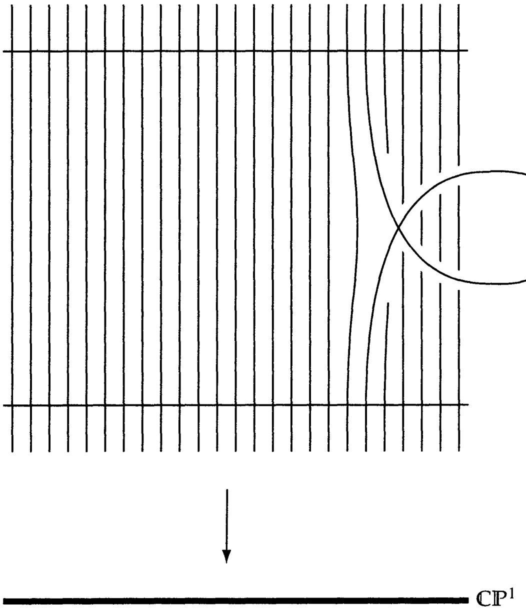
8.1. Pencil of curves

---

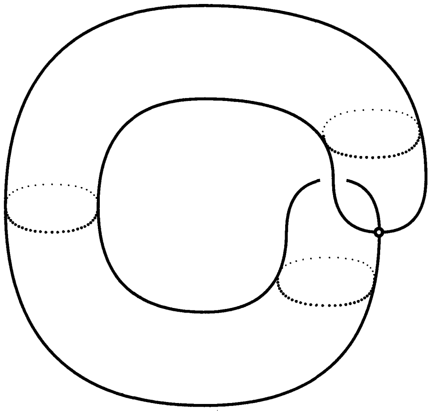
[2] From section 6.3 (page 281).



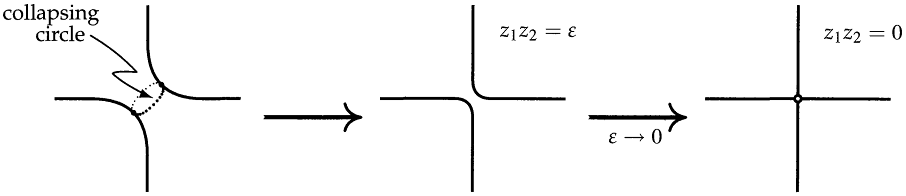
8.2. Pencil of curves, and its projection



8.3. Blown-up pencil



8.4. Fishtail fiber



8.5. Apparition of the singularity

Singular fibers. This map is a fibration, with generic fiber a torus. Among the fibers must be singular ones as well, and these can be of several types. The most common are the fishtail and the cusp singular fibers:

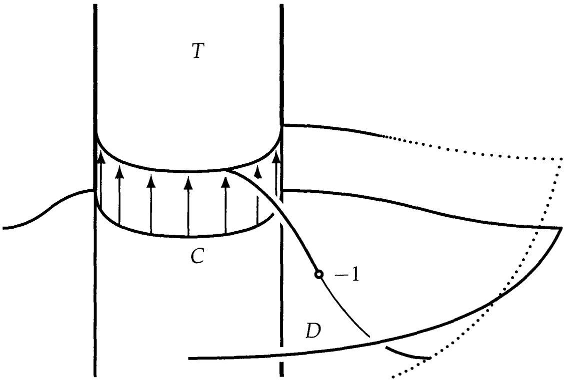
— The **fishtail fiber** is the most frequent. It consists of a sphere with a point of transverse self-intersection, as in figure 8.4.

The fishtail appears from the collapse of a homologically-nontrivial circle in the generic fiber, as outlined in figure 8.5. (Such a collapsing circle is called a **vanishing cycle**, *i.e.*, a circle that bounds a disk of self-intersection -1 embedded in the fiber's complement, as pictured in figure 8.6 on the next page.)

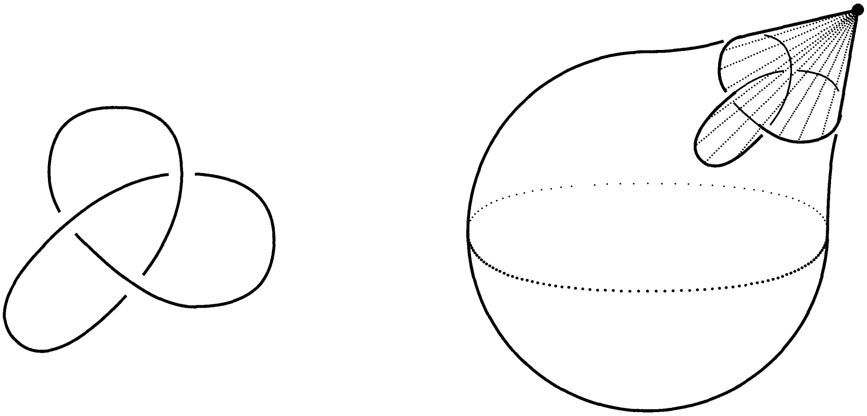
— The **cusp fiber** is a sphere as well. It is embedded with a singular point whose neighborhood looks like a cone over the trefoil knot, see figure 8.7 on the facing page.

A cusp appears from the collapse of two vanishing cycles, *i.e.*, two circles, generators of $H_1(torus)$ in the generic fiber (think "meridian" and "longitude").

The singular fibers that appear in a particular fibration $\mathbb{CP}^2 \# 9\overline{\mathbb{CP}}^2 \to \mathbb{CP}^1$ depend on the concrete choice of polynomials P and Q. For a generic



8.6. Vanishing cycle



8.7. Trefoil knot, and cusp fiber

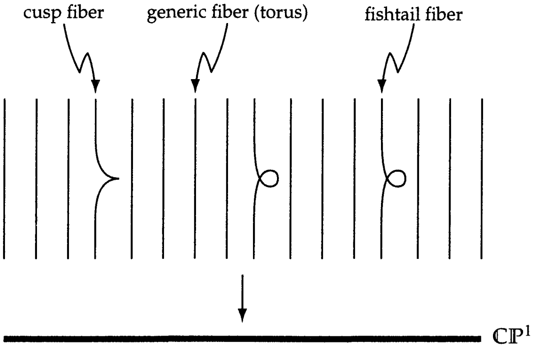
choice, only fishtail fibers appear. Nonetheless, besides fishtails and cusps, there are other singular fibers upon which one can stumble, and a complete list will be displayed in the end-notes of this chapter (page 319).

For most constructions to follow we will like to have a cusp fiber around—its presence will ensure that the result of various procedures is independent, up to diffeomorphisms, from the various choices made along the way.[3]

The manifold $\mathbb{CP}^2\,\#9\,\overline{\mathbb{CP}}^2$ is the simplest example of an elliptic surface. It is called the **rational elliptic surface** and is denoted by $E(1)$. (Keep in mind that $E(1)$ does not denote a specific elliptic fibration; it merely denotes the manifold $\mathbb{CP}^2\,\#9\,\overline{\mathbb{CP}}^2$ viewed as an elliptic surface, *i.e.*, endowed with some elliptic fibration over $\mathbb{CP}^1$.)

---

[3] Compare with **R. Gompf**'s *Nuclei of elliptic surfaces* [**Gom91a**], and with a statement on page 537.



8.8. An elliptic fibration

Homology. *The 2–homology of $E(1)$ has the natural basis $\{e_0, e_1, \ldots, e_9\}$ coming from copies of $\mathbb{CP}^1$ inside each of the ten projective planes. In this basis, the intersection form of $E(1)$ is*

$$Q_{E(1)} = [+1] \oplus 9\,[-1] \ .$$

If, instead, we choose the basis $\{3e_0 - e_1 - \cdots - e_9,\ e_9,\ e_1 - e_2,\ e_2 - e_3,\ \ldots,\ e_7 - e_8,\ \ - e_0 + e_6 + e_7 + e_8\}$, then the intersection form has matrix

$$Q_{E(1)} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \oplus -E_8 \ .$$

This rewriting of $Q_{E(1)}$ helps to see how the K3 surface can be obtained from two copies of $E(1)$, as we will see shortly. Observe that the first element $3e_0 - e_1 - e_2 - \cdots - e_9$ of the basis above is the class of $E(1)$'s fiber—a cubic torus, blown-up nine times.

All other simply-connected elliptic surfaces can be obtained (up to diffeomorphisms) starting from $E(1)$ and using two techniques: fiber-sums and logarithmic transformations.

8.2. Fiber-sums

Smooth construction. We start with a random elliptic fibration

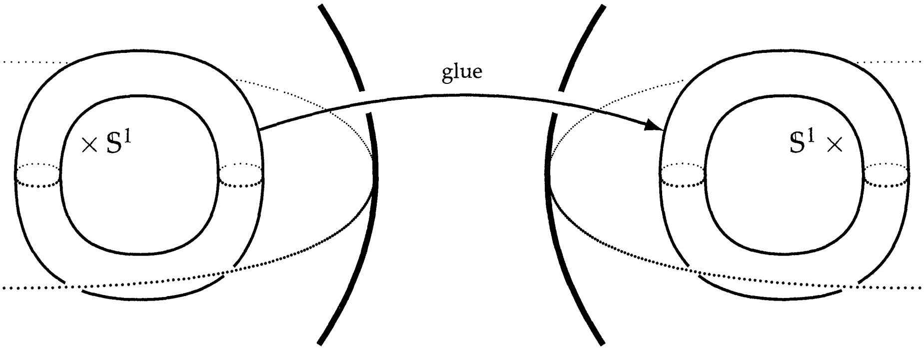
$$E(1)\colon\ \mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2 \xrightarrow{p} \mathbb{CP}^1$$

that contains a cusp fiber. We will cut out a tubular neighborhood of a generic fiber and glue together two copies of such a carved-out $E(1)$. Then, we repeat.

Pick a generic torus fiber F, sitting over some point $x = p[F]$ in $\mathbb{CP}^1$. Take a small disk D around x and cut out of $E(1)$ the fibered tubular neighborhood $E|_D = p^{-1}[D]$ of F. Since $E|_D \cong \mathbb{T}^2 \times \mathbb{D}^2$, the remainder is a 4–manifold with boundary $\mathbb{T}^2 \times \mathbb{S}^1$. We then glue two copies of such a butchered $E(1)$ along their boundaries,[4] as suggested in figure 8.9. The result, called the **fiber-sum** of $E(1)$ and $E(1)$, will be denoted by

$$E(2) = E(1) \#_{\mathit{fiber}} E(1) .$$

When building $E(2)$, we respected the fibered structure of its components, and therefore the result still fibers over $\mathbb{CP}^1$ and thus is itself an elliptic surface.



8.9. Fiber-summing two elliptic fibrations

It turns out that $E(2)$ is diffeomorphic to the $K3$ surface:

$$E(2) \cong K3 .$$

Thus, we now have two descriptions of the $K3$ manifold: the Kummer construction from section 3.3 (page 127) and the fibered-sum construction above. Of course, the two versions, while diffeomorphic, display quite different elliptic fibrations.[5]

Repeating the fiber-sum procedure yields a whole family of manifolds

$$E(n) = \#_{\mathit{fiber}}\, n\, E(1) ,$$

which all fiber over $\mathbb{CP}^1$ and are simply-connected elliptic surfaces. Starting with a fibration $E(1)$ that contained a cusp fiber ensures that, as smooth 4–manifolds, the resulting $E(n)$'s are uniquely determined by n and do not depend on the auxiliary choices made during construction.

The canonical class of $E(n)$ is

$$K_{E(n)} = (n-2)[F] ,$$

where F denotes a generic fiber. (In particular, $K_{K3} = 0$.)

---

[4] Using an orientation-reversing diffeomorphism, coming from an orientation-preserving identification of the torus fibers, times complex-conjugation on $\mathbb{S}^1$.

[5] For example, look at the singular fibers: $E(2)$ can be arranged to have only fishtails and cusps, while the Kummer $K3$ has singular fibers of type I_0^*, made from five spheres.

Homology. *For other classical invariants of the* $E(n)$*-manifolds, we have*

$$\operatorname{sign} E(n) = -8n ,$$

with $b_2(E(n)) = 12n - 2$ *and* $b_2^+(E(n)) = 2n - 1$. *Therefore, when* n *is even, the intersection form is even and can be written*

$$Q_{E(\text{even } n)} = \oplus\, n(-E_8) \oplus (2n-1)H ,$$

while when n *is odd we have*

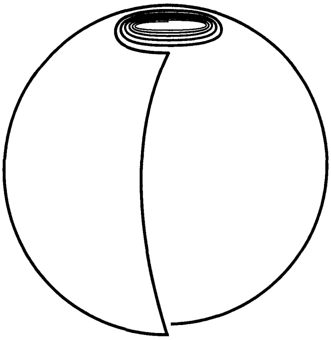
$$Q_{E(\text{odd } n)} = \oplus\, (2n-1)\left[+1\right] \oplus (10n-1)\left[-1\right] .$$

Holomorphic construction. It is certainly not at all clear that the result of a fiber-sum construction as described above would admit any complex structures. Indeed, a complex geometer would use an altogether different procedure to obtain a complex manifold diffeomorphic to our $E(n)$'s. Namely, in order to build $E(n)$, she would take an n–fold cyclic cover of $E(1)$, branched over two regular fibers.[6]

This can be described as follows: Consider the map

$$\varphi_n \colon \mathbb{CP}^1 \to \mathbb{CP}^1 \qquad \varphi_n(z) = z^n ,$$

with 0 and ∞ as fixed points. Visually, φ_n takes the 2–sphere $\mathbb{CP}^1$ and wraps it n times around itself, fixing the north and south poles, as pictured in figure 8.10.



8.10. Branched cover $\mathbb{CP}^1 \to \mathbb{CP}^1$, as an onion peel

Arrange that $E(1)\colon\ \mathbb{CP}^2 \# 9\overline{\mathbb{CP}}^2 \xrightarrow{p} \mathbb{CP}^1$ have regular fibers over 0 and ∞. We then build the pull-back $\varphi_n^* E(1)$ of the fibration $E(1)$. It turns out that the result is exactly the earlier n–fold fiber-sum:

$$E(n) = \varphi_n^* E(1) .$$

---

[6] An n–fold **branched cover** is a map $f\colon M \to N$ and a subset $B \subset M$ (called the *branch locus*) so that $f|_{M \setminus B}$ is an n–fold cover map, and around each $b \in B$ there are charts so that f acts by $f(z,w) = (z^p, w)$ for some integer $p > 0$ ($z \in \mathbb{C}$).

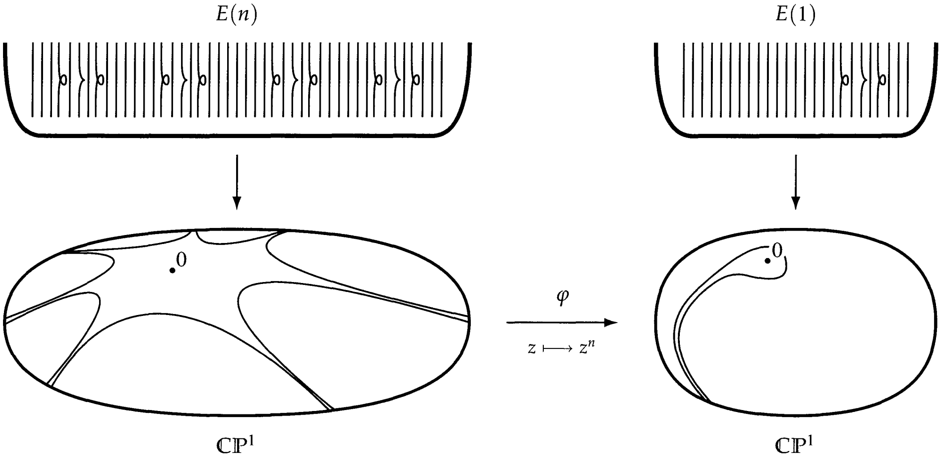
In other words, $E(n)$ can be defined directly as the set

$$E(n) = \{(z,e) \in \mathbb{CP}^1 \times E(1) \mid p(e) = z^n\} .$$

The latter is clearly an n–fold cover of $E(1)$, branched along the regular fibers over 0 and ∞, and it describes a complex surface.

To see that this construction is indeed diffeomorphic to the n–fold fiber-sum of $E(1)$ described earlier, arrange that all singular fibers of $E(1)$ sit over a big disk D in $\mathbb{CP}^1$ that does not contain either 0 or ∞. Then $\mathbb{CP}^1 \setminus D$ is a disk with no singular fibers over it, and so $E(1)|_{\mathbb{CP}^1 \setminus D}$ is diffeomorphic to the trivial fibration $\mathbb{T}^2 \times (\mathbb{CP}^1 \setminus D) \longrightarrow \mathbb{CP}^1 \setminus D$. We like to think of $\mathbb{CP}^1 \setminus D$ as a small disk and view $E(1)|_D$ as $E(1)$ from which a small tubular neighborhood $E(1)|_{\mathbb{CP}^1 \setminus D}$ of some regular fiber has been cut out. In other words, we view $E(1)|_D$ as containing most of $E(1)$.

On one hand, since D does not touch either 0 or ∞, its preimage under φ_n is made of n disjoint disks in $\mathbb{CP}^1$, and thus the pull-back $\varphi_n^*(E(1)|_D)$ is diffeomorphic to n disjoint copies of $E(1)|_D$. On the other hand, complementarily, the preimage of $\mathbb{CP}^1 \setminus D$ can only be $\mathbb{CP}^1$ with n disks cut out, and thus the pull-back $\varphi_n^*(E(1)|_{\mathbb{CP}^1 \setminus D})$ will be diffeomorphic to the trivial $\mathbb{T}^2$–fibration over $\mathbb{CP}^1$ with n disks cut out (trivial, because it is a genuine bundle, over a contractible base). In conclusion, we exhibited $\varphi_n^* E(1)$ as made from n copies of $E(1)|_D$, glued-up by an ambient trivial fibration. See figure 8.11.



8.11. Pulling-back the n–fold fiber-sum

Thinking of $E(1)|_D$ as the complement of a thin tubular neighborhood of a regular fiber of $E(1)$ and imagining that the disk $\mathbb{CP}^1 \setminus D$ (and thus also its preimage through φ) is made ever smaller, should help visualize that $E(n)$ is indeed the n–fold fiber-sum of $E(1)$.

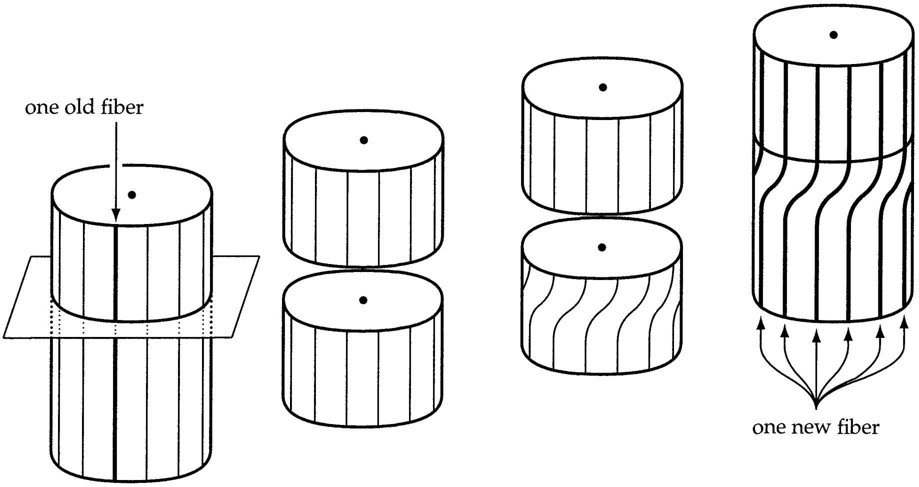
8.3. Logarithmic transformations

We will take an elliptic fibration

$$E \xrightarrow{p} \mathbb{CP}^1 ,$$

cut out a tubular neighborhood $E|_D$ of a generic torus fiber, then glue in its stead an m–times twisted version of $E|_D$. One effect is the creation of a **multiple fiber**, specifically, a fiber that (while still being a torus) is covered m times by surrounding generic fibers.

Smooth construction. Fix an integer $m > 1$. Choose a small disk D in $\mathbb{CP}^1$ and consider the fibered set $E|_D = p^{-1}[D]$. Assume that $E|_D$ contains only regular fibers of E. Choose some smooth parametrization of $E|_D$ as $\mathbb{D}^2 \times \mathbb{S}^1 \times \mathbb{S}^1$. Take $E|_D$ out of E, then cut open the second circle-factor of $E|_D$ to obtain a copy of $\mathbb{D}^2 \times \mathbb{S}^1 \times [0, 1]$. Rotate one end of this cylinder by an angle of $2\pi/m$, then glue back, as suggested in figure 8.12.



8.12. Logarithmic transformation

The result is a new fibration $E_m^D \to D$. Everywhere but at the center the new fibers of E_m^D are obtained from gluing p old fibers of $E|_D \to D$ "end-to-end". In fact, away from the center, this new fibration is isomorphic with the old $E|_D$:

$$E_m^D|_{D\setminus 0} \approx E|_{D\setminus 0} .$$

Therefore E_m^D can be plugged back into E and fill the gap created when we removed $E|_D$.

The elliptic fibration resulting from replacing $E|_D$ with E_m^D is called the **logarithmic transformation** of E with multiplicity m, and is denoted by

$$E_m \longrightarrow \mathbb{CP}^1 .$$

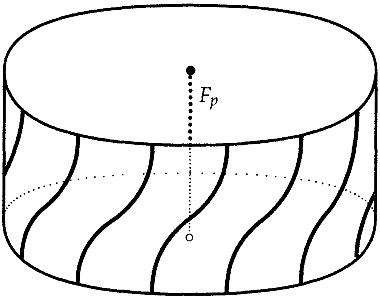
If the initial elliptic fibration E contained a cusp fiber, then the resulting E_m is independent, up to diffeomorphisms, of the various choices made along the way.

What is new in E_m is the central fiber F_m of E_m^D, which is now covered m times by the fibers around it. Indeed, the new fibers of E_m^D, which appeared from gluing m cut-open old fibers end-to-end, are twisting m times around the unchanged central fiber F_m. See figure 8.13.

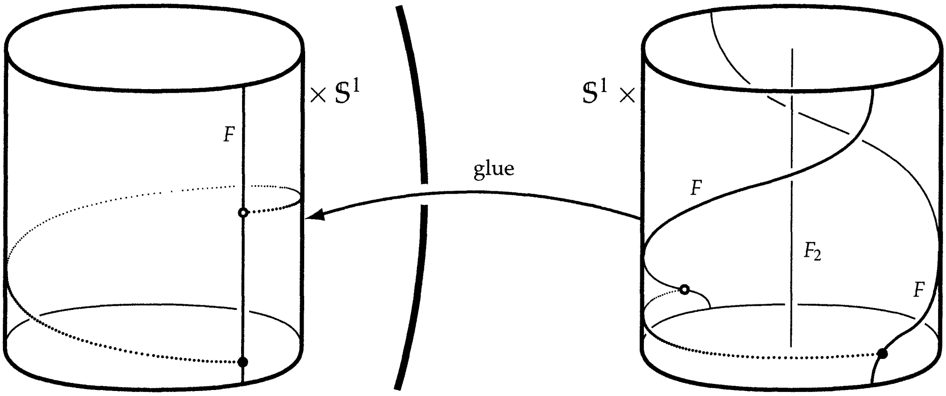
This F_m is called a **multiple fiber**, and we will say that it has *multiplicity* m. Homologically, we have

$$[F] = m\,[F_m]\,,$$

where F denotes a generic fiber. Look also at figure 8.14.



8.13. Replacement piece with multiple fiber



8.14. Logarithmic transformation (of multiplicity 2)

Notice that this whole construction is merely a 4*–dimensional version of creating a multiple fiber in a Seifert* 3*–manifold (i.e., a* 3*–manifold fibered by circles) by using Dehn surgery. The dimension is raised from* 3 *to* 4 *by crossing with* S^1.

In what follows,

$$E(n)_{p,q}$$

will denote the surface $E(n)$ to which two logarithmic transformations have been applied, of multiplicities p and q. While more than two logarithmic transformations can certainly be performed, they cease to yield simply-connected elliptic surfaces.

Furthermore, if one performs logarithmic transformations of multiplicities having some nontrivial common divisor, then the result again ceases to be simply-connected. Thus, when writing $E(n)_{p,q}$, we will *always* assume that p and q are *coprime* integers.

The order in which the logarithmic transformations are performed does not matter. Thus, when writing $E(n)_{p,q}$ we will also assume that $p < q$. Finally, since a logarithmic transformation of multiplicity 1 has no smooth effect (in the presence of a cusp fiber), we will write:

$$E(n)_{1,1} = E(n) \qquad \text{and} \qquad E(n)_{1,q} = E(n)_q \,.$$

Thus, the writing "$E(n)_{p,q}$" will include the cases when two, one, or no multiple fibers have been created in $E(n)$.

The surface $E(n)_{p,q}$ has canonical class

$$K_{E(n)_{p,q}} = (n-2)[F] + (p-1)[F_p] + (q-1)[F_q] \,.$$

> ***Homology.*** *We still have* $\operatorname{sign} E(n)_{p,q} = -8n$, *and* $b_2\big(E(n)_{p,q}\big) = 12n-2$, *and* $b_2^+\big(E(n)_{p,q}\big) = 2n-1$, *just as we had for* $E(n)$; *but the intersection form remains even if and only if* n *was even and both* p *and* q *are odd.*

It is important to note that the *smooth* logarithmic transformation can be generalized to more general gluings of $F \times D$ back into $E \setminus F$, which use various self-diffeomorphisms of the boundary 3–torus. These generalized logarithmic transformations will be presented in section 12.1 (page 536), in the context of their effect on the Seiberg–Witten invariants.

Finally, we should mention that logarithmic transformations admit an alternative smooth description using the *rational blow-down* construction of R. Fintushel and R. Stern. See the end-notes of chapter 12 (page 547).

Holomorphic construction. Of course, so far it is far from clear why the result of the above procedure would still be a complex surface. The problem is that, while the parametrization $E|_D \approx D \times S^1 \times S^1$ can be done by a diffeomorphism, it cannot be made holomorphically: the complex structure of the torus-fibers is not constant. To make clear the complex-holomorphic nature of logarithmic transformations, we must proceed with a bit more caution, and find a suitable holomorphic replacement for the parametrization of $E|_D$.

Start with a generic fiber F of $p\colon E \to \mathbb{CP}^1$ and pick a small disk D in $\mathbb{CP}^1$ around $p[F]$, parametrized as the unit-disk $\mathbb{D}^2$ in $\mathbb{C}$. Assume that every fiber over this $\mathbb{D}^2$ is regular. Each fiber is a torus, but its complex structure might change from fiber to fiber.

Nonetheless, if we choose a point in an elliptic curve and declare it to be the identity element, then we effectively organize our torus as an Abelian group (isomorphic to the group $\mathbb{S}^1 \times \mathbb{S}^1$). We pick such identity elements in each fiber over $\mathbb{D}^2$ through the use of a random holomorphic section

$$e\colon \mathbb{D}^2 \to E|_{\mathbb{D}^2} .$$

As a consequence, each fiber $E|_x$ is now viewed as a group with identity $e(x)$. Further, we locate the elements of order m in $E|_x$, *i.e.*, those $t \in E|_x$ for which we have

$$m \cdot t = e(x)$$

in the group structure of $E|_x$.

These order–m elements specify a m^2–fold trivial cover of $\mathbb{D}^2$ inside $E|_{\mathbb{D}^2}$. We choose a branch of this covering *via* a holomorphic section

$$\tau\colon \mathbb{D}^2 \to E|_{\mathbb{D}^2} ,$$

and hence we can write

$$m \cdot \tau(x) = e(x) .$$

Consider now the map

$$\varphi_m\colon \mathbb{D}^2 \to \mathbb{D}^2 \qquad \varphi_m(z) = z^m .$$

Away from the center of $\mathbb{D}^2$, φ_m is a simple m–fold cover of $\mathbb{D}^2 \setminus 0$. Then we pull-back the fibration of $E|_{\mathbb{D}^2}$ through φ_m to get a new holomorphic fibration

$$\varphi_m^*(E|_{\mathbb{D}^2}) \longrightarrow \mathbb{D}^2 .$$

To make holomorphy even more obvious, we can exhibit $W = \varphi_m^*(E|_{\mathbb{D}^2})$ as

$$W = \left\{(z,t) \in \mathbb{D}^2 \times E|_{\mathbb{D}^2} \mid p(t) = z^m\right\} ,$$

which fibrates over $\mathbb{D}^2$ through the projection on the first factor. Its fiber over any $z \in \mathbb{D}^2$ is identical with the fiber of $E|_{\mathbb{D}^2}$ over z^m. Specifically, for every $z \neq 0$, each of the fibers of W over the m points

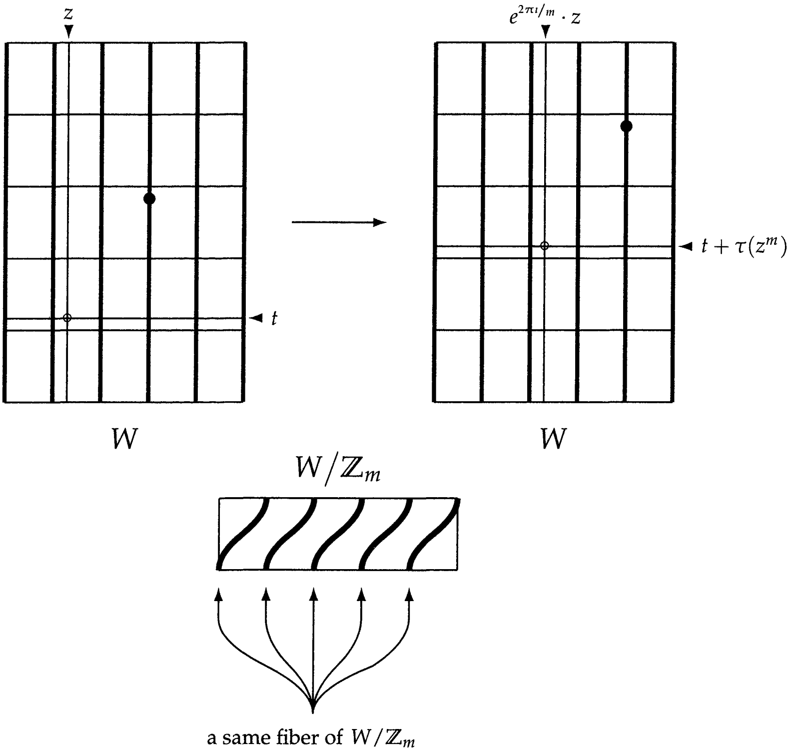
$$z, \quad e^{2\pi i/m} \cdot z, \quad e^{2 \cdot 2\pi i/m} \cdot z, \quad \ldots , \quad e^{(p\,-\,1) \cdot 2\pi i/m} \cdot z$$

is identifiable with the fiber of $E|_{\mathbb{D}^2}$ over z^m. Indeed, if we take the action of $\mathbb{Z}_m$ on W that is generated by the map sending $(z,t) \in W$ to $(e^{2\pi i/m} \cdot z,\ t)$, then the resulting quotient of W by this action would simply be $E|_{\mathbb{D}^2}$ again.

Instead, consider the action of $\mathbb{Z}_m$ generated by the map

$$(z,t) \longmapsto \left(e^{2\pi i/m} \cdot z,\ t + \tau(z^m)\right) .$$

With this action, at the same time with jumping from one copy of the fiber to the next (by going from z to $e^{2\pi i/m} \cdot z$), we also jump "up" inside the fibers (through translation by τ). See figure 8.15.



8.15. Quotient for logarithmic transformation

The quotient $W/\mathbb{Z}_m$ is a new fibration over $\mathbb{D}^2$, which has a multiple fiber on top of 0. More, the restriction of $W/\mathbb{Z}_m$ over $\mathbb{D}^2 \setminus 0$ is isomorphic to $E|_{\mathbb{D}^2\setminus 0}$. Therefore, if we remove the fiber $F = E|_0$ from E and identify $E|_{\mathbb{D}^2\setminus 0}$ with $(W/\mathbb{Z}_m)|_{\mathbb{D}^2\setminus 0}$, then we effectively replace F by the new multiple fiber in the center of $W/\mathbb{Z}_m$. The resulting elliptic fibration E_m thus described is clearly a complex surface.

8.4. Topological classification

Remember that we agreed that writing $E(n)_{p,q}$ implies that $p < q$ and p is coprime to q; further, multiplicity 1 is smoothly-irrelevant, and we have $E(n)_{1,q} = E(n)_q$ and $E(n)_1 = E(n)$. Also, we always assume that our $E(n)_{p,q}$'s contain a cusp fiber.

All these $E(n)_{p,q}$'s are sufficient to cover all the ground we need:

Theorem *(B. Moishezon). Any simply-connected elliptic surface that is minimal with respect to blow-downs is diffeomorphic to some $E(n)_{p,q}$.* □

Indeed, it is known from complex geometry that simply-connected minimal elliptic surfaces are classified, up to complex deformations, by their geometric genus $p_g = \frac{1}{2}(b_2^+ - 1)$ and by the multiplicities of their two multiple fibers. Since complex deformations induce diffeomorphisms, and since $p_g(E(n)_{p,q}) = n - 1$, it follows that the $E(n)_{p,q}$'s represent all smooth types of simply-connected elliptic surfaces.

Next, we divide these complex surfaces into types, first up to homeomorphisms, then up to diffeomorphisms.

The homeomorphic classification follows from computing their intersection forms and using Freedman's classification:

Homeomorphic Classification. *A simply-connected elliptic surface $E(n)_{p,q}$ is homeomorphic to some $E(m)_{r,s}$ if and only if*

— *$n = m$ and is odd;*

— *$n = m$ and is even, and further pq and rs have the same parity.*[7]

In particular, all the surfaces $E(\text{odd } n)_{p,q}$ are homeomorphic to $E(n)$, while the surfaces $E(\text{even } n)_{p,q}$ are homeomorphic either to $E(n)$ or to $E(n)_2$. □

The smooth classification of elliptic surfaces appears from combining the positive results coming from complex geometry (complex deformations), with the negative results yielded by differing gauge-theoretic invariants.[8] While the homeomorphic classification above might give hope that many of the $E(n)_{p,q}$'s, while distinct as complex manifolds, could be diffeomorphic to each other, they are in fact all smoothly distinct:

Diffeomorphic Classification. *An elliptic surface $E(n)_{p,q}$ with $n \geq 2$ is diffeomorphic to some $E(m)_{r,s}$ if and only if*

$$n = m \qquad \text{and} \qquad p = r \qquad \text{and} \qquad q = s\,.$$

In particular, $E(n)_p$ is diffeomorphic to $E(n)_r$ only when $p = r$, and coincides with $E(n)$ only when $p = 1$.

All the surfaces $E(1)_p$ are diffeomorphic to each other and to $E(1)$. Aside from multiplicity $p = 1$, a surface $E(1)_{p,q}$ is diffeomorphic to some $E(1)_{r,s}$ if and only if $p = r$ and $q = s$. □

Combining the above two classifications provides us with a startling series of examples:

---

[7] If pq is odd, then $E(\mathit{even})_{p,q}$ is spin; if pq is even, then it is not.

[8] See later, section 10.6 (page 413); compare with the gluing results from section 12.1 (page 532).

Corollary (*Exotic elliptic surfaces*).

- *For any odd* $n \geq 2$*, the* 4*–manifolds* $E(n)_{p,q}$ *are all homeomorphic but not diffeomorphic.*
- *For any even* $n \geq 2$*, the* 4*–manifolds* $E(n)_{p,q}$ *with* pq *even are all homeomorphic but not diffeomorphic; similarly for the manifolds* $E(n)_{p,q}$ *with* pq *odd.* □

Thus, each of the elliptic surfaces $E(n)_{p,q}$ is a topological 4–manifold that admits *infinitely* many distinct smooth structures. We didn't even travel too far into the wild: all of them are nice complex surfaces...

In particular, to obtain a simple example of a 4–manifold with two distinct smooth structures, take $\mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$, fibrate it, and choose two torus fibers in it. Then perform logarithmic transformations of multiplicities 2 and 3 on them: the resulting 4–manifold will be homeomorphic but not diffeomorphic to $\mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$. Other multiplicities yield other smooth structures.

Of course, each example of two homeomorphic manifolds that are non-diffeomorphic leads to a nontrivial h–cobordism, thus illustrating the failure of the smooth h–cobordism theorem in dimension 4.

8.5. Notes

Note: Decomposability

A somewhat older trend in the study of the topology of complex surfaces was the study of almost-complete decomposability. A 4–manifold M is called **almost-completely decomposable** if connect-summing with one copy of $\mathbb{CP}^2$ is enough to make $M \# \mathbb{CP}^2$ smoothly split ("dissolve") as $\# m\, \mathbb{CP}^2 \# n\, \overline{\mathbb{CP}}^2$.

A series of results of R. Mandelbaum and B. Moishezon have shown that many complex surfaces are indeed almost-completely decomposable. For example, all surfaces that appear as hypersurfaces in $\mathbb{CP}^3$ have this property, as proved in **R. Mandelbaum** and **B. Moishezon**'s ***On the topological structure of non-singular algebraic surfaces in*** $\mathbb{CP}^3$ **[MM76]**; and so do all simply-connected elliptic surfaces, as proved in **B. Moishezon**'s ***Complex surfaces and connected sums of complex projective planes*** **[Moi77b]**, as well as all complete intersections, see **R. Mandelbaum** and **B. Moishezon**'s ***On the topology of simply connected algebraic surfaces*** **[MM80]**. See also **R. Mandelbaum**'s expositions ***Decomposing analytic surfaces*** **[Man79]** and ***Four-dimensional topology: an introduction*** **[Man80]**. In particular,

$$E(n)_{p,q} \# \mathbb{CP}^2 \;\cong\; \# 2n\, \mathbb{CP}^2 \# (10n-1)\, \overline{\mathbb{CP}}^2 .$$

A similar result for non-simply-connected elliptic surfaces is obtained in **R. Mandelbaum**'s ***Lefschetz fibrations of Riemann surfaces and decompositions of complex elliptic surfaces*** **[Man85]**.

In a similar spirit, **R. Gompf**'s ***Sums of elliptic surfaces*** **[Gom91b]** proved that, given any two elliptic surfaces E' and E'', their connected sum $E' \# E''$ is diffeomorphic to either $\# m\, \mathbb{CP}^2 \# n\, \overline{\mathbb{CP}}^2$ or $K3 \# m\, S^2 \times S^2$. For splittings of manifolds $M' \# \overline{M}''$, with M' and M'' complex surfaces, see **R. Gompf**'s ***On sums of algebraic surfaces*** **[Gom88]**.

Of course, these results are to be contrasted with the above existence of infinitely many smooth structures on homeomorphic elliptic surfaces and to be compared with Wall's result about sum-stabilizations from section 4.2 (page 149). A conclusion is that the wildness of 4–manifolds is quite delicate: even a little more room is enough to make it all wither into triviality.

Note: Knots, complex singularities, sometimes spheres

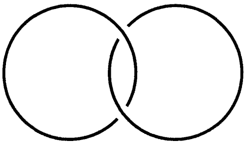
The fact that the singularity of a cusp fiber (inside an elliptic surface) is a cone on the trefoil knot is not an isolated case.

Let $P(z_0, z_1)$ be any two-variable complex polynomial whose zero-set

$$Z = \{(z_0, z_1) \in \mathbb{C}^2 \mid P(z_0, z_1) = 0\}$$

has an isolated singularity at $(0, 0)$. Then a neighborhood in Z of the singularity must be the cone on a link K in S^3. In fact, $K = Z \cap S^3$ (for small enough radius).

Therefore, every singularity of a complex curve inside a complex surface can be described topologically as the tip of a cone on a link. As a most simple example, the nodal singularity $z_0 z_1 = 0$ (two complex lines meeting at a point; equivalent to $z_0^2 + z_1^2 = 0$) is the cone over the simple Hopf link in figure[1] 8.16. The singularity described by $z_0^3 + z_1^2 = 0$ is the cusp singularity, a cone on the trefoil knot.



8.16. The Hopf link

Higher-dimensional cases. Singularities in higher dimensions exhibit higher-dimensional knotting and linking phenomena, and in certain cases the knotted manifolds are exotic spheres.

As a first example, the Poincaré homology sphere[2] Σ_P appears in $\mathbb{C}^3$ as $\{z_0^5 + z_1^3 + z_2^2 = 0\} \cap S^5$. Perturbing the equation to $z_0^5 + z_1^3 + z_2^2 = \varepsilon$ still draws Σ_P in S^5, but now the "interior" complex surface $\{z_0^5 + z_1^3 + z_2^2 = \varepsilon\} \cap \mathbb{D}^6$ is non-singular and diffeomorphic to the E_8-plumbing P_{E_8}.

Creatures obtained in this manner from simple equations of shape

$$z_0^{q_0} + \cdots + z_n^{q_n} = 0$$

in $\mathbb{C}^{n+1}$ are sometimes homology 3-spheres or exotic $(2n-1)$-spheres. In these cases, they are called **Brieskorn spheres** and are denoted by $\Sigma(q_0, \ldots, q_n)$.

Consider such a sphere-candidate

$$\Sigma^{2n-1} = \{z_0^{q_0} + \cdots + z_n^{q_n} = 0\} \cap S^{2n+1}$$

with all exponents $q_i \geq 2$. Its study starts by looking at the non-singular perturbation $V^{2n} = \{z_0^{q_0} + \cdots + z_n^{q_n} = \varepsilon\} \cap \mathbb{D}^{2n+2}$. This is a $2n$-manifold whose boundary is diffeomorphic to Σ. Having even dimension, V has an intersection form Q_V on $H_n(V; \mathbb{Z})$ (which is skew-symmetric when n is odd). Further, the interior of V is diffeomorphic to the entire affine hypersurface $\{z_0^{q_0} + \cdots + z_n^{q_n} = 1\}$ in $\mathbb{C}^{n+1}$, and thus is known to be $(n-1)$-connected. It follows that the boundary $\Sigma = \partial V$ is a homology/exotic sphere if and only if the form Q_V is unimodular.[3]

---

[1] Compare also with the standard desingularization of the double point back in section 3.1 (page 112), especially with figure 3.2 on page 113.

[2] Described back in section 2.3 (page 86).

[3] To see this, combine the $(n-1)$-connectedness of V with an adaptation of the argument from the end-notes of chapter 5 (homology spheres as boundaries, page 261), together with the generalized Poincaré conjecture when $\dim \Sigma \geq 5$.

One case when unimodularity is guaranteed is when two of the exponents q_i are coprime to all others. For example, in dimension 7 we have that

$$\Sigma(5,3,2,2,2)$$

coincides with the exotic 7–sphere Σ^7_P presented in the end-notes of chapter 2 (page 97). (Remember that Σ^7_P appeared as the boundary of the E_8–plumbing $P^8_{E_8}$, and that it generates the group $\Theta_7 = \mathbb{Z}_{28}$.)

Boundaries of plumbings. *Other Brieskorn manifolds also can be viewed as boundaries of plumbings. For example, the manifolds* $\Sigma(q,2,\ldots,2)$ *coincide with the boundaries of the* A_{q-1}*–plumbing* $P^{2n-1}_{A_{q-1}}$ *built from* $q-1$ *copies of* $\mathbb{D}T_{\mathbb{S}^n}$ *(again, see the notes on page 97). When the number* $n+1$ *of variables is even and* $q = 0 \pmod 8$*, then* $\Sigma(q,2,\ldots,2)$ *is diffeomorphic to a standard sphere* $\mathbb{S}^{4k+1}$*. On the other hand, when* $n+1$ *is even but* $q = 3$ *or* $5 \pmod 8$*, then* $\Sigma(q,2,\ldots,2)$ *is diffeomorphic to a Kervaire sphere* Σ^{4k+1}_K*, boundary of* $P^{4k+2}_{A_2}$*. The Kervaire spheres are sometimes trivial, but never when* k *is even. For example, while* $\Sigma(3,2,2,2,2,2)$ *is an exotic 9–sphere,* $\Sigma(3,2,2,2)$ *is the standard 5–sphere* $\mathbb{S}^5$*; nonetheless its embedding in* $\mathbb{S}^7 \subset \mathbb{C}^4$ *is knotted, just as* $\Sigma(3,2)$ *is a standard circle, but embedded in* $\mathbb{S}^3 \subset \mathbb{C}^2$ *as a trefoil knot. (On the other hand, when the number of 2's is even the Brieskorn manifold is not homeomorphic to a sphere.)*

The reference for the higher-dimensional material is **E. Brieskorn**'s ***Beispiele zur Differentialtopologie von Singularitäten*** **[Bri66]**, or **F. Hirzebruch** and **K. Mayer**'s monograph ***O(n)–Mannigfaltigkeiten, exotische Sphären und Singularitäten*** **[HM68]**.

For the general theory of hypersurface singularities, see **J. Milnor**'s classic ***Singular points of complex hypersurfaces*** **[Mil68]**. A more recent monograph is **A. Dimca**'s ***Singularities and topology of hypersurfaces*** **[Dim92]**.

Away from the complex realm, also take a look at **R. Fox** and **J. Milnor**'s ***Singularities of 2–spheres in 4–space and cobordism of knots*** **[FM66]**.

Note: Classification of singular fibers

Since one cannot but be curious about the types of singular fibers that can occur inside a (minimal) elliptic fibration, their complete classification, due to K. Kodaira, is gathered in table X on the following page.

Fiber I_0 is just a generic elliptic fiber, that is to say, a torus. Fiber I_1 is the fishtail fiber. Fiber I_k (with $k \geq 2$) is made from a "ring" of k spheres and is called the **necklace fiber.** Fiber *II* is the cusp fiber. Fiber *III* is made of two spheres with one common point where they are tangent. Fiber *IV* is made of three spheres meeting each other at a transverse triple point. The other fibers are obtained by connecting spheres at transverse double-points, as suggested in their diagrams. In particular, remember that we have already encountered fiber I_0^* as the singular fiber in the Kummer construction of $K3$ in section 3.3 (page 127).

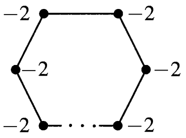
Multiplicities. *Assume that a singular fiber has irreducible components* S_k*. In the elliptic fibration we can imagine regular fibers* F *approaching our singular fiber, so that in homology we have* $[F] = \sum m_k [S_k]$ *for some appropriate (homological)* multiplicities m_k*. (For example, in the Kummer construction, we noted that the central sphere of* I_0^* *was covered twice by an approaching regular fiber; thus, it had multiplicity 2.) Since the regular fibers* F *are disjoint from our singular fiber, we must have* $[F] \cdot [S_k] = 0$ *for all* k*. In fact, these latter conditions (together with the intersections of the various* S_k*'s) determine completely the multiplicities* m_k*.*

X. Kodaira classification of fibers of minimal elliptic surfaces

I_0

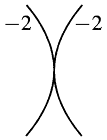
I_1

I_k

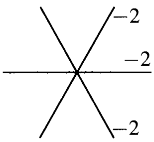


II

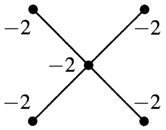
III



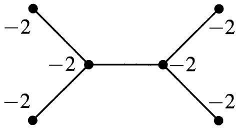
IV



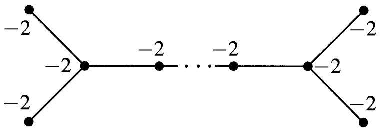
I_0^*



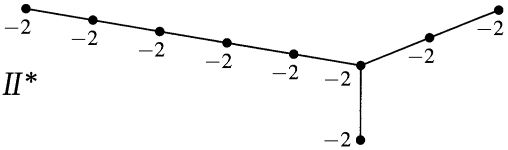
I_1^*



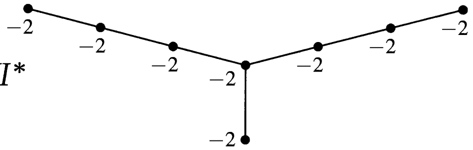
I_k^*



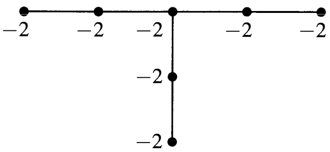
II^*



III^*



IV^*



Given the diagrams of some of the fibers in table X on the preceding page, notice that we can use them as a recipe for plumbing[4] disk bundles over spheres (with the specified Euler class) and obtain a copy of a neighborhood of the corresponding singular fiber inside its elliptic fibration.

Monodromies. Around each fiber there is a unique *monodromy.* That is, if the fiber F is projected by $E \to \mathbb{CP}^1$ to a point x, then we look at a small circle C around x in $\mathbb{CP}^1$: the restricted fibration $E|_C$ can be thought of as being built from $\mathbb{T}^2 \times [0,1]$ with the end-tori $\mathbb{T}^2 \times 0$ and $\mathbb{T}^2 \times 1$ identified by a twist. This twist is called the **monodromy** of E around the fiber F.

More precisely, one should be aware of orientations and consider the twist only up to isotopy, that is to say, as an element of the mapping class group of $\mathbb{T}^2$ (*i.e.*, the group of orientation-preserving self-diffeomorphisms of $\mathbb{T}^2$ up to isotopy). The mapping class group of the torus is simply $SL(2,\mathbb{Z})$—think action on $H_1(\mathbb{T}^2;\mathbb{Z})$. Thus the monodromies can be pictured as integral unimodular 2×2 matrices.

All the monodromies that can appear around fibers in a minimal elliptic surface are exhibited in table XI. Notice that the seemingly random nomenclature from table X on the preceding page seems to find some justification when inspecting the corresponding monodromies.

XI. Monodromies of singular fibers

$I_0 : \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$	$I_0^* : \begin{bmatrix} -1 & \\ & -1 \end{bmatrix}$
$I_1 : \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}$	$I_1^* : \begin{bmatrix} -1 & -1 \\ & -1 \end{bmatrix}$
$I_k : \begin{bmatrix} 1 & k \\ & 1 \end{bmatrix}$	$I_k^* : \begin{bmatrix} -1 & -k \\ & -1 \end{bmatrix}$
$II : \begin{bmatrix} 1 & 1 \\ -1 & \end{bmatrix}$	$II^* : \begin{bmatrix} & -1 \\ 1 & 1 \end{bmatrix}$
$III : \begin{bmatrix} & 1 \\ -1 & \end{bmatrix}$	$III^* : \begin{bmatrix} & -1 \\ 1 & \end{bmatrix}$
$IV : \begin{bmatrix} & 1 \\ -1 & -1 \end{bmatrix}$	$IV^* : \begin{bmatrix} -1 & -1 \\ 1 & \end{bmatrix}$

---

[4] Plumbing was described in section 2.3 (page 86).

See ***Compact complex surfaces*** [BPVdV84, BHPVdV04] if you crave more. For a topological discussion of these fibers, see **R. Kirby** and **P. Melvin**'s ***The*** E_8***–manifold, singular fibers and handlebody decompositions*** [KM99].

Bibliography

As always, the complete reference for the complex-geometric side is **W. Barth**, **C. Peters** and **A. Van de Ven**'s ***Compact complex surfaces*** [BPVdV84], or the second edition (with **K. Hulek**) [BHPVdV04].

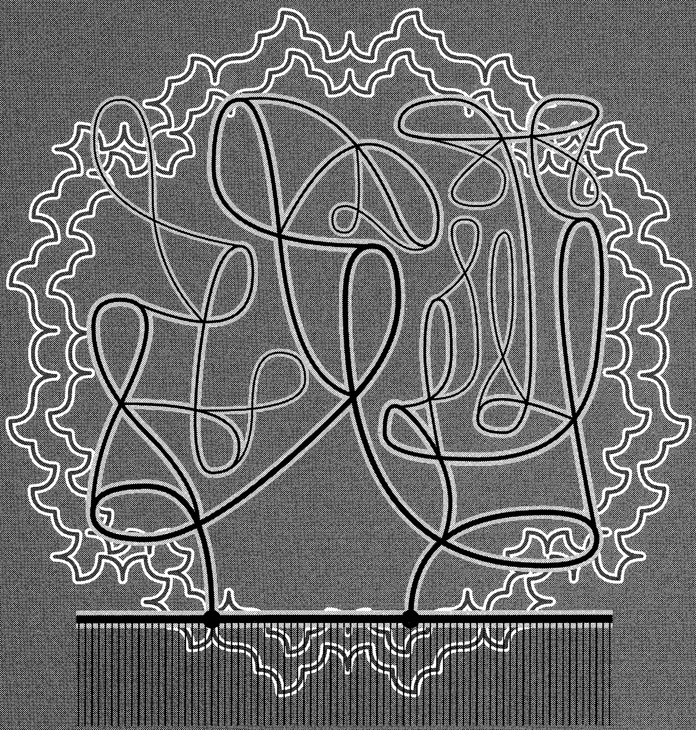
That all simply-connected elliptic surfaces are diffeomorphic to some $E(n)_{p,q}$ was shown in **B. Moishezon**'s ***Complex surfaces and connected sums of complex projective planes*** [Moi77b].

Many smooth classification results were obtained using Donaldson gauge theory. **S.K. Donaldson**'s ***Irrationality and the h–cobordism conjecture*** [Don87] proved that the manifold $E(1)$ admits two distinct smooth structures, thus providing the first example of a nontrivial h–cobordism. Later infinitely-many distinct smooth structures on $E(1)$ were uncovered independently in **R. Friedman** and **J. Morgan**'s ***On the diffeomorphism types of certain algebraic surfaces. I & II*** [FM88], and in **C. Okonek** and **A. Van de Ven**'s ***Stable bundles and differentiable structures on certain elliptic surfaces*** [OVdV86].

The fact that all $E(n)_{p,q}$'s are smoothly distinct was first proved by using Donaldson theory. It was done in **J. Morgan** and **T. Mrowka**'s ***On the diffeomorphism classification of regular elliptic surfaces*** [MM93], in **A. Stipsicz** and **Z. Szabó**'s ***The smooth classification of elliptic surfaces with*** $b^+ > 1$ [SS94], in **P. Lisca**'s ***On the Donaldson polynomials of elliptic surfaces*** [Lis94], and finished for $E(1)_{p,q}$'s by **R. Friedman**'s ***Vector bundles and*** $SO(3)$***–invariants for elliptic surfaces*** [Fri95]. Nowadays, one would use the Seiberg–Witten invariants to more quickly obtain the same classification. Most results of Donaldson theory on elliptic surfaces are presented in **R. Friedman** and **J. Morgan**'s book ***Smooth four-manifolds and complex surfaces*** [FM94a]. See also the references at the end of the previous chapter, page 299.

Using Seiberg–Witten theory, all these classification results are much easier to obtain. See **R. Brussee**'s ***The canonical class and the*** C^∞ ***properties of Kähler surfaces*** [Bru96] and **R. Friedman** and **J. Morgan**'s ***Obstruction bundles, semiregularity, and Seiberg–Witten invariants*** [FM99]. See also the references on page 475, at the end of chapter 10.

For a topological description of elliptic surfaces in terms of handle decompositions, see **J. Harer**, **A. Kas** and **R. Kirby**'s ***Handlebody decompositions of complex surfaces*** [HKK86], as well as **R. Gompf** and **A. Stpisicz**'s ***4–Manifolds and Kirby calculus*** [GS99, ch 8]. For more descriptions of complex surfaces diffeomorphic to $E(n)$, see section 7.3 of the latter work.



Part IV

Gauge Theory on 4–Manifolds

NOW has come the time to look at gauge theory, an approach to smooth topology that is founded on differential geometry. The more it was followed, the more it ruined any optimism in a near-future classification of 4–manifolds, and displayed in full view our current ignorance of smooth 4–manifolds.

One can safely say that, among all dimensions, smooth dimension 4 is currently the least understood. Historically, manifolds of dimensions 5 and higher became well-understood between 1950–1970, in dimension 3 great progress was made between 1950–1980 and steadily ever since (in 2003 even the Poincaré conjecture might have been proved), while topological 4–manifolds were essentially killed in 1981 by Freedman's classification. However, for smooth 4–manifolds, starting after 1982 and getting worse after 1994, we just seem to get dizzier in a whirlpool of amazing creatures. The lens that made the whole bestiary visible is gauge theory.

For this final part of the book, we start in chapter 9 by preparing the ground: we explain the general strategy of gauge theory, then we review a bit of differential geometry, and finally we take a glance at Donaldson's invariants.

In chapter 10 (starting on page 375), we move on to the Seiberg–Witten invariants and see how they apply to the smooth topology of 4–manifolds. These invariants are best understood in the special case of complex and symplectic manifolds. Furthermore, the farther away one moves from the complex realm, the less powerful gauge theory becomes.

Chapter 11 is devoted to the problem of determining the smallest genus needed for representing a homology class by smooth surfaces. Here, the celebrated Seiberg–Witten adjunction inequality has allowed great progress to be made. For contrast, we start the chapter with a glance at the pre-gauge approach stemming from Rokhlin's theorem.

Finally, in chapter 12 (starting on page 531) we present the disturbing surgery construction of R. Fintushel and R. Stern, which empowers one to easily produce infinite series of homeomorphic but non-diffeomorphic 4–manifolds.

Contents of Part IV

Chapter 9

Prelude, and the Donaldson Invariants

THE chapter starts with an outline of the general approach of gauge theory to smooth topology, followed by a review of a few basic differential-geometric notions. Finally, the setting of Donaldson theory is sketched, but only to leave it aside in favor of Seiberg–Witten theory, whose exposition will start in the next chapter and dominate the rest of the book.

We start by drafting the general strategy of gauge theory. Then, in section 9.2 (page 333) we present a smattering of differential geometry, quickly explaining the concepts of connection and curvature, in a perspective centered on parallel transports. In section 9.3 (page 350) we explain how dimension 4, from the point-of-view of a differential geometer, is special: 2–forms split into self-dual and anti-self-dual 2–forms. Finally, section 9.4 (page 353) takes the quickest of glances at the Donaldson invariants, which are based on anti-self-dual connections on bundles.

Besides a discussion of the equivalence between the Donaldson and Seiberg–Witten theories (page 370), the end-notes contain discussions of anti-self-dual connections on line bundles (page 357) and of the relation of anti-self-duality to holomorphic bundles (page 365), both of which will be important technically in the later study of Seiberg–Witten theory (the first will be used for dealing with reducible Seiberg–Witten solutions, the second for understanding the Seiberg–Witten equations on Kähler and symplectic manifolds.)

9.1. Prelude

Terminology. In physics, a **local gauge** is a local trivialization of a vector bundle. A *(local)* **change of gauge** is a change of local trivialization. The **group of** *(global)* **changes of gauge**, or in short the **gauge group**, is the group of symmetries (automorphisms) of the bundle. Two creatures are called **gauge equivalent** if they belong to the same orbit of the gauge group. **Gauge theory** is the study of gauge-independent creatures on vector bundles, such as connections, curvatures, sections, considered up to the action of the gauge group.

General strategy. Often, solutions of equations contain topological information. Consider the following two elementary examples:

(A) Let $f\colon X^m \to Y^n$ be a map between compact oriented manifolds. Then, for a generic $y \in N$, the space of solutions to the equation $f(x) = y$ is a compact submanifold of X, and it determines a class in $H_*(X;\mathbb{Z})$; this class is the pull-back $f^*[Y]$ and depends only on the homotopy class of f.

(B) Let $E \to X$ be an oriented vector bundle over a compact oriented manifold X. Then, for a generic section $f\colon X \to E$, the space of solutions to the equation $f(x) = 0$ is a compact submanifold of X, and it determines a class in $H_*(X;\mathbb{Z})$; this class is the Poincaré-dual of the Euler class $e(E)$, and depends only on the bundle E.

We wish to apply a similar approach to a more sophisticated setting, involving vector bundles on M and certain partial differential equations governing creatures on these bundles (connections, curvatures, sections). We will consider the solutions of the PDEs only up to automorphisms of the bundles, *i.e.*, only up to the action of the gauge group. The gauge classes of solutions will make up what we will call the **moduli space** $\mathfrak{M}$. We will wish to extract invariants of the base manifold M from this moduli space $\mathfrak{M}$, and a good candidate would be the homology class that $\mathfrak{M}$ might determine inside its ambient space.

To apply such a program, one typically needs to check several conditions:

- The situation is "like" a finite-dimensional one (*i.e.*, it involves elliptic operators and it has Fredholm linearizations; in particular, these conditions raise the chances that $\mathfrak{M}$ be a manifold, not to mention that the Atiyah–Singer index theorem can be applied).
- We have compactness for the moduli space $\mathfrak{M}$, or a reasonable compactification.
- We are able to orient $\mathfrak{M}$ (else we only get modulo 2 invariants).

— We manage to avoid **reducible solutions** (these are solutions with nontrivial stabilizer[1] under the action of the gauge group, which create singularities in the quotient space $\mathfrak{M}$).

If these conditions are met, then the oriented solution space $\mathfrak{M}$ determines a homology class $\mathfrak{m}$ of some larger configuration space $\mathfrak{C}$. If we manage to avoid reducible solutions, then when varying various choices made along the way (such as Riemannian metrics) we only change $\mathfrak{M}$ by a cobordism, and thus the class $\mathfrak{m}$ is unchanged: it is an invariant of the base manifold M. Then, by evaluating various natural cohomology classes of $\mathfrak{C}$ on $\mathfrak{m}$, we obtain numerical invariants of the smooth manifold M.

Of course, in certain instances information about M is obtained from a direct study of the space $\mathfrak{M}$ itself.[2] In those cases the reducible solutions are often the carriers of information.

Shortcut, anyone? Note that the rest of this chapter and the beginning of the next can be skipped at a first reading. By trusting that the strategy above can be used to define an invariant of 4–manifolds, called the Seiberg–Witten invariant, the reader can jump from here and start reading again with the main results governing the behavior of these Seiberg–Witten invariants, in section 10.4 (page 404). See you later.

9.2. Bundles, connections, curvatures

In this section we review a few concepts from differential geometry that will be used for defining both the Donaldson invariants (in section 9.4) and the Seiberg–Witten invariants (in the next chapter).

We present the concepts of cocycle defining a vector bundle,[3] of connection on a bundle, and of its curvature. Peculiarities of our exposition might be that we use cocycles as replacements for principal bundles and that we base the intuition for connections and curvatures on parallel transports. We start with reviewing a few classical Lie groups, keeping it all to a minimum:

Lie groups

Remember that a **Lie group** G is a smooth manifold organized as a group. The tangent space at the identity element $1 \in G$ is denoted by $\mathfrak{g}$ and called the **Lie algebra** of G.

---

[1] When a group G acts on a set S, the **stabilizer** of $x \in S$ is the subgroup of G made of those $g \in G$ that keep x fixed.

[2] For example, Donaldson's theorem and Furuta's $^{10}/_{8}$–theorem, see back section 5.3 (page 243); see also the Seiberg–Witten proof of Donaldson's theorem, in the end-notes of the next chapter (page 454).

[3] This was already encountered back in the end-notes of chapter 4 (definition of spin structures, page 174), but not in the main text; thus, we must review.

Real groups. The **general linear group**

$$GL(n)$$

is the group of automorphisms of the real vector space $\mathbb{R}^n$, that is to say, all $n \times n$ invertible matrices. Its Lie algebra

$$\mathfrak{gl}(n) = \operatorname{End}(\mathbb{R}^n)$$

is the algebra of endomorphisms of $\mathbb{R}^n$, *i.e.*, of all $n \times n$ matrices. The **special orthogonal group**

$$SO(n)$$

is the subgroup of $GL(n)$ consisting of all orientation-preserving Euclidean isometries of $\mathbb{R}^n$, *i.e.*, the rotations of $\mathbb{R}^n$. Its Lie algebra

$$\mathfrak{so}(n)$$

is made of all skew-symmetric endomorphisms of $\mathbb{R}^n$. It can also be identified with $\Lambda^2(\mathbb{R}^n)$ (thinking of $v \wedge w \in \Lambda^2$ both as defining the oriented plane $\mathbb{R}\{v, w\}$ and as being tangent to the path of rotations that rotates v toward w with unit angular speed).

Complex groups. On the complex side of the world, the **complex general linear group**

$$GL_{\mathbb{C}}(m)$$

is the group of $\mathbb{C}$–linear automorphisms of $\mathbb{C}^m$, *i.e.*, all $m \times m$ complex invertible matrices. The **unitary group**

$$U(m)$$

is the group of $\mathbb{C}$–linear automorphisms of $\mathbb{C}^m$ that preserve the standard Hermitian inner product[4] on $\mathbb{C}^m$. The complex determinant $\det_{\mathbb{C}} \rho$ of any $\rho \in U(m)$ belongs to the unit-circle $\mathbb{S}^1 \subset \mathbb{C}$. The **special unitary group**

$$SU(m)$$

is made of all elements of $U(n)$ of determinant $+1$. We will identify the complex space $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ in the standard way, and thus have inclusions

$$GL_{\mathbb{C}}(m) \subset GL(2m) \qquad \text{and} \qquad U(m) \subset SO(2m)\ .$$

Finally, two simple low-dimensional examples:

$$SO(2) = U(1) = \mathbb{S}^1 \subset \mathbb{C} \qquad \text{and} \qquad SU(2) = \mathbb{S}^3 \subset \mathbb{H}\,,$$

the latter acting by quaternion multiplication on the *right*.[5]

---

[4] The Hermitian product on $\mathbb{C}^m$ is defined by $\langle v, w\rangle_{\mathbb{C}} = \sum v_k \cdot \overline{w}_k$. It is $\mathbb{C}$–linear in the first argument and $\mathbb{C}$–anti-linear in the second.

[5] We identify $\mathbb{C}^2 \equiv \mathbb{H}$ by $(z_1, z_2) \equiv z_1 + z_2 j$, and thus complex-scalars quaternion-multiply on the left, while $SU(2)$ is modeled by unit-quaternions multiplying on the right. For more on Lie groups and quaternions, see the end-notes of the next chapter (quaternions and spinors, page 432).

Spin groups. Since for all $n \geq 3$ we have[6] $\pi_1 SO(n) = \mathbb{Z}_2$, this means that the double cover of $SO(n)$ will be simply-connected. The cover can be organized as a Lie group and is known as the **spin group**

$$Spin(n) .$$

A complexified version of this is the **complex spin group**

$$Spin^{\mathbb{C}}(n) ,$$

defined as $Spin^{\mathbb{C}}(n) = U(1) \times Spin(n)/\pm 1$. The complex spin group admits a copy of $U(n)$ as a subgroup. The complex spin group will be revisited in the next chapter.

Vector bundles

A **vector bundle** E of rank k over X^m (also called a **k–plane bundle** over X) is an open $(m+k)$–manifold E together with a map $p\colon E \to X$ such that the fibers $p^{-1}[x]$ are vector spaces isomorphic to $\mathbb{R}^k$, and p locally looks like projections $U \times \mathbb{R}^k \to U$. In other words, there is an open covering $\{U_\alpha\}$ of X and an atlas of maps

$$\left\{\varphi_\alpha\colon p^{-1}\left[U_\alpha\right] \cong U_\alpha \times \mathbb{R}^k\right\} ,$$

having $\mathrm{pr}_1 \circ \varphi_\alpha = p$, and so that every overlap $\varphi_\alpha \circ \varphi_\beta^{-1}$ is a map $(x, w) \longmapsto \left(x,\, g_{\alpha\beta}(x) \cdot w\right)$ for some change-of-coordinates function

$$g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow GL(k) ,$$

thus ensuring that the $\mathbb{R}^k$–factors are identified linearly.

Cocycles. The maps $g_{\alpha\beta}$ are in fact all that is needed to describe E: One can just glue E up from trivial patches $U_\alpha \times \mathbb{R}^k$, by identifying (x, w_α) from $U_\alpha \times \mathbb{R}^k$ with (x, w_β) from $U_\beta \times \mathbb{R}^k$ if and only if $w_\alpha = g_{\alpha\beta}(x) \cdot w_\beta$.

For an open covering $\{U_\alpha\}$ of X together with a random collection of maps

$$\left\{g_{\alpha\beta}\colon U_\alpha \cap U_\beta \longrightarrow GL(n)\right\}$$

to actually define a k–plane bundle on X, certain simple compatibility relations need to be satisfied. These are:

$$g_{\alpha\alpha}(x) = id , \qquad g_{\beta\alpha}(x) = g_{\alpha\beta}(x)^{-1} , \qquad g_{\alpha\delta}(x) = g_{\alpha\beta}(x) \cdot g_{\beta\delta}(x) ,$$

All three can be contracted in a single condition

$$g_{\alpha\beta}(x) \cdot g_{\beta\delta}(x) \cdot g_{\delta\alpha}(x) = id .$$

The latter is called the **cocycle condition.** Any collection $\left\{U_\alpha,\, g_{\alpha\beta}\right\}$ satisfying it will be called a **cocycle.** A more thorough discussion of this concept

---

[6] That $\pi_1 SO(n) = \mathbb{Z}_2$ was pictured in the end-notes of chapter 4 (spin structure definition, page 178).

was made in the two end-notes at the end of chapter 4, on page 174 (while defining spin structures) and on page 189 (*via* Čech cohomology).

G–bundles. For vector bundles E defined as above, $GL(k)$ is called the **structure group** of E, and we sometimes call E a $GL(k)$–bundle. If we manage to redefine E by using a cocycle $\{g_{\alpha\beta}\}$ whose values are all contained in a subgroup G of $GL(k)$, then we say that we **reduced the structure group** of E to G, or that E is now a **G–bundle.**

Orientations. A typical example is to define a bundle E by using a cocycle whose identifications $g_{\alpha\beta}(x)$ all have positive determinant. This is known as an **orientation** of E.

Metrics. Another example comes from endowing the fibers of E with a smoothly-varying inner product, called a **fiber-metric**; in this case, the structure group of E can immediately be reduced to $O(k)$. If further E can be oriented, then the group can further be reduced to $SO(k)$, and E becomes an $SO(k)$–bundle. In particular, as soon as we endow an oriented manifold M^4 with a Riemannian metric, its tangent bundle becomes an $SO(4)$–bundle, described by an $SO(4)$–valued cocycle.

Complex structures. If the vector bundle E can be endowed with a linear anti-involution $J\colon E \to E$, covering the identity of the base X and such that $J \circ J = -id$, then we think of this J as a proxy for multiplying with the imaginary-unit $i \in \mathbb{C}$. In effect, the fibers of E become complex vector spaces, with $i \cdot v = J(v)$. This reduces the group of E to $GL_{\mathbb{C}}$; such a bundle E is called a **complex bundle.**

If further E is endowed with a J–invariant fiber-metric (called a **Hermitian metric**), then E becomes an $U(m)$–bundle. In particular, if M^4 is endowed with an almost-complex structure $J\colon T_M \to T_M$, then, after picking a compatible Riemannian metric, its tangent bundle T_M can be described by using an $U(2)$–cocycle.

Parallel transport

Over any contractible piece D of X^m, the restriction $E|_D$ must be trivial, *i.e.*, $E|_D \approx D \times \mathbb{R}^n$. In particular, over any embedded curve $c\colon [0,1] \to X$, the bundle $E|_c$ is trivial. Given such a curve c, we could then choose a favorite trivialization $E|_c \approx [0,1] \times \mathbb{R}^n$. This, in particular, means choosing a distinguished isomorphism $E|_{c(0)} \approx E|_{c(1)}$ between the fibers at the ends of c.

Pushing things toward globalization, imagine that we choose trivializations of E over each curve[7] in X, make these choices in a coherent fashion, and then wonder whether we can patch these thin trivializations into a trivialization of the whole bundle E over X.

Parallel transports. Cleaning up this idea a bit, imagine that we somehow choose, for every curve $c\colon [0,1] \to X$, a favorite isomorphism

$$\tau_c\colon E|_{c(0)} \xrightarrow{\approx} E|_{c(1)} .$$

A minimal condition of coherence would be that, if $c'\colon [0,1] \to X$ and $c''\colon [1,2] \to X$ are two paths that follow each other (*i.e.*, $c'(1) = c''(1)$), then

$$\tau_{c' * c''} = \tau_{c''} \circ \tau_{c'} ,$$

where $c' * c''$ is the obvious join $[0,1] \cup [1,2] \to X$ of the two paths.

We call such a choice τ a **parallel transport** on E: once an element $e \in E|_x$ and a path c from x to y are chosen, we can now transport e to a well-determined element $\tau_c(e)$ in $E|_y$. In general, of course, the result depends heavily on the chosen path c. See figure 9.1 on the following page.

For any two paths c' and c'' from x to y, the difference $\tau_{c'}(e) - \tau_{c''}(e)$ will measure the failure of the parallel transport from corresponding to a global trivialization of $E \to X$. Thus, properly studied, the parallel transport can uncover the intrinsic nontriviality[8] of the bundle E.

Case of G–bundles. If E happens to be structured as a G–bundle, then we naturally wish to restrict to parallel transports that act by isomorphisms from G. For example, if our bundle is oriented and endowed with a fiber-metric (thus being an $SO(n)$–bundle), then we wish the parallel transport to preserve that structure, that is to say, act by isometries

$$\langle \tau_c(e'),\, \tau_c(e'') \rangle = \langle e', e'' \rangle ,$$

and preserve orientations. If our bundle E has a complex structure, then the parallel transport should be $\mathbb{C}$–linear.

Parallel sections. Given a parallel transport τ on $E \to X$, for every path c starting at x and every $e \in E|_x$, there is a *unique* section σ of E that is defined on c, starts with[9] $\sigma(0) = e$, and is drawn by the parallel transport of e over c. In other words, σ is defined by

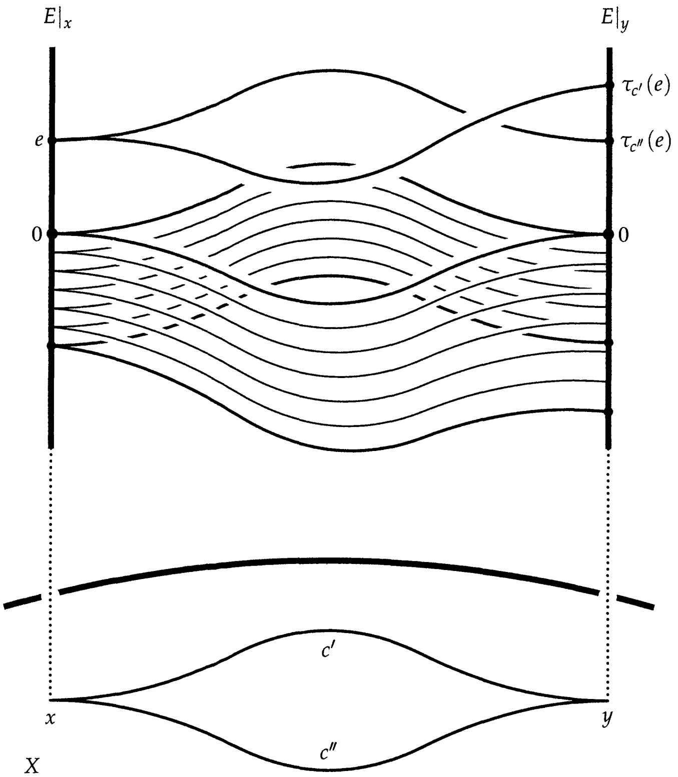
$$\sigma(t) = \tau_{c|_{[0,t]}}(e) .$$

---

[7] The curve can be embedded or not. When c is not embedded, we trivialize the pull-back $c^*E \to [0,1]$. In particular, for each loop at $x \in X$, we choose an automorphism of the fiber $E|_x$.

[8] This, of course, is just an obscure and vague way of referring to the Chern–Weil method of obtaining characteristic classes of E from the curvature of a random connection on E.

[9] For simplicity, we denote by $\sigma(t)$ the value of σ at the point $c(t)$ of the path c.



9.1. Parallel transport, depending on path

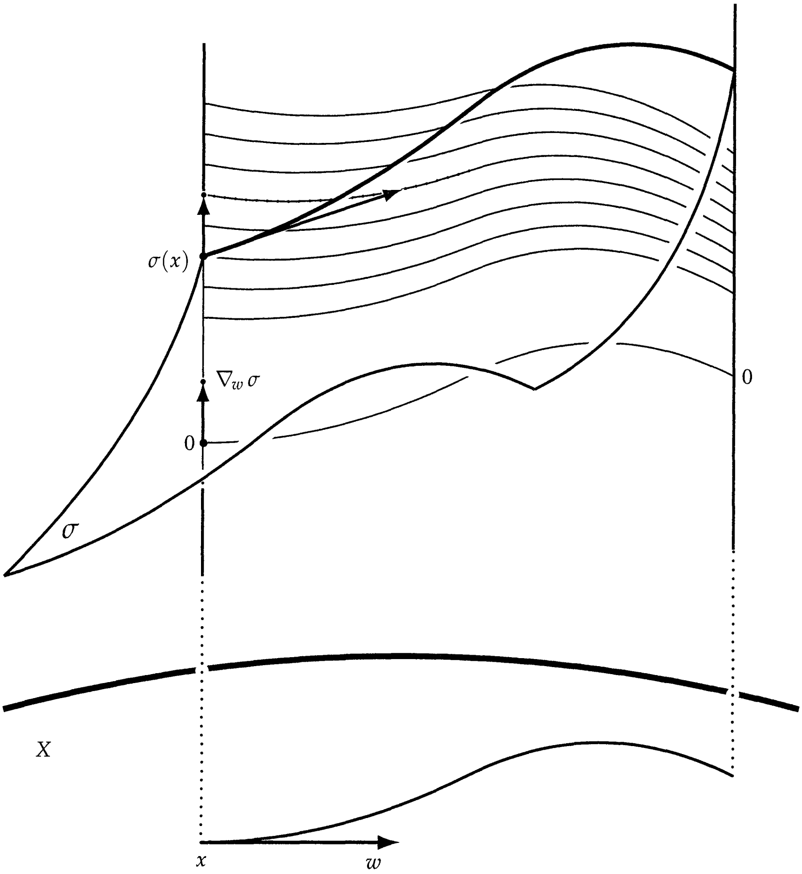
Such a section σ is called **parallel** over c, or the **parallel lift** of c at e.

Connections

Parallel transports are certainly not the most comfortable objects to play with. We wish to move toward infinitesimal versions, which will behave like derivatives and allow us to do calculus on bundles.

Covariant derivatives from parallel transport. Given a parallel transport τ, we wish to measure how much does a random section $\sigma\colon X \to E$ fail from being τ–parallel at $x \in X$, in the direction of the vector $w \in T_X|_x$. For that, we choose a random path $c\colon [0,1] \to X$ representing w, that is to say, a curve c so that $x = c(0)$ and $w = \frac{d}{dt}\big|_{t=0} c(t)$. Then we can measure the infinitesimal deviation of σ from being parallel by using the quantity

$$(\nabla_w \sigma)(x) = \frac{d}{dt}\Big|_{t=0} \tau^{-1}_{c|_{[0,t]}}\big(\sigma(t)\big) \quad \in E|_x \,.$$



9.2. Covariant derivative, from parallel transport

This is called the **covariant derivative** of σ in the direction w. It is simply the derivative of σ with respect to τ, in the w–direction. See figure 9.2. In particular, a section σ is *parallel* over a path c if and only if $\nabla_{\frac{d}{dt}c}\,\sigma = 0$ everywhere on c.

If $V \in \Gamma(T_X)$ is a vector field on X and $\sigma \in \Gamma(E)$ is a section of E, then the covariant derivative defines a new section

$$\nabla_V \sigma \in \Gamma(E) .$$

The covariant derivative can therefore be viewed as a map

$$\nabla \colon \Gamma(T_X) \times \Gamma(E) \longrightarrow \Gamma(E) .$$

Further, if we have the covariant derivative ∇ given to us, then the parallel transport can be recovered immediately, by integration. Therefore, from now on we will think of the covariant derivative as the primary creature, and start all over:

Covariant derivatives. We call **covariant derivative** any map

$$\nabla \colon \Gamma(T_X) \times \Gamma(E) \longrightarrow \Gamma(E)$$

satisfying the following properties:

- ∇ is $\mathbb{R}$–bilinear;[10]
- $\nabla_{fV}\sigma = f \cdot \nabla_V \sigma$ for all functions $f \colon X \to \mathbb{R}$;
- $\nabla_V \sigma = df(V) \cdot \sigma + f \cdot \nabla_V \sigma$ for all functions $f \colon X \to \mathbb{R}$.

In fact, there is a completely equivalent concept that manages to express all this in an even more succinct language:

Connections. A covariant derivative ∇ is equivalent to a map

$$d_\nabla \colon \Gamma(E) \longrightarrow \Gamma(E \otimes T_X^*)$$

satisfying the Leibnitz property

$$d_\nabla(f \cdot \sigma) = f \cdot d_\nabla \sigma + \sigma \otimes df$$

for all sections $\sigma \in \Gamma(E)$ and functions $f \colon X \to \mathbb{R}$; here $df \in \Gamma(T_X^*)$ is the usual exterior derivative of f, viewed as a 1–form. Setting

$$\nabla_V \sigma = \big(d_\nabla \sigma\big)(V)$$

immediately makes apparent the equivalence of d_∇ with ∇.

The operator d_∇ is called a **connection**[11] on the bundle E, since it can be used to "connect" the fibers of E through the associated parallel transport: after fixing a curve from x to y, one can essentially identify $E|_x$ with $E|_y$.

Often enough in this volume, we will get overenthusiastic and also call "connection" the covariant derivative ∇.[12] Be forewarned that the notation itself will sooner-or-later become confused, with ∇ and d_∇ used interchangeably.

Connections and plane fields. Of course, we can view the total space of the bundle E as a manifold. Then we should also glance at its tangent bundle T_E. Since E is split into vector-space fibers, we immediately notice a special field of n–planes in T_E, the field of **vertical planes**, the n–planes tangent to the fibers.

---

[10] The property that $\nabla_{V+W}\sigma = \nabla_V\sigma + \nabla_W\sigma$ does not follow from the properties of the parallel transport as defined here. Nevertheless, we postulate it for covariant derivatives.

[11] Physicists, for reasons of their own, like to call d_∇ a "gauge potential".

[12] As it happens, the terminology is unsettled, and often in the literature "connection" and "covariant derivative" are used as synonyms.

Since each fiber $E|_x$ is a vector space (and the tangent space $T_V|_v$ to any vector space V is canonically identified with the vector space V itself[13]), we can identify the vertical plane field with the pulled-back bundle p^*E. (In other words, for each $e \in E$ we can identify $\big(T_{(E|_{p(e)})}\big)|_e$ with $E|_{p(e)}$.) Thus, we have a *canonical* embedding

$$p^*E \subset T_E$$

that sets p^*E as the subbundle of T_E drawn by the vertical plane field.

Horizontal ambiguities. The vertical plane field p^*E does *not* have a canonical complement in T_E. Of course, any such complement would be isomorphic to p^*T_X, but there is no canonical embedding of p^*T_X in T_E, except along X itself:

$$T_E|_X = T_X \oplus E\,,$$

where X is understood as the submanifold of E drawn by the zero-section.

From connections to horizontal fields. On the other hand, if we choose a connection d_∇ on E, then d_∇ induces a unique global splitting

$$T_E = p^*E \oplus p^*T_X\,.$$

Indeed, for any $e \in E|_x$ and $w \in T_X|_x$, we can build a path $c\colon [0,1] \to X$ with $c(0) = x$ and tangent to w. We then lift c to a unique section $\sigma\colon [0,1] \to E$ that starts at $\sigma(0) = e$, covers c and is parallel for d_∇; in particular it has $\nabla_w \sigma = 0$. Seeing σ as a curve in the manifold E, we notice that its tangent vector $\widetilde{w} = \frac{d}{dt}\big|_{t=0} \sigma(t)$ at e is independent of choices and depends only on w. Further, if we vary $w \in T_X|_x$, then the map $w \mapsto \widetilde{w}$ defines a linear embedding $T_X|_x \subset T_E|_e$. Its image is complementary to the vertical plane $p^*E|_e$ at e. Globally, we obtain a bundle embedding

$$p^*T_X \subset T_E$$

whose image is complementary to p^*E, and thus exhibits a splitting $T_E = p^*E \oplus p^*T_X$.

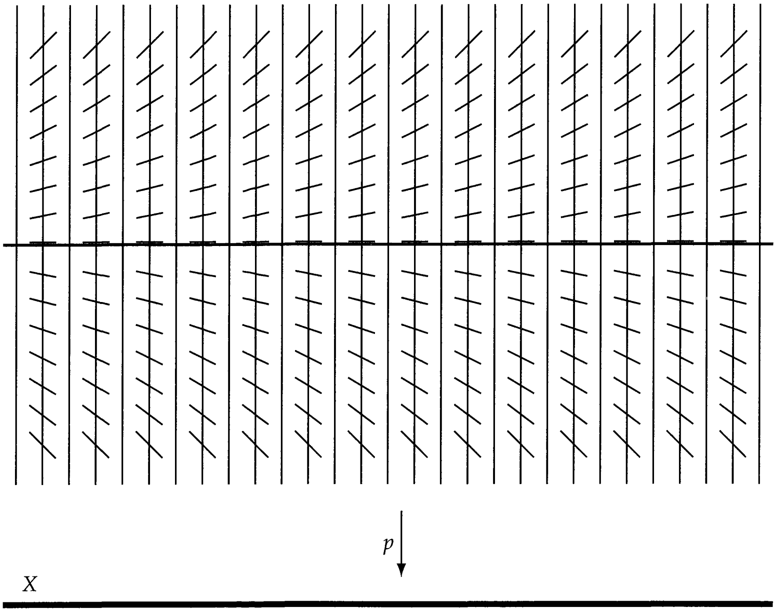
From horizontal fields to connections. Conversely, choose a *horizontal* plane field $\mathcal{H}$ in T_E, meaning a plane field $\mathcal{H}$ that is complementary to the vertical plane field p^*E. Require that $\mathcal{H}$ be invariant under scalar multiplications, as in figure 9.3 on the following page. Then it turns out that in effect you have chosen a connection d_∇ on E.

This can be seen as follows: The choice of $\mathcal{H}$ induces a splitting of T_E as

$$T_E = p^*E \oplus \mathcal{H}\,,$$

---

[13] For each $v \in V$, we have $V \approx T_V|_v$ by sending $w \in V$ to the tangent vector at v of the curve $t \mapsto v + t \cdot w$.



9.3. Connection, as horizontal plane field

and in particular determines a well-defined projection along $\mathcal{H}$, from T_E onto the vertical plane-field p^*E. Since each fiber $p^*E|_e$ can be identified with the fiber $E|_{p(x)}$, this vertical projection $T_E \rightarrow p^*E$ induces a bundle map pr: $T_E \rightarrow E$, fitting at the top of the diagram

$$\begin{tikzcd}
p^*E \arrow[d] \arrow[r, phantom, "\subset"] & T_E \arrow[r, "\mathrm{pr}"'] \arrow[d] & E \arrow[d, "p"] \\
E \arrow[r, equal] & E \arrow[r, "p"] & X .
\end{tikzcd}$$

This map pr is fiberwise-surjective and restricts to p^*E as a fiberwise-isomorphism, thus displaying p^*E anew as a pull-back of E.

Furthermore, the projection pr determines a connection on E: for every section $\sigma\colon X \rightarrow E$, its covariant derivative $\nabla_V\sigma$ is evaluated by first using the differential $d\sigma\colon T_X \rightarrow T_E$ to get a tangent vector $d\sigma(V)$ to E, then projecting the latter through pr to E:

$$\nabla_V\sigma = \mathrm{pr}\big(d\sigma(V)\big) .$$

Conclusions. We have argued that

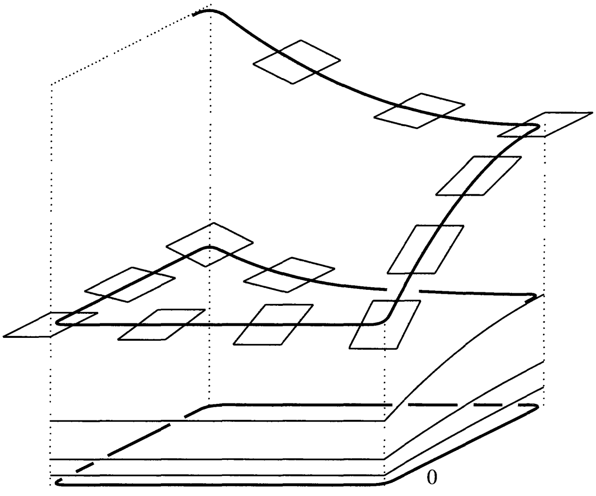
Lemma. *A choice of connection d_∇ on E is equivalent to a choice of horizontal plane field $\mathcal{H}$ in T_E that is scalar-multiplication invariant.* □

The parallel sections of d_∇ are exactly those sections of E that are tangent to d_∇'s horizontal plane field $\mathcal{H}_\nabla$ inside T_E. Clearly, we can find such parallel sections over any curve in X, and this yields d_∇'s parallel transport.

Finding parallel sections over more solid pieces of X runs into obstructions. In the extreme, the search for a *global* parallel section becomes the question of whether the horizontal plane field $\mathcal{H}_\nabla$ is *integrable.*[14]

Flat connections. If $\mathcal{H}_\nabla$ happens to be integrable, then the corresponding connection d_∇ is called **flat**, and the manifold E is foliated by leaves transverse to the fibers. Furthermore, if X happens to be simply-connected, then any bundle E on X that admits a flat connection must be trivial and is trivialized by the induced foliation.

In general, the obstruction to the horizontal plane field being integrable is measured by the *curvature* F_∇ of the connection, which we will encounter soon enough. See also figure 9.4.



9.4. This is not a flat connection

Difference of two connections. Imagine given two random connections d' and d'' on E. It follows from their properties that they must differ by a 1-form whose values are linear endomorphisms of the fibers of E. That is:

$$d'\sigma - d''\sigma = A \cdot \sigma \qquad \text{for some } A \in \Gamma\big(\mathrm{End}(E) \otimes T_X^*\big) ,$$

where $\mathrm{End}(E) = E \otimes E^*$ is the bundle of linear endomorphisms of the fibers of E. One can think of A simply as the derivative of the difference of the

---

[14] A k-plane field (or subbundle) P in T_Z is called **integrable** on Z if it is *everywhere* tangent to a well-defined family of k-dimensional submanifolds of Z, that is, if for each $z \in Z$ there is a k-submanifold L_z such that $P|_{L_z} = T_{L_z}$. A plane field P is integrable if and only if for every two sections V, $W \in \Gamma(P)$ their bracket $[V, W]$ takes values only from P. (A 1-dimensional plane field, *i.e.*, a line field, is always integrable.) An integrable plane field determines a foliation of Z (with leaves the L_z's above), while a nowhere-integrable plane field on a 3-manifold is called a **contact structure.** Foliations will be discussed again in section 11.3 (page 492) (for unrelated reasons).

two associated parallel transports τ' and τ'' (remember that $\mathrm{End}(\mathbb{R}^n) = \mathfrak{gl}(n)$).

The space $Conn(E)$ of connections on E is thus an affine space, modeled on the space of sections $\Gamma\big(\mathrm{End}(E) \otimes T^*_X\big)$. In other words,

$$Conn(E) = \{d_\nabla\} + \Gamma\big(\mathrm{End}(E) \otimes T^*_X\big) ,$$

for any particular connection d_∇ on E.

Case of G–bundles. If the bundle E happens to be a G–bundle, then the natural parallel transports to be considered on E are those that act through elements of G.

Therefore, since connections (covariant derivatives) are obtained by taking derivatives of the parallel transport, these must act by elements of the Lie algebra $\mathfrak{g}$. In particular, any two connections d' and d'', compatible with the G–structure of E, will differ by a 1–form whose values are endomorphisms of the fibers that act by elements[15] of $\mathfrak{g}$. We call **G–connections** those connections on a G–bundle E that respect its G–structure.

For example, if E is endowed with a metric and becomes an $SO(n)$–bundle, then we wish its connections to act by elements of $\mathfrak{so}(n)$, *i.e.*, by skew-symmetric endomorphisms. This means simply that a $SO(n)$–connection must satisfy

$$\tfrac{\partial}{\partial w}\langle\sigma', \sigma''\rangle = \big\langle\nabla_w \sigma', \sigma''\big\rangle + \big\langle\sigma', \nabla_w \sigma''\big\rangle .$$

Notice that this is just the derivative of the required metric-compatibility $\langle e', e''\rangle = \big\langle\tau_c(e'), \tau_c(e'')\big\rangle$ of the parallel transport. An $SO(n)$–connection is often called a **metric connection** on E.

Local expression. Imagine that our bundle is trivial, $E \approx X \times \mathbb{R}^n$. Then E admits global fiber-coordinates, and we can define its **flat connection** by

$$d_{flat}\,\sigma = \big(d\sigma_1, \ldots, d\sigma_n\big) ,$$

where on the right we have the usual exterior derivatives of the components $\sigma_k \colon X \to \mathbb{R}$ of the section σ with respect to the trivialization.

If our bundle is not trivial, then the flat connection d_{flat} cannot be defined, since there are no global coordinates on the fibers. Nonetheless, given a local trivialization $E|_{U_\alpha} \approx U_\alpha \times \mathbb{R}^n$, there is an associated *local* flat connection $d_{flat}|_{U_\alpha}$ on $E|_{U_\alpha}$. Further, every global connection d_∇ on E must differ

---

[15] Rigorously, by using a G–cocycle for E, one lets it act, instead of on $\mathbb{R}^n$, on the Lie algebra $\mathfrak{g}$ (by using the adjoint action of G on $\mathfrak{g}$). The resulting bundle is denoted $\mathfrak{g}(E)$ (or $\mathrm{ad}\,E$). The global difference between two G–connections is a section of $\mathfrak{g}(E) \otimes T^*_X$. (Notice that $\mathfrak{gl}(E) = \mathrm{End}(E)$, and that, if G is a subgroup of $GL(n)$, then $\mathfrak{g}(E)$ will be a subbundle of $\mathrm{End}(E)$; also notice that, if G is Abelian, then $\mathfrak{g}(E)$ is the trivial bundle $X \times \mathfrak{g}$.)

from $d_{\mathit{flat}}|_{U_\alpha}$ by a local $\mathrm{End}(\mathbb{R}^n)$–valued 1–form, and therefore we can always write in local coordinates

$$d_\nabla \sigma|_{U_\alpha} = d_{\mathit{flat}} \sigma^\alpha + A_\alpha \cdot \sigma^\alpha$$

for some suitable local 1–form

$$A_\alpha \in \Gamma(\mathfrak{gl}(n) \otimes T^*_X|_{U_\alpha}) .$$

One can think of the form A_α either as a matrix-valued 1–form or as a 1–form-valued matrix, acting on the local coordinates $\sigma^\alpha = (\sigma^\alpha_1, \ldots, \sigma^\alpha_n)$ of the section σ in the chosen local coordinates of E over U_α.

> ***Local patching.*** *The various local forms A_α are related by formulae of shape*
>
> $$A_\alpha = g_{\alpha\beta} \cdot dg_{\beta\alpha} + g_{\alpha\beta} \cdot A_\beta \cdot g_{\beta\alpha} ,$$
>
> *where $dg_{\alpha\beta}$ is the matrix made from the differentials of the matrix-components of the change-of-coordinates map $g_{\beta\alpha} = g^{-1}_{\alpha\beta}$, and the multiplication is multiplication of matrices.*

The case of G–bundles. If our bundle E is a G–bundle, then we prefer to choose a set of local trivializations compatible with the G–structure of E. In this case, every G–connection d_∇ can be written $d_\nabla \sigma|_{U_\alpha} = d_{\mathit{flat}} \sigma^\alpha + A_\alpha \cdot \sigma^\alpha$ for a suitable local form

$$A_\alpha \in \Gamma(\mathfrak{g} \otimes T^*_X|_{U_\alpha}) .$$

Like exterior derivatives. The visible kinship of a connection d_∇ with the usual exterior derivative d can be pursued further, by extending d_∇ to act on E–valued k–forms. Define inductively operators

$$d_\nabla \colon \Gamma(E \otimes \Lambda^k(T^*_X)) \longrightarrow \Gamma(E \otimes \Lambda^{k+1}(T^*_X))$$

by using the Leibnitz property[16]

$$d_\nabla(\sigma \otimes \alpha) = (d_\nabla \sigma) \wedge \alpha + \sigma \otimes (d\alpha)$$

for all $\alpha \in \Gamma(\Lambda^k)$ and $\sigma \in \Gamma(E)$.

Unlike the exterior derivative though, in general the operators d_∇ do not satisfy anything like $d\,d = 0$. Which is another way to stumble upon curvatures, as we will see shortly.

---

[16] Often in the literature one writes E–valued k–forms as sections of $\Lambda^k \otimes E$, which forces the Leibnitz relation to be written $d_\nabla(\alpha \otimes \sigma) = (d\alpha) \otimes \sigma + (-1)^k \alpha \wedge (d_\nabla \sigma)$, with an unpleasant sign-correction depending on the degree k of α.

Torsion-free connections. Before that, though, notice that the case $E = T^*_X$ stands apart. Indeed, each connection d_∇ on the cotangent bundle T^*_X of X is a map

$$d_\nabla \colon \Gamma(T^*_X) \longrightarrow \Gamma(T^*_X \otimes T^*_X) \ .$$

By skew-symmetrization, it induces an operator

$$\mathrm{Alt}\, d_\nabla \colon \Gamma(\Lambda^1(T^*_X)) \longrightarrow \Gamma(\Lambda^2(T^*_X)) \ .$$

It is then natural to ask a further compatibility condition from the connections that act on the cotangent bundle T^*_X: we ask that the induced operator $\Gamma(\Lambda^1) \to \Gamma(\Lambda^2)$ coincide with the usual exterior derivative d:

$$\mathrm{Alt}\, d_\nabla \ = \ d \ .$$

Such a nice connection is known as a **torsion-free** connection on T^*_X.

It is best to think of this torsion-free condition as ensuring that the connection d_∇ is compatible with the fact that, unlike most other bundles, T^*_X admits special local fiber-coordinates coming from chart-coordinates on the base-manifold X itself.

The Levi–Cività connection. A manifold X^m is called a **Riemannian** manifold if its tangent bundle T_X is endowed with an inner product on its fibers (a **Riemannian metric**), and thus becomes an $SO(m)$–bundle. In this case, for connections on T_X it is natural to restrict attention to $SO(m)$–connections.

However, the metric also establishes a bundle-isomorphism $T_X \approx T^*_X$, and thus any connection on T_X is a connection on T^*_X as well. Thus, it is also natural to ask that our connections be torsion-free.

In fact, for every Riemannian metric on X, there is exactly *one* torsion-free $SO(m)$–connection d_∇ on T_X. It is called the **Levi–Cività connection** of the metric. Specifically, the Levi–Cività connection must satisfy the two properties:

$$\begin{aligned} \tfrac{\partial}{\partial Z}\langle V, W\rangle &= \langle \nabla_Z V, W\rangle + \langle V, \nabla_Z W\rangle \,, \\ [V, W] &= \nabla_V W - \nabla_W V \,. \end{aligned}$$

This second condition, by invoking the bracket $[V, W]$ of the two vector fields V and W, encodes that ∇ is torsion-free.

The **bracket** $[V, W]$ is fully characterized by its bilinearity and its vanishing whenever V and W are induced from local coordinates $\{x_1, \ldots, x_m\}$ on X, so that, say, $V = \frac{\partial}{\partial x_i}$ and $W = \frac{\partial}{\partial x_j}$. If $E_1, \ldots, E_m$ is any local frame of vector fields in T_X induced from local coordinates on X, then the torsion-free condition becomes merely the symmetry $\nabla_{E_i} E_j = \nabla_{E_j} E_i$.

Curvatures

If we apply the connection operators d_∇ twice, as in

$$\Gamma(E) \xrightarrow{d_\nabla} \Gamma(E \otimes T_X^*) \xrightarrow{d_\nabla} \Gamma\big(E \otimes \Lambda^2(T_X^*)\big) ,$$

then the result is not trivial (as it would be for the usual exterior derivative d), but instead yields the curvature 2–form of ∇. That is, we have

$$d_\nabla(d_\nabla \sigma) = F_\nabla \cdot \sigma$$

for some suitable endomorphism-valued global 2–form

$$F_\nabla \in \Gamma\big(\mathrm{End}(E) \otimes \Lambda^2(T_X^*)\big) ,$$

acting on sections of E. The form F_∇ is called the **curvature form**[17] of ∇.

If E is a G–bundle, then the curvature of a G–connection will only act by endomorphisms that belong to[18] $\mathfrak{g}$.

Flat connections. If our bundle E is trivial and we are looking at the flat connection d_{flat} induced from some trivialization $E \approx X \times \mathbb{R}^n$, then of course we have $d_{flat}\, d_{flat} = 0$, and thus the curvature is $F_{flat} \equiv 0$.

In general, we call **flat** any connection d_∇ whose curvature vanishes identically. This is in fact equivalent to the connection being locally isomorphic to d_{flat}, or to the horizontal plane field $\mathcal{H}_\nabla$ in T_E being integrable. Moreover, if the base manifold X is simply-connected and the bundle E admits a flat connection, then E must be *trivial.*

> ***Characteristic classes.*** *The detection of triviality through curvatures goes further: By extracting various global differential forms on X from the curvature F_∇ of any random connection on E, one obtains closed forms that represent all the characteristic classes of E in de Rham cohomology.*[19] *This is called the* **Chern–Weil method.** *See for example* **S. Kobayashi** *and* **K. Nomizu***'s* ***Foundations of differential geometry*** [**KN69, KN96**, vol. II, ch. XII].

---

[17] Geometers would likely denote F_∇ by something like $\mathcal{R}_\nabla$ (with "$\mathcal{R}$" from "Riemann"), while the physicists would denote it by F_A, where $A = \{A_\alpha\}$ are the local 1-forms defining ∇ with respect to local trivializations. Just as they call the connection a "gauge potential", its curvature F_∇ would be known to physicists as a "gauge field".

[18] Rigorously, one builds the bundle of Lie algebras $\mathfrak{g}(E)$, and sees F_∇ as a section of $\mathfrak{g}(E) \otimes \Lambda^2(T_X^*)$. (If G is Abelian, then $\mathfrak{g}(E) = X \times \mathfrak{g}$, and F_∇ is simply a $\mathfrak{g}$–valued 2–form.)

[19] Cohomology with $\mathbb{R}$-coefficients, thus torsion and the Stiefel–Whitney classes are not visible.

Local expression. If the connection is expressed locally as

$$d_\nabla|_{U_\alpha} = d_{flat} + A_\alpha$$

for some suitable $\mathfrak{gl}(n)$–valued 1–form A_α over U_α, and its curvature F_∇ is written locally as

$$F_\nabla \cdot \sigma|_{U_\alpha} = F_\alpha \cdot \sigma^\alpha ,$$

where F_α is a local matrix-valued 2–form, acting on the local coordinates σ_α of σ over U_α; then these local curvature forms are described explicitly by

$$F_\alpha = dA_\alpha + A_\alpha \wedge A_\alpha ,$$

where dA_α is the $n \times n$ matrix built from the differentials of the components of A_α, while $A_\alpha \wedge A_\alpha$ is obtained by multiplying the matrix of 1–forms A_α with itself, in a combination of matrix multiplication and exterior product of forms.

> ***Local patching.*** *The various local representations F_α are related by $F_\alpha = g_{\alpha\beta} \cdot F_\beta \cdot g_{\beta\alpha}$. In particular, if the structure group G of E is Abelian, then F_∇ is simply a global matrix-valued 2–form on X.*

A consequence of the explicit formula above is that the curvature 2–form F_∇ satisfies the following

Bianchi Identity. $dF_\alpha = F_\alpha \wedge A_\alpha - A_\alpha \wedge F_\alpha$. □

This can be expressed coordinate-free as follows: The connection d_∇ on E induces a connection on E^*, and thus defines a connection[20] on $E \otimes E^* = \mathrm{End}(E)$. If we denote the latter still by d_∇, then the Bianchi identity can be written

$$d_\nabla F_\nabla = 0 .$$

> *In particular, it follows that, if the structure group G of E is Abelian, then F_∇ is a matrix of* closed *2–forms.*

Curvature from covariant derivative. The curvature can be defined directly in terms of the covariant derivative, through the formula:

$$F_\nabla(V, W) \cdot \sigma = \nabla_V \nabla_V \sigma - \nabla_W \nabla_W \sigma - \nabla_{[V,W]} \sigma ,$$

where $[V, W]$ is the bracket of the vector-fields V and W. The endomorphism

$$F_\nabla(V, W)|_x : E|_x \longrightarrow E|_x$$

depends only on the values of V and W at x.

In particular, if $E_1, \ldots, E_n$ is any local frame of vector fields in T_X induced from coordinates on X, then we can write simply

$$F_\nabla(E_i, E_j) \cdot \sigma = \nabla_{E_i} \nabla_{E_j} \sigma - \nabla_{E_j} \nabla_{E_i} \sigma .$$

---

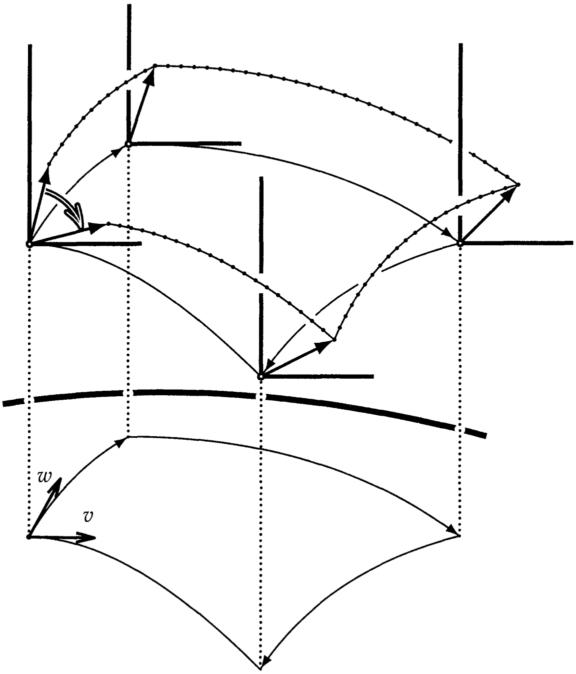
[20] On E^*, by $(d_\nabla f)(e) = d(f(e)) - f(d_\nabla e)$; then on $E \otimes E^*$ by $d_\nabla(f \otimes e) = (d_\nabla f) \otimes e + f \otimes (d_\nabla e)$.

Thus, in a certain sense the curvature measures the failure of mixed second derivatives from commuting.

Curvature from parallel transport. The curvature can also be expressed directly in terms of the parallel transport.

Given two vectors v, $w \in T_X|_p$ based at some $p \in X$, we pick some coordinates $\{x_1, \ldots, x_m\}$ on X around p such that, say, the coordinate x_1 goes in the v–direction while x_2 goes in the w–direction. Using these local coordinates, we build a little square loop along the v– and w–directions. Namely, we start at p and go a length of ε in the x_2–direction, then a length of ε in the x_1–direction, then backwards a length of ε in the $(-x_1)$–direction, and finally a length of ε in the $(-x_2)$–direction, and end up back at p.

We now pick our favorite vector e in the fiber $E|_p$ above p and use the parallel transport along our square path to move it to another vector $\tau_\square(e)$ of $E|_p$, as in figure 9.5. The difference $\tau_\square(e) - e$ measures the failure of the parallel transport to trivialize E over the (x_1, x_2)–plane in the neighborhood of p.



9.5. Curvature, from parallel transport

To get a quantity independent of the various choices, we push it to an infinitesimal one by shrinking the square, and thus obtain the curvature:

$$F_\nabla(v, w) \cdot e \;=\; \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon^2}\left(\tau_\square(e) \;-\; e\right) .$$

In particular, this shows again that, if we are dealing with a G–bundle, then the endomorphism-part of F_∇ must act through elements of $\mathfrak{g}$.

Examples. Consider a complex-line bundle L endowed with a Hermitian metric. Then the structure group of L is $U(1)$. Since $U(1) = \mathbb{S}^1$, its Lie algebra is simply $\mathfrak{u}(1) = i\mathbb{R}$. This implies that any $U(1)$–connection on L has as curvature a global imaginary-valued 2–form

$$F_\nabla \in i\Gamma(\Lambda^2(T_X^*)) .$$

From the Bianchi identity it follows that F_∇ must be a *closed* 2–form. Further, in de Rham cohomology F_∇ represents the Chern class of L:

$$[F_\nabla] = -2\pi i\, c_1(L) .$$

(This is the simplest instance of the Chern–Weil method.)

Aside from the neat case of line bundles, in general the curvature is a very complicated object. Indeed, one of the technical advantages of Seiberg–Witten theory over Donaldson's is that the former uses connections on complex-line bundles, while the latter deals with connections on $SU(2)$– or $SO(3)$–bundles.

When the manifold X is endowed with a Riemannian metric, the study of the curvature F_∇ of the Levi–Cività connection on T_X is made mostly after reducing it to various simpler creatures, by taking traces. Thus, one talks about sectional curvature, Ricci curvature, *etc.* The only one that we will encounter later is the **scalar curvature**, which is the function

$$scal\colon X \longrightarrow \mathbb{R} \qquad\qquad scal = \sum \langle F_\nabla(e_i, e_j) \cdot e_j,\ e_i \rangle$$

for any local orthonormal frame $e_1, \ldots, e_m$ in T_X. (The value $scal(p)$ is best imagined as an average of the sectional curvatures at p.)

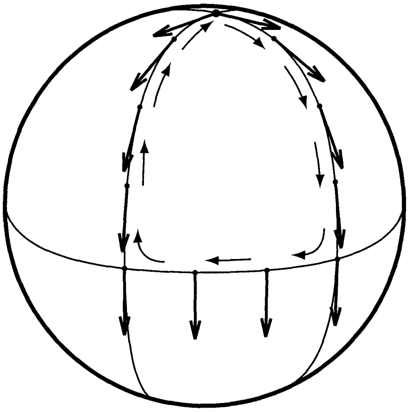
In figure 9.6 on the next page is pictured the parallel transport on the 2–sphere with its standard Riemannian metric. The curvature of the sphere (of unit radius) turns out to be constant, $F = \left[\begin{smallmatrix} & 1 \\ -1 & \end{smallmatrix}\right]$*, which leads to the constant scalar curvature* $scal = +2$*. Notice that the Gaußian curvature is* $+1$*.*

9.3. We are special: self-duality

When viewed from the perspective of differential geometry, 4–manifolds have a unique distinguishing property: the Hodge star-operator

$$*\colon \Lambda^k(T_M^*) \longrightarrow \Lambda^{4-k}(T_M^*)$$

takes 2–forms to 2–forms. This is remarkable for a differential geometer because curvatures are also differential 2–forms of various flavors.



9.6. The sphere is curved

In this section we will review a few of the consequences of this distinguishing property of 4–manifolds. As will later become apparent, self-duality and its ramifications thoroughly permeate the underpinnings of gauge theory in dimension 4, both in its Donaldson and Seiberg–Witten flavors.

The Hodge duality operator. If we endow an m–manifold X^m with a Riemannian metric, then the orientation of X endows it with a top differential form vol_X, called the **volume form** of X. Locally we have

$$vol_X|_x = e^1 \wedge \ldots \wedge e^m$$

for any orienting orthonormal frame $\{e^1, \ldots, e^m\}$ of $T^*_X|_x$. The **Hodge operator**

$$*\colon \Lambda^k(T^*_X) \longrightarrow \Lambda^{m-k}(T^*_X)$$

is defined by the equality

$$\alpha \wedge * \beta = \langle \alpha, \beta \rangle \cdot vol_X \, .$$

for all $\alpha, \beta \in \Gamma(\Lambda^k)$. In particular, we have $\alpha \wedge * \alpha = |\alpha|^2 \cdot vol_X$.

The meaning of the Hodge operator is, roughly, that if $\beta|_x \in \Lambda^k|_x$ is dual to an oriented $(m-k)$–plane P, then $* \beta|_x$ will be dual to its oriented orthogonal complement $P^\perp$. Not surprisingly, the Hodge operator can be used to prove the Poincaré duality for de Rham cohomology.

The case of dimension 4: self-dual / anti-self-dual. On a 4–manifold M, the most interesting part of $*$ is its action of 2–forms. We have

$$*\colon \Lambda^2(T^*_M) \longrightarrow \Lambda^2(T^*_M) \, ,$$

determined concretely by the property that $*(e^1 \wedge e^2) = e^3 \wedge e^4$ if and only if $\{e^1, e^2, e^3, e^4\}$ is an orienting orthonormal frame in T^*_M.

Since $**=id$, the operator $*$ on Λ^2 has eigenvalues ± 1 and splits $\Lambda^2(T_M^*)$ into its eigenbundles, denoted by

$$\Lambda^2_+(T_M^*) \qquad \text{and} \qquad \Lambda^2_-(T_M^*) \,.$$

A section in $\Lambda^2_+(T_M^*)$ is called a **self-dual** 2–form, while a section in $\Lambda^2_-(T_M^*)$ is called an **anti-self-dual** 2–form.

Every 2–form α splits into self-dual and anti-self-dual parts

$$\alpha = \alpha^+ + \alpha^- \,,$$

with $*\alpha^+ = \alpha^+$ and $*\alpha^- = -\alpha^-$. More, every self-dual 2–form is orthogonal, with respect to the wedge-product, to every anti-self-dual 2–form: for any $\alpha^+ \in \Gamma(\Lambda^2_+)$ and $\beta^- \in \Gamma(\Lambda^2_-)$ we have

$$\alpha^+ \wedge \beta^- = 0 \,.$$

This follows since $\alpha^+ \wedge \beta^- = (*\alpha^+) \wedge (-*\beta^-) = -\alpha^+ \wedge \beta^-$.

A 2–form α is self-dual at x if and only if there is some orienting orthonormal frame $\{e^1, e^2, e^3, e^4\}$ in $T_M^*|_x$ such that we can write

$$\alpha|_x = \rho\big(e^1 \wedge e^2 + e^3 \wedge e^4\big)$$

for some real constant ρ. In general, for a random orienting orthonormal frame $\{e^1, e^2, e^3, e^4\}$ in T_M^* at x, we have

$$\Lambda^2_+(T_M^*)|_x = \mathbb{R}\big\{e^1e^2 + e^3e^4, \quad e^1e^3 + e^4e^2, \quad e^1e^4 + e^2e^3\big\} \,.$$

De Rham cohomology. The splitting of bundles

$$\Lambda^2(T_M^*) = \Lambda^2_+(T_M^*) \oplus \Lambda^2_-(T_M^*)$$

descends in de Rham cohomology to a splitting

$$H^2(M;\mathbb{R}) = H^2_+(M;\mathbb{R}) \oplus H^2_-(M;\mathbb{R}) \,.$$

On one hand, since a self-dual 2–form α^+ has $*\alpha^+ = \alpha^+$, we deduce that

$$\int_M \alpha^+ \wedge \alpha^+ = \int_M \left|\alpha^+\right|^2 \cdot vol_X > 0 \,.$$

Similarly, for an anti-self-dual 2–form α^- with $*\alpha^- = -\alpha^-$, we have

$$\int_M \alpha^- \wedge \alpha^- = -\int_M \left|\alpha^-\right|^2 \cdot vol_X < 0 \,.$$

Furthermore, remember that we always have $\alpha^+ \wedge \alpha^- = 0$. Since the wedge-product on 2–forms becomes the intersection form Q_M on cohomology classes, we deduce that $H^2_+(M;\mathbb{R})$ is a maximal positive-definite subspace for Q_M, while $H^2_-(M;\mathbb{R})$ is a maximal negative-definite subspace. In particular, we recover the partial Betti numbers

$$b_2^+(M) = \dim H^2_+(M;\mathbb{R}) \qquad \text{and} \qquad b_2^-(M) = \dim H^2_-(M;\mathbb{R}) \,,$$

and hence we can write

$$\begin{aligned} \operatorname{sign} M &= b_2^+(M) - b_2^-(M) , \\ b_2(M) &= b_2^+(M) + b_2^-(M) . \end{aligned}$$

The partial Betti number b_2^+ plays an essential role in gauge theory. As preparation, more consequences of Hodge duality and the influence of $b_2^+(M)$ will be developed in the end-notes of this chapter (page 357).

Lie and bundles. The vector space $\Lambda^2(\mathbb{R}^n)$ is canonically isomorphic to the Lie algebra $\mathfrak{so}(n)$ of the rotation group $SO(n)$. In particular

$$\Lambda^2(\mathbb{R}^4) = \mathfrak{so}(4) .$$

The bundle splitting

$$\Lambda^2(T_M) = \Lambda^2_+(T_M^*) \oplus \Lambda^2_-(T_M^*)$$

corresponds fiberwise to the exceptional Lie algebra isomorphism

$$\mathfrak{so}(4) = \mathfrak{so}(3) \oplus \mathfrak{so}(3) .$$

At the level of Lie groups, the latter integrates to

$$Spin(4) = SU(2) \times SU(2)$$

and can thus be linked to both spin structures and complex geometry. This isomorphism doubly-covers a map $SO(4) \to SO(3) \times SO(3)$ that fits in the diagram:

$$\begin{tikzcd} Spin(4) \arrow[r, equal] \arrow[d] & SU(2) \arrow[d] & \times & SU(2) \arrow[d] \\ SO(4) \arrow[r] & SO(3) & \times & SO(3) . \end{tikzcd}$$

The latter map can be used to project the $SO(4)$–cocycle of T_M onto two $SO(3)$–cocycles, which in fact glue exactly the bundles Λ^2_+ and Λ^2_-.

On the other hand, if one can lift the $SO(4)$–cocycle of T_M to a $Spin(4)$–cocycle, *i.e.*, a spin structure, then this can be projected to two $SU(2)$–cocycles that will glue the bundles of spinors. We will discuss this again in section 10.2 (page 382).

9.4. The Donaldson invariants

After reviewing the previous basic differential-geometric notions, we can now briefly outline the setting of the invariants defined by S.K. Donaldson. They follow the general methodology suggested at the beginning of the chapter.

Minimize the energy. Let M be a closed oriented 4–manifold, endowed with a Riemannian metric. Take a complex-plane bundle

$$E_k \to M$$

with structure group $SU(2)$. Such bundles are fully classified by their second Chern class $k = c_2(E_k) \in \mathbb{Z}$. Consider an $SU(2)$–connection d_A on E: its curvature F_A is an $\mathrm{End}(E_k)$–valued 2–form. It is natural to consider its "energy", as given by the **Yang–Mills functional**

$$\mathcal{YM}(d_A) = \int_M |F_A|^2 \; vol_M \,,$$

and then try to minimize it.

> *One can think of this as a non-Abelian analogue of Hodge theory: in the latter, one tries to minimize the energy of an exterior form that represents a fixed cohomology class, and ends up with harmonic forms.*

The critical points of $\mathcal{YM}$ are described by the Euler–Lagrange equations

$$d_A F_A = 0 \qquad \text{and} \qquad d_A * F_A = 0\,,$$

where d_A denotes the induced connection on $\mathrm{End}(E_k)$–valued forms. The first equation $d_A F_A = 0$ is always satisfied, it being the Bianchi identity. While searching for solutions of the second equation, one can simplify the quest by focusing on the special cases when $* F_A$ is a multiple of F_A. This can happen only if either $* F_A = F_A$ (self-dual case) or $* F_A = -F_A$ (anti-self-dual case).

Furthermore, on one hand we can evaluate

$$\mathcal{YM}(A) = \int_M |F_A^+|^2 + |F_A^-|^2 \,,$$

while on the other hand we get

$$c_2(E) = \tfrac{1}{8\pi^2} \int_M - |F_A^+|^2 + |F_A^-|^2$$

from the Chern–Weil method.

Therefore, when $c_2(E) = 0$, the absolute minima of the Yang–Mills functional are the flat connections; when $c_2(E) < 0$, the absolute minima are the connections with $F_A^- = 0$; when $c_2(E) > 0$, the absolute minima are the connections with $F_A^+ = 0$. The latter equality means that $F_A = F_A^-$ and hence that the curvature of d_A is anti-self-dual.

The difference between the last two cases is merely a matter of orientations, so we pick the case when $c_2(E) = k > 0$ and look for connections with anti-self-dual curvature. These are called **k–instantons**, or more directly **anti-self-dual connections.**

The moduli space. We thus set ourselves to study the moduli space $\mathfrak{M}_k$ of all connections on E_k that satisfy the equation

$$F_A^+ = 0\,,$$

considered up to gauge-equivalence.

If $b_2^+(M) \geq 1$, then, for a generic Riemannian metric on M, the space $\mathfrak{S}$ of all anti-self-dual connections on E_k will contain no reducible solutions.[21]

Therefore, for a generic metric, the quotient of the solution space $\mathfrak{S}$ by the gauge group $\mathscr{G}(E_k)$ of automorphisms of E_k, *i.e.*, the moduli space $\mathfrak{M}_k$, will be a finite-dimensional orientable smooth manifold, of dimension

$$\dim \mathfrak{M}_k = 8k - 3(1 - b_1 + b_2^+)\,.$$

Certain further choices allow us to choose and fix an orientation of $\mathfrak{M}_k$.

The moduli space $\mathfrak{M}_k$ is almost never compact. However, it can be compactified in a reasonable way, such that the compactification $\widehat{\mathfrak{M}}_k$ carries a fundamental class $\big[\widehat{\mathfrak{M}}_k\big]$. It is worth noting that this compactification always involves the base-manifold M in a nontrivial way,[22] as well as the moduli spaces of lesser k's.

The invariants. The natural cohomology classes of $\widehat{\mathfrak{M}}_k$ are parametrized through certain maps

$$\mu_j\colon H_j(M;\mathbb{R}) \to H^{4-j}(\widehat{\mathfrak{M}}_k;\mathbb{R})\,.$$

If $b_1 + b_2^+$ is odd, then $\mathfrak{M}_k$ is even-dimensional, and one defines invariants

$$\mathfrak{q}_{p,r}(\alpha) = \int_{[\widehat{\mathfrak{M}}_k]} \mu_2(\alpha)^p\, \mu_0(point)^r$$

that depend on $\alpha \in H_2(M;\mathbb{R})$. If $b_1 + b_2^+$ is even, then we set all $\mathfrak{q}_{p,r} = 0$. These $\mathfrak{q}_{p,r}$ turn out to be polynomial functions on $H_2(M;\mathbb{R})$ and are called the **Donaldson** *(polynomial)* **invariants** of M.

A priori these polynomials depend on the choice of Riemannian metric, but, if we further require that $b_2^+(M) \geq 2$, then any two generic metrics can be connected through a path of metrics for which there are no reducible solutions.[23] The moduli spaces for various metrics will be cobordant, and therefore the invariants will not depend on the metric, but only on the base-manifold M.

---

[21] In this setting, a reducible solution is an anti-self-dual connection that preserves a splitting of E_k into a sum of two complex-line bundles. For the role of b_2^+ in avoiding reducibles, see the end-notes of this chapter (anti-self-dual connections on line bundles, page 357), especially page 364.

[22] Recall the earlier discussion from section 5.3 (page 243), where the compactification of $\mathfrak{M}_1$ was done by adding as its boundary exactly the manifold M.

[23] Again, see the end-notes of this chapter (page 364).

Finally, observe that not all applications of Donaldson theory come from these invariants $\mathfrak{q}_{p,r}$. Some come from a careful study of the moduli space itself, as we saw happen in section 5.3 (page 243), with Donaldson's theorem.

Epilogue, toward Seiberg–Witten theory. One can somehow intuitively understand why Donaldson theory might offer deep results about the topology of a 4–manifold M: for example, on one hand Donaldson theory is an analogue of Hodge theory and the whole construction is natural; and, on the other hand, the compactifications of its moduli spaces always involve the base-manifold M itself. In contrast, it is intuitively quite unclear why Seiberg–Witten theory, the simpler but much different equivalent, does give such deep insights into 4–dimensional topology.

The origins of the latter do not shed much light either: In 1994, following their work on $N = 2$ super-symmetric Yang–Mills theory, N. Seiberg and E. Witten proposed their invariant of 4–manifolds. They claimed that it would be equivalent to Donaldson's invariants based on physical arguments (the quantum field theory they were considering had a scale parameter t; when $t \to 0$ one got essentially the Donaldson invariants, while when $t \to \infty$ one got the Seiberg–Witten invariants).

Mathematically, there is a hard program due to V. Pidstrigach and A. Tyurin for proving the equivalence of the two theories, which has essentially been carried through in a long series of papers of P. Feehan and T. Leness. Thus, if one trusts the equivalence of the two theories, then maybe one should let Seiberg–Witten theory borrow from whatever intuitive meaning one had managed to lay upon Donaldson's...

In any case, all results obtained using Donaldson theory can be proved again, usually easier, by using Seiberg–Witten theory. For example, in the end-notes of the next chapter (page 454), we will explain the Seiberg–Witten proof of Donaldson's theorem on definite intersection forms.

More than that, Seiberg–Witten theory led to striking new results, which seemed out of reach while using instantons. Shortly after their appearance, most unanswered questions that were raised in the instanton period were quickly solved. Further, Seiberg–Witten theory has managed to completely transform our image of the world of smooth 4–manifolds, by bringing to light all over the place overflowing infinities of smooth structures on homeomorphic manifolds.

Thus, starting with the next chapter (starting on page 375) we leave Donaldson theory behind and focus exclusively on its Seiberg–Witten cousin.

9.5. Notes

Note: Anti-self-dual connections on line bundles

In what follows, we study the simplest case of instantons: the case of anti-self-dual connections on complex-line bundles. While these do not lead to any 4–manifold invariants (indeed, their moduli space is either empty or a single point[1]), they play a fundamental role in both Donaldson and Seiberg–Witten theories.

Anti-self-dual connections on line bundles are the place where b_2^+ enters gauge theory to never leave it again. Indeed, in both theories, anti-self-dual connections on line bundles correspond to *reducible solutions*. According to the general program outlined at the beginning of the chapter, we should try to avoid them. Thus, this whole note is concerned with finding, and then avoiding, these simple instantons. The techniques employed are elementary gymnastics with self-dual/anti-self-dual 2–forms.

Notations. In what follows, we will denote by

$$\Omega^k = \Gamma(\Lambda^k(T_M^*))$$

the space of global k–forms on M. In particular, $\Omega^0 = \{f\colon M \to \mathbb{R}\}$. We will also denote by $\Omega^k|_U$ the space of local k–forms defined on U.

The exterior derivative acts as a collection of maps $d\colon \Omega^k \to \Omega^{k+1}$. When we need to make obvious to which such map we are referring, we will write

$$d|_k\colon \Omega^k \longrightarrow \Omega^{k+1} .$$

If M is endowed with a Riemannian metric, then $\Lambda^2(T_M^*)$ splits into self-dual and anti-self-dual parts. We will correspondingly denote $\Omega^2_\pm = \Gamma(\Lambda^2_\pm(T_M^*))$, thus completing a splitting

$$\Omega^2 = \Omega^2_+ \oplus \Omega^2_- .$$

Accordingly, the operator $d|_1\colon \Omega^1 \to \Omega^2$ splits into $d|_1 = d^+ \oplus d^-$, with

$$\begin{array}{ccc} d^+\colon \Omega^1 \to \Omega^2_+ & \text{and} & d^-\colon \Omega^1 \to \Omega^2_- \\ d^+\alpha = (d\alpha)^+ & & d^-\alpha = (d\alpha)^- \end{array} .$$

---

[1] If M is not simply-connected and has $H^1(M;\mathbb{R}) \neq 0$, then the moduli space is either empty or a copy of the torus $H^1(M;\mathbb{R})/H^1(M;\mathbb{Z})$.

Finally, we adopt the slightly illogical notation-convention customary in gauge theory: we will denote a connection by[2] A, acting through the operator d_A and the covariant derivative ∇^A, and with curvature F_A.

Connections and curvatures. Let L be a complex-line bundle, endowed with a Hermitian fiber-metric or, equivalently, with a $U(1)$–structure. Further, assume that $\{g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to S^1\}$ is its $U(1)$–cocycle. A $U(1)$–connection A on L is an operator

$$d_A\colon \Gamma(L) \longrightarrow \Gamma(L)\ .$$

It is determined by a family of local imaginary-valued 1–forms $A_\alpha \in i\Omega^1|_{U_\alpha}$, with

$$d_A\,\sigma|_{U_\alpha} = d\sigma^\alpha + A_\alpha \cdot \sigma^\alpha\ ,$$

where $\sigma^\alpha\colon U_\alpha \to \mathbb{C}$ is the representation of σ in local coordinates of L over U_α.

The local 1–forms A_α are related to each other by $A_\alpha = A_\beta + g_{\alpha\beta}^{-1}\,dg_{\alpha\beta}$, where $dg_{\alpha\beta}$ is the imaginary-valued 1–form representing the differential[3] of $g_{\alpha\beta}$. If the U_α's are simply-connected, then we can represent the $g_{\alpha\beta}$'s as $g_{\alpha\beta} = e^{if_{\alpha\beta}}$ for suitable functions $f_{\alpha\beta}\colon U_\alpha \cap U_\beta \to \mathbb{R}$, and then the coordinate-change is written $A_\alpha = A_\beta + i\,df_{\alpha\beta}$.

The curvature F_A of a connection A is given locally by $F_A|_{U_\alpha} = dA_\alpha + A_\alpha \wedge A_\alpha$, but in our context that becomes simply

$$F_A|_{U_\alpha} = dA_\alpha\ .$$

More, since $A_\alpha = A_\beta + i\,df_{\alpha\beta}$, it follows that $dA_\alpha = dA_\beta$, and thus F_A is a *global* imaginary-valued 2–form, which furthermore is *closed.* In de Rham cohomology, the form F_A represents the Chern class of L. In review:

$$F_A \in i\Omega^2\ , \qquad dF_A = 0\ , \qquad [F_A] = -2\pi i\, c_1(L)\ .$$

Conversely, given any random closed 2–form $\Phi \in i\Omega^2$ so that $[\Phi] = -2\pi i\, c_1(L)$, there must exist a connection A on L with $F_A = \Phi$. Indeed, pick a random connection A^0 on L, with curvature F_{A^0}. Let Φ be any closed imaginary 2–form with $[\Phi] = -2\pi i\, c_1(L)$. Then $\Phi - F_{A^0}$ must be cohomologically-trivial, that is to say, there exists some 1–form $\varphi \in \Omega^1$ so that $\Phi - F_{A^0} = i\,d\varphi$, but then $A = A^0 + i\varphi$ is a connection on L whose curvature is exactly $F_A = \Phi$.

In conclusion, denoting by $Conn(L)$ the space of $U(1)$–connections of L, we have:

Lemma. *The map taking a connection to its curvature,*

$$\begin{aligned} Conn(L) &\longrightarrow \{\Phi \mid \Phi \in i\Omega^2,\ d\Phi = 0,\ [\Phi] = -2\pi i\, c_1(L)\} \\ A &\longmapsto F_A\ , \end{aligned}$$

is always surjective. □

Of course, this map is not injective. To remedy that, we need to consider our connections only up to the action of the gauge group $\mathcal{G}(L)$ of L.

---

[2] This has its origins in physics, where the connection is concretely understood as the family of local 1–forms $A = \{A_\alpha\}$ that describes the gauge potential of the field F_A.

[3] A map $g\colon U \to S^1$ has differential $dg\colon T_U \to T_{S^1}$. Since $T_{S^1} = S^1 \times i\mathbb{R}$, this dg can be viewed as a section in $T_U^* \otimes i\mathbb{R}$, in other words, as an imaginary-valued 1–form $dg \in i\Omega^1|_U$.

Gauge transformations. The gauge group, *i.e.*, the group of automorphisms of the $U(1)$–bundle L, can be described simply as

$$\mathscr{G}(L) = \{g\colon M \to S^1\}\,,$$

acting on L by complex-multiplication. We prefer to see its action on the sections of L as an action by pull-backs:[4] for every $\sigma \in \Gamma(L)$, the section $g \cdot \sigma$ is described by

$$(g \cdot \sigma)(x) = g(x)^{-1}\sigma(x)\,.$$

If M is simply-connected, then every $g \in \mathscr{G}(L)$ can be written as $g(x) = e^{if(x)}$ for some $f\colon M \to \mathbb{R}$. Up to translations by 2π, this identifies $\mathscr{G}(L)$ with Ω^0, that is to say, $\mathscr{G}(L) = \Omega^0/\{f\colon M \to 2\pi\mathbb{Z}\}$.

The gauge group has an induced action on connections; in terms of covariant derivatives, we have

$$g \cdot \nabla^A = g \circ \nabla^A \circ g^{-1}\,.$$

That is, the connection $g \cdot A$ is described by

$$\begin{aligned}\nabla_V^{(g\cdot A)}\sigma &= (g \cdot \nabla^A)_V\sigma = g^{-1}(\nabla_V^A(g\sigma)) = g^{-1}g\,(\nabla_V^A\sigma) + g^{-1}dg(V)\,\sigma \\ &= \nabla_V^A\sigma + g^{-1}dg(V)\,\sigma\,,\end{aligned}$$

and therefore

$$d_{(g\cdot A)} = d_A + g^{-1}dg\,.$$

In the simply-connected case, when we can write $g = e^{if}$, this becomes simply

$$(e^{if}) \cdot d_A = d_A + i\,df\,.$$

Simply-connected or not, every $g \in \mathscr{G}(L)$ can *locally* be written as $g = e^{if}$, and we deduce that $\mathscr{G}(L)$ always acts trivially on curvatures:

$$F_{(g\cdot A)} = F_A\,.$$

In conclusion, the curvature is a gauge-invariant of connections.

If two connections A' and A'' on L have the same curvature $F_{A'} = F_{A''}$, then their local forms must differ by a closed 1–form. Namely, there exists some $\varphi \in \Omega^1$ so that $A' = A'' + i\varphi$ and $d\varphi = 0$.

Assume that M is simply-connected, or at least that $H^1(M;\mathbb{R}) = 0$. Then φ must also be an exact 1–form; in other words, there exists a function $f\colon M \to \mathbb{R}$ so that $df = \varphi$. Then $g = e^{if}$ is a gauge transformation that takes A'' to A', *i.e.*, $A' = g \cdot A''$. Thus, two connections have the same curvature if and only if they are gauge-equivalent. We have proved:

Classification Lemma. *If M has $H^1(M;\mathbb{R}) = 0$ (e.g., when M is simply-connected), then the map taking a connection to its curvature,*

$$\begin{array}{ccc} Conn(L)\,/\,\mathscr{G}(L) & \xrightarrow{\approx} & \{\Phi \mid \Phi \in i\Omega^2,\ d\Phi = 0,\ [\Phi] = -2\pi i\, c_1(L)\} \\ [A] & \longmapsto & F_A\,, \end{array}$$

is a bijection. If M has $H^1(M;\mathbb{R}) \neq 0$, then this map is nonetheless a surjection. □

---

[4] This is merely an aesthetic choice, which will make certain formulae look more pleasant later on.

Non-simply-connected case. If M has $H^1(M;\mathbb{R}) \neq 0$, then a connection on L is completely determined up to gauge by its curvature form, together with its holonomy along a basis of cycles (circles) for $H^1(M;\mathbb{R})$. The **holonomy** *of a connection along a loop is the parallel transport along that loop. In the case of a $U(1)$–bundle, the holonomy is merely a rotation by some angle ϑ. Thus, a connection A on L is determined up to gauge-equivalence by its curvature F_A and by its holonomy angles $\vartheta_1, \ldots, theta_{b_1(M)} \in S^1$ along a basis of $H^1(M;\mathbb{R})$. In other words, for any fixed A^0 the solutions to the equation $F_A = F_{A^0}$ make up a torus in $Conn(L)/\mathscr{G}(L)$, isomorphic to $H^1(M;\mathbb{R})/H^1(M;\mathbb{Z})$.*

Anti-self-dual connections. A connection $A \in Conn(L)$ is called **anti-self-dual** if and only if its curvature F_A is an anti-self-dual 2–form, or in other words if

$$F_A^+ = 0\,.$$

To investigate the problem of finding (or avoiding) anti-self-dual connections, we will prove that the self-dual part of the curvature classifies connections just as well as the whole curvature form:

Classification Lemma II. *If M has $H^1(M;\mathbb{R}) = 0$ (e.g., when M is simply-connected), then the map taking a connection to the self-dual part of its curvature,*

$$\begin{array}{ccc} Conn(L) \,/\, \mathscr{G}(L) & \xrightarrow{\approx} & \{\Phi^+ \mid \Phi \in i\Omega^2,\ d\Phi = 0,\ [\Phi] = -2\pi i\, c_1(L)\} \\ [A] & \longmapsto & F_A^+\,, \end{array}$$

is a bijection. If M has $H^1(M;\mathbb{R}) \neq 0$, then this map is nonetheless a surjection.

Proof. The proof rests on the observation that, for any 1–form α, we have

$$d^+\alpha = 0 \qquad \Longleftrightarrow \qquad d\alpha = 0\,.$$

It is clear that $d\alpha = 0$ implies $d^+\alpha = 0$. For the converse, take any 1–form α and start with Stokes':

$$\begin{aligned} 0 = \int d(\alpha \wedge d\alpha) &= \int d\alpha \wedge d\alpha \\ &= \int (d^+\alpha + d^-\alpha) \wedge (d^+\alpha + d^-\alpha) \\ &= \int (d^+\alpha) \wedge (d^+\alpha) \ + \ (d^-\alpha) \wedge (d^-\alpha)\,, \end{aligned}$$

after using that Ω^2_+ and Ω^2_- are wedge-orthogonal. Further:

$$\begin{aligned} &= \int (d^+\alpha) \wedge (*(d^+\alpha)) \ + \ (d^-\alpha) \wedge (-*(d^-\alpha)) \\ &= \int \left|d^+\alpha\right|^2 - \int \left|d^-\alpha\right|^2\,. \end{aligned}$$

Therefore, $d^+\alpha = 0$ if and only if $d^-\alpha = 0$. In particular, if $d^+\alpha = 0$, then $d\alpha = 0$. By linearity it follows that, for any two 1–forms α and β, if $d^+\alpha = d^+\beta$, then $d\alpha = d\beta$.

If applied to the local forms A_α describing a connection A, this immediately shows that the map

$$F_A \longmapsto F_A^+$$

must be a bijection. Combining it with the earlier classification lemma, this concludes the proof. □

A bit of Hodge theory. As mentioned, once M is endowed with a metric, then one can define its Hodge star-operator $*\colon \Lambda^k(T^*_M) \to \Lambda^{4-k}(T^*_M)$. Acting on 2–forms, it splits Ω^2 into its (± 1)–eigenspaces, namely Ω^2_+ and Ω^2_-. Moreover, by using the star-operator we can define the (formal) adjoint to the exterior differential d, specifically

$$d^*\colon \Omega^k \longrightarrow \Omega^{k-1} \qquad \text{with} \qquad d^* = - * d * .$$

Its fundamental properties are:

$$\int_M \langle d\alpha,\, \beta\rangle = \int_M \langle \alpha,\, d^*\beta\rangle \qquad \text{and} \qquad d^*\, d^* = 0\,.$$

When we wish to specify which one of the operators $d^*\colon \Omega^k \to \Omega^{k-1}$ we have in mind, we will write $d^*|_k$.

Any exterior k–form α with both $d\alpha = 0$ and $d^*\alpha = 0$ is called a **harmonic k–form.** A harmonic k–form is uniquely characterized by its minimizing the length $\|\alpha\|^2 = \int_M |\alpha|^2\, vol_M$ among all closed k–forms α that represent the same cohomology class. In other words, a harmonic form offers a minimal representative for each cohomology class in $H^k(M;\mathbb{R})$.

Even more, a harmonic representative always exists and is unique. Therefore, if we denote by

$$\mathcal{H}^k(M) = \{\alpha \in \Omega^k \mid d\alpha = 0,\ d^*\alpha = 0\}$$

the space of all harmonic k–forms on M, then we can state

De Rham's Theorem. *On any Riemannian manifold M we have isomorphisms*

$$\mathcal{H}^k(M) \;\approx\; H^k(M;\mathbb{R})\,. \qquad \square$$

One of its consequences is the splitting

$$\Omega^k \;=\; \mathcal{H}^k(M) \,\oplus\, \operatorname{Im}(d|_{k-1}) \,\oplus\, \operatorname{Im}(d^*|_{k+1})\,,$$

which is orthogonal with respect to the chosen Riemannian metric on M.

In the special case of 2–forms on 4–manifolds, we denote by $\mathcal{H}^2_+(M)$ the space of all harmonic 2–forms that are self-dual, and by $\mathcal{H}^2_-(M)$ the space of all harmonic anti-self-duals. These $\mathcal{H}^2_+$ and $\mathcal{H}^2_-$ catch the whole 2–cohomology of M:

Lemma. *On any Riemannian 4–manifold M we have the splitting*

$$\mathcal{H}^2(M) = \mathcal{H}^2_+(M) \oplus \mathcal{H}^2_-(M)$$

and, correspondingly,

$$\Omega^2_+ \,=\, \mathcal{H}^2_+(M) \oplus \operatorname{Im} d^+ \qquad \textit{and} \qquad \Omega^2_- \,=\, \mathcal{H}^2_-(M) \oplus \operatorname{Im} d^-\,.$$

Proof. We prove that $\Omega^2_+ = \mathcal{H}^2_+ \oplus \operatorname{Im} d^+$. Assume $\beta \in \Omega^2_+$ is orthogonal to $\operatorname{Im} d^+$. This means that, for all $\alpha \in \Omega^1$, we have

$$\begin{aligned} 0 = \int \langle \beta,\, d^+\alpha\rangle \;&=\; \tfrac{1}{2}\int \langle \beta,\, d\alpha + * \, d\alpha\rangle \\ &= \tfrac{1}{2}\int \langle \beta,\, d\alpha\rangle + \tfrac{1}{2}\int \langle * \,\beta,\, d\alpha\rangle \\ &= \int \langle \beta, d\alpha\rangle \;=\; \int \langle d^*\beta,\, \alpha\rangle\,. \end{aligned}$$

Since this happens for all α's, it implies that $d^*\beta = 0$. Further, since β is self-dual, this implies that $d\beta = 0$. Therefore β is harmonic, and we have the orthogonal splitting $\Omega^2_+ = \mathcal{H}^2_+ \oplus \operatorname{Im} d^+$. Similarly, $\Omega^2_- = \mathcal{H}^2_- \oplus \operatorname{Im} d^-$.

Adding Ω^2_+ to Ω^2_- yields

$$\Omega^2 = \mathcal{H}^2_+(M) \oplus \mathcal{H}^2_-(M) \oplus \operatorname{Im} d^+ \oplus \operatorname{Im} d^- .$$

This compares with the earlier de Rham splitting $\Omega^2 = \mathcal{H}^2(M) \oplus \operatorname{Im} d \oplus \operatorname{Im} d^*$ through the formulae $d^+ = \frac{1}{2}(d + d^* *)$ and $d^- = \frac{1}{2}(d - d^* *)$, and therefore $\mathcal{H}^2 = \mathcal{H}^2_+ \oplus \mathcal{H}^2_-$. □

An important consequence of the splitting $\Omega^2_+ = \mathcal{H}^2_+(M) \oplus \operatorname{Im} d^+$ is that $\operatorname{Im} d^+$ is an (infinite-dimensional) subspace of Ω^2_+ of *codimension* $\dim \mathcal{H}^2_+(M) = b^+_2(M)$.

Self-dual parts of curvatures. We have seen from the classification lemmata that

$$\{F^+_A \mid A \in \mathcal{C}onn(L)\} = \{\Phi^+ \mid \Phi \in i\Omega^2,\ d\Phi = 0,\ [\Phi] = -2\pi i c_1(L)\} .$$

In what follows, we wish to better locate this space among all imaginary-valued self-dual 2–forms from $i\Omega^2_+$.

Every two connections A' and A'' on L differ by a global 1–form $\varphi \in \Omega^1$, so that $d_{A'} = d_{A''} + i\varphi$. The corresponding curvatures differ by $d\varphi$, as $F_{A'} = F_{A''} + i\,d\varphi$. Therefore

$$\{F_A \mid A \in \mathcal{C}onn(L)\} = \{F_{A^0}\} + \operatorname{Im} d ,$$

and hence

$$\{F^+_A \mid A \in \mathcal{C}onn(L)\} = \{F^+_{A^0}\} + \operatorname{Im} d^+ .$$

However, $\Omega^2_+ = \mathcal{H}^2_+(M) \oplus \operatorname{Im} d^+$, and thus the codimension of $\operatorname{Im} d^+$ inside Ω^2_+ must be $b^+_2(M)$. It follows that:

Lemma. *For any Hermitian complex-line bundle L over a Riemannian 4–manifold M, we have that*

$$\{F^+_A \mid A \in \mathcal{C}onn(L)\} \subset i\Omega^2_+$$

is an affine subspace of $i\Omega^2_+$ of codimension $b^+_2(M)$. □

Finding and avoiding anti-self-dual connections. Let us denote by

$$\mathscr{F}(L) = \{F_A \mid A \in \mathcal{C}onn(L)\}$$

the subspace of $i\Omega^2$ formed by the curvatures of all $U(1)$–connections on L. As we have seen, this is simply the set of all closed 2–forms $\Phi \in i\Omega^2$ that represent the class $-2\pi i c_1(L)$. Thus, it does not depend on the Riemannian metric on M.

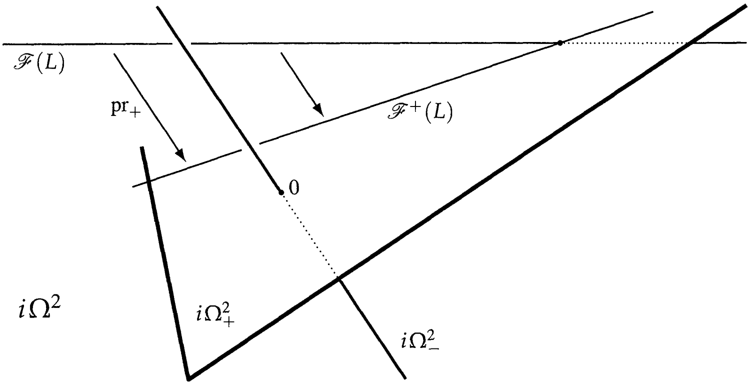
On the other hand, if we denote by

$$\mathscr{F}^+(L) = \{F^+_A \mid A \in \mathcal{C}onn(L)\}$$

the space of all self-dual parts of $U(1)$–curvatures on L, then this affine space space does depend on the choice of metric of M.

Indeed, the choice of metric induces the splitting $\Lambda^2 = \Lambda^2_+ \oplus \Lambda^2_-$, and thus determines $\mathscr{F}^+(L)$ as the projection of $\mathscr{F}(L)$ onto Ω^2_+:

$$\mathscr{F}^+(L) = \mathrm{pr}_+[\mathscr{F}(L)] \subset i\Omega^2_+ .$$



9.7. Yet another picture

We are now ready to investigate the solutions to the equation

$$F_A^+ = 0\,,$$

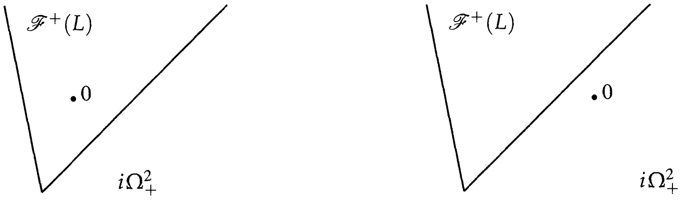
which describes the anti-self-dual connections on L.

A first remark is that, if M has $H^1(M;\mathbb{R}) = 0$, then either all the anti-self-dual connections on L are unique up to gauge-equivalence, or none exists.[5]

A second remark is that the equation $F_A^+ = 0$ has solutions if and only if the affine subspace $\mathscr{F}^+(L)$ is a linear subspace of $i\Omega^2_+$, in other words, if and only if

$$0 \in \mathscr{F}^+(L)\,,$$

where 0 denotes the constantly-zero 2–form. The actual space $\mathscr{F}^+(L)$ depends in an essential fashion on the choice of Riemannian metric on M.



anti-self-dual connections exist

anti-self-dual connections do not exist

9.8. Instanton, no instanton

A happy case. Assume that $c_1(L) \cdot c_1(L) > 0$. Then there cannot be any 2–forms α so that both $[\alpha] = c_1(L)$ and $\alpha^+ = 0$. Indeed, if such a form existed, we would have $\alpha = - * \alpha$, and hence

$$0 < c_1(L) \cdot c_1(L) = \int \alpha \wedge \alpha = \int \alpha \wedge (- * \alpha) = - \int |\alpha|^2 \leq 0\,,$$

which is impossible. Therefore:

---

[5] In general, when $H^1(M;\mathbb{R}) \neq 0$, the anti-self-dual gauge-classes are parametrized by the torus $H^1(M;\mathbb{R})/H^1(M;\mathbb{Z})$.

Lemma. *If $c_1(L) \cdot c_1(L) > 0$, then for every Riemannian metric on M there are no anti-self-dual connections on L.* □

Of course, for $c_1(L)$ to have positive self-intersection, it is first necessary that $b_2^+(M) \geq 1$.

Enters Mr Betti. For the general case, recall that $b_2^+(M)$ is the codimension in $i\Omega^2_+$ of the affine subspace $\mathscr{F}^+(L)$.

If $b_2^+(M) = 0$, then $\mathscr{F}^+(M) = i\Omega^2_+$ and in particular there *always* exist anti-self-dual connections on L, and they are all gauge-equivalent.

> ***A look back at Donaldson's theorem.*** *This explains why, in the proof of Donaldson's theorem,[6] for a simply-connected 4–manifold M with negative-definite intersection form, the count of singularities of $\mathfrak{M}_1$ was $\frac{1}{2}\{a \in H^2(M;\mathbb{Z}) \mid a \cdot a = -1\}$: Each such class a corresponds to a splitting $E = L_a \oplus L_a^*$ of the $SU(2)$–bundle E (with $c_2 E = 1$) into a sum of complex-line bundles with $c_1(L_a) = a$. Since $b_2^+(M) = 0$, the bundles L_a and L_a^* each admit an anti-self-dual connection, A_a and $-A_a$, unique up to gauge-equivalence. The two add up to a reducible anti-self-dual connection $A_a \oplus -A_a$ on E. The factor $\frac{1}{2}$ appears because of the symmetry $L_a \oplus L_a^* = L_{-a}^* \oplus L_{-a}$. Since a reducible anti-self-dual connection on E is defined to be one that splits into two connections as above, this shows that we have in fact detected all reducible 1–instantons on M.*

If $b_2^+(M) \geq 1$, then for a generic metric one expects for $\mathscr{F}^+(L)$ to miss $0 \in i\Omega^2_+$, and thus for anti-self-dual connections to be avoidable *via* perturbations of the metric. If further $b_2^+(M) \geq 2$, then one even expects to be able to avoid anti-self-dual connections on generic *paths* of metrics. Both these expectations are indeed met, and we state:

Reducible Solutions Lemma.

— *If $b_2^+(M) = 0$, then, for every Riemannian metric on M, the bundle L must admit anti-self-dual connections.*

— *If $b_2^+(M) \geq 1$, then, for a generic metric, L admits no anti-self-dual connections.*

— *If $b_2^+(M) \geq 2$, then, given any two Riemannian metrics g_0 and g_1 on M, each without anti-self-dual connections on L, they can be connected by a path of metrics g_t so that none of them admits anti-self-dual connections on L.*

Sketch of proof. Notice that the splitting

$$\Lambda^2(T_M^*) = \Lambda^2_+(T_M^*) \oplus \Lambda^2_-(T_M^*)$$

depends only on the *conformal class*[7] of the Riemannian metric.

Remarkably, the converse is also true: Consider a random splitting

$$\Lambda^2(T_M^*) = E^+ \oplus E^-$$

into a direct sum of a subbundle E^+, positive-definite for the wedge-product,[8] and a subbundle E^- that is negative-definite, and so that E^+ and E^- are

---

[6] See its outline back in section 5.3 (page 243).

[7] Two Riemannian metrics g' and g'' are called **conformally equivalent** if they measure angles in the same way; that is, if there is a positive function $f\colon M \to (0,\infty)$ so that we have $g'\langle v,w\rangle = f(x)\,g''\langle v,w\rangle$ for all $v, w \in T_M|_x$, $x \in M$.

[8] A subspace $Z \subset \Lambda^2(T_M^*)|_x$ is positive-definite for the wedge-product if, for every 2–form $\alpha \in Z$, we have that the 4–form $\alpha \wedge \alpha|_x$ orients $T_M|_x$ in the same way as the chosen orientation of M.

> wedge-orthogonal. Then there exists a Riemannian metric on M, unique up to conformal equivalence, for which $\Lambda^2_\pm(T^*_M) = E^\pm$.
>
> Therefore, the problem of finding a Riemannian metric on M for which L admits no anti-self-dual connections becomes the problem of finding a suitable splitting $\Lambda^2(T^*_M) = E^+ \oplus E^-$ so that the projection along $\Gamma(E^-)$ of $\mathscr{F}(L)$ onto $\Gamma(E^+)$ does not touch the zero-section $0 \in \Gamma(E^+)$. In fact, all we need is to find a suitable maximal positive-definite subbundle E^+ because then E^- will just be its wedge-complement. Of course, this problem needs only be solved fiberwise, in each $\Lambda^2|_x$.
>
> The full argument can be read from **C. Taubes**' ***Self-dual Yang–Mills connections on non-self-dual 4–manifolds*** [**Tau82**], or from **S.K. Donaldson** and **P. Kronheimer**'s ***The geometry of four-manifolds*** [**DK90**, ch 4]. □

Why do we care? Since anti-self-dual connections on complex-line bundles correspond to reducible solutions for both Donaldson and Seiberg–Witten theories (and thus justifies the name we gave to the previous lemma), their avoidance for generic metrics ensures that the moduli spaces are non-singular manifolds.

Further, being able to connect two metrics while avoiding reducible solutions implies that the corresponding moduli spaces are connected by cobordisms, and thus their cobordism class is independent of the Riemannian metric. This guarantees that we obtain invariants that depend only on the smooth structure on M, not on the particular choice of Riemannian metric.

Note: Donaldson theory and complex geometry

Donaldson theory is strongly related to complex geometry, and in what follows we will try to clarify this. The main point is that an anti-self-dual connection on a bundle will organize the bundle as a holomorphic bundle.

The notions outlined in what follows will also be of help for understanding the relationship between Seiberg–Witten theory and complex/symplectic geometry, which will be explained in the end-notes of the following chapter, on page 457 and on page 465.

Complex-valued forms. Let M be a complex surface. The bundles $\Lambda^k \otimes \mathbb{C}$ of complex-valued forms on M all split according to *type* into $\Lambda^{p,q}$'s, depending on the number of dz's and $d\bar{z}$'s needed. While this was already mentioned back on page 136, inside the notes at the end of chapter 3, it will be explained again here:

Forms of degree 0. We denote by $\Lambda^{0,0}$ the bundle of complex-valued 0–forms $\Lambda^0 \otimes \mathbb{C} = \Lambda^{0,0}$. Its sections are all smooth functions $f\colon M \to \mathbb{C}$. There is no splitting here, so we move on:

Splitting forms of degree 1. If we think of $\Lambda^1 \otimes \mathbb{C}$ as complex-valued 1–forms, *i.e.*, as maps $T_M \to \mathbb{C}$ that are fiberwise-linear, then, since both T_M and $\mathbb{C}$ are endowed with complex structures, it only makes sense to split $\Lambda^1 \otimes \mathbb{C}$ into $\mathbb{C}$–linear and $\mathbb{C}$–anti-linear forms:

$$\Lambda^1 \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1} \qquad \text{with} \qquad \begin{aligned} \Lambda^{1,0} &= \{\alpha\colon T_M \to \mathbb{C} \mid \alpha \text{ is } \mathbb{C}\text{–linear}\} \\ \Lambda^{0,1} &= \{\alpha\colon T_M \to \mathbb{C} \mid \alpha \text{ is } \mathbb{C}\text{–anti-linear}\} \,. \end{aligned}$$

Correspondingly, every complex-valued function $f\colon M \to \mathbb{C}$ has its exterior differential $df \in \Gamma(T^*_M \otimes \mathbb{C})$ split into a $(1,0)$–part $\partial f \in \Gamma(\Lambda^{1,0})$ and a $(0,1)$–part $\bar\partial f \in \Gamma(\Lambda^{0,1})$:

$$df = \partial f + \bar\partial f \,.$$

If $\bar\partial f = 0$, this means that f's derivative is complex-linear, and thus

$$\bar\partial f = 0 \qquad \Longleftrightarrow \qquad f \text{ holomorphic} \,.$$

In such cases, ∂f represents the complex derivative of f.

In effect we have split the (complexified) exterior differential $d\colon \Gamma(\Lambda^0 z \otimes \mathbb{C}) \to \Gamma(\Lambda^1 \otimes \mathbb{C})$ into two operators

$$d = \partial + \bar\partial \qquad \text{with} \qquad \begin{aligned} \partial&\colon \Gamma(\Lambda^{0,0}) \longrightarrow \Gamma(\Lambda^{1,0}) \\ \bar\partial&\colon \Gamma(\Lambda^{0,0}) \longrightarrow \Gamma(\Lambda^{0,1}) \,. \end{aligned}$$

The operator $\bar\partial$ is called the **Cauchy–Riemann operator** on M (but many other versions and generalizations are called the same, as we will see shortly).

Using local real coordinates $\{x_1, y_1, x_2, y_2\}$ on M, chosen such that $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are local complex coordinates on M, we can define the complex-valued 1–forms $dz_k = dx_k + i\,dy_k$ and $d\bar z_k = dx_k - i\,dy_k$. Then we have locally

$$\Lambda^{1,0} = \mathbb{C}\{dz_1, dz_2\} \qquad \text{and} \qquad \Lambda^{0,1} = \mathbb{C}\{d\bar z_1, d\bar z_2\} \,.$$

Splitting higher forms. The splitting $\Lambda^1 \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1}$ further leads to a splitting of all complex-valued forms into (p,q)–types. Specifically, we can define $\Lambda^{p,q}$ as made of all complex-valued forms that can be written locally by using exactly p of the dz_k's and q of the $d\bar z_k$'s. Then we have

$$\Lambda^k \otimes \mathbb{C} \;=\; \Lambda^{k,0} \oplus \Lambda^{k-1,1} \oplus \cdots \oplus \Lambda^{1,k-1} \oplus \Lambda^{0,k} \,,$$

by summing up all $\Lambda^{p,q}$'s with $p + q = k$.

For example, complex-valued 2–forms split as

$$\Lambda^2 \otimes \mathbb{C} \;=\; \Lambda^{2,0} \oplus \Lambda^{1,1} \oplus \Lambda^{0,2} \,.$$

These terms can be described coordinate-freely by thinking of any $\alpha \in \Gamma(\Lambda^2 \otimes \mathbb{C})$ as a bilinear skew-symmetric map $\alpha\colon T_M \times T_M \to \mathbb{C}$. Then we have:

$$\begin{aligned} \Lambda^{2,0} &= \{\alpha\colon T_M \times T_M \to \mathbb{C} \mid \alpha \text{ is } \mathbb{C}\text{–bilinear}\} \,, \\ \Lambda^{1,1} &= \{\alpha\colon T_M \times T_M \to \mathbb{C} \mid \alpha(iv,\, iw) = \alpha(v,w)\} \,, \\ \Lambda^{0,2} &= \{\alpha\colon T_M \times T_M \to \mathbb{C} \mid \alpha \text{ is } \mathbb{C}\text{–bi-anti-linear}\} \,. \end{aligned}$$

The $(0,2)$–forms are conjugate to the $(2,0)$–forms: $\Lambda^{0,2} = \overline{\Lambda^{2,0}}$, while the bundle of $(1,1)$–forms is self-conjugate.

If M is a Kähler surface, with Kähler form $\omega(v,w) = \langle iv,\, w\rangle_{\mathbb{R}}$, then always

$$\omega \in \Gamma(\Lambda^{1,1}) \,.$$

Splitting differentials. The higher-order differentials $d\colon \Gamma(\Lambda^k \otimes \mathbb{C}) \to \Gamma(\Lambda^{k+1} \otimes \mathbb{C})$ split accordingly:

$$d|_{(p,q)} = \partial + \bar\partial \qquad \text{with} \qquad \begin{aligned} &\partial\colon \Gamma\big(\Lambda^{p,q}\big) \longrightarrow \Gamma\big(\Lambda^{p+1,q}\big) \\ &\bar\partial\colon \Gamma\big(\Lambda^{p,q}\big) \longrightarrow \Gamma\big(\Lambda^{p,q+1}\big)\,, \end{aligned}$$

where $d|_{(p,q)}$ denotes here the restriction of d to sections of $\Lambda^{p,q} \subset \Lambda^{p+q} \otimes \mathbb{C}$. These operators $\bar\partial\colon \Gamma\big(\Lambda^{p,q}\big) \to \Gamma\big(\Lambda^{p,q+1}\big)$ can be defined directly by recursion through the formula

$$\bar\partial(\alpha \wedge \beta) = \big(\bar\partial\alpha\big) \wedge \beta \pm \alpha \wedge \big(\bar\partial\beta\big)\,,$$

which allows one to backtrack to the basic operator $\bar\partial\colon \Gamma\big(\Lambda^{0,0}\big) \to \Gamma\big(\Lambda^{0,1}\big)$ discussed earlier. Similarly for the ∂'s. Of particular importance for us will be the operator $\bar\partial\colon \Gamma\big(\Lambda^{0,1}\big) \to \Gamma\big(\Lambda^{0,2}\big)$.

Finally, notice that, just as we had $d\,d = 0$, we also have

$$\bar\partial\,\bar\partial = 0\,.$$

(This leads to the *Dolbeault cohomology* $H^{p,q}(M)$ of M, but not for us.)

Almost-complex case. *Note that the splitting* $\Lambda^1 \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1}$ *can be defined in the same way on any manifold* M *that is merely almost-complex.*[9] *In that case, since there are no complex-coordinates on* M, *the* $\Lambda^{p,q}$*'s should not be described using* dz_k*'s and* $d\bar z_k$. *Further, while* $d\colon \Gamma(\Lambda^0 \otimes \mathbb{C}) \longrightarrow \Gamma(\Lambda^1 \otimes \mathbb{C})$ *does split into* $\partial + \bar\partial$, *the higher-order exterior differentials are no longer exhausted by the corresponding sum* $\partial + \bar\partial$ *(there are extra terms involving the Nijenhuis tensor* $\mathcal{N}$, *which measures the failure of the almost-complex structure from being induced by holomorphic coordinates on* M*), and hence* $\bar\partial\,\bar\partial \neq 0$. *In fact, we have* $\bar\partial\,\bar\partial = 0$ *if and only if the almost-complex structure is integrable, i.e., if it corresponds to a complex structure on* M. *This latter statement is the celebrated Newlander–Nirenberg theorem. We will make use of the almost-complex case when discussing the Seiberg–Witten equations on symplectic 4-manifolds later, in the the end-notes of the next chapter (page 465).*

Relations with self-dual/anti-self-dual 2–forms. Assume that M is Kähler, endowed with a Riemannian metric both with $\langle iv, iw\rangle = \langle v, w\rangle$ and so that the 2–form $\omega(v,w) = \langle iv, w\rangle$ is closed. From the metric, on M appears another fundamental split of 2–forms, specifically into self-dual and anti-self-dual 2–forms

$$\Lambda^2 = \Lambda^2_+ \oplus \Lambda^2_-\,.$$

The two splits are related through:

Lemma. *On every Kähler surface M with Kähler form ω, we have:*

$$\begin{aligned} \Lambda^2_+ \otimes \mathbb{C} &= \Lambda^{2,0} \oplus \Lambda^{0,2} \oplus \mathbb{C}\,\omega\,, \\ \Lambda^2_- \otimes \mathbb{C} &= \Lambda^{1,1} \cap \omega^\perp\,. \end{aligned}$$

---

[9] Almost-complex structures will be properly discussed later, in section 10.1 (page 376).

Proof. We compute directly, in local coordinates:

$$\begin{aligned}
dz_1 \wedge dz_2 &= (dx_1 \wedge dx_2 - dy_1 \wedge dy_2) + i(dx_1 \wedge dy_2 - dx_2 \wedge dy_1) \\
dz_1 \wedge d\bar{z}_1 &= -2i\, dx_1 \wedge dy_1 \\
dz_1 \wedge d\bar{z}_2 &= (dx_1 \wedge dx_2 + dy_1 \wedge dy_2) - i\,(dx_1 \wedge dy_2 + dx_2 \wedge dy_1) \\
dz_2 \wedge d\bar{z}_1 &= (dx_1 \wedge dx_2 + dy_1 \wedge dy_2) + i\,(dx_1 \wedge dy_2 + dx_2 \wedge dy_1) \\
dz_2 \wedge d\bar{z}_2 &= -2i\, dx_2 \wedge dy_2 \\
d\bar{z}_1 \wedge d\bar{z}_2 &= (dx_1 \wedge dx_2 - dy_1 \wedge dy_2) - i(dx_1 \wedge dy_2 - dx_2 \wedge dy_1)
\end{aligned}$$

and, after noticing that $\omega = \frac{i}{2}(dz_1 \wedge d\bar{z}_1 + dz_2 \wedge d\bar{z}_2)$, the result follows. □

It is worth comparing this lemma with the Hodge signature theorem.[10]

Also notice that this result still holds for merely almost-complex manifolds, with ω still defined by $\omega(x,y) = \langle Jx, y\rangle$. One would not use dz_k's in its proof, but some proxy fiber-coordinates in $T^*_M \otimes \mathbb{C}$.

Holomorphic bundles. A **holomorphic bundle** E over a complex surface M is a bundle defined by a cocycle $\{g_{\alpha\beta}\}$ of transition functions $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to GL_{\mathbb{C}}(n)$ that are required to be *holomorphic.* For every holomorphic bundle E we can define the operator[11]

$$\bar{\partial}^E : \Gamma(E) \longrightarrow \Gamma(E \otimes \Lambda^{0,1})$$

through the two properties

— $\bar{\partial}^E(f \cdot \sigma) = \sigma \otimes (\bar{\partial} f) + f \cdot \bar{\partial}^E \sigma$, for every $f\colon M \to \mathbb{C}$ and $\ell \in \Gamma(E)$;

— $\bar{\partial}^E \sigma|_U = 0$ if and only if σ is a holomorphic section of $E|_U$.

This operator is well-defined precisely owing to the fact that the $g_{\alpha\beta}$'s are all holomorphic.

The same formulae (but allowing for complex-valued forms in place of the f's above) can be used to extend $\bar{\partial}$ to general E–valued exterior forms, as operators

$$\bar{\partial}^E : \Gamma(E \otimes \Lambda^{p,q}) \longrightarrow \Gamma(E \otimes \Lambda^{p,q+1}) \,,$$

and we always have

$$\bar{\partial}^E\, \bar{\partial}^E = 0\,.$$

Connections and holomorphic structures. For a *smooth* complex bundle $E \to M$ over a complex surface M, a complex connection A on E is a $\mathbb{C}$–linear operator[12]

$$d_A : \Gamma(E) \longrightarrow \Gamma(E \otimes T^*_M)\,.$$

The splitting $T^*_M \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1}$ induces a corresponding splitting of the connection as[13]

$$d_A = \partial_A + \bar{\partial}_A \qquad \text{with} \qquad \begin{aligned} \partial_A &\colon \Gamma(E) \longrightarrow \Gamma(E \otimes \Lambda^{1,0}) \\ \bar{\partial}_A &\colon \Gamma(E) \longrightarrow \Gamma(E \otimes \Lambda^{0,1})\,. \end{aligned}$$

---

[10] The Hodge signature theorem was stated back in section 6.2 (page 278).

[11] Here, $E \otimes \Lambda^{0,1} = E \otimes_{\mathbb{C}} \Lambda^{0,1}$.

[12] Here, $E \otimes T^*_M = E \otimes_{\mathbb{R}} T^*_M = E \otimes_{\mathbb{C}} (T^*_M \otimes_{\mathbb{R}} \mathbb{C})$.

[13] Here, $E \otimes \Lambda^{0,1} = E \otimes_{\mathbb{C}} \Lambda^{0,1}$.

If E were a holomorphic bundle, then the operator $\bar{\partial}^E$ would distinguish its special sections: the holomorphic sections. On our merely smooth complex bundle, the partial connection $\bar{\partial}_A$ also distinguishes certain sections of the smooth bundle E. This analogy can be taken further. In fact, a structure of holomorphic bundle on E is equivalent to a suitable choice of connection:

Integrability Theorem. *A connection A on a smooth complex bundle E over a complex surface M defines a holomorphic structure on E if and only if*

$$\bar{\partial}_A \bar{\partial}_A = 0 ,$$

and in this case the holomorphic structure of E has $\bar{\partial}^E = \partial_A$. □

The operators $\bar{\partial}$, $\bar{\partial}^E$ and $\bar{\partial}_A$ are all called **Cauchy–Riemann operators.**

The converse of this integrability theorem holds true as well:

If E is a holomorphic bundle, then there exist connections A on E such that $\bar{\partial}_A = \bar{\partial}^E$.

Such connections A are called *compatible* with the holomorphic structure of E. This remains true if we endow E with a Hermitian fiber-metric and require A to be a $U(n)$–connection.

Curvatures. Since the composite $\bar{\partial}_A \bar{\partial}_A$ is part of $d_A d_A = F_A$, one should expect that the integrability condition above be translatable as a condition on the curvature of A. Specifically, if we think of the curvature F_A as a section of $\mathrm{End}(E) \otimes_{\mathbb{R}} \Lambda^2 = \mathrm{End}(E) \otimes_{\mathbb{C}} (\Lambda^2 \otimes \mathbb{C})$, and since

$$\bar{\partial}_A \bar{\partial}_A : \Gamma(E) \longrightarrow \Gamma(E \otimes \Lambda^{0,2}) ,$$

then it follows that what we need is that the projection of F_A on $E \otimes \Lambda^{0,2}$ be trivial.

Corollary. *A connection A on the smooth complex bundle E over a complex surface M defines a holomorphic structure on E if and only if*

$$F_A^{0,2} = 0 .$$ □

Choosing a Hermitian fiber-metric on E and restricting to $U(n)$–connections on E preserves the truth of this statement.

Furthermore, since for a $U(n)$–connection A its curvature F_A must act by elements of $\mathfrak{u}(n)$, in that case we have $F_A^{2,0} = -\left(F_A^{0,2}\right)^*$. This implies:

Lemma. *For a $U(n)$–connection A on E to define a holomorphic structure on E, its curvature F_A must be a section of* $\mathrm{End}(E) \otimes \Lambda^{1,1}$. □

In other words, F_A must be a $(1,1)$–form.

Anti-self-duality and holomorphy. Assume again that M is a Kähler surface, with Kähler form ω. By comparing the splittings

$$\Lambda^2 \otimes \mathbb{C} = \Lambda^{2,0} \oplus \Lambda^{1,1} \oplus \Lambda^{0,2} \qquad \text{and} \qquad \Lambda^2 = \Lambda^2_+ \oplus \Lambda^2_- ,$$

as applied to the curvature 2–form F_A, we notice that

$$F_A^+ = F_A^{2,0} + (F_A^+)^{1,1} + F_A^{0,2} \qquad \text{and} \qquad F_A^- = (F_A^-)^{1,1} .$$

Therefore, the vanishing of F_A^+ would ensure that the curvature is a $(1,1)$–form.

Corollary. *Let E be a smooth Hermitian complex bundle over a Kähler surface. Then any anti-self-dual $U(n)$–connection A on E defines a holomorphic structure on E.* □

Since all self-dual $(1,1)$–forms are spanned by ω, it follows that

$$(F_A^+)^{1,1} = \tfrac{1}{2}\langle F_A, \omega\rangle \cdot \omega \, ,$$

and therefore we also have the converse:

Corollary. *Let E be a holomorphic Hermitian bundle E over a Kähler surface. Then a $U(n)$–connection A, compatible with the holomorphic structure of E, is anti-self-dual if and only if*

$$F_A \perp \omega \, . \qquad \square$$

Conclusion. These results show the very strong relation between complex geometry and Donaldson theory. Indeed, Donaldson invariants have been especially successful in exploring complex surfaces.

The alliance of gauge theory with complex geometry remains true for Seiberg–Witten theory and will be discussed later.[14] Remarkably, in the case of Seiberg–Witten theory, the alliance can even be stretched into the almost-complex domain, at least as long as we remain under the spell of a symplectic structure.[15]

References. The results of this note are due to **M. Atiyah, N. Hitchin** and **I. Singer** in their classic ***Self-duality in four-dimensional Riemannian geometry*** [AHS78], and can also be found explained in **S.K. Donaldson** and **P. Kronheimer**'s ***The geometry of four-manifolds*** [DK90]. The integrability theorem can be proved as a consequence of the Newlander–Nirenberg theorem (on the integrability of almost-complex structures) or can be proved directly, under the assumption of real-analyticity, by complexifying and using Frobenius's theorem.

For the Newlander–Nirenberg theorem, refer to the original paper ***Complex analytic coordinates in almost complex manifolds*** of **A. Newlander** and **J. Nirenberg** [NN57], or, in book-format, see **L. Hörmander**'s ***An introduction to complex analysis in several variables*** [Hör66, Hör90] or **G. Folland** and **J. Kohn**'s ***The Neumann problem for the Cauchy–Riemann complex*** [FK72]. A quick proof of the Newlander–Nirenberg theorem under the simplifying assumption that everything is real-analytic can be read from **S. Kobayashi** and **K. Nomizu**'s classic ***Foundations of differential geometry*** [KN69, KN96, vol. II, app. 8] .

Note: Equivalence between Donaldson and Seiberg–Witten

In what follows, we will very briefly outline the relation between the Donaldson and Seiberg–Witten theories, and their conjectured, and almost-fully proved, equivalence.

---

[14] See the argument in the end-notes of the next chapter (Seiberg–Witten on Kähler, page 457).

[15] See section 10.5 (page 409), and the argument in the end-notes of the next chapter (page 465).

The Donaldson side. While pursuing their work on the minimum genus problem, **P. Kronheimer** and **T. Mrowka** in their ***Recurrence relations and asymptotics for four-manifold invariants*** **[KM94b]** uncovered a deep symmetry of the Donaldson polynomial.

They called a 4–manifold M of *(Donaldson)* **simple type** if the polynomials satisfy

$$\mathfrak{q}_{p,r+2} = 4\,\mathfrak{q}_{p,r}\;.$$

A large class of 4–manifolds are of simple type, and it is conjectured that all simply-connected ones might be so.

In any case, if M is of simple type, then all the information of the Donaldson invariants is enclosed in $\mathfrak{q}_{p,0}$ and $\mathfrak{q}_{p,1}$. One can then define $\widetilde{\mathfrak{q}}_p = \mathfrak{q}_{p,0}$ if $p = b_2^+ + 1 \pmod 2$ and $\widetilde{\mathfrak{q}}_p = \frac{1}{2}\,\mathfrak{q}_{p,1}$ if $p = b_2^+ \pmod 2$ and build the **Donaldson** *(formal)* **power series:**

$$\mathfrak{D}_M\colon H_2(M;\mathbb{R}) \longrightarrow \mathbb{R} \qquad \mathfrak{D}_M(\alpha) = \sum \tfrac{1}{p!}\,\widetilde{\mathfrak{q}}_p(\alpha)\;.$$

Kronheimer and Mrowka showed that the whole information of the Donaldson polynomials is actually caught in a few coefficients and homology classes:

Kronheimer–Mrowka Structure Theorem. *If M is of Donaldson simple type and $b_2^+(M) \geq 2$, then the Donaldson series can be written*

$$\mathfrak{D}_M(\alpha) = e^{\frac{1}{2}\alpha\cdot\alpha} \sum a_i \cdot e^{\kappa_i\cdot\alpha}$$

for some rational coefficients $a_i \in \mathbb{Q}$, and finitely-many classes $\kappa_i \in H^2(M;\mathbb{Z})$. □

The classes κ_i are called *(Donaldson)* **basic classes.** Their remarkable relevance to smooth topology became apparent through their role in the adjunction inequality

$$\chi(S) + S\cdot S \leq \kappa_i\cdot S\,,$$

which holds true for all homologically-nontrivial embedded connected surfaces S with $S\cdot S \geq 0$ and offers lower bounds on the genus needed to represent a given homology class by an embedded surface. (The Seiberg–Witten counterpart of this result will be discussed in chapter 11.)

The equivalence. N. Seiberg and E. Witten created their invariant as a physically-equivalent method for evaluating the Donaldson invariants. The equivalence of the two theories is summed-up into

Witten's Conjecture (*almost proved*). *If M is of Donaldson simple type and has $b_2^+(M) \geq 2$, then*

$$\mathfrak{D}_M(\alpha) = 2^{\nu(M)}\, e^{\frac{1}{2}\alpha\cdot\alpha} \sum e^{\kappa\cdot\alpha}\, \mathcal{SW}_M(\kappa)\,,$$

where $\nu(M) = 2 + \frac{7}{4}\chi(M) + \frac{11}{4}\operatorname{sign} M$, the summation is over all $\kappa \in H^2(M;\mathbb{Z})$ with $\kappa\cdot\kappa = 2\chi(M) + 3\operatorname{sign} M$, and $\mathcal{SW}_M(\kappa) \in \mathbb{Z}$ denotes the value of the Seiberg–Witten invariant on κ.

In particular, this implies that all the Donaldson basic classes must satisfy

$$\kappa_i\cdot\kappa_i = 2\chi(M) + 3\operatorname{sign} M\,,$$

while the coefficients a_i are merely $a_i = 2^{\nu(M)}\,\mathcal{SW}_M(\kappa_i)$.

References. The mathematical program to prove this conjecture was outlined in **V. Pidstrigach** and **A. Tyurin**'s ***Localisation of the Donaldson's invariants along Seiberg–Witten classes*** **[PT95]**, and most of it has been carried through in a still-growing series of papers by **P. Feehan** and **T. Leness**, with the strategy outlined in ***$PU(2)$ monopoles and relations between four-manifold invariants*** **[FL98a]**, then developed in **[FL98b, FL01b, FL01a, FL02b, FL02a]**, and recently surveyed in *On Donaldson and Seiberg–Witten invariants* [FL03].

Even this partial proof of the Witten conjecture has yielded fruits, see for example **P. Kronheimer** and **T. Mrowka**'s *Witten's conjecture and property P* [KM04] where they prove the longstanding property–P conjecture for knots.

Seiberg–Witten theory will be explained in the next chapter, and then the Seiberg–Witten side of the above conjecture will become more clear.

Bibliography

For differential geometry, the classic reference is **S. Kobayashi** and **K. Nomizu**'s ***Foundations of differential geometry*** **[KN69, KN96]** . Among the many references for Riemannian geometry, one cannot help but mention **A. Besse**'s splendid ***Einstein manifolds*** **[Bes87]**, which, even though focused on Einstein manifolds, covers quite a lot of geometric topics. For a recent treatment, we also recommend **J. Jost**'s ***Riemannian geometry and geometric analysis*** **[Jos95]**, which in its third edition **[Jos02]** also discusses Floer homology and mentions Seiberg–Witten.

An old collection on 4–dimensional Riemannian geometry is the collection ***Géométrie riemannienne en dimension* 4** **[BBB81]**. Other references on Riemannian geometry are **M. do Carmo**'s friendly ***Riemannian geometry*** **[dC92]**, and **I. Chavel**'s ***Riemannian geometry—a modern introduction*** **[Cha93]**, as well as **P. Petersen**'s coordinate-oriented ***Riemannian geometry*** **[Pet98]**, and **T. Sakai**'s dense ***Riemannian geometry*** **[Sak96]**. A quick survey is **P. Petersen**'s *Aspects of global Riemannian geometry* **[Pet99]**, while a beautiful comprehensive one is **M. Berger**'s ***Riemannian geometry during the second half of the twentieth century***[16] **[Ber98, Ber00]**.

In 1982, just one year after M. Freedman's 1981 classification of topological 4–manifolds, instantons and gauge theory entered 4–dimensional topology with a bang, through **S.K. Donaldson**'s thesis *An application of gauge theory to four-dimensional topology* [Don83]. The latter contained the result discussed earlier in section 5.3 (page 243), about definite intersection forms. See also the references back on page 266, at the end of chapter 5.

The polynomial invariants were defined in **S.K. Donaldson**'s *Polynomial invariants for smooth four-manifolds* [Don90]. Many more mathematicians joined the instanton party, among them J. Morgan, P. Kronheimer, T. Mrowka, R. Fintushel, R. Stern...

A comprehensive reference for Donaldson theory is **S.K. Donaldson** and **P. Kronheimer**'s ***The geometry of four-manifolds*** **[DK90]**. For a focus on applying gauge

---

[16] Note that **M. Berger**'s more recent volume ***A panoramic view of Riemannian geometry*** **[Ber03]** is *not* an expansion of the text mentioned above.

theory to complex surface, read **R. Friedman** and **J. Morgan**'s ***Smooth four-manifolds and complex surfaces*** [FM94a].

In 1994, Seiberg–Witten theory entered the scene. Since this theory was similar in spirit but much easier to manipulate, the Donaldson theory experts were able to use their bag of tricks on the Seiberg–Witten invariants, and they very quickly obtained a series of astonishing results, which will be discussed in the next chapters. Nonetheless, after that explosion the rhythm began to slow down, and the front-lines moved rather toward 3–manifolds, with Seiberg–Witten–Floer homology theories, *etc.*

Chapter 10

The Seiberg–Witten Invariants

We introduce the Seiberg–Witten invariants. As background, we start by looking at almost-complex structures on 4–manifolds, and their stronger relatives, symplectic structures. Then, in section 10.2 (page 382), we present the common generalization of both almost-complex structures and spin structures, specifically $\mathrm{spin}^{\mathbb{C}}$ structures. On top of a choice of $\mathrm{spin}^{\mathbb{C}}$ structure, we then write in section 10.3 (page 396) the Seiberg–Witten equations and explain how one extracts a 4–manifold invariant from them. Section 10.4 (page 404) contains the main properties that govern the behavior of these invariants. After that, in section 10.5 (page 409) we detail the remarkable behaviour of the Seiberg–Witten invariants on symplectic 4–manifolds, then in section 10.6 (page 412) we make a few comments on the complex case. Indeed, gauge theory is most useful in the close-to-complex realm.

The main text of the chapter will detail only the Seiberg–Witten theory arguments that follow from applying the Lichnerowicz formula (stated on page 393); it is remarkable how many results follow from this invaluable formula. Proofs involving other techniques are exiled to the end-notes; for the contents of the end-notes, we refer to their introduction on page 415.

Note that a truly hasty reader can skip many technical details and just faithfully jump ahead to section 10.4 (page 404). In the next chapter (starting on page 481) we will explore the use of Seiberg–Witten theory in studying the problem of the minimum genus needed for representing a homology class by an embedded surface.

10.1. Almost-complex structures

If M is a complex surface, then T_M is a complex-plane bundle. For a general 4–manifold M, if T_M can be organized as a complex bundle,[1] we say that M was endowed with an **almost-complex structure.**

An almost-complex structure is called **integrable** if it corresponds to a complex structure on the manifold itself. Of course, most almost-complex structures are not integrable.[2] In fact, organizing T_M as a complex bundle is a simple problem in homology, while covering M with holomorphic coordinates is a most delicate and rigid problem.

The usual way to represent an almost-complex structure is to define in each fiber of T_M a multiplication by the complex-scalar i. This is achieved by specifying a fiber-preserving automorphism J of the tangent bundle that is an anti-involution:

$$J\colon T_M \xrightarrow{\approx} T_M \qquad\qquad J(J(v)) = -v\ .$$

Given such a J, the bundle T_M becomes a complex bundle through the complex-scalar action $(a+ib)\cdot w = aw + bJ(w)$. Intuitively, it is best to think of J as a field of prescribed $\pi/2$–rotations.

Being a complex bundle, (T_M, J) has Chern classes. While $c_2(T_M)$ is just the Euler class $e(T_M) = \chi(M)$, the first Chern class is quite important and is denoted by

$$c_1(J) = c_1(T_M, J)\ .$$

It is always a characteristic element of M, and finding a good candidate for $c_1(J)$ is all that is needed to ensure the existence of a corresponding J, as we will see shortly.

Mimicking complex geometry, we can define the **canonical bundle** $K_J = \det_{\mathbb{C}} T^*_M$. More useful in this chapter will be its dual, the **anti-canonical bundle**

$$K^*_J = \det_{\mathbb{C}}(T_M, J)\ .$$

This is a complex-line bundle of Chern class $c_1(K^*_J) = c_1(T_M, J)$. Often, we will denote the class $c_1(K^*_J)$ directly by K^*_J or even by K^*.

J–holomorphic curves

Given an almost-complex structure J on M, a surface S embedded in M is called a **J–holomorphic curve** (or **pseudo-holomorphic curve**) if its tangent bundle is J–invariant:

$$J[T_S] = T_S\ .$$

---

[1] Thinking of M as already oriented, we always require that the complex structure of T_M induce the chosen orientation of M.

[2] Whether an almost-complex structure is integrable is governed by the Newlander–Nirenberg theorem, stated ahead in footnote 42 on page 465.

J–holomorphic curves try to play the role that complex curves played for complex surfaces, but they are truly successful only on symplectic manifolds, where one can control their areas.

Given a J–holomorphic curve, we can always embed its normal bundle $N_{S/M}$ in $T_M|_S$ in such a way that $N_{S/M}$ is itself J–invariant. In this case both T_S and $N_{S/M}$ appear as complex-line bundles, and the proof of the complex adjunction formula transcribes verbatim from page 281, and we get:

Adjunction Formula. *If M is an almost-complex 4–manifold and S is a J–holomorphic curve in M, then we have:*

$$\chi(S) + S \cdot S = K^* \cdot S ,$$

where K^* denotes the Chern class $c_1(J)$. □

Thus, the genus of a J–holomorphic curve is completely determined by its homology class.

> *Be forewarned that later in this chapter we will also encounter an adjunction* inequality, *which applies to random embedded surfaces in 4–manifolds. To help avoid confusion, we will often call the above statement the "complex adjunction formula", and the latter the "Seiberg–Witten adjunction inequality".*

Finally, notice that the embedding of a J–holomorphic curve S in M endows S itself with an almost-complex structure $J|_{T_S}$, which is always integrable and thus organizes S as a complex curve. Often one thinks of a J–holomorphic curve as a map $f\colon S \subset M$, and quite often one allows f to be merely an immersion or have singularities, just like complex curves.

Existence of almost-complex structures

Almost-complex structures are very flexible, and their existence is a mere problem in homology:

Existence Theorem. *Assume M admits an almost-complex structure J. Then the Chern class of J must satisfy*

$$c_1(J) \cdot c_1(J) = 3 \operatorname{sign} M + 2\chi(M) ,$$

and $c_1(J)$ must be a characteristic element of M, i.e., an integral lift of $w_2(T_M)$. Furthermore, $b_2^+(M) + b_1(M)$ must be odd.

Conversely, if there exists an integral lift $\underline{w}$ of $w_2(T_M)$ so that

$$\underline{w} \cdot \underline{w} = 3 \operatorname{sign} M + 2\chi(M) ,$$

then M admits an almost-complex structure J with Chern class $c_1(J) = \underline{w}$. Moreover, assuming either that M is simply-connected or has indefinite intersection form, such a class $\underline{w}$ does exist whenever $b_2^+(M) + b_1(M)$ is odd. □

The proof is detailed in the end-notes of this chapter (page 420).

Even without satisfying the condition $\underline{w} \cdot \underline{w} = 3 \operatorname{sign} M + 2\chi(M)$, from any characteristic $\underline{w}$ one can nonetheless obtain a partial almost-complex structure $J|_3$, well-defined over the 3–skeleton of M (or, if you prefer, over $M \setminus \{point\}$) and having the Chern class[3] $c_1(J|_3) = \underline{w}$. As we will see in the next section, such an almost-complex structure on the 3–skeleton is nothing but a *spin*$^{\mathbb{C}}$ *structure* on the manifold M.

Almost-complex structures, Riemannian metrics, 2–forms

In what follows we will see that almost-complex structures, exterior 2–forms, and Riemannian metrics, when suitably compatible, in fact determine each other.

From almost-complex structures to 2–forms. A Riemannian metric on M is said to be **compatible** with the almost-complex structure J (or is called a **Hermitian metric** for J) if J acts by isometries,

$$\langle Jx, Jy\rangle = \langle x, y\rangle .$$

In particular, J becomes skew-symmetric, $\langle Jx, y\rangle = -\langle x, Jy\rangle$. A manifold endowed with a compatible pair of an almost-complex structure and a Riemannian metric is called **almost-Hermitian manifold.**[4] Such compatible metrics always exist.

The presence of a metric on M establishes an isomorphism $T_M \approx T_M^*$. This isomorphism transports the canonical identity[5] $\operatorname{Hom}_{\mathbb{R}}(T_M, T_M) = T_M^* \otimes T_M$ to an isomorphism $\operatorname{Hom}(T_M, T_M) \approx T_M^* \otimes T_M^*$. Specifically, an endomorphism $f \colon T_M \to T_M$ is identified with the bilinear form $\alpha_f(x, y) = \langle f(x), y\rangle$. Further, under the identification $\operatorname{Hom}(T_M, T_M) \approx T_M^* \otimes T_M^*$ the skew-symmetric endomorphisms of T_M correspond precisely to the elements[6] of $\Lambda^2(T_M^*) \subset T_M^* \otimes T_M^*$.

Applying this to the almost-complex structure $J \colon T_M \to T_M$ itself, we obtain the exterior 2–form

$$\omega(x, y) = \langle Jx, y\rangle ,$$

---

[3] Since the Chern class is an element of $H^2(M; \mathbb{Z})$, it is determined only by interactions inside the 3–skeleton of M, and thus $c_1(J|_3)$ is a well-defined element of $H^2(M; \mathbb{Z})$ even though $J|_3$ is not globally defined on M. In other words, take $c_1(J|_3) \in H^2(M|_3; \mathbb{Z})$ then pass it through the isomorphism $H^2(M|_3; \mathbb{Z}) \approx H^2(M; \mathbb{Z})$.

[4] If J were integrable, it would be called a **Hermitian manifold.** Kähler surfaces are particular instances of Hermitian manifolds.

[5] Given two finite-dimensional vector spaces V and W, the canonical isomorphism $\operatorname{Hom}(V, W) = V^* \otimes W$ is established by sending each linear map $f \colon V \to W$ to the vector $\sum e_k^* \otimes f(e_k)$ in $V^* \otimes W$, for some random basis $\{e_k\}$ of V.

[6] This can also be viewed as a consequence of the canonical isomorphism $\mathfrak{so}(4) \approx \Lambda^2(\mathbb{R}^4)$.

which is called the **fundamental 2–form** of J.

On the other hand, the presence of the Riemannian metric splits Λ^2 into self-dual/anti-self-dual parts. Since J is assumed to respect the orientation of M, the fundamental 2–form must always be a self-dual 2–form:

$$\omega \in \Gamma(\Lambda^2_+(T^*_M)) \ .$$

Indeed, at every point x of M, and for every choice of orthonormal basis $\{e_1, e_2, e_3, e_4\}$ of $T_M|_x$ so that $Je_1 = e_2$ and $Je_3 = e_4$, we have

$$\omega|_x = e^1 \wedge e^2 + e^3 \wedge e^4 \ ,$$

where $e^k = \langle \cdot , e_k \rangle$ make up the dual frame of T^*_M. The fundamental form ω has constant length $\sqrt{2}$.

The bundle Λ^2_+ splits as

$$\Lambda^2_+(T^*_M) = \mathbb{R}\,\omega \oplus K^* \ ,$$

where K^* is identified with the anti-canonical bundle of J.

From 2–forms to almost-complex structures. Conversely, consider a fixed Riemannian metric on some 4–manifold. Then each nowhere-zero section of Λ^2_+ determines an almost-complex structure on M. Specifically, after rescaling such a section to a 2–form ω of constant length $\sqrt{2}$, we define the corresponding almost-complex structure J by the equation

$$\langle Jx, y \rangle = \omega(x, y) \ .$$

Then ω becomes the fundamental form of J, and the anti-canonical bundle K^* of J appears as the complement of ω in Λ^2_+.

Therefore, in the presence of a metric, compatible almost-complex structures on M correspond perfectly to sections in the sphere-bundle

$$\mathbb{S}\Lambda^2_+(T^*_M) \ ,$$

made from spheres of radius $\sqrt{2}$. This creature is known as the **twistor bundle**[7] of M. Its $\mathbb{S}^2$–fiber can also be thought of as the quotient $SO(4)/U(2)$.

In light of all this, it follows that a good way to visualize an almost-complex structure J is as a global *field of rotations* of angle $\pi/2$ in the fibers of T_M.

---

[7] See **M. Atiyah, N. Hitchin** and **I. Singer**'s ***Self-duality in four-dimensional Riemannian geometry*** [**AHS78**] for a study of this bundle. Its total space itself admits an almost-complex structure, obtained by lifting the almost-complex structure of M and combining it with the natural complex structure of the fibers $\mathbb{S}^2 = \mathbb{CP}^1$. This almost-complex structure on the twistor space is sometimes integrable, and then allows to translate the smooth geometry of M into the complex geometry of the twistor 3–fold.

From almost-complex structures and 2–forms to metrics. In the absence of any metric, we say that a random 2–form $\omega \in \Gamma\big(\Lambda^2(T^*_M)\big)$ is **compatible** with a given almost-complex structure J if we have

$$\omega(Jx, Jy) = \omega(x, y) \ .$$

We say that ω **tames** J if ω is positive on any complex-plane of J:

$$\omega(x, Jx) > 0 \ .$$

If both above conditions are met, then ω and J define a Riemannian metric on M through the formula

$$\langle x, y \rangle = \omega(x, Jy) \ .$$

This metric is compatible with J, and ω becomes J's fundamental 2–form (in particular, ω becomes self-dual).

In conclusion, any two of J, ω and $\langle \cdot \, , \cdot \rangle$, when suitably compatible, determine uniquely the third.

Symplectic manifolds

Symplectic, between almost-complex and Kähler. Consider a 4–manifold M endowed with an almost-complex structure J, a compatible Riemannian metric, and the corresponding fundamental 2–form ω.

Almost-complex case. We already mentioned that, at every point x of M and for every choice of orthonormal basis $\{e_1, e_2, e_3, e_4\}$ of $T_M|_x$ so that $Je_1 = e_2$ and $Je_3 = e_4$, we have

$$\omega|_x = e^1 \wedge e^2 + e^3 \wedge e^4 \ .$$

The above is a pointwise equality. While it certainly can be extended locally by using some local frame field $\{E_1, E_2, E_3, E_4\}$ with $JE_1 = E_2$ and $JE_3 = E_4$ and $\omega = E^1 \wedge E^2 + E^3 \wedge E^4$, this frame field will usually not correspond to any coordinates on the manifold M itself. The closer we get to making such an extension correspond to coordinates on M, the more special the triplet $\big(\omega,\ J,\ \langle \cdot \, , \cdot \rangle\big)$ must be.

Symplectic case. If the basis $\{e^1, e^2, e^3, e^4\}$ of $T^*_M|_x$ can be extended around x as a local frame field $\{dx_1, dx_2, dx_3, dx_4\}$ that comes from some local coordinates $\{x_1, x_2, x_3, x_4\}$ on M and such that we have

$$\omega = dx_1 \wedge dx_2 + dx_3 \wedge dx_4 \ ,$$

then ω must be a closed 2–form around x. If we can do this around every point of M, then ω is globally closed,

$$d\omega = 0 \ .$$

In this case, ω is called a **symplectic structure** on M. Such a manifold M, together with the compatible metric and almost-complex structure, is known as an **almost-Kähler manifold.**

Kähler case. If local frame fields $\{dx_1, dx_2, dx_3, dx_4\}$ as above can be found so that, furthermore, we have

$$J(dx_1) = dx_2 \qquad \text{and} \qquad J(dx_3) = dx_4 \,,$$

then the Levi–Cività connection ∇ must in fact be $\mathbb{C}$–linear with respect to J; that is to say, we have $\nabla_v(Jw) = J(\nabla_v w)$. It follows that the fundamental form ω must be parallel:[8]

$$\nabla\omega = 0 \,.$$

Even more, it turns out that the almost-complex structure J must be *integrable.* In other words, our 4–manifold M is in fact nothing but a *Kähler surface.*

Symplectic, directly. Without reference to any metric or almost-complex structure, an exterior 2–form ω on a 4–manifold M is called a **symplectic structure** if it is closed and non-degenerate, that is to say, if both

$$d\omega = 0 \qquad \text{and} \qquad \omega \wedge \omega > 0 \,,$$

where the last relation means that the 4–form $\omega \wedge \omega$ is nowhere-zero and compatible with the orientation[9] of M.

Given any symplectic structure ω on M, there always exist both compatible Riemannian metrics and almost-complex structures J on M, unique up to homotopy. Thus, the previous almost-Kähler configuration can be rebuilt. In particular, every symplectic structure has a well-defined **Chern class**

$$c_1(\omega) = c_1(J) \,.$$

This Chern class could also be defined directly as the class of the subbundle of Λ^2_+ that is complementary to ω.

Notice that, if M admits any symplectic structure ω, then, since the class $[\omega]$ has positive self-intersection, we must have

$$b_2^+(M) \geq 1 \,.$$

Keep in mind that the classes $c_1(\omega)$ and $[\omega]$ *a priori* have no relationship.[10]

---

[8] Parallel for the connection induced on Λ^2.

[9] A 4–form φ is said to be *compatible* with the orientation of M if φ is everywhere a positive multiple of local forms $e^1 \wedge e^2 \wedge e^3 \wedge e^4$ for $\{e_1, e_2, e_3, e_4\}$ local *orienting* frames.

[10] Nonetheless, compare with the classification of Kähler surfaces, section 7.4 (page 295).

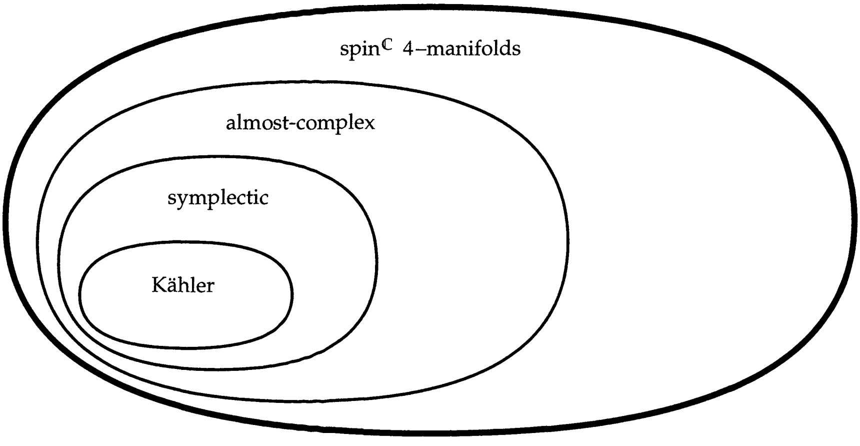
A symplectic structure on a M offers, for every compatible almost-complex structure and Riemannian metric, a perfect control over the area of all J-holomorphic curves:[11]

$$\mathrm{Area}(S) = \int_S \omega = [\omega] \cdot S .$$

This happens because of the equality $vol_S = e \wedge Je = \omega|_S$ for any random unit-length $e \in T_S^*$. Therefore, on one hand the area of a J-holomorphic curve is constant, on the other hand a J-holomorphic curve is never homologically-trivial.

In many respects, the almost-complex structures that are tamed by symplectic forms are as-close-as-it-gets to the complex world. Symplectic 4-manifolds have been a very fruitful subject of recent research. Nonetheless, their presence in this volume is meager.

In the following section, by moving in the opposite direction—away from Kähler surfaces—we will generalize almost-complex structures until they become available on *every* 4-manifold. Such structures are known as *spin*$^{\mathbb{C}}$ *structures.*



10.1. Cutting the 4-manifold cake, II

10.2. Spin$^{\mathbb{C}}$ structures and spinors

In this section we describe the ingredients that are needed for merely writing the Seiberg–Witten equations. The equations are built on top of a choice of spin$^{\mathbb{C}}$ structure. The latter is the offspring of crossing spin structures with almost-complex structures. An advantage is that, unlike any of its

---

[11] Controlling the area of J-holomorphic curves is essential in proving Gromov's compactness theorem, which is stated in the end-notes of this chapter (page 471).

parents, a spin$^{\mathbb{C}}$ structure always exists on *every* 4–manifold. On the other hand, every spin structure is a spin$^{\mathbb{C}}$ structure and every almost-complex structure is a spin$^{\mathbb{C}}$ structure. After having presented one of the parents above, we start with a very brief review of the other progenitor below:

Spin structures, hurriedly revisited

We have met spin structures before, in section 4.3 (page 162). There, we defined them in terms of partial trivializations of T_M. After that, in the extensive notes at the end of chapter 4, we explained the equivalence with a more common differential-geometric definition.[12] It is the latter, cocycle-based version, that we now favor, and we review it below:

Playing with cocycles. Pick a random Riemannian metric on M; this exhibits the tangent bundle T_M as an $SO(4)$–bundle, glued-up by an $SO(4)$–valued cocycle

$$\{g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to SO(4)\}\ ,$$

with maps defined on the overlaps of some open covering $\{U_\alpha\}$ of M. Such a collection of transition-functions satisfies the cocycle condition

$$g_{\alpha\beta}(x)\cdot g_{\beta\gamma}(x)\cdot g_{\gamma\alpha}(x) = id\ .$$

If M has $w_2(T_M) = 0$, then this $SO(4)$–cocycle of T_M can be lifted to a cocycle with values in the simply-connected group

$$Spin(4)$$

that doubly-covers $SO(4)$. That is to say, there will exist maps $\tilde{g}_{\alpha\beta}$ fitting in the diagram

$$\begin{tikzcd} U_\alpha \cap U_\beta \arrow[r, "\tilde{g}_{\alpha\beta}"'] \arrow[d, equal] & Spin(4) \arrow[d] \\ U_\alpha \cap U_\beta \arrow[r, "g_{\alpha\beta}"] & SO(4) \end{tikzcd}$$

and so that the $\tilde{g}_{\alpha\beta}$'s themselves still satisfy the cocycle condition

$$\tilde{g}_{\alpha\beta}(x)\cdot \tilde{g}_{\beta\gamma}(x)\cdot \tilde{g}_{\gamma\alpha}(x) = id\ .$$

Such a lifted cocycle is called a **spin structure** on M and is considered only up to cocycle-isomorphisms.

---

[12] The note on page 174 explained cocycles and the definition of spin structures in those terms. The note on page 181 proved the equivalence of the extendable-trivializations definition with the cocycle-based definition. The note on page 189 put cocycles (and spin structures) in their natural habitat, Čech cohomology. The note on page 204 described spin structures in terms of classifying spaces.

Bundles, connections, and Dirac operators. Using the isomorphism

$$Spin(4) = SU(2) \times SU(2)\,,$$

we can immediately project any $Spin(4)$–cocycle onto two $SU(2)$–valued cocycles. These can then be used to glue-up two $SU(2)$–bundles with complex-plane fibers, denoted by

$$\mathcal{S}^- \to M \qquad \text{and} \qquad \mathcal{S}^+ \to M\,.$$

These are called the **bundles of spinors** of our spin structure. A section in one of these bundles is called a **spinor field.**

It happens that these spinor bundles $\mathcal{S}^{\pm}$ always come from birth equipped with a bit of extra structure, namely a so-called **Clifford multiplication**

$$T_M \times \mathcal{S}^+ \xrightarrow{\bullet} \mathcal{S}^-\,.$$

For every fixed $v \in T_M$, if we look at the map $\mathcal{S}^+ \to \mathcal{S}^- \colon \varphi \mapsto v \bullet \varphi$ and combine it with its adjoint $\mathcal{S}^- \to \mathcal{S}^+$, then we will bring to light the defining property of this Clifford multiplication: we must always have

$$v \bullet (v \bullet \varphi) = -\left|v\right|^2 \cdot \varphi\,.$$

(Clifford multiplication can be modeled by quaternion multiplication.)

On the other hand, the Riemannian metric on M endows T_M with its Levi–Cività connection ∇. This connection lifts and then projects through the diagram[13]

$$\begin{tikzcd} SU(2) & Spin(4) \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"'] \arrow[d] & SU(2)\,, \\ & SO(4) & \end{tikzcd}$$

first from $SO(4)$ up to $Spin(4)$, and then to the two copies of $SU(2)$. Hence, it generates an $SU(2)$–connection $\nabla^{\pm}$ on each of the spinor bundles $\mathcal{S}^{\pm}$.

Combining this connection $\nabla^+ \colon \Gamma(\mathcal{S}^+) \to \Gamma(\mathcal{S}^+ \otimes T^*_M)$ with the Clifford multiplication $T_M \times \mathcal{S}^+ \to \mathcal{S}^- \colon (v, \varphi) \mapsto v \bullet \varphi$ yields the **Dirac operator**

$$\mathcal{D} \colon \Gamma(\mathcal{S}^+) \longrightarrow \Gamma(\mathcal{S}^-)\,.$$

Specifically, $\mathcal{D}$ can be described as

$$\mathcal{D}\varphi = \sum e_k \bullet \nabla^+_{e_k} \varphi$$

for any local orthonormal frame $\{e_1, e_2, e_3, e_4\}$ in T_M.

---

[13] In fact, ∇ lifts and then projects through the diagram of Lie algebras corresponding to the diagram of Lie groups that is exhibited above.

The Dirac operator is a linear first-order elliptic differential operator, whose symbol is the Clifford multiplication, and whose (complex) index[14] is

$$\operatorname{Index}\mathcal{D} = -\tfrac{1}{8}\operatorname{sign} M\ .$$

The Dirac operator (and its many versions) is, in many ways, *the* essential elliptic operator.

The importance of these notions in differential geometry can hardly be understated. For example, the Dirac operator played an essential role in the proof of the Atiyah–Singer index theorem. Look at **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** **[LM89]**, while we proceed to twist all the above with complex scalars:

$\mathbf{Spin^{\mathbb{C}}}$ structures

We briskly describe a complexified version of spin structures. These have two fundamental advantages: they exist on *every* 4–manifold, and they collaborate well with (almost-)complex structures.

Definition and existence. If $w_2(T_M) \neq 0$, then there are no global spin structures on M. In other words, every lift $\widetilde{g}_{\alpha\beta}$ of the maps $g_{\alpha\beta}\colon U_\alpha \cap U_\beta \to SO(4)$ to $Spin^{\mathbb{C}}(4)$ can merely be expected to satisfy

$$\widetilde{g}_{\alpha\beta}(x)\cdot\widetilde{g}_{\beta\gamma}(x)\cdot\widetilde{g}_{\gamma\alpha}(x) = \pm\, id\ ,$$

and the appearance of some minus signs is inevitable.

This can be corrected if, instead of attempting to lift to $Spin(4)$, we lift to its complexified version, to the group $Spin^{\mathbb{C}}(4)$. This **complex spin group** can be defined as

$$Spin^{\mathbb{C}}(4) = U(1)\times Spin(4)\ /\ \pm 1\ .$$

As such, it admits a natural double-cover projection

$$Spin^{\mathbb{C}}(4) \longrightarrow U(1)\times SO(4)\ ,$$

which over $SO(4)$ looks like $Spin(4) \to SO(4)$, while over $U(1) = S^1$ it looks like the squaring double-cover $U(1) \to U(1)\colon z \mapsto z^2$. These two parts can be separated into two projections

$$U(1) \xleftarrow{\ \det\ } Spin^{\mathbb{C}}(4) \longrightarrow SO(4)\ .$$

Now, given any $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ for T_M, we can lift it to some collection of $Spin^{\mathbb{C}}(4)$–valued maps $\widetilde{g}^{c}_{\alpha\beta}$ that will fit in the diagram

$$\begin{tikzcd}
U_\alpha \cap U_\beta \arrow[r, "\widetilde{g}^{c}_{\alpha\beta}"'] \arrow[d, equal] & Spin^{\mathbb{C}}(4) \arrow[d] \\
U_\alpha \cap U_\beta \arrow[r, "g_{\alpha\beta}"] & SO(4)\ .
\end{tikzcd}$$

---

[14] For every (complex-)linear differential operator $P\colon \Gamma(E) \to \Gamma(F)$, its **index** is defined as $\operatorname{Index} P = \dim_{\mathbb{C}}\operatorname{Ker} P - \dim_{\mathbb{C}}\operatorname{Coker} P$, or, if one prefers, $\dim_{\mathbb{C}}\operatorname{Ker} P - \dim_{\mathbb{C}}\operatorname{Ker} P^*$.

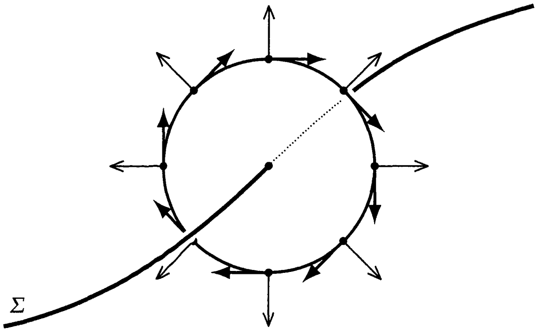
If, moreover, these lifted maps satisfy the cocycle condition

$$\tilde{g}^c_{\alpha\beta}(x) \cdot \tilde{g}^c_{\beta\gamma}(x) \cdot \tilde{g}^c_{\gamma\alpha}(x) = id ,$$

then the $Spin^{\mathbb{C}}(4)$–cocycle $\{\tilde{g}^c_{\alpha\beta}\}$, considered only up to cocycle-isomorphisms, will be called a **spin$^{\mathbb{C}}$ structure** on M. The beauty is that we have:

Lemma. *Every smooth 4–manifold admits spin$^{\mathbb{C}}$ structures.*

> ***Proof.*** Choose a characteristic element $\underline{w}$ (an integral lift of $w_2(T_M)$) and represent it as an embedded characteristic surface Σ in M. As was mentioned before,[15] the complement $M \setminus \Sigma$ always admits a spin structure, and such a spin structure cannot be extended across Σ.
>
> Think of such a spin structure as a trivialization of T_M over the 1–skeleton of $M \setminus \Sigma$ that extends across the 2–skeleton of $M \setminus \Sigma$, but not across disks normal to Σ. This outside trivialization can then be arranged to offer a trivialization of T_M over the fibers of the circle-bundle $\mathbb{S}N_{\Sigma/M}$ of Σ's normal bundle. Its non-extendability across Σ means that the trivialization of T_M over each such circle must describe the nontrivial element of $\pi_1 SO(4) = \mathbb{Z}_2$, that is to say, it must undergo a rotation of 2π, as suggested in figure 10.2.



10.2. Outside spin structure, not extending across a characteristic surface Σ

> The surface Σ is the obstacle to extending the outside spin structure on $M \setminus \Sigma$ across all M. To cross Σ, we need to somehow effect a 2π–twist. If we think in terms of spin$^{\mathbb{C}}$ structures, then we have a new degree of freedom coming from the $\mathbb{S}^1$–factor of $Spin^{\mathbb{C}}(4)$. Therefore we can now extend the outside spin structure of $M \setminus \Sigma$ as a spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ across all M, by passing across Σ through the use of a 2π–twist in the $\mathbb{S}^1$–factor. □

---

[15] In the end-note on page 179, and better argued in Čech-cohomological terms in the end-note on page 196, both inside chapter 4.

A proof of a different flavor, using cocycles and Čech cohomology, will be detailed in the end-notes of this chapter (page 423).

In any case, notice that the argument hinges on the fact that the Stiefel–Whitney class can always be lifted to some integral class $\underline{w}$. Even more, when $H^2(M;\mathbb{Z})$ has no 2–torsion, *e.g.*, when M is is simply-connected, then the resulting spin$^{\mathbb{C}}$ structure is *uniquely* determined by the homology class $\underline{w}$ of the chosen characteristic surface Σ.

Also notice that, whether $w_2(T_M)$ vanished or not has no bearing on the argument. Indeed, if $w_2(T_M) = 0$ and M admits a spin structure, then, by using the natural map

$$Spin(4) \subset U(1) \times Spin(4) \longrightarrow Spin^{\mathbb{C}}(4)$$

(which is in fact an inclusion), we can transport the spin structure $\{\widetilde{g}_{\alpha\beta}\}$ to a canonical spin$^{\mathbb{C}}$ structure $\{\widetilde{g}^c_{\alpha\beta}\}$. In other words, spin structures are exactly those spin$^{\mathbb{C}}$ structures $\{\widetilde{g}^c_{\alpha\beta}\}$ for which all maps[16] $\det \widetilde{g}^c_{\alpha\beta} \colon U_\alpha \cap U_\beta \to U(1)$ are constantly 1. Spin$^{\mathbb{C}}$ structures are *generalized spin structures.*

On the other hand, there is an alternative definition of spin$^{\mathbb{C}}$ structures, noticed by R. Gompf:[17] A spin$^{\mathbb{C}}$ structure on a manifold X is an almost-complex structure on the 2–skeleton of X that can be extended across the 3–skeleton of X. In particular, this implies that any almost-complex structure defines a canonical spin$^{\mathbb{C}}$ structure. Spin$^{\mathbb{C}}$ structures are *generalized almost-complex structures* as well.

We will explore the link to almost-complex structures in more detail later, but first we need to introduce a new creature:

The determinant line bundle. Remember the natural map

$$Spin^{\mathbb{C}}(4) \xrightarrow{\det} U(1)$$

induced by $U(1) \to U(1) \colon z \mapsto z^2$. Given a spin$^{\mathbb{C}}$ structure $\mathfrak{s} = \{\widetilde{g}^c_{\alpha\beta}\}$, it can be projected to a $U(1)$–valued cocycle $\{\det \widetilde{g}^c_{\alpha\beta}\}$, which can then be let act on $\mathbb{C}$ and thus glue-up a complex-line bundle

$$\mathcal{L} \to M\,.$$

This is called the **determinant line bundle** of the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$. Its Chern class $c_1(\mathcal{L})$ is called the **Chern class** of the spin$^{\mathbb{C}}$ structure, denoted by

$$c_1(\mathfrak{s}) = c_1(\mathcal{L})\,.$$

This class is always a characteristic element of M and coincides with the $\underline{w}$ used in the proof of the existence lemma.

---

[16] Remember that we denoted by det the natural projection $Spin^{\mathbb{C}}(4) \to U(1)$ induced by the squaring $U(1) \to U(1) \colon z \mapsto z^2$.

[17] See **R. Gompf**'s *Spin$^{\mathbb{C}}$–structures and homotopy equivalences* [**Gom97**, p 48]

Building the determinant line bundle. To better visualize $\mathcal{L}$ in relation to its spin$^{\mathbb{C}}$ structure $\mathfrak{s}$, we can proceed as follows:

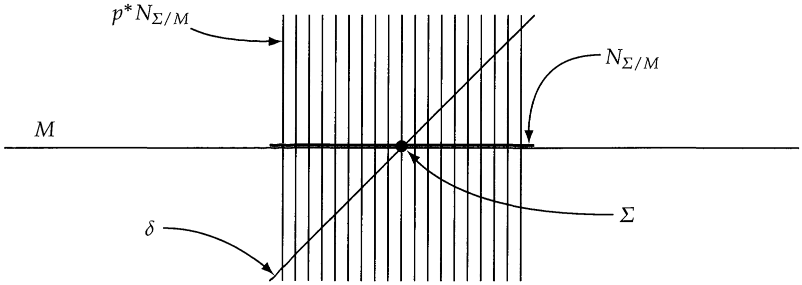
In the spirit of the existence proof above, first we represent the class $c_1(\mathfrak{s})$ by a characteristic surface Σ embedded in M, then we endow $M \setminus \Sigma$ with a non-extendable spin structure corresponding to $\mathfrak{s}$. Then we can view the determinant line bundle $\mathcal{L}$ of $\mathfrak{s}$ as that bundle which records the $U(1)$-twists used while extending the outside spin structure across Σ as the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$.

Moreover, the determinant bundle $\mathcal{L}$ also admits a direct description in terms of the embedding of the surface Σ in M: Start with its normal bundle $N_{\Sigma/M}$ and think of $N_{\Sigma/M}$ both as an oriented-plane bundle (and thus as a complex-line bundle) $p\colon N_{\Sigma/M} \to \Sigma$ and as a tubular neighborhood of Σ in M. Then pull-back the bundle $N_{\Sigma/M}$ over itself to get $p^*N_{\Sigma/M}$, as in

$$\begin{tikzcd}
 & p^*N_{\Sigma/M} \arrow[r] \arrow[d] & N_{\Sigma/M} \arrow[d, "p"'] \\
M \arrow[r, phantom, "\supset"] & N_{\Sigma/M} \arrow[r, "p"] & \Sigma .
\end{tikzcd}$$

We think of the resulting $p^*N_{\Sigma/M}$ as a complex-line bundle defined over the neighborhood $N_{\Sigma/M}$ of Σ in M.

Notice that this bundle $p^*N_{\Sigma/M}$ is trivial off Σ. Indeed, consider its canonical section $\delta\colon N_{\Sigma/M} \to p^*N_{\Sigma/M}$ given by $\delta(v) = (v, v)$: away from Σ, the section δ is nowhere-zero, and thus trivializes $p^*N_{\Sigma/M}$ over $N_{\Sigma/M} \setminus \Sigma$, as in figure 10.3.



10.3. Building a line bundle of Chern class $[\Sigma]$

We are thus looking at a line bundle $p^*N_{\Sigma/M}$ defined over $N_{\Sigma/M} \subset M$ and canonically trivialized away from Σ. We can then take the trivial bundle $(M \setminus \Sigma) \times \mathbb{C}$ and, above $N_{\Sigma/M} \setminus \Sigma$, glue it to $p^*N_{\Sigma/M}$ by using the latter's trivialization δ.

The result is a complex-line bundle defined over the whole manifold M, endowed with a section (the obvious extension of δ) that is zero only at Σ. This means that we have built a line bundle $\mathcal{L}$ of Chern class $c_1(\mathcal{L}) = [\Sigma]$.

Labeling spin$^{\mathbb{C}}$ structures. We need a comfortable way to refer to the various spin$^{\mathbb{C}}$ structures on a 4–manifold.

Chern classes. On one hand, if $H^2(M;\mathbb{Z})$ has no 2–torsion,[18] *e.g.*, when M is simply-connected, then the Chern class $c_1(\mathfrak{s})$ determines *uniquely* the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$. Since this volume is focused on simply-connected 4–manifolds, we will often refer to a spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ by its Chern class $c_1(\mathfrak{s})$ by using the canonical bijection between characteristic elements and spin$^{\mathbb{C}}$ structures

$$\{\text{spin}^{\mathbb{C}}\text{ structures}\} \approx \{\underline{w} \in H^2(M;\mathbb{Z}) \mid \underline{w} = w_2(T_M) \pmod 2\}\,,$$

which technically is available only in the 2–torsion-free case.

Transitive action of $H^2(M;\mathbb{Z})$. On the other hand, the set of spin$^{\mathbb{C}}$ structures can also be parametrized by the whole cohomology $H^2(M;\mathbb{Z})$:

$$\{\text{spin}^{\mathbb{C}}\text{ structures}\} \approx H^2(M;\mathbb{Z})\,.$$

This correspondence is *not* canonical, as it depends on the choice of a reference spin$^{\mathbb{C}}$ structure.[19] Nonetheless, this parametrization has the advantage of working just as well in the presence of 2–torsion.

The correspondence is established as follows: First fix a random spin$^{\mathbb{C}}$ structure $\mathfrak{s}_0 = \{\tilde{g}^c_{\alpha\beta}\}$. Then, for each class $\varepsilon \in H^2(M;\mathbb{Z})$, we can twist $\mathfrak{s}_0$ by the cocycle $\{\lambda_{\alpha\beta}\}$ of a complex-line bundle L_ε of Chern class ε, thus obtaining a new *Spin*$^{\mathbb{C}}(4)$–cocycle $\{\lambda_{\alpha\beta}\cdot\tilde{g}^c_{\alpha\beta}\}$. We denote this spin$^{\mathbb{C}}$ structure by

$$\mathfrak{s}_0 + \varepsilon\,.$$

Note that the corresponding cocycle for the determinant bundle is twisted by $\lambda^2_{\alpha\beta}$, and therefore that we have

$$\mathcal{L}_{\mathfrak{s}_0+\varepsilon} = \mathcal{L}_{\mathfrak{s}_0} \otimes L_\varepsilon^{\otimes 2}\,, \qquad \text{and hence} \qquad c_1(\mathfrak{s}_0+\varepsilon) = c_1(\mathfrak{s}_0) + 2\varepsilon\,.$$

Also observe how, if $H^2(M;\mathbb{Z})$ contains classes a with $2a = 0$, then $\mathfrak{s}_0 + a$ is a spin$^{\mathbb{C}}$ structure distinct from $\mathfrak{s}_0$, even though their Chern classes coincide: $c_1(\mathfrak{s}_0 + a) = c_1(\mathfrak{s}_0)$.

---

[18] Equivalent to saying that $H_1(M;\mathbb{Z})$ has no 2–torsion.

[19] One says that the set of spin$^{\mathbb{C}}$ structures is a $H^2(M;\mathbb{Z})$–**torsor**, or that $H^2(M;\mathbb{Z})$ acts freely on spin$^{\mathbb{C}}$ structures, or that the set of spin$^{\mathbb{C}}$ structures is an affine copy of $H^2(M;\mathbb{Z})$, or even that the set of spin$^{\mathbb{C}}$ structures is an $H^2(M;\mathbb{Z})$–principal bundle over a point. As long as we know what we are talking about...

Complex interactions. Besides their existence on every 4–manifold, spin$^{\mathbb{C}}$ structures have another advantage over spin structures: spin$^{\mathbb{C}}$ structures collaborate well with the complex world.

Indeed, any almost-complex structure J (together with a compatible Riemannian metric) reduces the cocycle of T_M to a $U(2)$–cocycle. The group $U(2)$ can be described as being $SU(2)$ enriched with the determinants from the unit-circle $\mathbb{S}^1$, as in

$$U(2) = \mathbb{S}^1 \times SU(2) \;/\; \pm 1 \,.$$

Furthermore, since we have a natural inclusion $\mathbb{S}^1 \subset SU(2)$, it follows that $U(2)$ can be embedded diagonally inside

$$Spin^{\mathbb{C}}(4) = \mathbb{S}^1 \times SU(2) \times SU(2) \;/\; \pm 1$$

by using the canonical inclusion map

$$U(2) \subset Spin^{\mathbb{C}}(4) \colon \qquad [\lambda, \rho] \longmapsto [\lambda, \rho, \lambda] \,.$$

Therefore, if M is endowed with an almost-complex structure J, then its induced $U(2)$–cocycle can be viewed directly as a $Spin^{\mathbb{C}}(4)$–cocycle, and thus defines a canonical spin$^{\mathbb{C}}$ structure, denoted by

$$\mathfrak{s}_J \,.$$

We will call such a spin$^{\mathbb{C}}$ structure an **almost-complex spin$^{\mathbb{C}}$ structure.**

The determinant line bundle of such a spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J$ is exactly the anti-canonical bundle K_J^* of the corresponding almost-complex structure J:

$$\mathcal{L}_{\mathfrak{s}_J} = K_J^* \,, \qquad \text{and hence} \qquad c_1(\mathfrak{s}_J) = c_1(J) \,.$$

Conversely, given a random spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ on M, we can check whether it corresponds to an almost-complex structure by merely verifying whether we have

$$c_1(\mathfrak{s}) \cdot c_1(\mathfrak{s}) = 3 \operatorname{sign} M + 2\chi(M) \,.$$

(In a similar vein, a spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ can be recognized as corresponding to a spin structure by checking whether $c_1(\mathfrak{s}) = 0$.)

Moreover, every spin$^{\mathbb{C}}$ structure on M induces a partial almost-complex structure $J|_3$, defined over the 3–skeleton of M and with $c_1(J|_3) = c_1(\mathfrak{s})$. This is further explained in the end-notes of this chapter (page 426).

If our 4–manifold happens to be a complex surface, or a symplectic manifold, or in some other way associated with a distinguished almost-complex structure J, then the induced spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J$ is a natural choice for parametrizing all other spin$^{\mathbb{C}}$ structures with respect to it, and thus writing all other spin$^{\mathbb{C}}$ structures as translates $\mathfrak{s}_J + \varepsilon$ by ε's from $H^2(M; \mathbb{Z})$.

Spinor bundles. Just as we had spinor bundles $\mathcal{S}^{\pm}$ associated to spin structures, so too their generalizations, spin$^{\mathbb{C}}$ structures, have their own associated bundles.

On one hand the complex spin group is

$$Spin^{\mathbb{C}}(4) = \mathbb{S}^1 \times SU(2) \times SU(2) \,/\, \pm 1 \,,$$

while on the other hand we have

$$U(2) = \mathbb{S}^1 \times SU(2) \,/\, \pm 1 \,.$$

It follows that there are two natural projections

$$\begin{tikzcd} U(2) & Spin^{\mathbb{C}}(4) \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"'] & U(2) \\ {[z, \xi_-]} & {[z, \xi_+, \xi_-]} \arrow[l, mapsto] \arrow[r, mapsto] & {[z, \xi_+]} \,. \end{tikzcd}$$

Using these projections, the cocycle $\mathfrak{s} = \{\tilde{g}^c_{\alpha\beta}\}$ can be projected onto two $U(2)$–cocycles $\{\rho_{\pm}(\tilde{g}^c_{\alpha\beta})\}$, which can then be used to build two complex-plane bundles, denoted by

$$\mathcal{W}^- \to M \qquad \text{and} \qquad \mathcal{W}^+ \to M \,.$$

These bundles are called the **bundles of complex spinors** or the **bundles of coupled spinors** (coupled, as it were, with $\mathcal{L}$).

More specifically, we will call $\mathcal{W}^+$ the bundle of **self-dual** spinors and $\mathcal{W}^-$ the bundle of **anti-self-dual** spinors. (The more customary name in the literature is bundle of **positive/negative** spinors.) Their structure group being $U(2)$ means that $\mathcal{W}^+$ and $\mathcal{W}^-$ naturally come equipped with a Hermitian fiber-metric. The sections of $\mathcal{W}^{\pm}$ are called *(coupled)* **spinor fields.**

Alternatively, one can write directly $\mathcal{W}^{\pm} = \mathcal{S}^{\pm} \otimes \mathcal{L}^{1/2}$. *Even though they do not exist globally on* M, *the bundles* $\mathcal{S}^{\pm}$ *and* $\mathcal{L}^{1/2}$ *do exist locally on* M. *This formula exhibits* $\mathcal{L}$ *as the determinant line bundle of the spinor bundles, i.e.,* $\mathcal{L} = \det_{\mathbb{C}} \mathcal{W}^+ = \det_{\mathbb{C}} \mathcal{W}^-$, *thus justifying* $\mathcal{L}$*'s name. In particular, we have* $c_1(\mathcal{W}^{\pm}) = c_1(\mathfrak{s})$.

It is worth noticing that, when the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ corresponds to some almost-complex structure J, then the associated spinor bundles can be described as

$$\mathcal{W}^+ = \underline{\mathbb{C}} \oplus K_J^* \qquad \text{and} \qquad \mathcal{W}^- = (T_M, J) \,.$$

Clifford multiplication. Just as for spin structures, every spin$^{\mathbb{C}}$ structure comes equipped with a **Clifford multiplication**

$$T_M \times \mathcal{W}^+ \xrightarrow{\bullet} \mathcal{W}^- \,.$$

A concrete quaternion-based description of Clifford multiplication will be presented in the end-notes of this chapter (page 432).

The Clifford multiplication and its adjoint $T_M \times \mathcal{W}^- \xrightarrow{\bullet} \mathcal{W}^+$ combine in the defining property that, for every $v \in T_M|_x$ and $\varphi \in \mathcal{W}^+|_x$, we have

$$v \bullet (v \bullet \varphi) = -|v|^2 \cdot \varphi .$$

All complex morphisms from $\mathcal{W}^+$ to $\mathcal{W}^-$ are caught by this Clifford action:

$$\mathrm{Hom}_{\mathbb{C}}(\mathcal{W}^+, \mathcal{W}^-) \approx T_M \otimes \mathbb{C} ,$$

meaning that, for every complex-linear morphism $f \colon \mathcal{W}^+ \to \mathcal{W}^-$, there exists a unique field $v \in \Gamma(T_M \otimes \mathbb{C})$ so that $f(\varphi) = v \bullet \varphi$. Turning this around, we could write $T_M \otimes_{\mathbb{R}} \mathbb{C} = (\mathcal{W}^+)^* \otimes_{\mathbb{C}} \mathcal{W}^-$ and think of the spinor bundles as a decomposition of T_M into complex bundles.

The Clifford multiplication extends in the obvious way to all tensor-powers of T_M, and in particular to exterior forms. In particular, there is an induced action $\Lambda^2 \times \mathcal{W}^+ \xrightarrow{\bullet} \mathcal{W}^+$ of 2–forms on self-dual spinors. This action turns out to be trivial for anti-self-dual forms, and thus it remains relevant only in its self-dual part

$$\Lambda^2_+ \times \mathcal{W}^+ \xrightarrow{\bullet} \mathcal{W}^+ .$$

The latter action catches all *trace-free* endomorphisms of $\mathcal{W}^+$: we thus have

$$\mathrm{End}_0(\mathcal{W}^+) \approx i\Lambda^2_+(T^*_M) ,$$

where $\mathrm{End}_0(\mathcal{W}^+) = \{ f \in \mathrm{Hom}_{\mathbb{C}}(\mathcal{W}^+, \mathcal{W}^+) \mid \mathrm{trace}\, f = 0 \}$. This means that for every trace-free complex-linear endomorphism $f \colon \mathcal{W}^+ \to \mathcal{W}^+$ there exists a self-dual 2–form $\alpha \in \Gamma(\Lambda^2_+)$ so that $f(\varphi) = i\alpha \bullet \varphi$.

Connections and Dirac operators. Unlike the case of spin structures, for endowing the complex spinor bundles $\mathcal{W}^\pm$ with connections, the Levi–Cività connection ∇ alone is not enough. Indeed, our data is now twisted by the line bundle $\mathcal{L}$.

Connections. We choose a random $U(1)$–connection A on $\mathcal{L}$. Then, by travelling along the diagram[20]

$$\begin{tikzcd} & U(1) & \\ U(2) & Spin^{\mathbb{C}}(4) \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"] \arrow[u, "\det"] \arrow[d] & U(2) \\ & SO(4) & \end{tikzcd}$$

we can combine A with ∇ inside $Spin^{\mathbb{C}}(4)$, and then project the result onto two $U(2)$–connections ∇^A on $\mathcal{W}^+$ and $\mathcal{W}^-$.

---

[20] In fact, by traveling along the associated diagram of Lie algebras.

If one likes to think in terms of the local equality $\mathcal{W}^{\pm} = \mathcal{S}^{\pm} \otimes \mathcal{L}^{1/2}$, then we have $\nabla^A = \nabla^{\pm} + A^{1/2}$, where $\nabla^{\pm}$ are the connections induced by ∇ on $\mathcal{S}^{\pm}$.

Dirac operators. The spinor connection $\nabla^A : \Gamma(\mathcal{W}^+) \to \Gamma(\mathcal{W}^+ \otimes T^*_M)$ combines with the Clifford multiplication $T_M \times \mathcal{W}^+ \to \mathcal{W}^-$ to yield the *coupled*[21] **Dirac operator**

$$\mathcal{D}^A : \Gamma(\mathcal{W}^+) \longrightarrow \Gamma(\mathcal{W}^-) .$$

Locally, it is described by the formula

$$\mathcal{D}^A \varphi = \sum e_k \bullet \nabla^A_{e_k} \varphi$$

for any local orthonormal frame $\{e_1, e_2, e_3, e_4\}$ in T_M.

Again, writing locally $\mathcal{W}^{\pm} = \mathcal{S}^{\pm} \otimes \mathcal{L}^{1/2}$, we can also define $\mathcal{D}^A = \mathcal{D} \otimes A^{1/2}$, where $\mathcal{D}$ is the canonical Dirac operator of the spinor bundles $\mathcal{S}^{\pm}$.

The Dirac operator is a first-order linear elliptic operator and satisfies:

Unique Continuation Property. *If $\varphi \in \Gamma(\mathcal{W}^+)$ is zero on an open set and we have $\mathcal{D}^A \varphi \equiv 0$, then φ must vanish globally.* □

The similarity with the Cauchy–Riemann operators does not stop here. In fact, a good way of thinking about Dirac operators is exactly as generalized Cauchy–Riemann operators.

The symbol of $\mathcal{D}^A$ is the Clifford multiplication, and its (complex) index is

$$\operatorname{Index} \mathcal{D}^A = \tfrac{1}{8}\big(c_1(\mathfrak{s}) \cdot c_1(\mathfrak{s}) - \operatorname{sign} M\big) .$$

Since $c_1(\mathfrak{s})$ is a characteristic element, this formula always yields an integer.

The Lichnerowicz formula. We now state a truly fundamental formula:

Lichnerowicz Formula (coupled). *For every spin$^{\mathbb{C}}$ structure on M, and any connection A on its determinant line bundle $\mathcal{L}$, we have:*

$$(\mathcal{D}^A)^* \, \mathcal{D}^A \varphi = (\nabla^A)^* \, \nabla^A \varphi + \tfrac{1}{4} \operatorname{scal} \cdot \varphi + \tfrac{1}{2} F_A^+ \bullet \varphi ,$$

where $(\mathcal{D}^A)^*$ is the formal adjoint of the Dirac operator $\mathcal{D}^A$; $(\nabla^A)^*$ is the formal adjoint of the connection ∇^A on $\mathcal{W}^+$ induced by A; scal denotes the scalar curvature of M; and $F_A^+ \bullet \varphi$ denotes the Clifford action of the curvature 2–form F_A of A on φ. □

Adjoint, you say. *Given two bundles E and F on M, endowed with fiber-metrics, and given an operator $P : \Gamma(E) \to \Gamma(F)$, we say that an operator $P^* : \Gamma(F) \to \Gamma(E)$ is the* **formal adjoint** *of P if and only if*

$$\int_M \langle P\alpha, \beta \rangle \operatorname{vol}_M = \int_M \langle \alpha, P^*\beta \rangle \operatorname{vol}_M$$

---

[21] Coupled, as it were, with A.

for all sections $\alpha \in \Gamma(E)$ *and* $\beta \in \Gamma(F)$. *The word "formal" is used only because the spaces of smooth sections* $\Gamma(E)$, $\Gamma(F)$ *are not complete (not Hilbert).*

The Lichnerowicz formula is one of many similar formulae known under the name of **Weitzenböck-type formulae.** Their most frequent use is for proving vanishing results by using what is known as the **Bochner technique.**[22]

All gauge-theoretic proofs that will be presented in the main text of this chapter will be consequences of the above Lichnerowicz formula. Indeed, the Lichnerowicz formula has a remarkable range of applications to Seiberg–Witten theory. As we will see, it is used to prove that the Seiberg–Witten moduli space is compact and that it is non-empty for at most finitely-many spin$^{\mathbb{C}}$ structures. The same formula will also show that the Seiberg–Witten invariants vanish on manifolds of positive scalar curvature and will be used to argue that the invariants vanish on connected sums. Another consequence is the behavior of the invariants under blow-ups. The formula is even used in proving the celebrated adjunction inequality that controls the genus of embedded surfaces, as we will see in the next chapter.

Of course, at this early moment we still need to define the Seiberg–Witten invariants. For that, we introduce a few more creatures:

Leftovers. We need to briefly discuss the squaring map $\sigma\colon \mathcal{W}^+ \to i\Lambda^2_+$, the curvature form F_A of the connection A on $\mathcal{L}$, and the gauge group $\mathcal{G}(\mathcal{L})$ of $\mathcal{L}$, with its action both on $\mathcal{C}onn(\mathcal{L})$ and on $\Gamma(\mathcal{W}^+)$.

The squaring map. The **squaring map** is the unique fiber-preserving map

$$\sigma\colon \mathcal{W}^+ \longrightarrow i\Lambda^2_+(T^*_M)$$

that for every $\varphi \in \mathcal{W}^+$ satisfies the equality

$$\sigma(\varphi) \bullet \varphi = \tfrac{1}{2}\,|\varphi|^2 \cdot \varphi\,.$$

While the squaring map has gained prominence especially after the birth of Seiberg–Witten theory, it was known and used before.[23]

An alternative for defining the squaring map is encoded in the formula

$$\sigma(\varphi) = \varphi \otimes \varphi^* - \tfrac{1}{2}\,|\varphi|^2\, id\,.$$

Specifically, a fixed spinor $\varphi \in \mathcal{W}^+$ *can be viewed as acting on other spinors through the* $\mathbb{C}$*-linear endomorphism* $\varphi \otimes \varphi^*\colon \mathcal{W}^+ \to \mathcal{W}^+$ *described by* $\psi \mapsto \langle \psi, \varphi \rangle_{\mathbb{C}} \cdot \varphi$. *Its trace-free part is* $\varphi \otimes \varphi^* - \frac{1}{2}\,|\varphi|^2\, id$. *The squaring map* σ *points to the imaginary 2–form* $\sigma(\varphi)$ *whose Clifford action on* $\mathcal{W}^+$ *is the same as this trace-free action of* φ.

---

[22] See the end-notes of this chapter (page 474) for more Weitzenböck-type formulae and a few vanishing results.

[23] See for example **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** [LM89, sec IV.10].

It is easy to check that

$$\sigma(e^{i\vartheta}\varphi) = \sigma(\varphi) .$$

If we restrict σ to the 3–sphere of radius 2 inside the fiber $\mathcal{W}^+|_x$, then its image will be the 2–sphere of radius $\sqrt{2}$ inside $i\Lambda^2_+|_x$. Hence, if we ignore scaling constants 2 and $\sqrt{2}$, then we see that the restriction of the squaring map to sphere-fibers is exactly the *Hopf map.*[24] Moreover, since we have

$$\sigma(r\varphi) = r^2\sigma(\varphi)$$

for every $r \in \mathbb{R}$, we can think of the whole σ as a "squared-cone" on fiber-wise Hopf maps.

A more concrete, quaternion-based, description of σ will be presented in the end-notes of this chapter (page 432).

The curvature form. Another object relevant for the Seiberg–Witten equations is the curvature form F_A of the $U(1)$–connection A on $\mathcal{L}$.

Since the Lie algebra of $U(1) = S^1$ is simply $i\mathbb{R}$, the curvature F_A must be an imaginary-valued 2–form. Furthermore, the Bianchi identity dictates that F_A be closed. In de Rham cohomology, the form F_A represents the Chern class $c_1(\mathfrak{s})$. In review, we have:

$$F_A \in \Gamma(i\Lambda^2(T^*_M)) , \qquad dF_A = 0 , \qquad [F_A] = -2\pi i\, c_1(\mathcal{L}) .$$

The gauge group. Finally, the **gauge group** $\mathscr{G}(\mathcal{L})$ of $\mathcal{L}$ is merely the space of all $U(1)$–valued functions on M:

$$\mathscr{G}(\mathcal{L}) = \{g \colon M \to S^1\} .$$

It has the technical advantage (over Donaldson theory) of being Abelian. Its action on $\mathcal{L}$ induces both an action on the connections of $\mathcal{L}$ and an action on the spinor fields of $\mathcal{W}^+$.

We prefer to work our way backwards, from $\mathcal{W}^+$ toward $\mathcal{L}$. The action of some $g \colon M \to S^1$ on a spinor field $\varphi \in \Gamma(\mathcal{W}^\pm)$ is expressed directly by scalar multiplication:

$$(g \cdot \varphi)(x) = g(x)^{-1}\varphi(x) .$$

(We define this action by anti-multiplication for purely aesthetic reasons. It makes the formula for the action on connections look a bit better.)

Since $\mathcal{L} = \det_{\mathbb{C}} \mathcal{W}^\pm$ (or since $\mathcal{W}^\pm = \mathcal{S}^\pm \otimes \mathcal{L}^{1/2}$), the corresponding action of $\mathscr{G}(\mathcal{L})$ on sections s of $\mathcal{L}$ must be described by

$$(g \cdot s)(x) = g(x)^{-2} s(x) .$$

---

[24] The Hopf map was recalled in footnote 34 on page 129.

This induces a pull-back action on the $U(1)$–connections A on $\mathcal{L}$, given by $g \cdot d_A = g \circ d_A \circ g^{-1}$, and written explicitly as

$$g \cdot A = A + 2g^{-1}dg\,,$$

where $dg\colon M \to i\mathbb{R}$ is the differential of $g\colon M \to S^1$. This can be seen directly by computing:

$$\begin{aligned}(g \cdot d_A)(s(x)) &= (g \circ d_A \circ g^{-1})(s(x)) = g(x)^{-2} \cdot d_A(g(x)^2 \cdot s(x)) \\ &= g(x)^{-2} \cdot \big(d(g(x)^2) \cdot s(x) + g(x)^2 \cdot d_A s(x)\big) \\ &= g(x)^{-2} \cdot 2g(x)\, dg(x) \cdot s(x) + d_A s(x)\,.\end{aligned}$$

In particular, if $g = e^{if}$ for some $f\colon M \to \mathbb{R}$, then $(e^{if}) \cdot A = A + 2i\,df$.

A more detailed study of connections and curvatures on complex-line bundles, and of the action of the gauge group on them, was made in the endnotes of the preceding chapter (page 357).

10.3. Definition of the Seiberg–Witten invariants

In this section we discuss the Seiberg–Witten equations and the method through which they yield numerical invariants of smooth 4–manifolds.

After choosing a spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ on M, with associated spinor bundles $\mathcal{W}^{\pm}$ and determinant line bundle $\mathcal{L}$, the objects of interest in what follows will be pairs

$$(\varphi, A)\,,$$

where $\varphi \in \Gamma(\mathcal{W}^+)$ is a self-dual spinor field and $A \in \mathcal{C}onn(\mathcal{L})$ is a $U(1)$–connection on $\mathcal{L}$. Namely, we will look at solutions (φ, A) to a couple of mildly-non-linear elliptic partial differential equations and consider such solutions only up to the action of the gauge group $\mathscr{G}(\mathcal{L})$ described earlier.

The Seiberg–Witten equations are:

$$\begin{cases} \mathcal{D}^A \varphi = 0 \\ F_A^+ = \sigma(\varphi)\,. \end{cases}$$

As explained before, $\mathcal{D}^A$ denotes the Dirac operator induced by A, while F_A is the imaginary-valued curvature 2–form of A, and $F_A^+ = \frac{1}{2}(F_A + *F_A)$ is its self-dual part. Finally, $\sigma\colon \mathcal{W}^+ \to i\Lambda^2_+$ is the squaring map.

The moduli space

The solutions (φ, A) to the Seiberg–Witten equations are called *(Seiberg–Witten)* **monopoles**. The monopoles form a subspace $\mathfrak{S}$ inside the infinite-dimensional configuration space $\Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L})$.

It is easy to check that the Seiberg–Witten equations are invariant under the action of the gauge group $\mathscr{G}(\mathcal{L}) = \{g\colon M \to \mathbb{S}^1\}$. Therefore $\mathfrak{S}$ is a $\mathscr{G}(\mathcal{L})$–invariant slice of $\Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L})$. It thus makes sense to factor everything by the action of the gauge group, hence obtaining the **moduli space**

$$\mathfrak{M} = \mathfrak{S} \,/\, \mathscr{G}\,.$$

Obviously, this moduli space depends on the choices of spin$^{\mathbb{C}}$ structure and Riemannian metric.

Reducible solutions. The stabilizer[25] of (φ, A) under the action of $\mathscr{G}$ is trivial, unless $\varphi \equiv 0$. Since every *constant* function $g(x) \equiv e^{i\vartheta}$ acts trivially on all connections on $\mathcal{L}$, it is not hard to see that the stabilizer of a solution $(0, A)$ is isomorphic to $\mathbb{S}^1$. Such monopoles $(0, A)$ create singularities in the moduli space, and, in analogy with Donaldson theory, are called **reducible solutions.**

When $\varphi = 0$, the Seiberg–Witten equations simply become the equation

$$F_A^+ = 0\,,$$

whose solutions are all anti-self-dual connections on $\mathcal{L}$. As such, this equation was studied at length in the end-notes of the preceding chapter (page 357). We have seen there that, if $b_2^+(M) \geq 1$, then all solutions can be avoided for a generic metric. If further $b_2^+(M) \geq 2$, then all solutions can be avoided over any generic path of metrics. (As explained there, the role of b_2^+ is that it represents the codimension of an affine subspace of $\Gamma(\Lambda_+^2)$ that can be missed by perturbing the metric.)

In particular for us, if $b_2^+ \geq 1$, then reducible monopoles do not appear at all for a generic choice of Riemannian metric. In this case the action of $\mathscr{G}(\mathcal{L})$ on the solution space $\mathfrak{S}$ is free, and hence the orbit space $\mathfrak{M}$ has a better chance of being well-behaved. Indeed:

They are manifolds. One proves that, away from the reducible solutions, the moduli space is a manifold. Therefore:

Theorem.

— *If $b_2^+(M) \geq 1$, then, for a generic Riemannian metric, the Seiberg–Witten moduli space $\mathfrak{M}$ is either empty or is a smooth manifold of dimension*

$$\dim \mathfrak{M} = \tfrac{1}{4}\big(c_1(\mathfrak{s})^2 - 2\chi(M) - 3\,\mathrm{sign}\, M\big)\,.$$

— *If $b_2^+(M) \geq 2$, then, for every two generic metrics g_0 and g_1 and every generic path g_t connecting them, all corresponding moduli spaces $\mathfrak{M}_t$ are smooth manifolds (maybe empty) and draw a smooth cobordism between $\mathfrak{M}_0$ and $\mathfrak{M}_1$.* □

---

[25] The **stabilizer** of (φ, A) is the subgroup $\{g\colon M \to \mathbb{S}^1 \mid g\cdot\varphi = \varphi \;\;\&\;\; g\cdot A = A\}$ of $\mathscr{G}(\mathcal{L})$.

The proof of this theorem will be discussed in detail in the end-notes of this chapter (page 439).

> ***Alternatives.*** *For proving that $\mathfrak{M}$ is a manifold, the standard development of the theory usually involves, instead of perturbing the metric alone, also perturbing the second Seiberg–Witten equation to $F_A^+ = \sigma(\varphi) + i\eta^+$ for some parameter $\eta^+ \in \Gamma(\Lambda_+^2)$. In fact, this is the approach that we will take in the end-notes. This perturbative approach is also fruitful for the application of Seiberg–Witten theory to symplectic manifolds, where a suitable η^+ is grown to infinity, as will be seen in the end-note on page 465. A proof that it is sufficient to perturb only the Riemannian metric and not the equations can be read from* **T. Friedrich***'s* ***Dirac operators in Riemannian geometry*** **[Fri00**, app A]*, after some background help from* **D. Ebin***'s* ***On the space of Riemannian metrics*** **[Ebi68]**.

The main consequence of the above theorem is that, when $b_2^+(M) \geq 2$, the Seiberg–Witten moduli space $\mathfrak{M}$ determines a well-defined bordism class inside $\Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L}) \ / \ \mathcal{G}$, which depends only on the manifold M and on the choice of spin$^{\mathbb{C}}$ structure $\mathfrak{s}$, but not on the Riemannian metric. Therefore, by evaluating various cohomology classes on $\mathfrak{M}$, we can obtain numerical invariants of the 4–manifold M, which will depend only on the Chern class $c_1(\mathfrak{s})$.

Most of them are empty. At most finitely-many spin$^{\mathbb{C}}$ structures are actually worth investigating:

Finiteness Theorem. *The Seiberg–Witten moduli space is non-empty for at most finitely-many spin$^{\mathbb{C}}$ structures.*

Proof. We first obtain a bound on the curvature F_A of any solution (φ, A), then remember that $[F_A] = -2\pi i\, c_1(\mathfrak{s})$ and restrict to positive-dimensional moduli spaces to conclude that $c_1(\mathfrak{s})$ must be confined to a finite subset of $H^2(M; \mathbb{Z})$.

Integral Curvature Bound. *If (φ, A) is a solution to the Seiberg–Witten equations, then we must have:*

$$2\sqrt{2}\ \left\|F_A^+\right\| \ \leq \ \left\|scal\right\| \ ,$$

where $\|\cdot\|$ denotes the L^2–norm $\|\alpha\|^2 = \int_M |\alpha|^2 \, vol_M$.

Proof of the integral curvature bound. Choose any solution (φ, A) of the Seiberg–Witten equations. Then $\mathcal{D}^A\varphi = 0$ and $F_A^+ = \sigma(\varphi)$. Plugging this into the Lichnerowicz formula

$$(\mathcal{D}^A)^*\, \mathcal{D}^A\, \varphi \ = \ (\nabla^A)^*\, \nabla^A \varphi \ + \ \tfrac{1}{4}\, scal \cdot \varphi \ + \ \tfrac{1}{2} F_A^+ \bullet \varphi$$

yields immediately

$$0 \ = \ (\nabla^A)^*\, \nabla^A \varphi \ + \ \tfrac{1}{4}\, scal \cdot \varphi \ + \ \tfrac{1}{4}\, |\varphi|^2 \cdot \varphi \, .$$

Take the inner-product with φ to get

$$0 = \left\langle (\nabla^A)^* \, \nabla^A \varphi, \, \varphi \right\rangle + \tfrac{1}{4} \left\langle scal \cdot \varphi, \, \varphi \right\rangle + \tfrac{1}{4} \left| \varphi \right|^4 .$$

Then integrate over M (using the Riemannian volume vol_M, which we drop from notation) and use that $(\nabla^A)^*$ is adjoint to ∇^A:

$$0 = \int \left| \nabla^A \varphi \right|^2 + \tfrac{1}{4} \int scal \cdot \left| \varphi \right|^2 + \tfrac{1}{4} \int \left| \varphi \right|^4 .$$

We rearrange this equality by separating the *scal*–term, then use the Cauchy–Schwarz inequality on the right and something-not-worth-a-name on the left:

$$\begin{aligned} \tfrac{1}{4} \int \left| \varphi \right|^4 &\leq \int \left| \nabla^A \varphi \right|^2 + \tfrac{1}{4} \int \left| \varphi \right|^4 = \\ &= \tfrac{1}{4} \int (-\, scal) \left| \varphi \right|^2 \leq \tfrac{1}{4} \Big(\int (scal)^2 \Big)^{1/2} \Big(\int \left| \varphi \right|^4 \Big)^{1/2} . \end{aligned}$$

Dropping the middle and canceling gives

$$\Big(\int \left| \varphi \right|^4 \Big)^{1/2} \leq \Big(\int (scal)^2 \Big)^{1/2} .$$

However, $\left| \sigma(\varphi) \right| = \frac{1}{2\sqrt{2}} \left| \varphi \right|^2$ and $\sigma(\varphi) = F_A^+$, so we write:

$$2\sqrt{2} \Big(\int \left| F_A^+ \right|^2 \Big)^{1/2} \leq \Big(\int (scal)^2 \Big)^{1/2} ,$$

which, when written in terms of L^2–norms is exactly the statement claimed above. □

We now relate $\| F_A^+ \|$ to $c_1(\mathcal{L})$.

Lemma Z. *Let α be any closed 2–form on a 4–manifold. Then:*

$$[\alpha] \cdot [\alpha] = \left\| \alpha^+ \right\|^2 - \left\| \alpha^- \right\|^2 .$$

Proof of lemma Z. Using that $\alpha^+ \wedge \alpha^- = 0$, that $* \, \alpha^+ = \alpha^+$ and $* \, \alpha^- = -\alpha^-$, and that $\beta \wedge * \, \beta = |\beta|^2$, we compute

$$\begin{aligned} [\alpha] \cdot [\alpha] &= \int \alpha \wedge \alpha = \int \left(\alpha^+ + \alpha^- \right) \wedge \left(\alpha^+ + \alpha^- \right) \\ &= \int \alpha^+ \wedge \alpha^+ + \int \alpha^- \wedge \alpha^- \\ &= \int \alpha^+ \wedge \left(* \, \alpha^+ \right) + \int \alpha^- \wedge \left(- * \alpha^- \right) \\ &= \int \left| \alpha^+ \right|^2 - \int \left| \alpha^- \right|^2 , \end{aligned}$$

which yields the promised formula. □

We apply this lemma to F_A and, since $\frac{i}{2\pi}[F_A] = c_1(\mathcal{L})$, we get

$$4\pi^2 c_1(\mathfrak{s})^2 = \left\|F_A^+\right\|^2 - \left\|F_A^-\right\|^2 \leq \left\|F_A^+\right\|^2 \leq \tfrac{1}{8}\left\|scal\right\|^2 .$$

We have obtained an upper bound on $c_1(\mathfrak{s}) \cdot c_1(\mathfrak{s})$, depending only on the geometry of M.

For a lower bound on $c_1(\mathcal{L}) \cdot c_1(\mathcal{L})$, we notice that it only makes sense to look at those moduli spaces that are expected to be of positive dimension. Using the formula $\dim \mathfrak{M} = \frac{1}{4}\left(c_1(\mathfrak{s})^2 - 2\chi(M) - 3\operatorname{sign} M\right)$, we conclude that

$$2\chi(M) - 3\operatorname{sign} M \leq c_1(\mathfrak{s}) \cdot c_1(\mathfrak{s}) \leq \tfrac{1}{32\pi^2}\left\|scal\right\|^2 .$$

Therefore only finitely-many choices of $c_1(\mathfrak{s})$ from the integral lattice $H^2(M;\mathbb{Z})$ have any chance to yield non-empty moduli spaces. □

Remember that, after we stated that $\mathfrak{M}$ was a manifold, we also claimed that we will obtain numerical invariants of $(M, \mathfrak{s})$ by evaluating cohomology classes on $\mathfrak{M}$. That requires a couple more properties for $\mathfrak{M}$:

They are compact. Of course, for anything like "evaluate cohomology classes on $\mathfrak{M}$" to work, we need $\mathfrak{M}$ to be compact. The miracle is that:

Compactness Theorem. *The moduli space $\mathfrak{M}$ is always compact.*

Idea of proof. First we obtain a pointwise *a priori* bound on $|\varphi|$:

Pointwise Curvature Bound. *If (φ, A) is a solution of the Seiberg–Witten equations, then either we have*

$$|\varphi|^2 \leq \max_{x \in M}\{-scal(x)\}$$

or φ is identically-zero.

Proof of the pointwise curvature bound. First off, notice that if a function $f\colon M \to \mathbb{R}$ has a local maximum at some $p \in M$, then it must have $(\Delta f)(p) \geq 0$, where $\Delta = -\sum \partial_{e_k}\partial_{e_k}$ is the Laplace operator.[26]

We apply this to the function $x \mapsto |\varphi(x)|^2$. We choose a orthonormal local frame $\{e_1, e_2, e_3, e_4\}$ in T_M, and we compute:

$$\begin{aligned}\Delta\left(|\varphi|^2\right) &= -\sum \partial_{e_k}\partial_{e_k}\langle\varphi, \varphi\rangle_{\mathbb{R}} \\ &= -\sum \partial_{e_k} 2\langle\nabla^A_{e_k}\varphi, \varphi\rangle \\ &= -\sum 2\langle\nabla^A_{e_k}\nabla^A_{e_k}\varphi, \varphi\rangle - \sum 2\langle\nabla^A_{e_k}\varphi, \nabla^A_{e_k}\varphi\rangle ,\end{aligned}$$

---

[26] Since $\Delta f = -\operatorname{trace}\operatorname{Hessian}(f)$, and a maximum at p implies that all the eigenvalues of $\operatorname{Hessian}(f)$ are non-positive, it follows that $(\Delta f)(p) \geq 0$.

where we have used that ∇^A is compatible with the fiber-metric of $\mathcal{W}^+$. We rearrange to

$$\Delta(|\varphi|^2) + 2\sum \left|\nabla^A_{e_k}\varphi\right|^2 = -\sum 2\langle \nabla^A_{e_k}\nabla^A_{e_k}\varphi,\ \varphi\rangle\,.$$

Assume now that $p \in M$ is the absolute maximum point of $|\varphi|^2$. Then $\Delta\big(|\varphi|^2\big)\,|_p \geq 0$. Hence at p we must have

$$-2\sum\langle \nabla^A_{e_k}\nabla^A_{e_k}\varphi,\ \varphi\rangle|_p \geq 0\,.$$

On the other hand, one can check directly (by inner-product with test-spinors ψ, integration over M, and using that on compact manifolds divergences integrate to zero) the following fact: If the chosen local frame $\{e_1, e_2, e_3, e_4\}$ in T_M is such that at p the Levi–Civitā connection has $\nabla_{e_i}e_j|_p = 0$ (so-called *geodesic coordinates* at p), then we have

$$(\nabla^A)^*\,\nabla^A\varphi = -\sum \nabla^A_{e_k}\nabla^A_{e_k}\varphi\,.$$

Therefore, at the maximum point p, we must have

$$2\langle(\nabla^A)^*\,\nabla^A\varphi,\ \varphi\rangle_{\mathbb{R}}\,|_p \geq 0\,.$$

On the other hand, starting again with the Lichnerowicz formula

$$(\mathcal{D}^A)^*\,\mathcal{D}^A\,\varphi = (\nabla^A)^*\,\nabla^A\varphi + \tfrac{1}{4}\,scal\cdot\varphi + \tfrac{1}{2}F_A^+\bullet\varphi$$

applied to a Seiberg–Witten solution (φ, A) exactly as in the proof of the integral curvature bound (page 398), we are led to

$$0 = \langle(\nabla^A)^*\,\nabla^A\varphi,\ \varphi\rangle + \tfrac{1}{4}\,scal\cdot|\varphi|^2 + \tfrac{1}{4}\,|\varphi|^4\,.$$

At the maximum point p, the first term is positive and that forces

$$\tfrac{1}{4}\,scal(p)\,\,|\varphi(p)|^2 + \tfrac{1}{4}\,|\varphi(p)|^4 \leq 0\,.$$

Assume now that φ is not everywhere-zero. Then $\varphi(p) \neq 0$ and we can cancel, obtaining

$$|\varphi(p)|^2 \leq -\,scal(p)\,.$$

Since $-\,scal(p) \leq \max\{-\,scal(x)\}$ and $|\varphi(x)| \leq |\varphi(p)|$, the result follows. □

Once this pointwise bound on $|\varphi|$ has been obtained, one uses a standard (but too long to explain here) "elliptic bootstrapping" argument to bound all higher derivatives of both φ and A, and thus deduces the compactness of the moduli space. □

Therefore, $\mathfrak{M}$ determines a well-defined homology class in its ambient.

They are orientable. Since we crave for more than merely a homology class modulo 2, we should be happy that:

Orientability Theorem. *The manifold $\mathfrak{M}$ is orientable. Its orientations correspond to orientations of the vector space $H^1(M;\mathbb{R}) \otimes H^2_+(M;\mathbb{R})$.* □

This result will be proved in the end-notes of this chapter (page 447).

The arbitrary nature of a choice of orientation of $\mathfrak{M}$ will make its presence felt in a lot of sign ambiguities in later formulae.

The invariant

Having thus obtained from the Seiberg–Witten equations some very nice moduli spaces, it can be further shown that the natural ambient of $\mathfrak{M}$, the space of all connection-and-spinor pairs modulo gauge-equivalence, has the homotopy type of $\mathbb{CP}^\infty$ (when M is simply-connected). Therefore, the cohomology ring of this ambient is $\mathbb{Z}[u]$ for a degree–2 class u. Hence, if $\mathfrak{M}$ is even-dimensional, then we can evaluate the appropriate class $u \cup \cdots \cup u$ on it and obtain a numerical invariant of M. Thus, we call

$$\mathcal{SW}_M(\mathfrak{s}) = \int_{\mathfrak{M}} u^k$$

the **Seiberg–Witten invariant** of the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$. It will depend only on M and $c_1(\mathfrak{s})$.

If the dimension of $\mathfrak{M}$ is odd, then all we can do is define

$$\mathcal{SW}_M(\mathfrak{s}) = 0\,,$$

and no information is obtained. Notice that the moduli space is odd-dimensional if and only if b_2^+ is *even,* and then the Seiberg–Witten invariants are blind.

> *The discussion in the non-simply-connected case is similar. The moduli spaces are either all even- or all odd-dimensional, depending only on whether*
>
> $$b_2^+(M) + b_1(M) + 1$$
>
> *is even or odd. In particular, if $b_2^+(M) + b_1(M)$ is* even, *then the homology class of $\mathfrak{M}$ is trivial and the Seiberg–Witten invariants tell us nothing.*

A bit more about this cohomology-evaluation procedure is explained in the end-notes of this chapter (page 452).

Simple type. This whole issue of evaluating cohomology classes on $\mathfrak{M}$ in order to obtain numerical invariants of M might prove to be rather moot:

Simple Type Conjecture (*open*). *For any simply-connected 4–manifold with $b_2^+ \geq 2$, if the the Seiberg–Witten moduli space $\mathfrak{M}$ is non-empty, then it must be zero-dimensional, and thus consist of finitely-many isolated points.*

If this conjecture were somehow proved, then for obtaining a numerical invariant from the Seiberg–Witten equations it would be enough to merely count (with signs) their solutions.

A large class of 4–manifolds for which the above conjecture is proved to hold are all the symplectic manifolds. Further, there is no known example of a simply-connected manifold with $b_2^+ \geq 2$ that has higher-dimensional moduli spaces.

On the other hand, there are plenty of examples, both of non-simply-connected manifolds and of manifolds with $b_2^+ = 1$, that each have Seiberg–Witten moduli spaces of *arbitrarily-high* dimensions.

A 4–manifold for which only zero-dimensional moduli spaces appear is said to be of *(Seiberg–Witten)* **simple type.**

It is interesting to notice that the zero-dimensional moduli spaces occur exactly for those spin$^{\mathbb{C}}$ structures for which

$$c_1(\mathfrak{s}) \cdot c_1(\mathfrak{s}) = 2\chi(M) + 3 \operatorname{sign} M \ .$$

These are exactly the spin$^{\mathbb{C}}$ structures $\mathfrak{s}_J$ that are induced from almost-complex structures. Notice that in particular $b_2^+(M) + b_1(M)$ would again have to be *odd* for any information to be obtainable.[27]

If the above conjecture turns out to be correct, then one could think of the Seiberg–Witten invariant as an invariant not of spin$^{\mathbb{C}}$ structures on 4–manifolds, but rather of their almost-complex structures.

Conclusion. In what follows, we will assume that the simple type conjecture is true. We will entirely restrict the discussion to simply-connected manifolds M with $b_2^+(M) \geq 2$. (While the invariants are quite useful and well-understood when $b_2^+(M) = 1$, we will not discuss that case here.)

Specifically, our assumptions imply that, for every spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ on M, the corresponding moduli space $\mathfrak{M}$ is either empty or a finite set of points. It can be oriented, and then the algebraic count of its points is independent of the auxiliary Riemannian metric. Hence this count, denoted by

$$\mathcal{SW}_M(\mathfrak{s}) \ ,$$

depends only on the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ and on the smooth topology of M. For us, this will be the Seiberg–Witten invariant.

We are ready to start investigating its properties:

---

[27] Remember that if M admits an almost-complex structure, then $b_2^+(M) + b_1(M)$ must be odd.

10.4. Main results and properties

As announced earlier, we tried to make it possible to start reading this chapter right here, and in this manner skip all technicalities related to the definition of the Seiberg–Witten invariants. Indeed, all one needs is trust that something called "the Seiberg–Witten invariant" exists as described below. The rest of this chapter, up to the start of the end-notes, could be viewed as a survey of Seiberg–Witten theory.

The invariants. The Seiberg–Witten invariant is the map

$$\mathcal{SW}_M\colon \{\text{spin}^{\mathbb{C}} \text{ structures on } M\} \longrightarrow \mathbb{Z}\,,$$

with $\mathcal{SW}_M(\mathfrak{s})$ defined by counting with signs the solutions of the Seiberg–Witten equations for the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$, considered up to gauge-equivalence:

$$\mathcal{SW}_M(\mathfrak{s}) = \#\big(\,\{\text{solutions } (\varphi, A) \text{ of } \mathcal{D}^A\varphi = 0 \text{ and } F_A^+ = \sigma(\varphi)\}\;/\;\mathcal{G}(\mathcal{L})\,\big)\,.$$

This count is well-defined when $b_2^+(M) \geq 2$. (It is still quite manageable when $b_2^+(M) = 1$, but for simplicity we avoid that case.)

Characteristic version. In case $H^2(M;\mathbb{Z})$ has no 2–torsion (for example if M is simply-connected), then we can uniquely identify each spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ on M by its Chern class $c_1(\mathfrak{s})$, which is always an integral lift of $w_2(T_M)$. Thus we can think of the Seiberg–Witten invariant as a map defined on the characteristic elements of M:

$$\mathcal{SW}_M\colon \{\underline{w} \in H^2(M;\mathbb{Z}) \mid \underline{w} = w_2(M) \pmod 2\} \longrightarrow \mathbb{Z}\,.$$

(Even when $H^2(M;\mathbb{Z})$ has 2–torsion, and thus several spin$^{\mathbb{C}}$ structures correspond to a same Chern class, one can still obtain an invariant defined on characteristic elements, by summing over all corresponding spin$^{\mathbb{C}}$ structures.)

Parametrized version. An alternative version is to use the parametrization of spin$^{\mathbb{C}}$ structures by the elements of $H^2(M;\mathbb{Z})$, once an "origin" is chosen. That is to say, after choosing your favorite spin$^{\mathbb{C}}$ structure $\mathfrak{s}_0$, you can view the Seiberg–Witten invariant as a map

$$\mathcal{SW}_M(\mathfrak{s}_0 + \,\cdot\,)\colon\; H^2(M;\mathbb{Z}) \longrightarrow \mathbb{Z}$$

by using the free transitive action of $H^2(M;\mathbb{Z})$ on the set of all spin$^{\mathbb{C}}$ structures. The latter uniquely lists all spin$^{\mathbb{C}}$ structures on M as $\mathfrak{s}_0 + \varepsilon$ for $\varepsilon \in H^2$. Recall that we have $c_1(\mathfrak{s}_0 + \varepsilon) = c_1(\mathfrak{s}_0) + 2\varepsilon$.

Basic classes. A class $\kappa \in H^2(M;\mathbb{Z})$ such that $\mathcal{SW}_M(\kappa) \neq 0$ is called a **basic class.** Since at most finitely many spin$^{\mathbb{C}}$ structures have non-empty moduli spaces, things are pretty tight: *Every manifold M has at most finitely-many basic classes.*

We are now ready to review the main results that govern these invariants:

Involution. Immediately, by complex-conjugating the spin$^{\mathbb{C}}$ structure and everything it might support, we get:

Involution Lemma. *If $b_2^+(M) \geq 2$, we have the symmetry:*

$$\mathcal{SW}_M(-\kappa) = \pm\, \mathcal{SW}_M(\kappa)\,. \qquad \square$$

Scalar curvature. The possible Riemannian geometries of M have influence on the invariants:

Vanishing Theorem for Positive Scalar Curvature. *If the 4–manifold M has $b_2^+(M) \geq 2$ and it admits a Riemannian metric of everywhere-positive scalar curvature, then*

$$\mathcal{SW}_M \equiv 0\,.$$

Proof. Once again, we use the Lichnerowicz formula (page 393),

$$(\mathcal{D}^A)^*\,\mathcal{D}^A\,\varphi = (\nabla^A)^*\,\nabla^A\varphi + \tfrac{1}{4}\,scal\cdot\varphi + \tfrac{1}{2}F_A^+\bullet\varphi\,.$$

The argument will be an instance of what is usually called the Bochner technique, *i.e.*, obtaining vanishing results from Weitzenböck-type formulae.[28]

Exactly as in the proof of the integral curvature bound (page 398), we assume that (φ, A) is a Seiberg–Witten monopole, so that $\mathcal{D}^A\varphi = 0$ and $F_A^+ = \sigma(\varphi)$. We plug into the formula, use that $\sigma(\varphi)\bullet\varphi = \frac{1}{2}\left|\varphi\right|^2\cdot\varphi$, then take the inner-product with φ and integrate over M:

$$0 = \int_M \left|\nabla^A\varphi\right|^2 + \tfrac{1}{4}\int_M scal\cdot\left|\varphi\right|^2 + \tfrac{1}{4}\int_M \left|\varphi\right|^4\,.$$

After staring at this formula we see that, if there are any monopoles (φ, A) on M, then the scalar curvature *scal* must be somewhere negative. (If *scal* were everywhere-zero, then the only monopoles would be those with $\varphi \equiv 0$, but those are reducibles and were avoided in the first place.) Of course, this vanishing result also follows directly from the pointwise curvature bound on page 400. □

---

[28] More examples of Bochner-technique results are quoted in the end-notes of this chapter (page 474).

Connected sums. From a stretching argument, we get:

Vanishing Theorem for Connected Sums. *Assume that the* 4*–manifold* M *smoothly splits as a connected sum*

$$M = N' \# N'' ,$$

with both $b_2^+(N') \geq 1$ *and* $b_2^+(N'') \geq 1$. *Then*

$$\mathcal{SW}_M \equiv 0 .$$

> ***Sketch of proof.*** This is proved by metrically-stretching the length of the connecting cylinder
>
> $$S^3 \times [0,1]$$
>
> between N' and N'' in $N' \# N''$. When stretching, the geometry of $N' \# N''$ becomes on average dominated by the connecting cylinder $S^3 \times [0,1]$, which is easily arranged to have positive scalar curvature. This implies that all Seiberg–Witten solutions must vanish on this cylinder.
>
> Therefore any solution on $N' \# N''$ must come from a solution on N' and a solution on N''. In other words, for every spin$^{\mathbb{C}}$ structure on M we have
>
> $$\mathfrak{M}_{N' \# N''} = \mathfrak{M}_{N'} \times \mathfrak{M}_{N''} .$$
>
> When writing this equality, we use that every spin$^{\mathbb{C}}$ structure $\mathfrak{s}_{N' \# N''}$ on $N' \# N''$ can be nicely split as $\mathfrak{s}_{N' \# N''} = \mathfrak{s}_{N'} \# \mathfrak{s}_{N''}$ for some spin$^{\mathbb{C}}$ structures $\mathfrak{s}_{N'}$ on N' and $\mathfrak{s}_{N''}$ on N''. Their Chern classes add in the obvious manner.
>
> Since $\chi(N' \# N'') = \chi(N') + \chi(N'') - 2$, the dimension formula (page 397) leads to
>
> $$\dim \mathfrak{M}_{N' \# N''} = \dim \mathfrak{M}_{N'} + \dim \mathfrak{M}_{N''} + 1 .$$
>
> Assume for simplicity that $N' \# N''$ is of simple type, and thus that the only interesting moduli spaces are those of dimension 0. However, $\mathfrak{M}_{N' \# N''}$ having dimension zero implies that one of $\mathfrak{M}_{N'}$ or $\mathfrak{M}_{N''}$ must have (virtual) dimension -1, which means that it is empty. This is ensured by the fact that both N' and N'' have $b_2^+ \geq 1$, and thus the moduli spaces are either empty or manifolds of the expected dimension.
>
> The conclusion is that $\mathfrak{M}_{N' \# N''}$ must be empty for every spin$^{\mathbb{C}}$ structure on M. (Compare also with the stretching argument in the proof of the adjunction inequality in the end-notes of the next chapter, on page 496.) □

This vanishing result is often used to show that manifolds with nontrivial Seiberg–Witten invariants are indecomposable as smooth connected sums. For example, if M has nontrivial invariants, then it cannot split into a sum of two symplectic manifolds (which must both have $b_2^+ \geq 1$ owing to their symplectic class).

A drawback of this vanishing property is that the Seiberg–Witten invariants are powerless on most connected sums, and thus no consequences (like minimum genus for embedded surfaces) can be obtained from them. Nonetheless, S. Bauer and M. Furuta[29] *have obtained a refinement of the Seiberg–Witten invariant in terms of equivariant cohomotopy that does survive connected sums and can thus be used in such cases.*

Close to complex. The Seiberg–Witten invariants collaborate well with the complex realm. For instance, we have:

Blow-Up Formula. *Let M be simply-connected, with $b_2^+(M) \geq 2$, and of Seiberg–Witten simple type.*[30] *Let $\{\kappa_i\}$ be the basic classes of M. Then the (topological) blow-up*

$$M \# \overline{\mathbb{CP}}^2$$

has basic classes $\{\kappa_i \pm E\}$, where E is the class of the (-1)-curve $\overline{\mathbb{CP}}^1 \subset \overline{\mathbb{CP}}^2$. We have

$$\mathcal{SW}_{M\#\overline{\mathbb{CP}}^2}(\kappa_i \pm E) = \pm\, \mathcal{SW}_M(\kappa_i)\ .$$

Sketch of proof. Similarly to the strategy used for the vanishing theorem for connected sums, we start by stretching the connecting cylinder $\mathbb{S}^3 \times [0,1]$ between M and $\overline{\mathbb{CP}}^2$, and thus reduce the moduli space to

$$\mathfrak{M}_{M\#\overline{\mathbb{CP}}^2} = \mathfrak{M}_M \times \mathfrak{M}_{\overline{\mathbb{CP}}^2}\ .$$

However, $\overline{\mathbb{CP}}^2$ is simply-connected and has $b_2^+ = 0$. Therefore, from results proved in the end-notes of the preceding chapter (page 357), up to gauge-equivalence there is exactly one reducible solution $(0, A)$ on $\overline{\mathbb{CP}}^2$. Furthermore, since $\overline{\mathbb{CP}}^2$ admits a metric of positive scalar curvature, no other monopoles appear. Therefore

$$\mathfrak{M}_{\overline{\mathbb{CP}}^2} = \{point\}$$

and hence $\mathfrak{M}_{M\#\overline{\mathbb{CP}}^2} \approx \mathfrak{M}_M$.

Finally, if all non-empty moduli spaces of M (corresponding to the κ_i's) were 0–dimensional, then the only moduli spaces of $M\#\overline{\mathbb{CP}}^2$ with expected non-negative dimension are those corresponding to $\kappa_i \pm E$, for which $\dim \mathfrak{M}_{\overline{\mathbb{CP}}^2} = -1$ and $\dim \mathfrak{M}_{M\#\overline{\mathbb{CP}}^2} = 0$. □

---

[29] See **S. Bauer** and **M. Furuta**'s ***A stable cohomotopy refinement of Seiberg–Witten invariants: I & II*** **[BF04]** and **S. Bauer**'s exposition ***Refined Seiberg–Witten invariants*** **[Bau03]**.

[30] Simple type means that M has non-zero Seiberg–Witten invariants only for spin$^{\mathbb{C}}$ structures induced from almost-complex structures. There are no known examples of simply-connected manifolds with $b_2^+ \geq 2$ that are not of simple type.

When M is not of simple type, a similar blow-up formula holds, but its proof is more delicate.[31]

Seiberg–Witten theory also collaborates well with the "close-to-complex" realm of symplectic manifolds (keep in mind that all Kähler surfaces are symplectic manifolds). A first instance of this is:

Non-Vanishing Theorem for Symplectic. *Let M be a simply-connected 4–manifold with $b_2^+(M) \geq 2$. If M admits a symplectic structure ω, then $K_M^* = c_1(\omega)$ is a basic class, and*

$$\mathcal{SW}_M(\pm K_M^*) = \pm 1 . \qquad \square$$

See the next section for a more thorough discussion of the symplectic case, and the end-notes of this chapter (page 465) for a proof.

Combining the non-vanishing result above with the earlier vanishing for connected sums yields:

Corollary (*Symplectics are irreducible*). *If M is any 4–manifold admitting a symplectic structure, then it cannot smoothly split as a connected sum $M = N' \# N''$ with $b_2^+(N') \geq 1$ and $b_2^+(N'') \geq 1$.* □

Genus bounds. A most remarkable feature of the Seiberg–Witten equations is that they give minimum genus bounds for all embedded surfaces that represent a given homology class.

Adjunction Inequality. *Let M be a smooth 4–manifold with $b_2^+(M) \geq 2$. Let S be any connected surface embedded in M such that either:*

— *$S \cdot S \geq 0$ and S is homologically nontrivial;*

— *M is of simple type (e.g., symplectic), and S is not a sphere.*

Then, for every basic class κ of M, we must have:

$$\chi(S) + S \cdot S \leq -\left|\kappa \cdot S\right| . \qquad \square$$

Since both $S \cdot S$ and $\kappa \cdot S$ depend only on the homology class of S, this means that the inequality offers upper bounds on the Euler–Poincaré characteristic $\chi(S)$ of a surface representing a fixed homology class. Notice that $\chi(S) = 2 - 2\,\text{genus}(S)$, and hence the inequality offers a lower bound on the genus needed for representing a fixed homology class. A more thorough discussion of this result will be taken up in the next chapter (starting on page 481), which is devoted to the problem of the minimum genus of surfaces.

---

[31] See for example **R. Fintushel** and **R. Stern**'s outline in ***Immersed spheres in 4–manifolds and the immersed Thom conjecture*** **[FS95]**.

Backwards usage. The adjunction inequality can also be applied backwards: If one knows the genus of some embedded surfaces in M, then one might determine the basic classes of M, or at least get restrictions on them.

For example, if M contains a torus T of self-intersection 0, then for every basic class κ we must have $\kappa \cdot T = 0$. By looking at characteristic elements orthogonal to T, certain exclusions should appear. Such an argument can be used together with a good understanding of the $K3$ surface to show that its only basic class is its canonical class.[32]

Another example of such backwards use of the adjunction inequality is:

Corollary *(Vanishing from spheres). If M contains a homologically-nontrivial embedded sphere S with $S \cdot S \geq 0$, then $\mathcal{SW}_M \equiv 0$.* □

Gluing formulae. Finally, note that the Seiberg–Witten invariants satisfy various gluing formulae for cutting and gluing 4–manifolds along 3–manifolds. The simplest cases, when we glue along a 3–torus, will be stated in section 12.1 (page 532). More general cases involve the Seiberg–Witten–Floer homology of the 3–manifold and will not be discussed in this volume.

10.5. Invariants of symplectic manifolds

The Dirac operator, as all elliptic operators, is closely related to the Cauchy–Riemann operators. Thus, it should be of no surprise that the best use of the Seiberg–Witten invariants occurs close to the complex world. (As an example, we have already seen the blow-up formula above.)

The Seiberg–Witten invariants are very well understood on Kähler surfaces, but one can actually extend this control a bit farther from the complex realm, namely to all symplectic 4–manifolds.

The results of this section are due to E. Witten in the Kähler case, and have been painfully extended to the symplectic realm by C. Taubes.

> ***Review.*** *For the readers who skipped over the beginning of this chapter, we rapidly review a few notions that were explained in section 10.1 (page 376).*
>
> *An* **almost-complex structure** *on M is any fiber-preserving automorphism $J\colon T_M \to T_M$ that mimics multiplication with i by satisfying $J(Jv) = -v$. We require that J orients T_M in the same way as the chosen orientation of M. Denote by K_M^* both the* **anti-canonical bundle**[33] *$K_M^* = \det_{\mathbb{C}}(T_M, J)$ and its Chern class $c_1(K_M^*) = c_1(T_M, J)$.*

---

[32] Explained for example in **T. Lawson**'s survey ***The minimal genus problem*** [**Law97**].

[33] It is worth noticing that, while complex geometry prefers to deal with K_M, gauge theory is better written in terms of its dual bundle K_M^*.

> *A* **symplectic structure** *on M is any exterior 2–form $\omega \in \Gamma(\Lambda^2(T_M^*))$ with both $\omega \wedge \omega > 0$ and $d\omega = 0$. Notice that all symplectic manifolds automatically have $b_2^+ \geq 1$, since the class of ω has positive self-intersection.*
>
> *An almost-complex structure J is called* compatible *with a symplectic form ω if $\omega(x,y) = \omega(Jx, Jy)$ and $\omega(x, Jx) > 0$. Any symplectic structure admits such compatible almost-complex structures, and they are unique up to homotopies. In particular, the anti-canonical bundle K_M^* of any such compatible almost-complex structure depends only on ω.*

Preparation. Assume that M has a fixed symplectic structure ω and pick some compatible almost-complex structure J. The almost-complex structure J determines a distinguished spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J$ with determinant line-bundle $\mathcal{L} = K_M^*$. Using this $\mathfrak{s}_J$, we can parametrize all other spin$^{\mathbb{C}}$ structures $\mathfrak{s}$ on M by writing them as translations $\mathfrak{s} = \mathfrak{s}_J + \varepsilon$, with ε running over $H^2(M;\mathbb{Z})$. In terms of their Chern classes, this becomes $c_1(\mathfrak{s}_J + \varepsilon) = K_M^* + 2\varepsilon$. Thus the set of all spin$^{\mathbb{C}}$ structures on M can be written $\{K_M^* + 2\varepsilon \mid \varepsilon \in H^2(M;\mathbb{Z})\}$, and we consider the latter as the domain of the Seiberg–Witten invariant $\mathcal{SW}_M$:

$$\mathcal{SW}_M \colon \{K_M^* + 2\varepsilon \mid \varepsilon \in H^2(M;\mathbb{Z})\} \longrightarrow \mathbb{Z}\,.$$

The expected dimension of the corresponding moduli spaces is

$$\dim \mathfrak{M}_{K_M^*+2\varepsilon} = K_M^* \cdot \varepsilon + \varepsilon \cdot \varepsilon\,.$$

First results. An important result for symplectic 4–manifolds is:

Simple Type Theorem. *All symplectic 4–manifolds having $b_2^+ \geq 2$ are of Seiberg–Witten simple type.* □

Specifically, this means that $\mathcal{SW}_M(K_M^* + 2\varepsilon)$ can be nonzero only when

$$\varepsilon \cdot \varepsilon = -K_M^* \cdot \varepsilon\,.$$

Further, we have already stated C. Taubes' result that

The class K_M^* is a basic class, and we have[34] $\mathcal{SW}_M(\pm K_M^*) = \pm 1$.

This is strengthened by the following restriction:

Lemma. *If $\mathcal{SW}_M(K_M^* + 2\varepsilon) \neq 0$, then ε must satisfy the inequalities*

$$0 \leq \varepsilon \cdot [\omega] \leq -K_M^* \cdot [\omega]\,,$$

with equality allowed only for $\varepsilon = 0$ and $\varepsilon = -K_M^*$. □

In particular, since the $K3$ surface has $K_{K3}^* = 0$, these results show that

Corollary (*Seiberg–Witten on K3*). *The only Seiberg–Witten basic class of the $K3$ manifold is the trivial class 0, with value $\mathcal{SW}_{K3}(0) = \pm 1$.* □

---

[34] This result will be proved in the end-notes of this chapter (page 465).

Example. *This can be put to use to exhibit two homeomorphic but non-diffeomorphic manifolds: Build the manifold* $K3\,\#\,\overline{\mathbb{CP}}^2$ *(the blow-up of* $K3$*). Its intersection form is odd indefinite and isomorphic to the intersection form of* $\#3\,\mathbb{CP}^2\,\#20\,\overline{\mathbb{CP}}^2$*. Then Freedman's classification implies that the manifolds*

$$K3\,\#\,\overline{\mathbb{CP}}^2 \qquad \text{and} \qquad \#3\,\mathbb{CP}^2\,\#20\,\overline{\mathbb{CP}}^2$$

must be homeomorphic. However, the Seiberg–Witten invariants show that they are not diffeomorphic: Indeed, $K3\,\#\,\overline{\mathbb{CP}}^2$ *has basic classes* $\pm E$*, while* $\#3\,\mathbb{CP}^2\,\#20\,\overline{\mathbb{CP}}^2$ *must have trivial Seiberg–Witten invariants, since it can be easily split into a connected sum of two manifolds with* $b_2^+ \geq 1$*.*

Seiberg–Witten and J–holomorphic curves. We have reached some of the most remarkable results about the Seiberg–Witten invariants on symplectic manifolds: they can be interpreted as Gromov-type invariants that count the J–holomorphic curves of M. Recall that a **J-holomorphic curve** is any surface S whose T_S is J–invariant.

The stunning relation of J–holomorphy to Seiberg–Witten theory is:

Taubes' Theorem. *Let* M *be a symplectic* 4*-manifold with symplectic form* ω *and with* $b_2^+(M) \geq 2$*. Assume that, for some* $\varepsilon \in H^2(M;\mathbb{Z})$*, we have* $\mathcal{SW}_M(K_M^* + 2\varepsilon) \neq 0$*. Then, for any generic choice of almost-complex structure* J *compatible with* ω*, only finitely-many* J*–holomorphic curves will represent the class* ε*. Further, we have:*

$$\mathcal{SW}_M(K_M^* + 2\varepsilon) \;=\; \#\{\,J\text{–holomorphic curves of class } \varepsilon\,\}\;.$$

Of course, this count of J–holomorphic curves is an algebraic count,[35] and thus one needs first to assign the appropriate sign to each curve. This is a delicate issue (especially for counting multiply-covered tori) which we will not pursue.

We should mention though that, in case M is actually Kähler, then all curves have *positive* sign. The argument in the Kähler case is detailed in the end-notes of this chapter (page 457). We should emphasize though that the requirement that J be *generic* is important. A Kähler structure is rather special and therefore the above result might not apply directly to that case.

As far as arguing for the general symplectic case, the best we can do is provide a vague outline:

Idea of proof. Taubes uses a deformation of the Seiberg–Witten equations, pushing them closer and closer toward a Cauchy–Riemann operator on the line bundle L_ε of Chern class ε. In the end, a Seiberg–Witten

---

[35] For a bit more on this Gromov-type invariant, see the end-notes of this chapter (page 471).

solution corresponds through this deformation to an (almost-)holomorphic section of L_ε, and thus the section's zero-set will in the limit be a J–holomorphic curve representing ε. The analysis involved is heroic, and the full proof fills up a 400–page book.[36] □

At the very least, from Taubes' theorem we know that, if $\mathcal{SW}_M(K_M^* + 2\varepsilon) \neq 0$, then the class ε can be represented by *at least one* J–holomorphic curve.

In particular, we notice that both K_M^* and $-K_M^*$ can always be represented by J–holomorphic curves. With a bit of care, this leads to

Corollary. *Let M be a symplectic manifold with $b_2^+(M) \geq 2$ and assume that*

$$K_M^* \cdot K_M^* < 0\,.$$

Then, for a generic almost-complex structure J, there exists an embedded J–holomorphic sphere S with $S \cdot S = -1$. Therefore S can be blown down[37] *and yields a decomposition*

$$M = N \# \overline{\mathbb{CP}}^2$$

involving some other symplectic 4–manifold N. □

Finally, since the general symmetry $\mathcal{SW}_M(-\kappa) = \pm\, \mathcal{SW}_M(\kappa)$ can in our case be written $\mathcal{SW}_M(K_M^* + 2\varepsilon) = \pm\, \mathcal{SW}_M\big(K_M^* + 2(K_M - \varepsilon)\big)$, we can combine it with Taubes' theorem and obtain a symplectic analogue of Serre duality:

Corollary. *If $\mathcal{SW}_M(K_M^* + 2\varepsilon) \neq 0$, then for a generic J we have:*

$$\begin{aligned} &\#\{\,J\text{–holomorphic curves of class } \varepsilon\,\} \\ &\qquad = \pm\,\#\{\,J\text{–holomorphic curves of class } K_M - \varepsilon\,\}\,. \end{aligned}$$ □

The proof of some of the results above in the special case of Kähler surfaces is contained in the end-notes of this chapter (page 457), followed by an argument that $\mathcal{SW}_M(\pm K^*) = \pm 1$ for general symplectic manifolds (page 465).

10.6. Invariants of complex surfaces

The Seiberg–Witten invariants are completely understood on complex surfaces. However, to quote from R. Friedman and J. Morgan's **[FM99]**: *the geometric interest of the Seiberg–Witten moduli spaces of a surface X is in a certain sense inversely proportional to the interest in X itself as an abstract surface.* In other words, the more well-understood X is, the richer the structure of

---

[36] C. Taubes' ***Seiberg–Witten and Gromov invariants for symplectic 4–manifolds*** [Tau00a].

[37] Blowing-down a J–holomorphic (-1)–curve can always be done so as to preserve symplectic structures. Similarly, a blow-up of a symplectic manifold is still symplectic.

the Seiberg–Witten invariants is; and the more arcane X is, the less light is shed by them.

If the complex surface is Kähler, then it is also symplectic, and all the considerations from the preceding section automatically apply. However, observe that the suitable "generic J" often does not correspond to the actual complex structure of the surface.

We certainly have

$$\mathcal{SW}_M(\pm K_M) = \pm 1 \, ,$$

as well as the blow-up formula:

Let M be simply-connected, with $b_2^+ \geq 2$, and of Seiberg–Witten simple type. Let $\{\kappa_i\}$ be the basic classes of M. Then the blow-up $M \# \overline{\mathbb{CP}}^2$ has basic classes $\{\kappa_i \pm E\}$, where E is the class of the (-1)-curve.

In light of this, it enough to concern ourselves with complex surfaces that are minimal with respect to blow-downs. Among these, the rational and ruled are not very interesting. We are left with the elliptic surfaces and the mysterious surfaces of general type.

Elliptic surfaces. First, the rational elliptic surface $E(1) = \mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$ has $b_2^+ = 1$ and admits a metric of positive scalar curvature.

Past this case, the basic classes of the elliptic surfaces $E(n)$ are all multiples of the generic torus fiber F, specifically

$$\{k[F] \mid k = n \pmod 2; \quad |k| \leq n-2\} \, .$$

The corresponding values of the Seiberg–Witten invariant are

$$\mathcal{SW}_{E(n)}(k[F]) = \pm \binom{n-2}{|k|} \, .$$

In particular, again, the only basic class of a $K3$ surface is its canonical class $K_{K3} = 0$, with invariant ± 1.

For $E(n)_{p,q}$, first recall[38] that a logarithmic transformation of multiplicity m creates a multiple fiber F_m with $[F_m] = \frac{1}{m}[F]$. Consider the elliptic surface $E(n)_{p,q}$, with $n \geq 2$ and $\gcd(p,q) = 1$. Denote by $[F_{pq}]$ the integral class $[F_{pq}] = \frac{1}{pq}[F]$. Then the basic classes of $E(n)_{p,q}$ can be listed as

$$\{k[F_{pq}] \mid k \equiv npq - p - q \pmod 2; \quad |k| \leq npq - p - q\} \, ,$$

and we have

$$\mathcal{SW}_{E(n)_{p,q}}(k[F_{pq}]) = \pm \binom{n-2}{|k|} \, .$$

---

[38] From section 8.3 (page 310).

Notice that, in terms of the class $[F_{pq}]$, the canonical class of $E(n)_{p,q}$ can be written

$$K_{E(n)_{p,q}} = (npq - p - q)[F_{pq}] .$$

Thus, all basic classes are rational multiples ρK_E with $|\rho| \leq 1$. A first step toward the proof of this result will be made in the end-notes of this chapter (page 463).

These results can be used to show that the various $E(n)_{p,q}$ are non-diffeomorphic even while many are homeomorphic, and thus exhibit infinite families of homeomorphic but non-diffeomorphic 4–manifolds, as we stated back in section 8.4 (page 314).

We will revisit these results in section 12.1, in the context of generalized fiber sums (page 534) and generalized logarithmic transformations (page 536), along with other gluing results for the Seiberg–Witten invariants. In particular, the above formulae for the Seiberg–Witten invariants of elliptic surfaces will get a concise rewriting, which is much easier to remember.

General type surfaces. For complex surfaces of general type, the Seiberg–Witten invariants see little: the *only* basic classes of such a surface M are $\pm K_M$, with values[39] ± 1. Nonetheless, this still implies that $\pm K_M$ is a diffeomorphism invariant, and in this manner have been found examples of homeomorphic but non-diffeomorphic surfaces of general type, smoothly distinguished by the divisibility of their canonical class.[40]

---

[39] We will prove this in the end-notes of this chapter (page 462).

[40] See **W. Ebeling**'s ***An example of two homeomorphic, nondiffeomorphic complete intersection surfaces*** [**Ebe90**].

10.7. Notes

Introduction

The first note (page 416) is a brief presentation of Lefschetz pencils and fibrations. These are singular fibrations on 4–manifolds, somewhat similar to elliptic fibrations, whose existence is essentially equivalent to a symplectic structure. We also quote the recent extension of such structures to *all* 4–manifolds with $b_2^+ \geq 1$.

The second note (page 420) proves the existence results for almost-complex structures that were stated in section 10.1, namely that the existence of an almost-complex structure is equivalent to finding a candidate for its Chern class, and that in most cases this is equivalent to $b_2^+(M) + b_1(M)$ being odd.

The third note (page 423) presents a few more details on spin$^\mathbb{C}$ structures. It starts with a Čech cohomology argument that every 4–manifold admits a spin$^\mathbb{C}$ structure (reading the earlier note about Čech cohomology, on page 189, is a requisite). Then it shows how the spinor bundles can be viewed as creatures parallel to the bundles of self-dual/anti-self-dual 2–forms (thus justifying the names of "self-dual/anti-self-dual spinor" that we have preferred over the more customary "positive/negative spinor"). After that, it notes that every spin$^\mathbb{C}$ structure induces an almost-complex structure on the 3–skeleton of M, thus complementing the induction of a spin$^\mathbb{C}$ structure by an almost-complex structure.

The fourth note (page 427) contains a description of the spin$^\mathbb{C}$ cobordism group (reading the note about cobordism groups on page 227 should help). It is identified with the so-called characteristic cobordism group (of 4–manifolds and characteristic surfaces), and the latter is evaluated using an argument that relies on Wall's theorems. The latter cobordism group will be used in the end-notes of the next chapter (page 502) to prove Rokhlin's theorem and its generalizations.

The fifth note (page 432) uses the division algebra of quaternions to obtain concrete descriptions of the Lie groups, bundle cocycles, and connections associated with spin$^{\mathbb{C}}$ structures. It ends with comments on the case when the spin$^{\mathbb{C}}$ structure comes from an almost-complex structure, and in particular proves that, in the Kähler case, there is a connection on $\mathcal{L}$ so that the Dirac operator is just $\sqrt{2}(\bar{\partial} \oplus \bar{\partial}^*)$.

The sixth note (page 439) proves many of the statements made in section 10.3 (page 396) about the Seiberg–Witten moduli space. Notice that a detailed discussion of the reducible Seiberg–Witten solutions was included back in the end-notes of the preceding chapter (page 357): indeed, a reducible monopole is simply an anti-self-dual connection on $\mathcal{L}$. Also, the proof of compactness was outlined in the main text. In this note, after taking on faith that $\mathfrak{M}$'s natural ambient $\Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L}) \,/\, \mathcal{G}(\mathcal{L})$ is an infinite-dimensional manifold, we prove that $\mathfrak{M}$ is itself a finite-dimensional manifold, compute its dimension, and show that it is orientable. We also comment on how to obtain invariants when $\mathfrak{M}$ is not zero-dimensional.

The seventh note (page 454) presents the Seiberg–Witten-based proof of Donaldson's theorem.[1] Since the proof involves the study of the moduli space $\mathfrak{M}$ around the inevitable reducible solution, reading the preceding note is a requisite.

The eighth note (page 457) proves that, on a Kähler manifold, the Seiberg–Witten monopoles correspond to complex curves representing a given homology class. A requisite is reading the end-notes of the preceding chapter (page 365), about complex-valued form, $\bar{\partial}$–operators, and holomorphic bundles.

The ninth note (page 465) extends the discussion to the case of symplectic manifold. It only proves the simplest result, specifically that $\mathcal{SW}_M(K^*) = \pm 1$.

The tenth note (page 471) describes the better setting for Taubes' theorem, which relates the Seiberg–Witten invariants on a symplectic manifold with a count of J–holomorphic curves.

The eleventh and last note (page 474) presents a few general differential-geometric results coming from Weitzenböck-type formulae through the use of the Bochner technique. These formulae are similar to the coupled Lichnerowicz formula (page 393) that we used extensively inside the chapter.

Note: Lefschetz pencils and fibrations

Definitions. A *(topological[2])* **Lefschetz fibration** on a simply-connected 4–manifold M is a smooth map $f\colon M \to S^2$ whose generic fiber $f^{-1}[x]$ is a surface S. The map f is allowed to have isolated critical points, modeled in local complex coordinates by

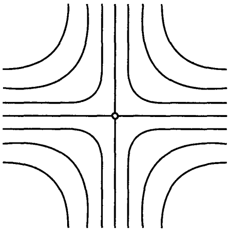
$$f(z_1, z_2) = z_1^2 + z_2^2$$

(or, if you prefer, by $f(w_1, w_2) = w_1 w_2$), see figure 10.4 on the facing page. Thus, while the generic fiber is a surface of fixed genus, some fibers of f are surfaces immersed in M with double-points. It can be assumed that each fiber of f has at most one singularity. Examples of Lefschetz fibrations with torus fibers are elliptic fibrations with fishtail singular fibers.

---

[1] Donaldson's theorem was presented in section 5.3 (page 243); it states that the only definite intersection forms of a smooth manifold are the diagonal ones.

[2] "Topological", as opposed to "holomorphic".

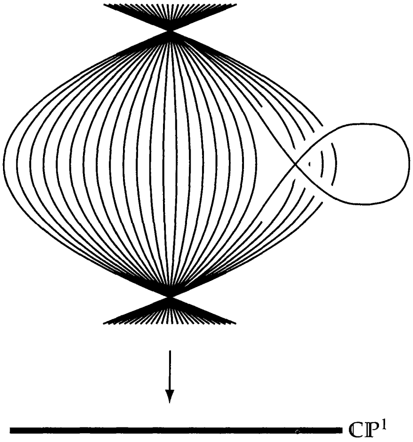


10.4. Singularity in a Lefschetz fibration

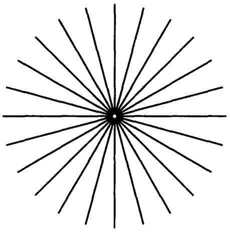
A **Lefschetz pencil** on M is a map $f \colon M \to \mathbb{S}^2$, not defined at a number of **base-points** $b_1, \ldots, b_m$, such that f defines a Lefschetz fibration on its domain and so that, around each base-point, f is modeled in local complex coordinates by

$$f(z_1, z_2) = z_1 / z_2$$

(or, if you prefer, $f(z_1, z_2) = [z_1 : z_2]$, thinking $\mathbb{S}^2 = \mathbb{CP}^1$). The fibers of f are punctured surfaces, to which one adds the base-points to obtain closed surfaces, called the fibers of the pencil. See figure 10.5. Near a base-point, the fibration looks like the slicing of $\mathbb{C}^2$ into the complex planes passing through the origin, as suggested in figure 10.6. Notice that if one blows-up a lefschetz pencil at all its base-points, then one obtains a Lefschetz fibration.

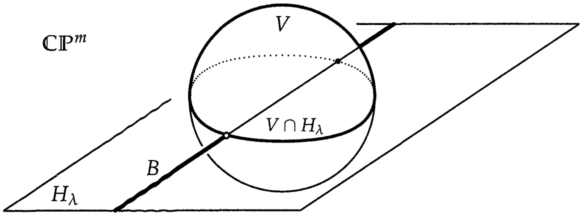


10.5. Lefschetz pencil



10.6. Lefschetz pencil around a base-point

The origin of Lefschetz pencils lies with **S. Lefschetz***'s study of algebraic varieties in* ***L'Analysis situs and et la géometrie algébrique*** [**Lef24**]*. Namely, for a complex algebraic variety V embedded in a projective space $\mathbb{CP}^m$, one takes a fixed projective subspace $B \approx \mathbb{CP}^{m-2}$ of $\mathbb{CP}^m$ and considers all hyperplanes H_λ^{m-1} in $\mathbb{CP}^m$ that contain B. This family of hyperplanes cuts V into slices $V \cap H_\lambda$, all passing through the base-locus $V \cap B$, as in figure 10.7. Since the hyperplanes $\{H_\lambda\}$ are naturally parametrized by $\lambda \in \mathbb{CP}^1$, this defines a map $f\colon V \to \mathbb{CP}^1$ (not defined at $V \cap B$) by sending all of $V \cap H_\lambda$ to λ. If V is non-singular and B is generic, then the fibers $V \cap H_\lambda$ only exhibit simple singularities and f is a Lefschetz pencil. The class of each $V \cap H_\lambda$ is Poincaré-dual to the Kähler class $[\omega]$ of V, for the Kähler structure inherited from $\mathbb{CP}^m$. For an exposition of S. Lefschetz's geometric approach to algebraic geometry, see* **K. Lamotke***'s* ***The topology of complex projective varieties after S. Lefschetz*** [**Lam81**].



10.7. Lefscetz pencil from hyperplane sections

A nice introduction to Lefschetz pencils can be found in **R. Gompf** and **A. Stipsicz**'s ***Kirby calculus and 4–manifolds*** [**GS99**, ch 8].

Lefschetz on symplectics. The relevance of Lefschetz pencils for 4–dimensional topology is that they are essentially equivalent to symplectic structures.

First, we have:

Theorem *(R. Gompf). If M admits a Lefschetz pencil whose fiber is homologically nontrivial, then it admits a symplectic structure ω, positive on all fibers.* □

If the Lefschetz pencil is *not* a Lefschetz fibration, *i.e.*, it has at least one base-point, then the fibers are automatically homologically-nontrivial: blow-up the base point and notice that the exceptional sphere has intersection $+1$ with each fiber. The symplectic form ω is built by pulling back the volume form of S^2 and perturbing it with a 2–form positive on the fibers. The argument can be read from **R. Gompf** and **A. Stipsicz**'s ***Kirby calculus and 4–manifolds*** [**GS99**, sec 10.2].

It is much more difficult to prove the converse:

Theorem *(S.K. Donaldson). Let M be a symplectic manifold with $[\omega]$ integral. Then, for sufficiently big k, there exists a Lefschetz pencil on M whose fibers represent the class $k[\omega]$ and on which ω is positive.*[3]

Since every symplectic form can be slightly perturbed so that $[\omega]$ is rational and then replace ω by a multiple $m\omega$, the condition on the integrality of $[\omega]$ is not very restrictive.

---

[3] In particular, there exists an almost-complex structure J compatible with ω for which all fibers are J-holomorphic.

Once $[\omega]$ is integral, one can build a line bundle $L \to M$ with $c_1(L) = [\omega]$ and take on L a connection with curvature form $-2\pi i\omega$. Together with an almost-complex structure J on M, this connection induces a Cauchy–Riemann operator $\bar\partial$ on the sections of L. However, since J is not integrable, holomorphic sections of this bundle (*i.e.*, $s \in \Gamma(L)$ with $\bar\partial s = 0$) are almost certainly non-existent.[4] Nonetheless, with a lot of hard work, one proves that, once k is large enough, there must exist two global sections s', s'' of $L^{\otimes k}$ that are *approximately holomorphic, i.e.,* with $\bar\partial s \ll \partial s$. Then the zeros of the family of sections $\{t's' + t''s'' \mid [t' : t''] \in \mathbb{CP}^1\}$ can be used to describe the fibers of a Lefschetz pencil on M.

The almost-holomorphic techniques were developed in **S.K. Donaldson**'s ***Symplectic submanifolds and almost-complex geometry*** **[Don96b]**, with the existence of the pencils announced in ***Lefschetz fibrations in symplectic geometry*** **[Don98]** and achieved in ***Lefschetz pencils on symplectic manifolds*** **[Don99]**. It is worth noting that, as k increases, the corresponding Lefschetz pencil becomes *unique* up to isotopy, as was showed in **D. Auroux**'s ***Asymptotically holomorphic families of symplectic submanifolds*** **[Aur97]**.

Beyond symplectic. When M does not admit a symplectic structure, then one can use a singular symplectic structure in its stead. Specifically, if $b_2^+(M) \geq 1$, then for a generic Riemannian metric on M, there exists a closed self-dual 2–form ω that is zero only along a few circles embedded in M. Since ω defines a symplectic structure on $M \setminus \{circles\}$, we call it a **near-symplectic form.**

The study of such structures, in connection with the Seiberg–Witten invariants and their interpretation in terms of J–holomorphic curves, is currently being pursued by C. Taubes.[5] Independently, **D. Auroux, S.K. Donaldson** and **L. Katzarkov,** in their recent ***Singular Lefschetz pencils*** **[ADK04]**, use such structures to build fibrations on general 4–manifolds.

Namely, define a **singular Lefschetz pencil** to be a map $f\colon M \to S^2$ (not defined at a few base-points) together with a 1–submanifold Z of M, so that: (1) f is a Lefschetz pencil away from Z; and (2) around each point of Z there are local real coordinates (x_1, x_2, x_3, t) on M with $Z = \{(0,0,0,t)\}$ so that f is described by

$$f(x_1, x_2, x_3, t) = x_1^2 - x_2^2 - x_3^2 + it$$

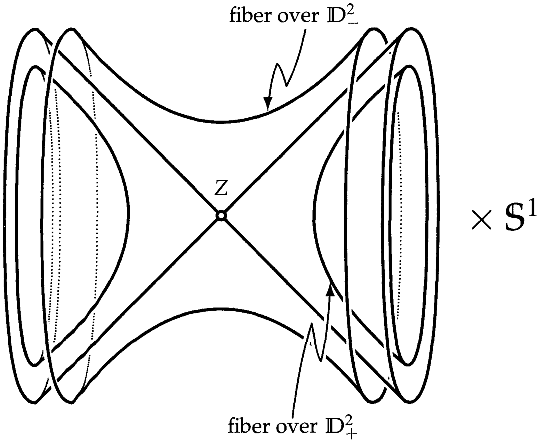
for some suitable complex coordinate on S^2. In fact, it can be arranged that f takes each circle from Z to the equator of S^2.

One can describe geometrically such a singular Lefschetz pencil $f\colon M \to S^2$ as follows: split S^2 into two hemispheres $\mathbb{D}_+^2$ and $\mathbb{D}_-^2$ and a thin neighborhood $S^1 \times [-\varepsilon, \varepsilon]$ of the equator. Then f describes two genuine Lefschetz pencils on $\mathbb{D}_\pm^2$ that are linked through $f^{-1}\big[S^1 \times [-\varepsilon, \varepsilon]\big]$. When crossing in M from over $\mathbb{D}_+^2$ to over $\mathbb{D}_-^2$, the fibers of one pencil are modified into the fibers of the other by the addition of a few 1–handles, one for each circle of Z that is crossed. The picture of this crossing is the one from figure 10.8 on the following page.

---

[4] An *integrable* almost-complex structure can be characterized as being one for which, around each point, there exist two holomorphic functions $f_1, f_2 : U \to \mathbb{C}$ with df_1, df_2 linearly-independent.

[5] See the references on page 473.



10.8. Singular Lefschetz pencil around a zero-circle

Every 4–manifold with $b_2^+ \geq 1$ admits such a structure:

Theorem (*D. Auroux and S.K. Donaldson and L. Katzarkov*). *Let M be a 4–manifold endowed with a closed self-dual 2–form ω that is zero along a family of embedded circles Z. Assume that $[\omega]$ is integral. Then, for k odd large enough, there exists a singular Lefschetz pencil $f \colon M \to \mathbb{S}^2$ whose fibers represent $k[\omega]$.*

Conversely, given a singular Lefschetz pencil $f \colon M \to \mathbb{S}^2$ so that there exists a cohomology 2–class positive on its fibers, one can find a near-symplectic form with zero-locus Z and positive on the fibers of f. □

This is a very recent result,[6] and its impact on 4–dimensional topology cannot be gauged at the time of this writing.

Note: Existence of almost-complex structures

In what follows we prove the results stated in section 10.1 (page 376), showing that almost-complex structures are very flexible and their existence is a mere problem in homology.

The good Chern class. An immediate property:

Lemma. *Assume M admits an almost-complex structure J. Then we must have*

$$c_1(J) \cdot c_1(J) = 3 \operatorname{sign} M + 2\chi(M) ,$$

and $c_1(J)$ is a characteristic element of M, i.e., an integral lift of $w_2(T_M)$.

Proof. Let E be any complex-plane bundle. Then we have $w_2(E) = c_1(E) \pmod 2$ and also $p_1(E) = -c_2(E \otimes_{\mathbb{R}} \mathbb{C}) = -c_2(E \oplus E^*) = c_1(E) \cdot c_1(E) - 2c_2(E)$. In the case $E = (T_M, J)$, we have $p_1(T_M) = 3 \operatorname{sign} M$ (by Hirzebruch's signature theorem[7]) and $c_2(T_M) = e(T_M) = \chi(M)$. □

---

[6] October 2004.

[7] Hirzebruch's signature theorem was stated in section 4.3 (page 166).

The converse is true as well. It was proved in **W-t. Wu**'s ***Sur les classes caractéristiques des structures fibrées sphériques*** **[Wu52**, ch IV], and in **F. Hirzebruch** and **H. Hopf**'s ***Felder von Flächenelementen in 4–dimensionalen Mannigfaltigkeiten*** **[HH58]**:

Lemma *(W-t. Wu; F. Hirzebruch and H. Hopf). Let M be an oriented 4–manifold. If there is characteristic element $\underline{w}$ such that*

$$\underline{w} \cdot \underline{w} = 3 \operatorname{sign} M + 2\chi(M) \,,$$

then M admits an almost-complex structure J such that $c_1(J) = \underline{w}$.

> ***Sketch of proof.*** We build a complex-plane bundle $E \to M$ with $c_1(E) = \underline{w}$ and $c_2(E) = \chi(M)$. A good way to build E is to start with the complex-line bundle $L_{\underline{w}}$ of Chern class $c_1(L_{\underline{w}}) = \underline{w}$, then modify $E_0 = L_{\underline{w}} \oplus \underline{\mathbb{C}}$ by picking a small 4–ball $\mathbb{D}^4 \subset M$, cutting $E_0|_{\mathbb{D}^4}$ out of E and gluing it back in using a suitable $SU(2)$–twist on the fibers of $E_0|_{\partial\mathbb{D}^4}$. This leaves c_1 unchanged but modifies c_2 at pleasure. The full argument is reviewed in **[Sco03]**.
>
> Having obtained such a complex bundle E, its characteristic classes are the same as those of T_M: $w_2(E) = w_2(T_M)$, $e(E) = e(T_M)$ (since $c_2(E) = \chi(M)$), and $p_1(E) = p_1(T_M)$ (since $p_1(E) = c_1(E) \cdot c_1(E) - 2c_2(E)$ and $\underline{w} \cdot \underline{w} = p_1(T_M) + 2\chi(M)$). Then, by using the Dold–Whitney theorem,[8] we conclude that, as real bundles, E and T_M must in fact be isomorphic. Through any such isomorphism, one can transport the complex structure from the fibers of E to an almost-complex structure on M. □

In particular, this lemma implies that, if M has indefinite intersection form and admits almost-complex structures, then the almost-complex structure on M is never alone: By Meyer's lemma,[9] there are classes α with $\alpha \cdot \alpha = 0$; then, if $\underline{w}$ is a good characteristic element, then every $\underline{w} + 2m\alpha$ will also be the Chern class of some new almost-complex structure on M.

> *As another testimony to the flexibility of almost-complex structures, we also have the following converse to the adjunction formula: If $\underline{w}$ is a good candidate for the Chern class of an almost-complex structure and S is any embedded surface that satisfies the adjunction formula for $\underline{w}$, i.e., if it has $\chi(S) + S \cdot S = \underline{w} \cdot S$, then there must exist an almost-complex structure J both with $c_1(T_M, J) = \underline{w}$ and so that S becomes J–holomorphic. See* **C. Bohr**'s ***Embedded surfaces and almost complex structures*** **[Boh00]**.

Notice also that, even if the condition $\underline{w} \cdot \underline{w} = 3 \operatorname{sign} M + 2\chi(M)$ fails, in the proof above one still gets an isomorphism between E and T_M over the 3–skeleton of M (or, if you prefer, over $M \setminus \{\text{point}\}$). In other words:

Lemma. *For every characteristic element $\underline{w}$ of M, there is a partial almost-complex structure $J|_3$ over the 3–skeleton of M, with $c_1(J|_3) = \underline{w}$.* □

Such an almost-complex structure, even though only partially-defined, does offer enough data to be lifted and extended to a unique spin$^{\mathbb{C}}$ structure across all M.

---

[8] The Dold–Whitney theorem (1959) was stated on page 167.

[9] Meyer's lemma was stated on page 238.

Finding Chern classes. We now explore what conditions need to be met for finding good candidates $\underline{w}$ for the Chern class of an almost-complex structure.

Noether's Lemma. *If the 4–manifold M admits an almost-complex structure, then it must be that $b_2^+(M) + b_1(M)$ is odd.*

Often, this lemma is written[10] $c_1(M)^2 + c_2(M) = 0 \pmod{12}$.

Proof. If M admits an almost-complex structure J, then we must have

$$c_1(J) \cdot c_1(J) = 3 \operatorname{sign} M + 2\chi(M) ,$$

and $c_1(J)$ must be a integral lift of $w_2(T_M)$ (a characteristic element of M). Furthermore, by van der Blij's lemma (page 170), we have

$$c_1(J) \cdot c_1(J) = \operatorname{sign} M \pmod 8 ,$$

and thus $\operatorname{sign} M = 3 \operatorname{sign} M + 2\chi(M) \pmod 8$, which rearranges as

$$\operatorname{sign} M + \chi(M) = 0 \pmod 4 .$$

Further, we have

$$\operatorname{sign} M = b_2^+(M) - b_2^-(M) ,$$
$$\chi(M) = b_2^+(M) + b_2^-(M) - 2b_1(M) + 2 ,$$

and hence

$$2b_2^+(M) - 2b_1(M) + 2 = 0 \pmod 4 .$$

Therefore $b_2^+(M) - b_1(M) + 1 = 0 \pmod 2$. □

Lemma. *Assume M is a smooth simply-connected 4–manifold. Then M admits almost-complex structures if and only if $b_2^+(M)$ is odd.*

Proof. Since M is smooth, the intersection form must be of shape either $Q_M = \oplus m \, [-1] \oplus n \, [+1]$ or $Q_M = \oplus 2m \, E_8 \oplus n \, H$. In both cases, the parity of $b_2^+(M)$ coincides with the parity of n. Thus, assuming b_2^+ is odd, we write $n = 2n' + 1$.

If Q_M can be written

$$Q_M = \oplus m \, [-1] \oplus (2n' + 1) \, [+1]$$

for some basis $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_{2n'+1}$, then the element

$$\underline{w} = \sum \alpha_k + 3\beta_1 + \sum (3\beta_{2k} + \beta_{2k+1})$$

is characteristic (since $\underline{w} \cdot x = x \cdot x \pmod 2$), and its self-intersection is $\underline{w} \cdot \underline{w} = -m + 9 + 9n' + n'$. Since $\operatorname{sign} M = 2n' + 1 - m$ and $\chi(M) = 2 + 2n' + 1 + m$, it follows that $\underline{w} \cdot \underline{w} = 3 \operatorname{sign} M + 2\chi(M)$.

If Q_M can be written

$$Q_M = \oplus \pm 2m \, E_8 \oplus (2n' + 1) \, H ,$$

then we notice that, Q_M being even, any class with even self-intersection will be a characteristic element. We have $\operatorname{sign} M = \pm 16m$ and $\chi(M) = 2 + 16m +$

---

[10] The formula known as *Noether's formula* states that, for a complex manifold M, we have $\chi(\mathcal{O}) = \frac{1}{12}(K_M \cdot K_M + \chi(M))$. Here $\chi(\mathcal{O})$ denotes the Euler characteristic of the cohomology $H^*(M; \mathcal{O})$ of the sheaf of holomorphic functions on M; or, if you prefer, of the Čech cohomology $\check{H}^*(M; \mathcal{O}(\mathbb{C}))$ in the notation of the end-notes of chapter 4 (page 193). We encountered $\chi(\mathcal{O})$ in the statement of the Riemann–Roch theorem, in section 6.3 (page 282).

$4n' + 2$. Therefore $3 \operatorname{sign} M + 2\chi(M)$ must be a multiple of 8, say $3 \operatorname{sign} M + 2\chi(M) = 8p$. Pick a basis $\{\alpha, \overline{\alpha}\}$ for your favorite H in Q_M. Then

$$\underline{w} = 2p\,\alpha + 2\overline{\alpha}$$

has self-intersection $\underline{w} \cdot \underline{w} = 8p$ and is characteristic. □

This last statement can be generalized to the non-simply-connected case as a partial converse of Noether's lemma above. For example, if Q_M is indefinite, then $b_1(M) + b_2^+(M)$ being odd ensures the existence of an almost-complex structure.

On the other hand, the manifolds $\#k\,S^1 \times S^3$ with k odd all have $b_1 + b_2^+$ odd, but if $k \geq 3$ there are no good candidates for $c_1(J)$ and thus these manifolds admit no almost-complex structures.

Note: Spin$^{\mathbb{C}}$ structures—existence and other extras

In this note we gathered together a few leftover details concerning spin$^{\mathbb{C}}$ structures. We start with a rigorous Čech-flavored argument proving that every 4–manifold admits a spin$^{\mathbb{C}}$ structure. Then we underline the parallelism between spinors and 2–forms, with, for example, self-dual spinors sitting atop self-dual 2–forms. Finally, we show that a spin$^{\mathbb{C}}$ structure induces a partial almost-complex structure over the 3–skeleton.

Spin$^{\mathbb{C}}$ structures always exist. We will explain an alternative proof that all 4–manifolds admit spin$^{\mathbb{C}}$ structures. The proof still rests essentially on the fact that $w_2(T_M)$ can be lifted to an integral class. This was proved to always be possible back in section 4.4 (page 168). The argument that follows is based on cocycle-presentations of bundles, and understanding some details of it requires first reading the note on Čech cohomology on page 189 at the end of chapter 4.

Failure to spin. Imagine we attempt to build a spin structure on M. We endow M with a Riemannian metric, take the resulting $SO(4)$–cocycle $\{g_{\alpha\beta} \colon U_\alpha \cap U_\beta \longrightarrow SO(4)\}$ of T_M and lift it to some maps

$$\widetilde{g}_{\alpha\beta} \colon U_\alpha \cap U_\beta \longrightarrow Spin(4)$$

so that $\widetilde{g}_{\alpha\beta} = \widetilde{g}_{\beta\alpha}^{-1}$. In general, of course, these maps fail from being a cocycle, and their failure is measured by the Čech 2–chain

$$w_{\alpha\beta\gamma} = \widetilde{g}_{\alpha\beta} \cdot \widetilde{g}_{\beta\gamma} \cdot \widetilde{g}_{\gamma\alpha}$$

with values in $\mathbb{Z}_2 = \{-1, +1\}$. In Čech cohomology, the cochain $\{w_{\alpha\beta\gamma}\}$ represents the Stiefel–Whitney class $w_2(T_M) \in H^2(M; \mathbb{Z}_2)$, as was explained in the end-notes of chapter 4 (page 189).

Failure to square-root. Now, pick some integral lift $\underline{w} \in H^2(M; \mathbb{Z})$ of $w_2(T_M)$ (a characteristic element). We can express this class by building a complex-line bundle $\mathcal{L}$ with Chern class

$$c_1(\mathcal{L}) = \underline{w}\,.$$

If we endow $\mathcal{L}$ with some Hermitian inner product on its fibers, then we transform $\mathcal{L}$ into a $U(1)$–bundle, which can thus be defined using a $U(1)$–valued cocycle

$$\ell_{\alpha\beta} \colon U_\alpha \cap U_\beta \longrightarrow U(1)\,.$$

Remember that $U(1)$ is merely the group S^1 of unit complex numbers.

Imagine now that we try to build a square-root of $\mathcal{L}$, *i.e.*, that we are looking for a complex-line bundle $\mathcal{L}^{1/2}$ such that

$$\mathcal{L}^{1/2} \otimes \mathcal{L}^{1/2} = \mathcal{L} \,.$$

Such a square-root exists if and only if we can lift the cocycle $\{\ell_{\alpha\beta}\}$ against the double cover $U(1) \to U(1) \colon z \mapsto z^2$ to a collection of maps $\sqrt{\ell}_{\alpha\beta} \colon U_\alpha \cap U_\beta \to U(1)$ that themselves satisfy the cocycle condition. Of course, in general all we get for our effort is a collection of maps that satisfy

$$\sqrt{\ell}_{\alpha\beta}(x) \cdot \sqrt{\ell}_{\beta\gamma}(x) \cdot \sqrt{\ell}_{\gamma\alpha}(x) = \pm\, id \,.$$

We can measure the failure of the $\sqrt{\ell}_{\alpha\beta}$'s from being a cocycle by using the $\mathbb{Z}_2$–valued Čech 2–cochain

$$c_{\alpha\beta\gamma} = \sqrt{\ell}_{\alpha\beta} \cdot \sqrt{\ell}_{\beta\gamma} \cdot \sqrt{\ell}_{\gamma\alpha} \,.$$

This cocycle represents in $H^2(M;\mathbb{Z}_2)$ the modulo 2 reduction of $c_1(\mathcal{L}) = \underline{w}$, in other words, $w_2(T_M)$ itself.

The class of $\{c_{\alpha\beta\gamma}\}$ also measures the obstruction to the existence of a square-root $\mathcal{L}^{1/2}$. Indeed, a bundle $\mathcal{L}^{1/2}$ exist if and only if there exists a candidate α for its Chern class, in other words, if we have $c_1(\mathcal{L}) = \alpha + \alpha$ for some α. However, that simply means that $c_1(\mathcal{L}) = \underline{w}$ be an *even* class, and thus that $w_2(T_M) = 0$.

Match them. In conclusion, the obstruction $\{w_{\alpha\beta\gamma}\}$ to the existence of a spin structure and the obstruction $\{c_{\alpha\beta\gamma}\}$ to the existence of a square root of $\mathcal{L}$ are cohomologous in $\check{H}^2(M;\mathbb{Z}_2)$. This implies that, with a little care, we can match the failures of $\{\tilde{g}_{\alpha\beta}\}$ and $\{\sqrt{\ell}_{\alpha\beta}\}$. Namely, we can arrange that the minus signs appearing in the failed cocycle test of $\{\tilde{g}_{\alpha\beta}\}$ correspond exactly to the minus signs from the failed cocycle test of $\{\sqrt{\ell}_{\alpha\beta}\}$.

Indeed, since $\{w_{\alpha\beta\gamma}\}$ and $\{c_{\alpha\beta\gamma}\}$ both represent the class $w_2(T_M)$ in $\check{H}^2(M;\mathbb{Z}_2)$, they differ by at most a Čech coboundary. That is, there is a family of functions $\varepsilon_{\alpha\beta} \colon U_\alpha \cap U_\beta \longrightarrow \mathbb{Z}_2 = \{-1,+1\}$ so that

$$w_{\alpha\beta\gamma} = c_{\alpha\beta\gamma} \cdot \varepsilon_{\alpha\beta}\, \varepsilon_{\beta\gamma}\, \varepsilon_{\gamma\alpha} \,.$$

Then we can modify the failed-cocycle $\{\sqrt{\ell}_{\alpha\beta}\}$ of $\mathcal{L}^{1/2}$ to

$$\sqrt{\ell'}_{\alpha\beta} = \varepsilon_{\alpha\beta}\sqrt{\ell}_{\alpha\beta} \,,$$

and then we will have $\tilde{g}_{\alpha\beta} \cdot \tilde{g}_{\beta\gamma} \cdot \tilde{g}_{\gamma\alpha} = \sqrt{\ell'}_{\alpha\beta} \cdot \sqrt{\ell'}_{\beta\gamma} \cdot \sqrt{\ell'}_{\gamma\alpha}$, as claimed.

Cancel them. In this case, we can multiply the $Spin(4)$–valued family $\{\tilde{g}_{\alpha\beta}\}$ and the $U(1)$–valued family $\{\sqrt{\ell'}_{\alpha\beta}\}$ inside the group

$$Spin^{\mathbb{C}}(4) = U(1) \times Spin(4) \;/\; \pm 1$$

to get a true $Spin^{\mathbb{C}}(4)$–cocycle

$$\mathfrak{s} = \{\sqrt{\ell'}_{\alpha\beta} \cdot \tilde{g}_{\alpha\beta}\} \,.$$

Such a cocycle, considered up to isomorphisms, is called a **spin$^{\mathbb{C}}$ structure** on M, and $\mathcal{L}$ is its determinant line bundle. We have just proved that every 4–manifold admits spin$^{\mathbb{C}}$ structures. □

2–*torsion.* Notice that, given a line bundle $\mathcal{L}$ with $c_1(\mathcal{L}) = w_2(T_M) \pmod 2$, the failed-cocycle for $\mathcal{L}^{1/2}$ is unique up to isomorphisms, unless $H^2(M;\mathbb{Z})$ has 2–torsion. (Think for a second that $\mathcal{L}$ actually admits a square-root bundle $\mathcal{L}^{1/2}$; such a square-root is characterized by $2c_1(\mathcal{L}^{1/2}) = c_1(\mathcal{L})$. In the absence of 2–torsion, this condition determines $c_1(\mathcal{L}^{1/2})$, and thus $\mathcal{L}^{1/2}$ itself, uniquely.)

Cohomological view. The double-cover $Spin^{\mathbb{C}}(4) \to \mathbb{S}^1 \times SO(4)$ fits into the exact sequence

$$0 \longrightarrow \mathbb{Z}_2 \longrightarrow Spin^{\mathbb{C}}(4) \longrightarrow \mathbb{S}^1 \times SO(4) \longrightarrow 0$$

with $\mathbb{Z}_2$ included as $\{[-1,-id],\ [1,id]\}$. In Čech cohomology, this leads to an exact sequence (of sets)

$$\cdots \longrightarrow H^1(M;\mathbb{Z}_2) \longrightarrow \check{H}^1(M;\ C^\infty Spin^{\mathbb{C}}(4)) \longrightarrow$$
$$\longrightarrow \check{H}^1(M;\ C^\infty \mathbb{S}^1) \oplus \check{H}^1(M;\ C^\infty SO(4)) \longrightarrow H^2(M;\mathbb{Z}_2)\,.$$

The last map can be determined to be the map that sends a pair (L,E), made of an $U(1)$–bundle L and an $SO(4)$–bundle E, to the class $c_1(L) + w_2(E) \pmod 2$.

Therefore the oriented bundle E admits a spin$^{\mathbb{C}}$ structure if and only if its Stiefel–Whitney class $w_2(E)$ can be lifted to an integral class $c_1(L)$. On the other hand, if E and L are thus compatible, then the various $Spin^{\mathbb{C}}(4)$–cocycles to which the pair (E,L) can be lifted are classified by the part of $H^1(M;\mathbb{Z}_2)$ coming from the 2–torsion of $H^1(M;\mathbb{Z})$, as can be seen after a careful study of the beginning of the exact sequence.

Relations with self-duality. There is a Lie-group diagram that is worth mentioning, since it exhibits the bundles of spinors as parallel to the bundles of 2–forms.

The diagram stems partly from the isomorphism

$$SO(3) = SU(2) \;/\; \pm 1\,.$$

Since $U(2) = \mathbb{S}^1 \times SU(2) \;/\; \pm 1$, there is a natural map $U(2) \to SO(3)$. On the other hand, the exceptional Lie algebra isomorphism

$$\mathfrak{so}(4) = \mathfrak{so}(3) \oplus \mathfrak{so}(3)$$

integrates to a map $SO(4) \to SO(3) \times SO(3)$, and thus yields two natural projections $SO(4) \to SO(3)$. These projections fit into the commutative diagram

$$\begin{tikzcd}
U(2) \arrow[d] & Spin^{\mathbb{C}}(4) \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"'] \arrow[d] & U(2) \arrow[d] \\
SO(3) & SO(4) \arrow[l, "\lambda_-"'] \arrow[r, "\lambda_+"] & SO(3)\,.
\end{tikzcd}$$

Notice that, by lifting the $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ of T_M to a $Spin^{\mathbb{C}}(4)$–cocycle $\{\tilde{g}^c_{\alpha\beta}\}$ and then projecting the latter to the two copies of $U(2)$ above, we obtain the two cocycles $\{\rho_\pm(\tilde{g}^c_{\alpha\beta})\}$ of the spinor bundles $\mathcal{W}^\pm$.

On the other hand, by projecting the $SO(4)$–cocycle $\{g_{\alpha\beta}\}$ directly to the two copies of $SO(3)$ above, we obtain two $SO(3)$–cocycles $\{\lambda_\pm(g_{\alpha\beta})\}$. These can be used to glue-up two 3–plane bundles, which turn up to be none other than the bundles $\Lambda^2_\pm(T^*_M)$ of self-dual/anti-self-dual 2–forms. This last fact stems from the fact that $\Lambda^2(\mathbb{R}^4) = \mathfrak{so}(4)$ and that the bundle $\Lambda^2(T^*_M)$ can naturally be viewed as glued through the adjoint action of $SO(4)$ on its Lie algebra.

In conclusion, we have the diagram

$$\begin{tikzcd}
\mathcal{W}^- & \circlearrowleft & U(2) \arrow[d] & Spin^{\mathbb{C}}(4) \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"'] \arrow[d] & U(2) \arrow[d] & \circlearrowright & \mathcal{W}^+ \arrow[d, "\sigma"] \\
\Lambda^2_- & \circlearrowleft & SO(3) & SO(4) \arrow[l, "\lambda_-"'] \arrow[r, "\lambda_+"] & SO(3) & \circlearrowright & \Lambda^2_+ \\
 & & & \circlearrowright & & & \\
 & & & T_M & & &
\end{tikzcd}$$

where we used "$\circlearrowright$" to show which bundles are induced from whose cocycles. Notice how the squaring map $\sigma \colon \mathcal{W}^+ \to i\Lambda^2_+$ completes the diagram in a rather pleasing way.

In light of the above diagram, one can think of the spinor bundles $\mathcal{W}^\pm$ as creatures parallel to the bundles $\Lambda^2_\pm(T^*_M)$ of self-dual/anti-self-dual 2–forms. Owing to this peculiarity of dimension[11] 4, we prefer to call $\mathcal{W}^\pm$ the bundles of *self-dual/anti-self-dual* spinors, instead of the more customary name of positive/negative spinors.

From spin$^{\mathbb{C}}$ structures to partial almost-complex structures. Let φ be a generic self-dual spinor field. Being a section of a 4–plane bundle over a 4–manifold, its zeros are isolated. Then the map

$$T_M \longrightarrow \mathcal{W}^- \colon \qquad v \longmapsto v \bullet \varphi$$

determines an isomorphism away from the zeros of φ:

$$T_M|_{\text{off zeros}} \approx \mathcal{W}^- \ .$$

Using this isomorphism, the complex structure of $\mathcal{W}^-$ can be pulled-back to a complex structure on T_M, well-defined off the zeros of φ. In other words, it defines an almost-complex structure on $M \setminus \{\textit{zeros of } \varphi\}$, which is essentially the same thing as an almost-complex structure $J|_3$ over the 3–skeleton of M.

The Chern class of this partial almost-complex structure is exactly[12]

$$c_1(J|_3) = c_1(\mathfrak{s}) \ .$$

More, its fundamental 2–form $\omega|_3$ essentially coincides with $\sigma(\varphi)$, but rescaled:

$$\tfrac{1}{\sqrt{2}}\,\omega|_3 = \sigma(\varphi) / |\sigma(\varphi)| \ .$$

It is well-defined, just as $J|_3$ is, away from the zeros of φ.

Further, by computing $e(\mathcal{W}^+)$ one notices that there exist nowhere-zero spinor fields $\varphi \in \Gamma(\mathcal{W}^+)$ if and only if the spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ corresponds to a genuine almost-complex structure, and the latter can be fully recovered as above (up to

---

[11] As well as following the terminology of **M. Atiyah**, **N. Hitchin** and **I. Singer**'s ***Self-duality in four-dimensional Riemannian geometry*** **[AHS78]**.

[12] Technically, $c_1(J|_3) \in H^2(M|_3;\mathbb{Z})$ while $c_1(\mathfrak{s}) \in H^2(M;\mathbb{Z})$. Nonetheless, the inclusion $M|_3 \subset M$ induces a natural isomorphism on H^2. This is the succinct way to say that 2–co/homology is influenced only by 1–, 2–, and 3–cells, not by 4–cells.

homotopy). Indeed, as we mentioned before, in the almost-complex case, we have the complex isomorphism $\mathcal{W}^- \approx (T_M, J)$.

Note: The spin$^{\mathbb{C}}$ and characteristic cobordism groups

In the end-notes of chapter 4 (page 227), we defined the oriented cobordism group Ω_4^{SO} and the spin cobordism group Ω_4^{Spin}. We can now also define a spin$^{\mathbb{C}}$ version

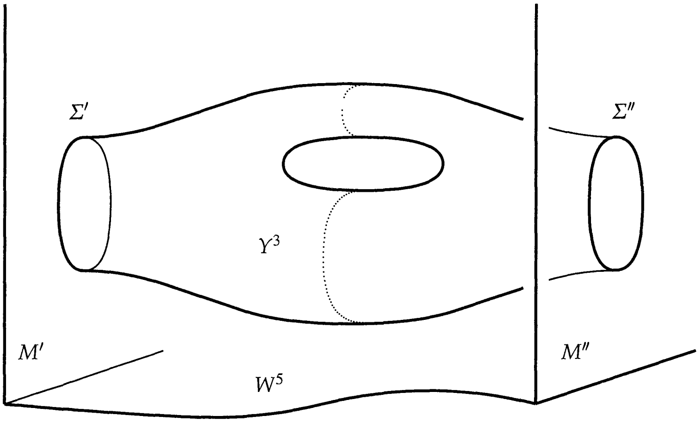
$$\Omega_4^{Spin^{\mathbb{C}}} .$$

It is generated by 4–manifolds endowed with spin$^{\mathbb{C}}$ structures. Two such spin$^{\mathbb{C}}$ manifolds M' and M'' are considered equivalent if there is a 5–manifold W with $\partial W = \overline{M'} \cup M''$ and such that W is endowed with a spin$^{\mathbb{C}}$ structure that restricts to the chosen spin$^{\mathbb{C}}$ structures on each of M' and M''.

In the end-notes of the next chapter (page 502), this cobordism group and its evaluation will be used to prove Rokhlin's theorem and its generalizations.

Characteristic cobordisms. Using the interpretation of spin$^{\mathbb{C}}$ structures in terms of outside spin structures extended across a characteristic surface Σ, we remark that the spin$^{\mathbb{C}}$ cobordism group can be defined alternatively as follows:

The group $\Omega_4^{Spin^{\mathbb{C}}}$ is also the cobordism group generated by 4–manifolds M with a chosen characteristic surface Σ and a spin structure on the complement of Σ. Two such pairs (M', Σ') and (M'', Σ'') are considered equivalent if they can be linked by a 5–manifold W^5 with a chosen unoriented 3–submanifold Y^3 Poincaré-dual to $w_2(T_W)$, so that $\partial W = \overline{M'} \cup M''$ and $\partial Y = \overline{\Sigma'} \cup \Sigma''$, and so that there exists a spin structure on $W \setminus Y$ which restricts to the chosen spin structures on $M' \setminus \Sigma'$ and $M'' \setminus \Sigma''$. See figure 10.9.



10.9. Characteristic cobordism

The latter description defines what is better known under the name of **characteristic cobordism group** and is usually denoted by

$$\Omega_4^{char} .$$

As such, it was evaluated in ***A geometric proof of Rochlin's theorem*** [FK78]:

Theorem *(M. Freedman and R. Kirby). In dimension 4, the correspondence given by* $(M,\Sigma) \mapsto \bigl(\operatorname{sign} M,\ \frac{1}{8}(\Sigma\cdot\Sigma - \operatorname{sign} M)\bigr)$ *establishes an isomorphism*

$$\Omega_4^{char} = \mathbb{Z} \oplus \mathbb{Z}\,.$$

The generators of Ω_4^{char} *are* $(\mathbb{CP}^2, \mathbb{CP}^1)$ *and* $\bigl(\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2,\ \# 3\,\mathbb{CP}^1 \# \overline{\mathbb{CP}}^1\bigr)$.

Described directly in terms of spin$^{\mathbb{C}}$ structures, the isomorphism between the spin$^{\mathbb{C}}$ cobordism group and $\mathbb{Z}\oplus\mathbb{Z}$ is $(M,\mathfrak{s}) \mapsto \bigl(\operatorname{sign} M,\ \frac{1}{8}(c_1(\mathfrak{s})^2 - \operatorname{sign} M)\bigr)$.

The rest of this note (till page 432) is taken by the proof of the above theorem.

Proof that $\Omega_4^{Spin^{\mathbb{C}}} = \mathbb{Z}\oplus\mathbb{Z}$. We think in terms of characteristic cobordisms and define the morphism

$$\vartheta\colon \Omega_4^{char} \longrightarrow \mathbb{Z} \qquad \vartheta(M,\Sigma) = \bigl(\operatorname{sign} M,\ \tfrac{1}{8}(\Sigma\cdot\Sigma - \operatorname{sign} M)\bigr)\ .$$

Recall that a characteristic surface is any oriented surface Σ embedded in M that is dual to $w_2(T_M)$, or, equivalently, that satisfies

$$\Sigma\cdot x = x\cdot x \pmod 2$$

for all homology 2–classes x of M. Owing to van der Blij's lemma, every characteristic surface satisfies $\Sigma\cdot\Sigma = \operatorname{sign} M \pmod 8$.

The map ϑ is *well-defined:* Assume (M,Σ) is the characteristic boundary of (W,Y). We want to show that $\Sigma\cdot\Sigma = 0$. Perturb Y to Y' such that Y and Y' are transverse, and such that Σ and $\Sigma' = \partial Y'$ are transverse as well. The intersection of Y with Y' consists of a family of 1–dimensional arcs and circles; each arc must connect two intersection points of Σ and Σ' with opposite signs. Therefore $\Sigma\cdot\Sigma' = 0$. Of course, the latter is just the self-intersection number of Σ. On the other hand, we proved earlier[13] that the signature of a boundary always vanishes, and therefore $\operatorname{sign} M = 0$. In conclusion,

$$\vartheta\bigl(\partial(W,Y)\bigr) = (0,0)\ .$$

It immediately follows that, if (M',Σ') and (M'',Σ'') are characteristically cobordant, then $\vartheta(M',\Sigma') = \vartheta(M'',\Sigma'')$.

The map ϑ is certainly *additive,* in other words, it is a group morphism: indeed, the group operation of a cobordism group is disjoint union (or connected sum), and it is easy to check ϑ's additivity.

The map ϑ is *surjective,* since

$$\vartheta(\mathbb{CP}^2,\mathbb{CP}^1) = (1,0)\,,$$
$$\vartheta\bigl(\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2,\ \# 3\,\mathbb{CP}^1 \# \overline{\mathbb{CP}}^1\bigr) = (0,1)\ .$$

To prove that ϑ is an isomorphism of groups, all we still need to prove is that ϑ is *injective.* In other words, we need only argue that

---

[13] See section 3.2 (page 120).

Lemma *(M. Freedman and R. Kirby). If* (M, Σ) *is a characteristic pair such that* $\operatorname{sign} M = 0$ *and* $\Sigma \cdot \Sigma = 0$*, then there is a pair* (W^5, Y^3) *such that*

$$(M, \Sigma) = \partial(W, Y) ,$$

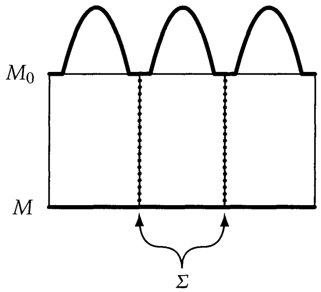
with Y *dual to* $w_2(T_W)$ *and with* $W \setminus Y$ *endowed with a spin structure that restricts to the chosen spin structure of* $M \setminus \Sigma$*.*

Proof. First, if M is not simply-connected, we will add 2–handles and reduce to the case of simply-connected manifolds. Then we will connect-sum with enough copies of $\mathbb{S}^2 \times \mathbb{S}^2$ till our manifold simplifies to either $\# m\, \mathbb{S}^2 \times \mathbb{S}^2$ or $\# m\, \mathbb{S}^2 \times \mathbb{S}^2 \# \mathbb{S}^2 \tilde{\times} \mathbb{S}^2$. We will then correspondingly cap it off with either $\# m\, \mathbb{D}^3 \times \mathbb{S}^2$ or $\# m\, \mathbb{D}^3 \times \mathbb{S}^2 \# \mathbb{D}^3 \tilde{\times} \mathbb{S}^2$. During all of this, we will also track what happens with Σ, and in fact it is the latter that will dictate how the capping is to be done. The proof rests on the four theorems of C.T.C. Wall discussed back in section 4.2 (page 149).

Make it simply-connected. Pick a handle decomposition of M that contains exactly one 0–handle and one 4–handle. Since $M \setminus \Sigma$ is endowed with a spin structure, this means that T_M is trivialized over the 1–skeleton of $M \setminus \Sigma$, and this trivialization extends to the 2–skeleton, and even further to the 3–skeleton. Thus, we consider $T_M|_{M \setminus \Sigma \cup \{point\}}$ as trivialized.

Choose a set of generators for $\pi_1(M)$, represented as embedded disjoint circles $\ell_1, \ldots, \ell_m$ in $M \setminus \Sigma \cup \{point\}$. Since T_M is trivialized over each ℓ_k, it follows that the normal bundle of each ℓ_k is trivialized; in other words, each ℓ_k has a favorite framing in M^4.

Thicken M to $M \times [0, \varepsilon]$. Attach 5–dimensional 2–handles $\mathbb{D}^2 \times \mathbb{D}^3$ to each $\ell_k \times \varepsilon$, respecting their framing in $M \times \varepsilon$. The result is a cobordism between M and a simply-connected 4–manifold M_0. Since the 2–handles were attached away from Σ, we observe that the surface Σ survives untouched into M_0. See figure 10.10.

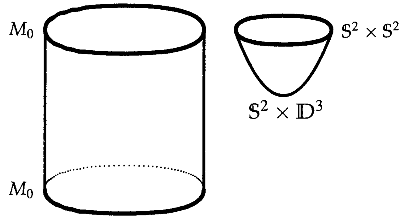


10.10. Make it simply-connected

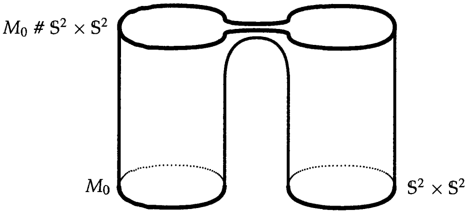
This trivialization of $T_M|_{M \setminus \Sigma \cup \{point\}}$ extends to a trivialization of $T_{M \times [0,\varepsilon]}$ away from $\Sigma \times [0, \varepsilon]$ and $point \times [0, \varepsilon]$. Moreover, because of our choice of framing for the attached 2–handles, this trivialization extends across these 2–handles as well, and yields a trivialization of $T_{M_0}|_{M_0 \setminus \Sigma \cup \{point\}}$. Since Σ in M_0 is still the obstruction to extending the trivialization of T_{M_0} across all

M_0, it follows that it is still a characteristic surface in M_0. The signature and self-intersection of Σ are, of course, unchanged.

Add spheres. Now, take M_0 and thicken it to a trivial cobordism $M_0 \times [0,1]$. Put this alongside the 5-manifold $\mathbb{S}^2 \times \mathbb{D}^3$, thought of as a cobordism between $\varnothing$ and $\mathbb{S}^2 \times \mathbb{S}^2$. This yields a cobordism between M_0 and the disjoint union $M_0 \sqcup \mathbb{S}^2 \times \mathbb{S}^2$, as in figure 10.11. This cobordism, of course, can be continued with a cobordism that connect-sums the two components, as suggested in figure 10.12. We end up with a cobordism[14] between M_0 and $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$.



10.11. Adding a copy of $\mathbb{S}^2 \times \mathbb{S}^2$



10.12. Cobordism between $M_0 \sqcup \mathbb{S}^2 \times \mathbb{S}^2$ and $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$

The intersection form of $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$ is $Q_{M_0} \oplus H$, and since the new term is even, Σ is still a characteristic surface in $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$.

The even case. Assume first that the intersection form of M_0 is *even*. Then, since its signature is zero, we must have

$$Q_{M_0} = \oplus\, n\, H \qquad \text{and hence} \qquad Q_{M_0 \# \mathbb{S}^2 \times \mathbb{S}^2} = \oplus\, (n+1)\, H\,.$$

Thus, $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$ has the same intersection form as $\#(n+1)\, \mathbb{S}^2 \times \mathbb{S}^2$. Therefore Wall's theorem on h-cobordisms, combined with Wall's theorem on stabilizations (*i.e.*, the corollary on page 155), implies that we have a diffeomorphism

$$M_0 \# k \mathbb{S}^2 \times \mathbb{S}^2 \;\cong\; \# m \mathbb{S}^2 \times \mathbb{S}^2$$

for some suitably large k and corresponding $m = n + k + 1$.

Adding the extra $\mathbb{S}^2 \times \mathbb{S}^2$-terms needed is not a problem: we proceed as we did for the first $\mathbb{S}^2 \times \mathbb{S}^2$, and end up with a cobordism between M_0 and $M_0 \# k\mathbb{S}^2 \times \mathbb{S}^2$. Owing to Wall's diffeomorphism above, we like to view the latter

---

14. Of course, the whole resulting cobordism between M_0 and $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$ is nothing but, succinctly, the boundary sum $(M_0 \times [0,1]) \natural\, \mathbb{S}^2 \times \mathbb{D}^3$. Boundary sums were briefly recalled on page 13.

as a cobordism between M_0 and $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$. The surface Σ survives as a characteristic surface embedded in the top boundary $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$.

Now choose a basis $\{\alpha_1, \overline{\alpha}_1, \ldots, \alpha_m, \overline{\alpha}_m\}$ for the homology of $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$, with each pair $\{\alpha_k, \overline{\alpha}_k\}$ coming from an $\mathbb{S}^2 \times \mathbb{S}^2$–term, and with the only non-zero intersections being $\alpha_k \cdot \overline{\alpha}_k = +1$. Every characteristic element of $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$ can be written as a sum of even multiples of these generators, and hence, since Σ is characteristic, we can write in homology

$$\Sigma = 2a_1\alpha_1 + 2\overline{a}_1\overline{\alpha}_1 + \cdots + 2a_m\alpha_m + 2\overline{a}_m\overline{\alpha}_m$$

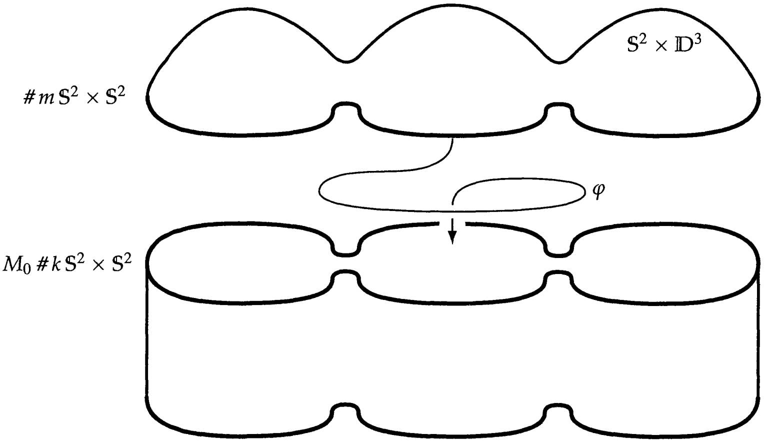
for some integers $a_1, \overline{a}_1, \ldots, a_m, \overline{a}_m$.

Using Wall's theorem on automorphisms (page 152), we deduce that the class of Σ can be taken through automorphisms of $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$'s intersection form to any other characteristic element with the same self-intersection. Since $\Sigma \cdot \Sigma = 0$, there must exist, for example, an automorphism φ_* such that

$$\varphi_*(\Sigma) = 2\alpha_1 \,.$$

Further, Wall's theorem on diffeomorphisms (page 153) implies that this algebraic automorphism can be realized through a self-diffeomorphism φ of $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$.

We use this diffeomorphism φ to attach[15] $\natural\, m\,\mathbb{D}^3 \times \mathbb{S}^2$ to the top of our earlier cobordism between M_0 and $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$, as in figure 10.13. Think of the class α_1 as represented by $\partial\,\mathbb{D}^3 \times 1$ of the first term of $\natural\, m\,\mathbb{D}^3 \times \mathbb{S}^2$.



10.13. Closing it up

Since the class $2\alpha_1$ bounds in $\natural\, m\,\mathbb{D}^3 \times \mathbb{S}^2$, the surface $\varphi[\Sigma]$ embedded in $\partial(\natural\, m\,\mathbb{D}^3 \times \mathbb{S}^2)$ must itself bound some oriented 3–manifold Z inside $\natural\, m\,\mathbb{D}^3 \times \mathbb{S}^2$. This follows from a mild refinement of the argument used for representing

---

[15] Here, $X \natural Y$ denotes the boundary sum of X and Y, which was explained on page 13. If you do not like boundary sums, you can always imagine starting with m copies of $\mathbb{D}^3 \times \mathbb{S}^2$ and continuing them with a small cobordism that connects their boundary $\mathbb{S}^2 \times \mathbb{S}^2$'s into $\# m\,\mathbb{S}^2 \times \mathbb{S}^2$.

homology 2–classes by surfaces in 4–manifolds. Further, Z^3 is dual to w_2 of its ambient.

Stacking, one on top of the other, all the small cobordisms that we encountered, we end up with a characteristic pair (W^5, Y^3) whose boundary is our initial (M, Σ).

The odd case. The case when the intersection form of M_0 is *odd* proceeds similarly, but literally with a twist. Since $\operatorname{sign} M_0 = 0$, we must have

$$Q_{M_0} = \oplus\, m\,[-1] \oplus m\,[+1] \;,$$

and hence $Q_{M_0 \# \mathbb{S}^2 \times \mathbb{S}^2} = \oplus\, m\,[-1] \oplus m\,[+1] \oplus H$. We change basis and rewrite the latter as

$$Q_{M_0 \# \mathbb{S}^2 \times \mathbb{S}^2} = \oplus\, m\, H \oplus \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} .$$

(This is possible because of the classification of indefinite forms: the right-hand term has the same rank, parity and signature.)

Thus, $M_0 \# \mathbb{S}^2 \times \mathbb{S}^2$ has the same intersection form as $\# m\, \mathbb{S}^2 \times \mathbb{S}^2 \# \mathbb{S}^2 \tilde{\times} \mathbb{S}^2$, where $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$ is the nontrivial 2–sphere bundle over $\mathbb{S}^2$. After adding enough extra copies of $\mathbb{S}^2 \times \mathbb{S}^2$, we end up with a cobordism between M_0 and $\# n\, \mathbb{S}^2 \times \mathbb{S}^2 \# \mathbb{S}^2 \tilde{\times} \mathbb{S}^2$, with Σ surviving as a characteristic surface.

Choose a basis $\{\alpha_1, \bar{\alpha}_1, \ldots, \alpha_m, \bar{\alpha}_m, \beta_0, \beta_1\}$ for the homology of $\# m\, \mathbb{S}^2 \times \mathbb{S}^2 \# \mathbb{S}^2 \tilde{\times} \mathbb{S}^2$, with the α's as before, and with β_0 the class of the fiber of $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$ and β_1 the class of its base; in other words, $\beta_0 \cdot \beta_0 = 0$, $\beta_0 \cdot \beta_1 = +1$, and $\beta_1 \cdot \beta_1 = +1$. In homology, we must have

$$\Sigma = 2a_1\alpha_1 + 2\bar{a}_1\bar{\alpha}_1 + \cdots + 2a_m\alpha_m + 2\bar{a}_m\bar{\alpha}_m + 2(b_0 + 1)\beta_0 + 2b_1\beta_1 \;.$$

Using Wall's theorems, we choose an automorphism φ_* such that

$$\varphi_*(\Sigma) = \beta_0 \;,$$

then find a diffeomorphism φ corresponding to it, and use it to attach the cap $\natural\, m\, \mathbb{D}^3 \times \mathbb{S}^2 \natural\, \mathbb{D}^3 \tilde{\times} \mathbb{S}^2$ to the top of our cobordism, where $\mathbb{D}^3 \tilde{\times} \mathbb{S}^2$ is the nontrivial 3–disk bundle over $\mathbb{S}^2$ and $\beta_0 = \partial\, \mathbb{D}^3 \tilde{\times} 1$. Since the class β_0 bounds, the surface $\varphi[\Sigma]$ must also bound a 3–manifold, and the conclusion of the theorem follows. □

Note: Modeling spin$^{\mathbb{C}}$ structures by using quaternions

If complex numbers model rotations and homotheties of the plane $\mathbb{R}^2$, the quaternions model the geometry of both $\mathbb{R}^3$ and $\mathbb{R}^4$. Indeed, this is how they were discovered by Hamilton. In what follows, we will use them to model the ingredients of a spin$^{\mathbb{C}}$ structure, *i.e.*, the Lie groups, maps, spinors, Clifford multiplication, and squaring map.

Meet the quaternions. Remember that the division algebra of **quaternions** is defined as the 4–dimensional real algebra

$$\mathbb{H} = \mathbb{R}\{1, i, j, k\}$$

with the multiplication governed by the relations $i^2 = -1$, $j^2 = -1$, $k^2 = -1$, and $ij = k$, $jk = i$, $ki = j$. The quaternions are not commutative.

The span of i, j, k is called the *imaginary* part of $\mathbb{H}$ and is denoted by $\operatorname{Im}\mathbb{H} = \mathbb{R}\{i, j, k\}$. Given a random quaternion $x = a + bi + cj + dk$, its *conjugate* is $\bar{x} = a - bi - cj - dk$. The natural metric of $\mathbb{R}^4$ can be recovered through

$$|x| = x \cdot \bar{x} .$$

In particular, for every unit-length quaternion $\xi \in S^3$, we have $\xi^{-1} = \bar{\xi}$. More generally, we have $\langle x, y \rangle = \operatorname{Re}(x \cdot \bar{y})$.

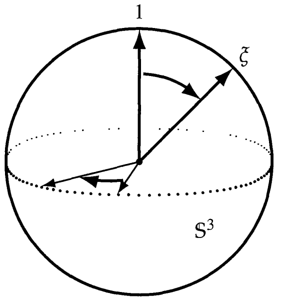
Also, if we restrict to $\operatorname{Im}\mathbb{H}$, then quaternion multiplication in $\operatorname{Im}\mathbb{H}$ coincides exactly with the cross-product of $\mathbb{R}^3$: we have $x \times y = x \cdot y$. The identification $\mathbb{R}^3 = \operatorname{Im}\mathbb{H}$ is in fact the reason why the directions in $\mathbb{R}^3$ are traditionally[16] labeled i, j, k.

Quaternion multiplication is best understood as follows: if one picks a unit-length quaternion ξ, then the map

$$x \longmapsto \xi \cdot x$$

is the isometry of $\mathbb{R}^4$ that acts on the plane $\mathbb{R}\{1, \xi\}$ as the rotation of angle ϑ that sends 1 to ξ, and acts on the complementary plane $\mathbb{R}\{1, \xi\}^\perp$ by rotating with the same angle ϑ, in the direction compatible with the orientation of $\mathbb{R}^4$, as in figure 10.14.

A choice of a favorite direction of rotation in a plane is the same thing as an orientation of that plane. Orienting a plane P in $\mathbb{R}^4$, together with the standard orientation of $\mathbb{R}^4$, induces a preferred orientation of its complement $P^\perp$, and thus a favorite direction of rotation in $P^\perp$.



10.14. Quaternion multiplication

The map $x \longmapsto x \cdot \xi$ acts similarly, only that the complement $\mathbb{R}\{1, \xi\}^\perp$ is spun in the opposite direction.

---

[16] In fact, quaternions were historically so fashionable that their supporters vehemently opposed the modern view on dot-product and cross-product; thus, when the latter inevitably prevailed, quaternions fell into a disgrace that lingers to the present day.

Lie groups. The quaternions can model all the isometries of $\mathbb{R}^4$. Indeed, the action of $SO(4)$ on $\mathbb{R}^4$ coincides with the action of $\mathbb{S}^3 \times \mathbb{S}^3$ on $\mathbb{H}$ given by

$$(\xi_+, \xi_-) \colon \; x \longmapsto \xi_+ \cdot x \cdot \xi_-^{-1} .$$

Since (ξ_+, ξ_-) and $(-\xi_+, -\xi_-)$ act the same, this exhibits

$$SO(4) = \mathbb{S}^3 \times \mathbb{S}^3 \,/\, \pm 1 ,$$

and therefore its double-cover can only be

$$Spin(4) = \mathbb{S}^3 \times \mathbb{S}^3 .$$

In particular, one can readily see the homotopy groups

$$\pi_1 SO(4) = \mathbb{Z}_2 , \qquad \pi_2 SO(4) = 0 , \qquad \pi_3 SO(4) = \mathbb{Z} \oplus \mathbb{Z} ,$$

and identify their generators, *etc.*

At the level of Lie algebras, since $\mathfrak{g} = T_G|_1$, we have

$$\mathfrak{so}(4) = \operatorname{Im}\mathbb{H} \oplus \operatorname{Im}\mathbb{H} ,$$

and the action of $\mathfrak{so}(4)$ on $\mathbb{R}^4$, as skew-symmetric endomorphisms, is modeled by

$$(a_+, a_-) \colon \; x \longmapsto a_+ \cdot x \, - \, x \cdot a_- .$$

On the other hand, if we look at the action of $\mathbb{S}^3$ by conjugation

$$\xi \colon \; x \longmapsto \xi \cdot x \cdot \xi^{-1} ,$$

the first thing to notice is that it fixes 1, and thus the whole real line $\mathbb{R} \subset \mathbb{H}$. If we look at its complement $\operatorname{Im}\mathbb{H}$ and identify the latter with $\mathbb{R}^3$, we recover the action of $SO(3)$ on $\mathbb{R}^3$. Thus

$$SO(3) = \mathbb{S}^3 \,/\, \pm 1 \qquad \text{and} \qquad Spin(3) = \mathbb{S}^3 .$$

In particular, the group $SO(3)$ is diffeomorphic to $\mathbb{RP}^3$.

The Lie algebras and adjoint action are

$$\mathfrak{so}(3) = \operatorname{Im}\mathbb{H} \qquad \text{with} \qquad a \colon \; x \longmapsto a \cdot x \, - \, x \cdot a .$$

The splitting $\mathfrak{so}(4) = \mathfrak{so}(3) \oplus \mathfrak{so}(3)$ is now obvious, and appears as the differential of the natural map

$$SO(4) \longrightarrow SO(3) \times SO(3) \colon \qquad [\xi_+, \xi_-] \longmapsto [\xi_+] \oplus [\xi_-] .$$

Complex groups. On the complex side of the world, we can identify $\mathbb{C}^2$ with $\mathbb{H}$. There are two ways of doing this, and one of them is wrong. We will work with

$$\mathbb{C}^2 \equiv \mathbb{H} \colon \qquad (z_1, z_2) \equiv z_1 + z_2 j .$$

This has the advantage, over the alternative $(z_1, z_2) \equiv z_1 + j z_2$, of preserving the natural orientations of $\mathbb{C}^2$ and $\mathbb{H} = \mathbb{R}^4$.

In this identification, the complex structure of $\mathbb{H}$ appears by multiplying with complex scalars on the *left.* Multiplication by quaternions on the *right* is then $\mathbb{C}$-linear. In fact, the action of $\mathbb{S}^3$ on $\mathbb{H}$ on the right as

$$\xi \colon \; x \longmapsto x \cdot \xi^{-1}$$

is modeling the action of $SU(2)$ on $\mathbb{C}^2$. Therefore

$$SU(2) = \mathbb{S}^3 \,.$$

The corresponding Lie algebra and adjoint action are

$$\mathfrak{su}(2) = \operatorname{Im}\mathbb{H} \qquad \text{with} \qquad a\colon\, x \longmapsto -x \cdot a \,.$$

Also, the double cover $\mathbb{S}^3 \to \mathbb{RP}^3$ can be viewed as a natural map

$$SU(2) \longrightarrow SO(3)\colon \qquad \xi \longmapsto [\xi] \,.$$

We can enrich the action of $SU(2)$ with determinants from $U(1)$, by coupling the right-action of $\mathbb{S}^3$ with a left action of $\mathbb{S}^1$, as in

$$(\lambda,\, \xi)\colon\, x \longmapsto \lambda \cdot x \cdot \xi^{-1} \,.$$

This exhibits the unitary group $U(2)$ as

$$U(2) = \mathbb{S}^1 \times \mathbb{S}^3 \,/\, \pm 1 \,,$$

with Lie algebra and adjoint action

$$\mathfrak{u}(2) = i\,\mathbb{R} \oplus \operatorname{Im}\mathbb{H} \qquad \text{with} \qquad (ir, a)\colon\, x \longmapsto irx \,-\, x \cdot a \,.$$

The inclusion of $U(2)$ into $SO(4)$ is simply

$$U(2) \subset SO(4)\colon \qquad [\lambda,\, \xi] \longmapsto [\lambda,\, \xi] \,,$$

as induced by $\mathbb{S}^1 \subset \mathbb{S}^3$.

Finally, the complex spin group is

$$Spin^{\mathbb{C}}(4) = \mathbb{S}^1 \times \mathbb{S}^3 \times \mathbb{S}^3 \,/\, \pm 1$$

and fits in the happy diagram:

$$\begin{tikzcd}
 & U(1) & \\
U(2) \arrow[d] & Spin^{\mathbb{C}}(4) \arrow[u, "\det"'] \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"] \arrow[d] & U(2) \arrow[d] \\
SO(3) & SO(4) \arrow[l, "\lambda_-"'] \arrow[r, "\lambda_+"] & SO(3)
\end{tikzcd}$$

$$\begin{tikzcd}
 & \lambda^2 & \\
{[\lambda,\, \xi_-]} \arrow[d] & {[\lambda,\, \xi_+,\, \xi_-]} \arrow[u, "\det"'] \arrow[l, "\rho_-"'] \arrow[r, "\rho_+"] \arrow[d] & {[\lambda,\, \xi_+]} \arrow[d] \\
{[\xi_-]} & {[\xi_+,\, \xi_-]} \arrow[l, "\lambda_-"'] \arrow[r, "\lambda_+"] & {[\xi_+]}
\end{tikzcd} \,.$$

This diagram was already mentioned earlier (page 425).

Cocycles and spin$^{\mathbb{C}}$ structures. To make notations readable, we will drop the subscripts $\alpha\beta$ from all cocycles discussed.

Let $\{[\xi_+,\, \xi_-]\}$ be the $SO(4)$–cocycle of T_M, gluing T_M through its action as

$$[\xi_+,\, \xi_-]\colon\ v \longmapsto \xi_- \cdot v \cdot \xi_+^{-1}\ .$$

Then $\{[\xi_+]\}$ is the $SO(3)$–cocycle that glues $\Lambda^2_+(T^*_M)$ by acting on $\operatorname{Im}\mathbb{H}$ through

$$[\xi_+]\colon\ \alpha \longmapsto \xi_+ \cdot \alpha \cdot \xi_+^{-1}\ .$$

Let $\mathfrak{s} = \{[\lambda,\, \xi_+,\, \xi_-]\}$ be a spin$^{\mathbb{C}}$ structure on M. Then its determinant line bundle $\mathcal{L}$ has cocycle $\{\lambda^2\}$, while the spinor bundles $\mathcal{W}^\pm$ are glued by the $U(2)$–cocycles $\{[\lambda,\, \xi_\pm]\}$, acting by

$$[\lambda,\, \xi_\pm]\colon\ w \longmapsto \lambda \cdot w \cdot \xi_\pm^{-1}\ .$$

Clifford multiplication is the map

$$T_M \times \mathcal{W}^+ \xrightarrow{\ \bullet\ } \mathcal{W}^-\colon \qquad v \bullet w = w \cdot v\ .$$

This local definition can be checked to be compatible with the cocycles of T_M, $\mathcal{W}^+$ and $\mathcal{W}^-$, and thus to be globally well-defined. Its adjoint $T_M \times \mathcal{W}^- \xrightarrow{\ \bullet\ } \mathcal{W}^+$ is given by $v \bullet w = -w \cdot \bar{v}$, and we have indeed $v \bullet (v \bullet w) = -\left|v\right|^2 w$. The induced action $\Lambda^2_+ \times \mathcal{W}^+ \xrightarrow{\ \bullet\ } \mathcal{W}^+$ is given by $a \bullet w = w \cdot a$, where a is the quaternion coordinate of a 2–form[17] from Λ^2_+ and w is a spinor field from $\mathcal{W}^+$.

The squaring map $\sigma\colon \mathcal{W}^+ \longrightarrow i\Lambda^2_+$, which is characterized by $\sigma(\varphi) \bullet \varphi = \frac{1}{2}\left|\varphi\right|^2 \varphi$, can be described as sending the self-dual spinor of coordinate w to i times the self-dual 2–form of quaternion coordinate $\frac{1}{2}\,\overline{w}\cdot i \cdot w$:

$$\sigma(w) = i\boldsymbol{\cdot}\left(\tfrac{1}{2}\,\overline{w}\cdot i\cdot w\right)\ .$$

(The outside i from $i\Lambda^2_+$ needs to stay apart from i as quaternion coordinate inside Λ^2_+; thus, the cumbersome phrase "i times *blah*", or the symbol "$\boldsymbol{\cdot}$".)

If one prefers complex coordinates, then by writing $w = z_1 + z_2 j$ with $z_1, z_2 \in \mathbb{C}$, one gets

$$\sigma(z_1 + z_2 j) = i\boldsymbol{\cdot}\left(i(|z_1|^2 - |z_2|^2) + 2i\bar{z}_1 z_2 j\right)\ .$$

Connections. Assume that the Levi–Cività connection ∇ of the Riemannian metric on M is locally described as

$$\nabla|_U = d_{\mathit{flat}} + \Theta$$

for some local $\mathfrak{so}(4)$–valued 1–form $\Theta \in \Gamma\bigl(\mathfrak{so}(4) \otimes T^*_M|_U\bigr)$. Here $d_{\mathit{flat}}v$ is the differential of the local-coordinate components of the vector field v.

Using quaternion coordinates, we write $\mathfrak{so}(4) = \operatorname{Im}\mathbb{H} \oplus \operatorname{Im}\mathbb{H}$, and we split $\Theta = (\Theta_+, \Theta_-)$ with $\Theta_\pm \in \Gamma\bigl(\operatorname{Im}\mathbb{H} \otimes T^*_M|_U\bigr)$. The Levi–Cività connection is then described by

$$\nabla v|_U = d_{\mathit{flat}}\, v + \Theta_+\cdot v - v\cdot\Theta_-$$

for any vector field locally described by $v\colon U \to \mathbb{H}$.

---

[17] For example, the 2–form $1 \wedge i + j \wedge k$ (written using quaternion coordinates on $T_M \approx T^*_M$) has quaternion coordinate $2i$.

> *The Levi–Cività connection induces a natural connection* ∇^+ *on* $\Lambda^2_+(T^*_M)$*, given by*
>
> $$\nabla^+\alpha|_U = d_{flat}\,\alpha + \Theta_+\cdot\alpha - \alpha\cdot\Theta_+$$
>
> *for any* 2*–form locally described by* $\alpha\colon U \to \operatorname{Im}\mathbb{H}$*. Since the Levi–Cività is torsion-free, a fundamental property of* ∇^+ *is that it can be used to compute exterior derivatives:*
>
> $$d\alpha = \sum e_k \wedge \nabla^+_{e_k}\alpha \qquad \text{and} \qquad d^*\alpha = -\sum e_k \lrcorner \nabla^+_{e_k}\alpha$$
>
> *for any local orthonormal frame* $\{e_1, e_2, e_3, e_4\}$ *in* $T_M \approx T^*_M$*. Here,* $d^*\colon \Gamma(\Lambda^2) \longrightarrow \Gamma(\Lambda^1)$ *denotes the formal adjoint of* d*, given by* $d^* = -*d*$*.*

Let A denote some $U(1)$–connection on $\mathcal{L}$, which acts on a section locally described by $\ell\colon U \to \mathbb{C}$ through the formula

$$d_A\ell|_U = d_{flat}\,\ell + i\vartheta_A\cdot\ell$$

for some suitable local 1–form $\vartheta_A \in \Gamma(T^*_M|_U)$. Then the induced connection ∇^A on $\mathcal{W}^+$ acts as

$$\nabla^A\varphi|_U = d_{flat}\,\varphi + \tfrac{1}{2}i\,\vartheta_A\cdot\varphi - \varphi\cdot\Theta_+$$

for every self-dual spinor field locally described by $\varphi\colon U \to \mathbb{H}$. (The 1/2–factor appears because the map $Spin^{\mathbb{C}}(4) \to U(1)$ is the map $\mathbb{S}^1 \times Spin(4)/\pm 1 \longrightarrow \mathbb{S}^1$ given by $[\lambda,\ \xi] \mapsto \lambda^2$.)

The complex case. Assume that M is endowed with an almost-complex structure J, and that the corresponding $U(2)$–cocycle of T_M is $\{[\lambda,\ \xi_-]\}$. The anti-canonical line bundle K^* has $U(1)$–cocycle $\{\lambda^2\}$.

The bundle $\Lambda^2_+(T^*_M)$ has $SO(3)$–cocycle $\{[\lambda]\}$. Writing $\alpha \in \operatorname{Im}\mathbb{H}$ as $\alpha = ai + bj + ck$, we have $\lambda\alpha\lambda^{-1} = ai + \lambda^2(bj + ck)$, and thus we see that $\Lambda^2(T^*_M)$ splits as

$$\Lambda^2_+(T^*_M) = \mathbb{R}\omega \oplus K^*\ .$$

In this writing, the fundamental form ω has quaternion coordinate i.

The canonical spin$^{\mathbb{C}}$ structure induced by J is the $Spin^{\mathbb{C}}(4)$–cocycle $\{[\lambda,\ \xi_-,\ \lambda]\}$, whose determinant bundle is indeed $\mathcal{L} = K^*$. Further, the way the spinor bundles are glued-up can easily be used to establish the complex-bundle isomorphisms

$$\mathcal{W}^+ = \underline{\mathbb{C}} \oplus K^* \qquad \text{and} \qquad \mathcal{W}^- = (T_M, J)\ .$$

Indeed, on one hand $\mathcal{W}^-$ and T_M have the same cocycle, acting in the same way. On the other hand, if $w = z_1 + z_2 j$ is the quaternion coordinate in $\mathcal{W}^+$, then the action of its cocycle $\{[\lambda, \lambda]\}$ is $\lambda(z_1 + z_2 j)\lambda^{-1} = z_1 + \lambda^2 z_2$, thus exhibiting $\mathcal{W}^+$'s splitting.

In the almost-complex case, the earlier description of the squaring map $\sigma\colon \mathcal{W}^+ \to i\Lambda^2_+$ can be rearranged according to the splittings of both $\mathcal{W}^+$ and Λ^2_+: If $z_1 + z_2 j$ is the quaternion coordinate of a self-dual spinor from $\mathcal{W}^+ = \underline{\mathbb{C}} \oplus K^*$, with z_1 being the coordinate in $\underline{\mathbb{C}}$ and z_2 the coordinate in K^*, then $\sigma(z_1 + z_2 j)$ is the imaginary-valued 2–form made by adding $i\cdot\left(|z_1|^2 - |z_2|^2\right)\omega$ from $i\mathbb{R}\omega$ together with $i\cdot(2i\,\bar{z}_1 z_2)$ from iK^*.

Using complex-valued forms. The isomorphisms $\mathcal{W}^+ = \underline{\mathbb{C}} \oplus K^*$ and $\mathcal{W}^- = (T_M, J)$ are most often written as

$$\mathcal{W}^+ = \Lambda^{0,0} \oplus \Lambda^{0,2} \qquad \text{and} \qquad \mathcal{W}^- = \Lambda^{0,1} ,$$

where $\Lambda^{p,q}$ is the bundle of complex-valued forms of type[18] (p,q). Indeed, $K^* = \det_{\mathbb{C}}(T_M, J)$ can on one hand be identified with $\Lambda^{0,2}$, and on the other included in Λ^2_+ as the complement of ω. The composition $K^* \subset \Lambda^2_+ \subset \Lambda^2_+ \otimes \mathbb{C} \to \Lambda^{0,2}$ is an isomorphism of complex-line bundles.[19]

In this case, the representation of the Clifford multiplication $T_M \times \mathcal{W}^+ \to \mathcal{W}^-$ becomes the map $T_M \times \left(\Lambda^{0,0} \oplus \Lambda^{0,2}\right) \longrightarrow \Lambda^{0,1}$ given by

$$v \bullet (f + \alpha) = \sqrt{2} f \cdot v^{0,1} - \sqrt{2}(v \lrcorner \alpha)^{0,1} .$$

Here $v \mapsto v^{0,1}$ denotes the identification of T_M with $\Lambda^{0,1}$ through

$$v^{0,1} = \tfrac{1}{2}\left(v^* - i(Jv)^*\right) ,$$

where $w^*(\cdot) = \langle \cdot , w \rangle_{\mathbb{R}}$. Also, $(v \lrcorner \alpha)^{0,1}$ denotes the projection to $\Lambda^{0,1}$ of the contraction (interior product) $v \lrcorner \alpha \in \Lambda^1 \otimes \mathbb{C}$ of the 2–form α by v. The 1–form $v \lrcorner \alpha$ is defined by the formula

$$(v \lrcorner \alpha)(\cdot) = \alpha(v, \cdot) ,$$

while the projection $\Lambda^1 \otimes \mathbb{C} \to \Lambda^{0,1} \colon \beta \mapsto \beta^{0,1}$ is given by

$$\beta^{0,1} = \tfrac{1}{2}\left(\beta + i\beta(J\cdot)\right) .$$

The factor $\sqrt{2}$ merely ensures that the Clifford multiplication preserves lengths.

By writing $\mathcal{W}^+ = \Lambda^{0,0} \oplus \Lambda^{0,2}$ and $\Lambda^2_+ = \mathbb{R}\omega \oplus \Lambda^{0,2}$, we end up with the following characterization of σ:

$$\langle \sigma(f+\alpha), i\omega \rangle = |f|^2 - |\alpha|^2 \qquad \text{and} \qquad \sigma(f+\alpha)^{0,2} = 2\bar{f} \cdot \alpha .$$

The Kähler case. Assume now that M is a Kähler manifold. This means that the Levi-Cività connection is in fact $\mathbb{C}$–linear, and therefore, in its local description as

$$\nabla v|_U = d_{\mathit{flat}}\, v + \Theta_+ \cdot v - v \cdot \Theta_- ,$$

all the local forms Θ_+ must take values in $i\,\mathbb{R}$. Further, the operators

$$d_{A_0} \ell|_U = d_{\mathit{flat}}\, \ell + 2\Theta_+ \cdot \ell$$

match up to define a $U(1)$–connection A_0 on the anti-canonical bundle K^*. This A_0 is a special connection, and we explore it further:

Take the spin$^{\mathbb{C}}$ structure determined by K^* and use d_{A_0} as a fixed connection on $\mathcal{L} = K^*$. Then the induced connection on $\mathcal{W}^+$ acts through

$$\nabla^{A_0} \varphi|_U = d_{\mathit{flat}}\, \varphi + \Theta_+ \cdot \varphi - \varphi \cdot \Theta_+ .$$

---

[18] These were mentioned in the end-notes of chapter 3 (complex duals, page 136), and explained in the end-notes of the preceding chapter (connections and holomorphic bundles, page 365).

[19] Remember also the decomposition $\Lambda^2_+ \otimes \mathbb{C} = \mathbb{C}\omega \oplus \Lambda^{0,2} \oplus \Lambda^{0,2}$, where $\Lambda^{2,0} \oplus \Lambda^{0,2} = K^* \otimes \mathbb{C}$.

We now split $\mathcal{W}^+ = \underline{\mathbb{C}} \oplus K^*$, and correspondingly write $\varphi = z_1 + z_2 j$ for some local $z_1, z_2 \colon U \to \mathbb{C}$ that describe, respectively, a complex-valued function on M and a section in K^*. We have

$$\nabla^{A_0}(z_1 + z_2 j)|_U = d_{flat} z_1 + d_{flat} z_2 j + \Theta_+ \cdot z_2 j - z_2 j \cdot \Theta_+ ,$$

since $\Theta_+ \cdot z_1 = z_1 \cdot \Theta_+$. On one hand, we can identify $d_{flat} z_1$ with the complex-valued differential dz_1 of $z_1 \colon U \to \mathbb{C}$. On the other hand, we can identify z_2 through the inclusion $K^* \subset \mathbb{R}\omega \oplus K^* = \Lambda^2_+$ with a self-dual 2–form; then we read the action of ∇^{A_0} on the spinor z_2 as identical to the action of ∇^+ on the 2–form z_2, where ∇^+ is the connection on Λ^2_+ induced from the Levi–Cività connection.

Going global, we think of our $\varphi = z_1 + z_2 j$ as the local description of the sum $f + \alpha$, with $f \in \Gamma(\Lambda^{0,0})$, *i.e.*, $f \colon M \to \mathbb{C}$, and $\alpha \in \Gamma(K^*) = \Gamma(\Lambda^{0,2})$. Then we write

$$\nabla^{A_0}(f + \alpha) = df + \nabla^+ \alpha .$$

Using the earlier description of Clifford multiplication as $v \bullet (f + \alpha) = \sqrt{2} f \cdot v^{0,1} - \sqrt{2}\,(v \lrcorner \alpha)^{0,1}$, as well as the description of the Dirac operator as locally $D^{A_0}\varphi = \sum e_k \bullet \nabla^{A_0}_{e_k}\varphi$, we deduce that

$$\mathcal{D}^{A_0}(f + \alpha) = \sqrt{2}\,(df)^{0,1} + \sqrt{2}\,(d^*\alpha)^{0,1}$$

or, in other words,

$$\tfrac{1}{\sqrt{2}}\mathcal{D}^{A_0}(f + \alpha) = \bar\partial f + \bar\partial^* \alpha ,$$

where $\bar\partial^* \colon \Gamma(\Lambda^{0,2}) \to \Gamma(\Lambda^{0,1})$ is the formal adjoint to $\bar\partial \colon \Gamma(\Lambda^{0,1}) \to \Gamma(\Lambda^{0,2})$. The Cauchy–Riemann operators $\bar\partial$ have been encountered in the end-notes of the preceding chapter (page 365).

Using these last observations, in the end-note on page 457 ahead, we will rewrite the Seiberg–Witten equations so as to explore the complex geometry of Kähler surfaces, and extend the discussion to symplectic manifolds on page 465. Indeed, symplectic manifolds are close enough to the complex realm that there exists a connection A_0 on K^* so that the induced Dirac operator is still $\frac{1}{\sqrt{2}}\mathcal{D}^{A_0} = \bar\partial + \bar\partial^*$. If the manifold is merely almost-complex, this does not happen.

Note: The Seiberg–Witten moduli space

In this note we will argue that the Seiberg–Witten moduli space is a smooth orientable manifold, as well as lay the foundation for proving in the next note (page 454) Donaldson's theorem by using Seiberg–Witten theory.

The analytic setup. We start by outlining some technical details that ensure that we can reasonably manipulate the infinite-dimensional creatures involved in the definition of the Seiberg–Witten invariants. We will not go deeper than a mere sketch, and afterwards will just assume that the various spaces encountered can be handled similarly to finite-dimensional manifolds.

While we all know and love smooth functions, they are not best suited for the study of infinite-dimensional beasts. Indeed, completeness of the various spaces becomes an important issue. For example, the space of all smooth sections in $\mathcal{W}^+$ can, of course, be endowed with the L^2 inner product

$$(\varphi, \psi) = \int_M \langle \varphi, \psi \rangle_{\mathbb{R}} \, \mathrm{vol}_M \,,$$

but the resulting topology of $\Gamma(\mathcal{W}^+)$ is not complete: there are Cauchy sequences with no limit.

One possibility for dealing with this is to refine the inner product to control *all* derivatives.[20] Another possibility is to use a suitable *completion* of $\Gamma(\mathcal{W}^+)$ by accepting, alongside sections with components locally described by functions from C^∞, sections locally described by elements of some suitable *Sobolev spaces* L^2_k.

Definition of Sobolev spaces. *Let $f\colon U \to \mathbb{R}$ be a locally integrable function, with $U \subset \mathbb{R}^n$. If f had a partial derivative $g = \partial_k f$, then, by using a simple integration by parts, one would have*

$$\int_U g \cdot v = -\int_U f \cdot \partial_k v$$

for every smooth compactly-supported test function $v \in C_0^\infty(U)$. Without assuming that f has any partial derivatives, we call **weak derivative** *of f any function g satisfying the above property for all smooth v's, and even write $\partial_k f = g$.*

A measurable function $f\colon U \to \mathbb{R}$ is an element of $L^2(U)$ if it is square-integrable, i.e., if $\int_U |f|^2 < \infty$. A measurable function $f\colon U \to \mathbb{R}$ is an element of $L^2_k(U)$ if it is square-integrable and has weak derivatives up to order k that are square-integrable.

The **Sobolev space** $$L^2_k(U) = \{ f \in L^2(U) \mid \partial_\alpha f \in L^2(U),\ |\alpha| \leq k \}$$

(using multi-index notation[21]) is the Hilbert space that appears as the completion of the space $C^\infty(U)$ with respect to the inner product

$$\langle f, g \rangle = \sum_{|\alpha| \leq k} \int_U \partial_\alpha f \cdot \partial_\alpha g \,.$$

Other notations often used for $L^2_k(U)$ are $H^{k,2}(U)$ and $W^{k,2}(U)$.

If one controls enough derivatives, the elements of $L^2_k(U)$ begin to be continuous and have actual derivatives. For example, if $U \subset \mathbb{R}^4$, then the closure of $C_0^\infty(U)$ in $L^2_k(U)$ is in fact contained in $C^m(\overline{U})$, for m with $0 \leq m \leq k-2$. In particular, the closure of $C_0^\infty(U)$ in $L^2_2(U)$ contains only continuous functions, while its closure in $L^2_3(U)$ is made of C^1-functions.

Thus, we extend the setting of the Seiberg–Witten invariants to allow for L^2_k-sections, L^2_k-connections, L^2_k-gauge-transformations, and so on. This means that in local coordinates everybody is described by elements from[22] $L^2_k(U)$'s. We pick k big enough so that we control enough of the derivatives involved in connections, curvature, gauge actions, *etc.* For example, $k = 4$ will do.

---

[20] For example, use the inner product $(\varphi, \psi) = \sum \frac{1}{2^k} \int_M \langle (\nabla^A)^k \varphi,\ (\nabla^A)^k \psi \rangle \, \mathrm{vol}_M$, where A is a random connection on $\mathcal{L}$, and $(\nabla^A)^k\colon \Gamma(\mathcal{W}^+) \to \Gamma(\mathcal{W}^+ \otimes (T^*_M)^{\otimes k})$ is the operator obtained by composition of k times ∇^A combined with the Levi-Cività connection on T^*_M; of course, $(\nabla^A)^0 \varphi = \varphi$. The factor $\frac{1}{2^k}$ is there simply to ensure the convergence of the series.

[21] That is, $\alpha = (k_1, \dots, k_n)$ with $|\alpha| = k_1 + \cdots + k_n$, and $\partial_\alpha f = \partial_1^{k_1} \cdots \partial_n^{k_n} f$.

[22] Of course, for a coordinate-free alternative, one can build a global theory of spaces like $L^2_k(\mathcal{W}^+)$, made of all measurable sections $\varphi\colon M \to \mathcal{W}^+$ such that $\sum_{j \leq k} \int_M \left| (\nabla^A)^j \varphi \right|^2 \mathrm{vol}_M < \infty$ for some random choice of connection ∇^A on $\mathcal{W}^+$ and Riemannian metric on M, with $(\nabla^A)^j \varphi$ being the $\mathcal{W}^+$-valued j-form that appears after applying ∇^A for j times (weak derivatives, of course). Cute.

The fundamental fact is that the moduli space $\mathfrak{M}$ does not become bigger after all this, only its surroundings get nicer. Indeed, the Seiberg–Witten equations are elliptic, and we have

Weyl's Lemma. *Let P be an elliptic partial differential operator. If, when given $v \in C^\infty$ and using weak derivatives, we have $Pu = v$, then $u \in C^\infty$, and further we also have $Pu = v$ when using genuine derivatives.* □

Rigorously, in all that follows in this note, $\Gamma(E)$ should be understood to denote the space of all L^2_4–sections of E, while $Conn(E)$ should denote the space of all L^2_4–connections on E, and so on. The Seiberg–Witten equations should be understood in terms of weak derivatives. In the same spirit, the gauge group $\mathscr{G}(\mathcal{L})$ will be the space of all L^2_4–maps $M \to \mathbb{S}^1$. It can be proved that $\mathscr{G}(\mathcal{L})$ is an infinite-dimensional Lie group, modeled on Hilbert spaces. Its action on $\Gamma(\mathcal{W}^+) \times Conn(\mathcal{L})$ is smooth (in the appropriate sense).

The **configuration space**

$$\mathfrak{B} = \Gamma(\mathcal{W}^+) \times Conn(\mathcal{L}) \;/\; \mathscr{G}(\mathcal{L})$$

is the natural ambient for the Seiberg–Witten moduli space $\mathfrak{M}$. At the outset, since $\mathscr{G}(\mathcal{L})$ is not compact, it is not obvious that this orbit space is even a Hausdorff space, in other words, that there cannot exist any orbits of $\mathscr{G}(\mathcal{L})$'s action that keep getting closer-and-closer inside $\Gamma(\mathcal{W}^+) \times Conn(L)$. Thus, one first needs to prove that $\mathfrak{B}$ is Hausdorff. It has been done.

Since the action of $\mathscr{G}(\mathcal{L})$ on $\Gamma(\mathcal{W}^+) \times Conn(\mathcal{L})$ fails to be free on the pairs (φ, A) with $\varphi \equiv 0$, we can cut these out and look at the space

$$\mathfrak{B}_* = \big(\Gamma(\mathcal{W}^+) \setminus 0\big) \times Conn(\mathcal{L}) \;/\; \mathscr{G}(\mathcal{L})\,.$$

This open subset $\mathfrak{B}_* = \mathfrak{B} \setminus \big\{[0, A]\big\}$ of $\mathfrak{B}$ is the natural ambient of the moduli space $\mathfrak{M}_*$ of irreducible solutions. If there are no reducible Seiberg–Witten solutions,[23] then $\mathfrak{B}_*$ is the ambient of the whole $\mathfrak{M}$. It can be proved that $\mathfrak{B}_*$ is a smooth infinite-dimensional manifold, modeled on Hilbert spaces.

Everything proceeds from here on as if we were dealing with smooth manifolds of finite dimension. The complete proofs detailing this similarity are quite extensive. See **J. Morgan**'s ***The Seiberg–Witten equations and applications to the topology of four-manifolds*** **[Mor96]** for a better outline, or **L. Nicolaescu**'s ***Notes on Seiberg–Witten theory*** **[Nic00]** for complete and detailed arguments.

As far as we are concerned, in what follows we will accept that everything can be checked to be wonderful in the infinite-dimensional realm, and we will treat such creatures as if they were finite-dimensional smooth manifolds. In particular, we will drop any mention of L^2_k's altogether.

Even so, there is plenty of work ahead for proving that $\mathfrak{M}$ is a smooth orientable manifold.

---

[23] For example, if $b_2^+(M) \geq 1$ and we pick a generic Riemannian metric.

Differential of the Seiberg–Witten map. For studying the Seiberg–Witten moduli space, we define the map

$$\begin{aligned} \mathfrak{sw}\colon \Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L}) &\longrightarrow \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+) \\ \mathfrak{sw}(\varphi, A) &= \big(\mathcal{D}^A\varphi,\ \ F_A^+ - \sigma(\varphi)\big)\ . \end{aligned}$$

The Seiberg–Witten solution space $\mathfrak{S}$ is the zero-set of this function, while $\mathfrak{M}$ is the quotient of $\mathfrak{S}$ by $\mathcal{G}(\mathcal{L})$.

We compute the derivative

$$d\mathfrak{sw}\colon\ T_{\Gamma(\mathcal{W}^+)\times\mathcal{C}onn(\mathcal{L})}\big|_{(\varphi,A)} \longrightarrow T_{\Gamma(\mathcal{W}^-)\times i\Gamma(\Lambda^2_+)}\big|_{(0,0)}$$

of this function, at a Seiberg–Witten solution (φ, A). Since $\Gamma(\mathcal{W}^\pm)$ and $\Gamma(\Lambda^2_+)$ both are vector spaces, they are their own tangent spaces. Since $\mathcal{C}onn(\mathcal{L})$ is an affine space modeled on $i\Gamma(\Lambda^1)$, the latter will be its tangent space.

Therefore, we pick some $\psi \in \Gamma(\mathcal{W}^+)$ and $\vartheta \in \Gamma(\Lambda^1)$ and consider the variation

$$t \longmapsto \mathfrak{sw}(\varphi + t\psi,\ \ A + it\vartheta)\ .$$

Then we compute $d\mathfrak{sw}$ by using

$$d\mathfrak{sw}\big|_{(\varphi,A)}(\psi, i\vartheta) = \tfrac{d}{dt}\Big|_{t=0}\mathfrak{sw}(\varphi + t\psi,\ \ A + it\vartheta)\ .$$

We have:

$$\mathfrak{sw}(\varphi + t\psi,\ A + it\vartheta) = \big(\mathcal{D}^{A+it\vartheta}(\varphi + t\psi),\ \ F_A^+ + it\,d^+\vartheta - \sigma(\varphi + t\psi)\big)\ .$$

On one hand, we get

$$\begin{aligned} \mathcal{D}^{A+it\vartheta}(\varphi + t\psi) &= \mathcal{D}^A\varphi + t\mathcal{D}^A\psi + \tfrac{1}{2}it\vartheta\bullet\varphi + \tfrac{1}{2}it^2\vartheta\bullet\psi \\ &= t\big(\mathcal{D}^A\psi + \tfrac{1}{2}i\vartheta\bullet\varphi\big) + \tfrac{1}{2}it^2\vartheta\bullet\psi\ , \end{aligned}$$

by using that $\mathcal{D}^A\varphi = 0$, that $\mathcal{D}^A$ is linear, and that it satisfies[24] $\mathcal{D}^{A+i\vartheta}\psi = \mathcal{D}^A\psi + \frac{1}{2}i\vartheta\bullet\psi$. On the other hand, we have

$$\begin{aligned} \sigma(\varphi + t\psi) &= (\varphi + t\psi)\otimes(\varphi + t\psi)^* - \tfrac{1}{2}\left|\varphi + t\psi\right|^2\cdot id \\ &= \varphi\otimes\varphi^* + t\psi\otimes\varphi^* + t\varphi\otimes\psi^* + t^2\psi\otimes\psi^* \\ &\qquad - \tfrac{1}{2}\left|\varphi\right|^2\cdot id - t\tfrac{1}{2}2\langle\varphi,\psi\rangle_{\mathbb{R}}\cdot id - t^2\tfrac{1}{2}\left|\psi\right|^2\cdot id\ . \end{aligned}$$

Since $\sigma(\varphi) = \varphi\otimes\varphi^* - \frac{1}{2}\left|\varphi\right|^2\cdot id$ and $F_A^+ = \sigma(\varphi)$, we conclude that

$$\begin{aligned} &F_A^+ + it\,d^+\vartheta - \sigma(\varphi + t\psi) = \\ &\qquad = t\big(i\,d^+\vartheta - \psi\otimes\varphi^* - \varphi\otimes\psi^* - \langle\varphi,\psi\rangle\cdot id\big) - t^2\tfrac{1}{2}\left|\psi\right|^2\cdot id\ . \end{aligned}$$

Therefore

$$d\mathfrak{sw}\big|_{(\varphi,A)}(\psi,\ i\vartheta) = \big(\mathcal{D}^A\psi + i\tfrac{1}{2}\vartheta\bullet\varphi,\ \ i\,d^+\vartheta - \psi\otimes\varphi^* - \varphi\otimes\psi^* - \langle\varphi,\psi\rangle_{\mathbb{R}}\cdot id\big).$$

---

[24] Proof: $\mathcal{D}^{A+i\vartheta}\psi = \sum e_k\bullet\nabla^{A+i\vartheta}_{e_k}\psi = \sum e_k\bullet\big(\nabla^A_{e_k}\psi + \frac{1}{2}i\vartheta(e_k)\,\psi\big) = \mathcal{D}^A\psi + \frac{1}{2}i\vartheta\bullet\psi$.

Differential of the gauge-action. Fix a random solution (φ, A) and consider the gauge action of $\mathscr{G}(\mathcal{L})$ on (φ, A) as a map

$$\begin{aligned} \mathfrak{g} \colon \mathscr{G}(\mathcal{L}) &\longrightarrow \Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L}) \\ \mathfrak{g}(g) &= \left(g^{-1}\varphi, \quad A + 2g^{-1}dg\right) . \end{aligned}$$

Locally (or, if M is simply-connected, globally) $g = e^{if}$ and then

$$\mathfrak{g}(e^{if}) = \left(e^{-if}\varphi, \quad A + 2i\,df\right) .$$

The tangent space to $\mathscr{G}(\mathcal{L})$ at 1 is the space $\{if \colon M \to i\mathbb{R}\} = i\Gamma(\underline{\mathbb{R}})$. Consider the variation

$$t \longmapsto \mathfrak{g}(e^{itf}) .$$

We will compute the derivative of $\mathfrak{g}$ at 1,

$$d\mathfrak{g}|_1 \colon T_{\mathscr{G}(\mathcal{L})}|_1 \longrightarrow T_{\Gamma(\mathcal{W}^+)\times\mathcal{C}onn(\mathcal{L})}|_{(\varphi,A)} ,$$

as the derivative at $t = 0$ of this variation. We have

$$\mathfrak{g}(e^{itf}) = \left(e^{-itf}\varphi, \quad A + 2it\,df\right) ,$$

and thus

$$d\mathfrak{g}|_1(if) = \tfrac{d}{dt}\Big|_{t=0}\mathfrak{g}(e^{itf}) = \left(-if\varphi, \quad 2i\,df\right) .$$

The setup. Combining the action of the gauge group and the evaluation of the Seiberg–Witten functional, consider the composition

$$T_{\mathscr{G}(\mathcal{L})}|_1 \xrightarrow{d\mathfrak{g}} T_{\Gamma(\mathcal{W}^+)\times\mathcal{C}onn(\mathcal{L})}|_{(\varphi,A)} \xrightarrow{d\mathfrak{sw}} T_{\Gamma(\mathcal{W}^-)\times i\Gamma(\Lambda^2_+)}|_{(0,0)} \cdot$$

Since the Seiberg–Witten equations are gauge-invariant, we expect this composition to be trivial, and indeed we can check it:

$$\begin{aligned} &d\mathfrak{sw}\big(d\mathfrak{g}(if)\big) = \\ &\quad = d\mathfrak{sw}\big(-if\varphi, \;\; 2i\,df\big) \\ &\quad = \big(\mathcal{D}^A(-if\varphi) + i\,df \bullet \varphi, \\ &\qquad\qquad i\,d^+df - (-if\varphi)\otimes\varphi^* - \varphi\otimes(-if\varphi)^* - \langle\varphi, -if\varphi\rangle_{\mathbb{R}}\cdot id\big) \\ &\quad = \big(-i\,df\bullet\varphi - if\mathcal{D}^A\varphi + i\,df\bullet\varphi, \\ &\qquad\qquad i\,(ddf)^+ - (-if\varphi)\otimes\langle\,\cdot\,,\varphi\rangle_{\mathbb{C}} - \varphi\otimes\langle\,\cdot\,,-if\varphi\rangle_{\mathbb{C}} - \langle\varphi, -if\varphi\rangle_{\mathbb{R}}\cdot id \\ &\quad = (0,0) , \end{aligned}$$

where we used that[25] $\mathcal{D}^A(f\varphi) = df\bullet\varphi + f\,\mathcal{D}^A\varphi$, and that $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathbb{C}}$ is complex-linear in the first, but complex-anti-linear in the second argument.

---

[25] Proof: $\mathcal{D}^A(f\varphi) = \sum e_k \bullet \nabla^A_{e_k}(f\varphi) = \sum e_k \bullet \big(df(e_k)\,\varphi\big) + e_k \bullet \big(f\,\nabla^A_{e_k}\varphi\big) = df\bullet\varphi + f\,\mathcal{D}^A\varphi$.

The tangent space. If (φ, A) were a regular point of $\mathfrak{sw}$, that is, if $d\mathfrak{sw}|_{(\varphi,A)}$ were surjective, then $\operatorname{Ker} d\mathfrak{sw}|_{(\varphi,A)}$ would represent the tangent space to the space of solutions $\mathfrak{S} = \mathfrak{sw}^{-1}\big[(0,0)\big]$ at (φ, A).

On the other hand, if (φ, A) were an irreducible solution, then $\mathfrak{g}$ would be an embedding of $\mathscr{G}(\mathcal{L})$ onto the orbit of (φ, A), and hence $\operatorname{Im} d\mathfrak{g}|_1$ would represent the tangent space to this orbit at (φ, A).

Therefore, in the orbit space $\mathfrak{M}$, the tangent space to $\mathfrak{M}$ at $[\varphi, A]$ would be exactly

$$T_{\mathfrak{M}}|_{[\varphi,A]} = T_{\mathfrak{S}}|_{(\varphi,A)} \,/\, T_{\mathscr{G}(\mathcal{L})\cdot(\varphi,A)} = \operatorname{Ker} d\mathfrak{sw}|_{(\varphi,A)} \,/\, \operatorname{Im} d\mathfrak{g}|_1 \,.$$

To better catch this, we formalize the setting as follows:

The complex. For every Seiberg–Witten solution (φ, A), we have checked that we have $d\mathfrak{sw} \circ d\mathfrak{g} = 0$. Hence we can write

$$0 \longrightarrow T_{\mathscr{G}(\mathcal{L})}|_1 \xrightarrow{\;d\mathfrak{g}\;} T_{\Gamma(\mathcal{W}^+)\times Conn(\mathcal{L})}|_{(\varphi,A)} \xrightarrow{\;d\mathfrak{sw}\;} T_{\Gamma(\mathcal{W}^-)\times i\Gamma(\Lambda^2_+)}|_{(0,0)} \longrightarrow 0$$

and consider it as a differential complex. Its zeroth cohomology group,

$$\mathcal{H}^0_{(\varphi,A)} = \operatorname{Ker} d\mathfrak{g} \,,$$

is trivial if and only if (φ, A) is an irreducible solution. The first cohomology,

$$\mathcal{H}^1_{(\varphi,A)} = \operatorname{Ker} d\mathfrak{sw} \,/\, \operatorname{Im} d\mathfrak{g} \,,$$

is called the **Zariski tangent space** of $\mathfrak{M}$, and, as suggested above, it has a good chance of actually being the tangent space of $\mathfrak{M}$ at $[\varphi, A]$. Finally, the second cohomology group,

$$\mathcal{H}^2_{(\varphi,A)} = \operatorname{Coker} d\mathfrak{sw} \,,$$

measures the failure of (φ, A) from being a regular point of $\mathfrak{sw}$; it is called the **obstruction space** at (φ, A).

Smoothness. We say that $\mathfrak{M}$ is **smooth** at $[\varphi, A]$ if both $\mathcal{H}^0_{(\varphi,A)}$ and $\mathcal{H}^2_{(\varphi,A)}$ vanish.

Smoothness Lemma. *If for all Seiberg–Witten solutions (φ, A) we have*

$$\mathcal{H}^0_{(\varphi,A)} = 0 \qquad \text{and} \qquad \mathcal{H}^2_{(\varphi,A)} = 0 \,,$$

then the space $\mathfrak{M}$ is either empty or is a smooth compact submanifold[26] *of $\mathfrak{B}_*$. In this case, its tangent space at $[\varphi, A]$ is precisely*

$$T_{\mathfrak{M}}|_{[\varphi,A]} = \mathcal{H}^1_{(\varphi,A)} \,.$$

Proof, or review. The vanishing of all $\mathcal{H}^2$'s means that $(0,0)$ is a regular value of the smooth map $\mathfrak{sw}$. Therefore the solution space $\mathfrak{S} = \mathfrak{sw}^{-1}\big[(0,0)\big]$ is either empty or is a smooth (infinite-dimensional) submanifold of $\Gamma(\mathcal{W}^+) \times Conn(\mathcal{L})$.

The vanishing of all $\mathcal{H}^0$'s means that there are no reducible solutions. In this case, the gauge group $\mathscr{G}(\mathcal{L})$ acts freely on $\mathfrak{S}$, and thus its quotient $\mathfrak{M}$ will be a smooth submanifold of the configuration space $\mathfrak{B}_*$.

The argument for the compactness of $\mathfrak{M}$ was already outlined in the main text, back on page 400. □

---

[26] Remember that $\mathfrak{B}_* = \big(\Gamma(\mathcal{W}^+) \setminus 0\big) \times Conn(\mathcal{L}) \,/\, \mathscr{G}(\mathcal{L})$.

Reducible monopoles. To ensure that all the $\mathcal{H}^0$'s are trivial means to avoid all reducible Seiberg–Witten solutions, *i.e.*, solutions with $\varphi \equiv 0$. When $\varphi = 0$, the Seiberg–Witten equations read simply $F_A^+ = 0$, and hence avoiding reducible solutions means avoiding anti-self-dual connections on $\mathcal{L}$. In the end-notes of the preceding chapter (page 357), we argued that, by perturbing the Riemannian metric, this is always possible if $b_2^+(M) \geq 1$. Further, if $b_2^+(M) \geq 2$, then any two such good metrics can be linked by a path of good metrics.

Hence, if we have $b_2^+(M) \geq 2$ and we manage to ensure that all $\mathcal{H}^2$'s also vanish, then the moduli spaces $\mathfrak{M}$ obtained for various good metrics can always be linked by a cobordism of *smooth* moduli spaces.

At the other extreme, if $b_2^+(M) = 0$, then reducibles are unavoidable, and $\mathfrak{M}$ will always have singularities. If M is simply-connected, there is exactly one such singularity,[27] and a neighborhood of it looks like a cone on $\mathbb{CP}^m$, with $2m = \dim \mathfrak{M} - 1$.

We will argue later that the obstruction spaces $\mathcal{H}^2$ can *always* be made to disappear. Accepting this last claim, it follows that, no matter what $b_2^+(M)$ is, it can always be arranged that $\mathfrak{M}$ be either empty or a smooth manifold with singularities at the reducible monopoles.

Before discussing the obstruction spaces $\mathcal{H}^2$, let us first compute the dimension of the Zariski tangent spaces (the virtual dimension of $\mathfrak{M}$), and see how we can orient these (and thus orient the smooth manifold $\mathfrak{M}$).

***Computing the virtual dimension of* $\mathfrak{M}$.** Assume that (φ, A) is an irreducible solution, and hence that $\mathcal{H}^0_{(\varphi,A)} = 0$. We compute in two ways the Euler characteristic $\chi_{(\varphi,A)}$ of the complex

$$0 \longrightarrow i\Gamma(\underline{\mathbb{R}}) \xrightarrow{d\mathfrak{g}} \Gamma(\mathcal{W}^+) \times i\Gamma(\Lambda^1) \xrightarrow{d\mathfrak{sw}} \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+) \longrightarrow 0 .$$

On one hand, $\chi_{(\varphi,A)}$ is the difference of dimensions

$$\chi_{(\varphi,A)} = -\dim \mathcal{H}^1_{(\varphi,A)} + \dim \mathcal{H}^2_{(\varphi,A)} .$$

On the other hand, $\chi_{(\varphi,A)}$ can be computed from the indices of the differential operators involved.

Since the index is strongly invariant and depends only on the symbol (*i.e.*, on the highest-derivative terms), we can deform the above complex to get rid of all zeroth order terms, and end up with the complex

$$0 \longrightarrow i\Gamma(\underline{\mathbb{R}}) \xrightarrow{(0,2d)} \Gamma(\mathcal{W}^+) \times i\Gamma(\Lambda^1) \xrightarrow{\mathcal{D}^A \oplus d^+} \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+) \longrightarrow 0 .$$

The Euler characteristic of this complex is still $\chi_{(\varphi,A)}$, but this complex can now be viewed as the sum of two complexes,

$$\begin{tikzcd}
0 \arrow[r] & 0 \arrow[r] & \Gamma(\mathcal{W}^+) \arrow[r, "\mathcal{D}^A"] & \Gamma(\mathcal{W}^-) \arrow[r] & 0 \\
0 \arrow[r] & i\Gamma(\underline{\mathbb{R}}) \arrow[r, "2d"] & \Gamma(\Lambda^1) \arrow[r, "d^+"] & \Gamma(\Lambda^2_+) \arrow[r] & 0 .
\end{tikzcd}$$

---

[27] In general, the singular set of $\mathfrak{M}$ can be identified with the torus $H^1(M;\mathbb{R}) \,/\, H^1(M;\mathbb{Z})$.

Both these complexes are elliptic and can be analyzed using standard Atiyah–Singer technology.[28] The Euler characteristic of the first complex is just the opposite of the *real*[29] index of the Dirac operator, and hence

$$\chi_1 = \tfrac{1}{4}\left(\operatorname{sign} M - c_1(\mathfrak{s}) \cdot c_1(\mathfrak{s})\right) .$$

The Euler characteristic of the second complex is easily computed to be

$$\chi_2 = 1 - b_1(M) + b_2^+(M) .$$

Indeed, inside the end-notes of the preceding chapter (connections on line bundles, page 357), we showed on page 360 that $\operatorname{Ker} d = \operatorname{Ker} d^+$, and on page 361 that $\Gamma(\Lambda^2_+) \approx H^2_+(M;\mathbb{R}) \oplus \operatorname{Im} d^+$; it follows that the cohomology groups of the second complex are $H^0(M;\mathbb{R})$, $H^1(M;\mathbb{R})$ and $H^2_+(M;\mathbb{R})$.

Since Euler characteristics are additive, we have $\chi_{(\varphi,A)} = \chi_1 + \chi_2$, and we conclude that

$$\begin{aligned} -\dim \mathcal{H}^1_{(\varphi,A)} + \dim \mathcal{H}^2_{(\varphi,A)} &= \\ &= 1 - b_1(M) + b_2^+ + \tfrac{1}{4}\left(\operatorname{sign} M - c_1(\mathfrak{s})^2\right) \\ &= \tfrac{1}{4}\left(4 - 4b_1(M) + 4b_2^+(M) + b_2^+(M) - b_2^-(M) - c_1(\mathfrak{s})^2\right) \\ &= \tfrac{1}{4}\left(3 \operatorname{sign} M + 2\chi(M) - c_1(\mathfrak{s})^2\right) , \end{aligned}$$

since $\operatorname{sign} M = b_2^+ - b_2^-$ and $\chi(M) = 2 - 2b_1 + b_2^+ + b_2^-$. Therefore:

Lemma. *If (φ, A) is an irreducible solution and its obstruction space $\mathcal{H}^2_{(\varphi,A)}$ vanishes, then the dimension of the Zariski tangent space of $\mathfrak{M}$ at $[\varphi, A]$ is*

$$\operatorname{vdim} \mathfrak{M} = \tfrac{1}{4}\left(c_1(\mathfrak{s})^2 - 3 \operatorname{sign} M - 2\chi(M)\right) . \qquad \square$$

The quantity $\operatorname{vdim} \mathfrak{M}$ is the expected dimension of $\mathfrak{M}$ and is called the **virtual dimension** of $\mathfrak{M}$.

If $\mathcal{H}^2$ vanishes for all solutions (φ, A), this means that $(0,0)$ is a regular value of $\mathfrak{sw}$, and therefore its preimage $\mathfrak{S} = \mathfrak{sw}^{-1}\left[(0,0)\right]$ is a smooth (infinite-dimensional) manifold in $\Gamma(\mathcal{W}^+) \times \mathit{Conn}(\mathcal{L})$. Nonetheless, $(0,0)$ often is a regular value of $\mathfrak{sw}$ without being being an actual value[30] of $\mathfrak{sw}$. Hence the solution space $\mathfrak{S}$ could be empty. Certainly, if the virtual dimension of $\mathfrak{M}$ is negative and we somehow ensure the vanishing of all $\mathcal{H}^2$'s, then $\mathfrak{M}$ will necessarily be empty (or contain *only* reducible monopoles). In conclusion:

Dimension Lemma. *If the Riemannian metric on M is such that there are no reducible solutions, and if the obstruction spaces vanish for all solutions, then the moduli space $\mathfrak{M}$ is either empty or a smooth compact manifold of dimension*

$$\dim \mathfrak{M} = \tfrac{1}{4}\left(c_1(\mathfrak{s})^2 - 3 \operatorname{sign} M - 2\chi(M)\right) . \qquad \square$$

---

[28] See for example, **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** [**LM89**].

[29] The *complex* index of $\mathcal{D}^A$ is the difference of complex dimensions of the kernel and cokernel of $\mathcal{D}^A$. The *real* index is the difference of their real dimensions, and thus is twice the complex index.

[30] As M. Hirsch put it: *This is in accordance with the principle that in mathematics a "red herring" does not have to be either red or a herring.* [**Hir76, Hir94**, p 22].

Orientations for $\mathfrak{M}$. Assume that all is good, namely that at every Seiberg–Witten solution (φ, A) the spaces $\mathcal{H}^0$ and $\mathcal{H}^2$ both vanish, and thus that

$$T_{\mathfrak{M}}\big|_{[\varphi,A]} = \operatorname{Ker} d_{\mathfrak{sw}} \,/\, \operatorname{Im} d_{\mathfrak{g}} .$$

This tangent space appears as the first cohomology $\mathcal{H}^1$ of the complex

$$0 \longrightarrow i\Gamma(\underline{\mathbb{R}}) \xrightarrow{d_{\mathfrak{g}}} \Gamma(\mathcal{W}^+) \times i\Gamma(\Lambda^1) \xrightarrow{d_{\mathfrak{sw}}} \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+) \longrightarrow 0 . \qquad (\diamond)$$

and it is the only non-vanishing cohomology group of this complex.

As before, this complex can be deformed by eliminating the zeroth order terms into the simpler complex

$$0 \longrightarrow i\Gamma(\underline{\mathbb{R}}) \xrightarrow{(0,2d)} \Gamma(\mathcal{W}^+) \times i\Gamma(\Lambda^1) \xrightarrow{\mathcal{D}^A \oplus d^+} \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+) \longrightarrow 0 . \quad (\heartsuit)$$

The cohomology groups of this deformed complex are

$$\begin{aligned} \mathcal{H}^0_\heartsuit &= H^0(M;\mathbb{R}) \\ \mathcal{H}^1_\heartsuit &= \operatorname{Ker} \mathcal{D}^A \oplus H^1(M;\mathbb{R}) \\ \mathcal{H}^2_\heartsuit &= \operatorname{Coker} \mathcal{D}^A \oplus H^2_+(M;\mathbb{R}) , \end{aligned}$$

where, again, we use that $d^+\alpha = 0$ if and only if $d\alpha = 0$ (see page 360), and that $\Gamma(\Lambda^2_+)/\operatorname{Im} d^+ = H^2_+(M;\mathbb{R})$ (see page 361).

Orient them. The line $H^0(M;\mathbb{R})$ has a canonical orientation, given by the class of the constant function $x \mapsto 1$. Since $\mathcal{D}^A$ is elliptic, both $\operatorname{Ker}\mathcal{D}^A$ and $\operatorname{Coker}\mathcal{D}^A$ are finite-dimensional vector spaces. Further, since $\mathcal{D}^A$ is $\mathbb{C}$-linear, both $\operatorname{Ker}\mathcal{D}^A$ and $\operatorname{Coker}\mathcal{D}^A$ are complex spaces, and thus have natural orientations of their own.

Thus, to endow each of $\mathcal{H}^0_\heartsuit$, $\mathcal{H}^1_\heartsuit$ and $\mathcal{H}^2_\heartsuit$ with an orientation, we need only pick orientations of $H^1(M;\mathbb{R})$ and of $H^2_+(M;\mathbb{R})$.

Assume we choose some orientations of $H^1(M;\mathbb{R})$ and $H^2_+(M;\mathbb{R})$. Then in effect we have oriented the three cohomology spaces $\mathcal{H}^0_\heartsuit$, $\mathcal{H}^1_\heartsuit$, and $\mathcal{H}^2_\heartsuit$, and in particular oriented their determinant lines $\Lambda^{top}\mathcal{H}^0_\heartsuit$, $\Lambda^{top}\mathcal{H}^1_\heartsuit$, and $\Lambda^{top}\mathcal{H}^2_\heartsuit$. This becomes an orientation of the line

$$(\Lambda^{top}\mathcal{H}^0_\heartsuit)^* \otimes_{\mathbb{R}} (\Lambda^{top}\mathcal{H}^1_\heartsuit) \otimes_{\mathbb{R}} (\Lambda^{top}\mathcal{H}^2_\heartsuit)^* .$$

Transport them. The fundamental point is that, by standard Fredholm technology, the above real line can be identified across the deformation of our initial Seiberg–Witten complex $(\diamond)$ into its zeroth-order-free version $(\heartsuit)$, and hence its orientation can be transported back to $(\diamond)$.

Fredholm's line. *A linear operator $F\colon V \longrightarrow W$ is called* **Fredholm** *if it is bounded and has finite-dimensional kernel and cokernel; all elliptic operators are examples of Fredholm operators. Given a Fredholm operator, we define its* **determinant line** $\det F$ *as the line*

$$\det F = (\Lambda^{top} \operatorname{Ker} F) \otimes_{\mathbb{R}} (\Lambda^{top} \operatorname{Coker} F)^* .$$

This determinant line can be identified across continuous deformations (or families). Indeed, let $F_t\colon V \longrightarrow W$ be a homotopy through Fredholm operators. Then $\det F_t \mapsto t$ determines a real-line bundle over $[0,1]$, which can only be trivial and thus offers an identification of $\det F_0$ with $\det F_1$.

This is remarkable because across such homotopies the dimensions of both $\operatorname{Ker} F_t$ *and* $\operatorname{Coker} F_t$ *can change,*[31] *but nonetheless the determinant line stays the same.*

The situation of interest to us is when we are dealing with the line $L = (\Lambda^{top} H^0)^* \otimes (\Lambda^{top} H^1) \otimes (\Lambda^{top} H^2)^*$*, appearing from a complex*

$$0 \longrightarrow A \xrightarrow{a} B \xrightarrow{b} C \longrightarrow 0$$

of Fredholm operators. We have $H^0 = \operatorname{Ker} a = \operatorname{Coker} a^*$ *(where* $a^*\colon B \to A$ *is the adjoint of* a*), and* $H^1 = \operatorname{Ker} b / \operatorname{Im} a = \operatorname{Ker} b \cap \operatorname{Coker} a = \operatorname{Ker} b \cap \operatorname{Ker} a^*$*, as well as* $H^2 = \operatorname{Coker} b$*. Therefore*

$$(\Lambda^{top} H^0)^* \otimes (\Lambda^{top} H^1) \otimes (\Lambda^{top} H^2)^* = \det(a^* \oplus b) \ ,$$

for $a^* \oplus b\colon B \to A \oplus C$*. Since deformations of complexes translate into deformations of the corresponding Fredholm operators* $a^* \oplus b$*, the transport of the line* L *across homotopies follows.*

Using this identification of the line $\left(\Lambda^{top}\mathcal{H}^0_\heartsuit\right)^* \otimes \left(\Lambda^{top}\mathcal{H}^1_\heartsuit\right) \otimes \left(\Lambda^{top}\mathcal{H}^2_\heartsuit\right)^*$ with the line $\left(\Lambda^{top}\mathcal{H}^0_{(\varphi,A)}\right)^* \otimes \left(\Lambda^{top}\mathcal{H}^1_{(\varphi,A)}\right) \otimes \left(\Lambda^{top}\mathcal{H}^2_{(\varphi,A)}\right)^*$ across the homotopy of $(\diamondsuit)$ into $(\heartsuit)$, together with the assumption that the only non-zero cohomology of $(\diamondsuit)$ is the Zariski tangent space $\mathcal{H}^1_{(\varphi,A)}$, we end up with an orientation of

$$\Lambda^{top}\, \mathcal{H}^1_{(\varphi,A)} \ .$$

This is in effect an orientation of the Zariski tangent space $\mathcal{H}^1_{(\varphi,A)} = T_{\mathfrak{M}}|_{[\varphi,A]}$ itself, and thus of the moduli space $\mathfrak{M}$ at $[\varphi, A]$. We still need to argue that repeating this for all solutions (φ, A) yields an orientation of the whole moduli space $\mathfrak{M}$:

Compatibility. The orientations induced through the above procedure on the various tangent spaces $T_{\mathfrak{M}}|_{[\varphi,A]} = \mathcal{H}^1_{(\varphi,A)}$ fit into a global orientation of $\mathfrak{M}$.

Consider the real-line bundle $\widetilde{L}_\heartsuit$ over $\Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L})$, given by:

$$\widetilde{L}_\heartsuit\colon \ \left(\Lambda^{top}\mathcal{H}^0_\heartsuit\right)^* \otimes \left(\Lambda^{top}\mathcal{H}^1_\heartsuit\right) \otimes \left(\Lambda^{top}\mathcal{H}^2_\heartsuit\right)^* \longmapsto (\varphi, A) \ .$$

The bundle $\widetilde{L}_\heartsuit$ has its fibers determined by constant operators (d and d^+) and complex vector spaces ($\operatorname{Ker} \mathcal{D}^A$ and $\operatorname{Coker} \mathcal{D}^A$), and hence is trivial.[32] Since $\widetilde{L}_\heartsuit$ is $\mathcal{G}(\mathcal{L})$–equivariant, its triviality descends to the triviality of its quotient bundle $L_\heartsuit$ over $\mathfrak{B} = \Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L}) \ / \ \mathcal{G}(\mathcal{L})$ given by

$$L_\heartsuit\colon \ \left(\Lambda^{top}\mathcal{H}^0_\heartsuit\right)^* \otimes \left(\Lambda^{top}\mathcal{H}^1_\heartsuit\right) \otimes \left(\Lambda^{top}\mathcal{H}^2_\heartsuit\right)^* \longmapsto [\varphi, A] \ .$$

This $L_\heartsuit \to \mathfrak{B}$ has fibers that can be identified back across the deformation of $(\diamondsuit)$ into $(\heartsuit)$ with the fibers of the line bundle $L_\diamondsuit \to \mathfrak{B}$ given by

$$L_\diamondsuit\colon \ \left(\Lambda^{top}\mathcal{H}^0_{(\varphi,A)}\right)^* \otimes \left(\Lambda^{top}\mathcal{H}^1_{(\varphi,A)}\right) \otimes \left(\Lambda^{top}\mathcal{H}^2_{(\varphi,A)}\right)^* \longmapsto [\varphi, A] \ .$$

It follows that the latter line bundle is a global trivial line-bundle on $\mathfrak{B}$. Since its fibers over $\mathfrak{M}$ coincide with the tangent spaces of the moduli space, it follows that $\mathfrak{M}$ can be coherently oriented. We have proved:

Orientation Lemma. *If the Riemannian metric on* M *is such that there are no reducible solutions, and if the obstruction spaces vanish for all solutions, then the moduli space*

---

[31] Observe that nonetheless the dimensional jumps are balanced: the index $\dim \operatorname{Ker} F_t - \dim \operatorname{Coker} F_t$ must stay constant.

[32] It is a real-line bundle, so it is either trivial or non-orientable.

$\mathfrak{M}$ is either empty or an orientable smooth manifold. An orientation of $\mathfrak{M}$ is uniquely determined by a choice of orientations of the vector spaces

$$H^1(M;\mathbb{R}) \qquad \text{and} \qquad H^2_+(M;\mathbb{R})\,. \qquad \square$$

Notice that in fact the orientation argument above works just as well if there happen to be reducible solutions in $\mathfrak{M}$, and yields an orientation of the irreducible part $\mathfrak{M}_*$ of $\mathfrak{M}$.

Transversality. Finally, we need to argue that all $\mathcal{H}^2$'s can be made to vanish. Recall that $\mathcal{H}^2$ at (φ, A) measures whether the Seiberg–Witten map $\mathfrak{sw}\colon \Gamma(\mathcal{W}^+) \times \mathcal{C}onn(\mathcal{L}) \to \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+)$ has (φ, A) as a regular point. In other words, it detects whether its differential

$$d\mathfrak{sw}|_{(\varphi,A)}\colon \Gamma(\mathcal{W}^+) \times i\Gamma(\Lambda^1) \longrightarrow \Gamma(\mathcal{W}^-) \times i\Gamma(\Lambda^2_+)$$
$$d\mathfrak{sw}|_{(\varphi,A)}(\psi,\, i\vartheta) \;=\; \big(\mathcal{D}^A\psi + \tfrac{i}{2}\vartheta \bullet \varphi,\;\; i\,d^+\vartheta - \psi\otimes\varphi^* - \varphi\otimes\psi^* - \langle\varphi,\psi\rangle\cdot id\big)$$

is surjective.

First component. The differential $d\mathfrak{sw}$ is always onto the $\Gamma(\mathcal{W}^-)$–factor:

Lemma. *If (φ, A) is an irreducible Seiberg–Witten solution, then the map*

$$\Gamma(\mathcal{W}^+) \times i\Gamma(\Lambda^1) \longrightarrow \Gamma(\mathcal{W}^-)\colon \qquad (\psi,\, i\vartheta) \longmapsto \mathcal{D}^A\psi + \tfrac{i}{2}\vartheta \bullet \varphi$$

is always surjective.

Proof. Pick a section $\xi \in \Gamma(\mathcal{W}^-)$ that is L^2–orthogonal to the image of our map above. In other words, assume that for every $(\psi,\, i\vartheta)$ we have

$$\int_M \langle \mathcal{D}^A\psi + \tfrac{i}{2}\vartheta \bullet \varphi,\;\; \xi\rangle_{\mathbb{R}} = 0\,.$$

In particular $\int_M \langle \mathcal{D}^A\psi,\, \xi\rangle = 0$ for all ψ. This means that $\xi \perp \operatorname{Im} \mathcal{D}^A$. Then we must have $\xi \in \operatorname{Coker} \mathcal{D}^A$, that is, $\xi \in \operatorname{Ker}(\mathcal{D}^A)^*$, in other words,

$$(\mathcal{D}^A)^*\,\xi = 0$$

for the adjoint Dirac operator $(\mathcal{D}^A)^*\colon \Gamma(\mathcal{W}^-) \to \Gamma(\mathcal{W}^+)$. However, this adjoint is an elliptic operator, and therefore it satisfies the unique continuation property (stated in section 10.2, on page 393). Hence, either ξ is constantly-zero, or there is no open set on which ξ vanishes identically.

Assume that $\xi \in \Gamma(\mathcal{W}^-)$ is not trivial. Then there are no open sets on which ξ is identically-zero. On the other hand, $\varphi \in \Gamma(\mathcal{W}^+)$ itself has $\mathcal{D}^A\varphi = 0$. Therefore, since (φ, A) was assumed not reducible, φ cannot be trivial, and hence there are no open sets on which φ is identically-zero either. It follows that there must be some point $x \in M$ where neither φ nor ξ vanish. Since both φ and ξ are continuous, choose a small round neighborhood U of x so that both φ and ξ are nowhere-zero on U.

Clifford multiplication exhibits an isomorphism $T_M \otimes \mathbb{C} \approx \operatorname{Hom}_{\mathbb{C}}(\mathcal{W}^+, \mathcal{W}^-)$ (indeed, Clifford multiplication is modeled on quaternion-multiplication, see the previous note on page 432). In particular, there must be some local vector field $\widetilde{\vartheta}_0 \in \Gamma(T_M|_U)$ so that

$$i\widetilde{\vartheta}_0 \bullet \varphi = \xi$$

(the role of the "i" is purely aesthetic, for later use). Multiply $\widetilde{\vartheta}_0$ with your favorite bump-function on U to obtain a vector field ϑ_0 that approaches 0 as you get close to ∂U. We still have $\langle i\vartheta_0 \bullet \varphi,\ \xi\rangle_{\mathbb{R}} > 0$ over all U. The vector field ϑ_0 can be extended by 0 over the whole manifold M and has the property that

$$\int_M \langle i\vartheta_0 \bullet \varphi,\ \xi\rangle_{\mathbb{R}} > 0\,.$$

Since $\int_M \langle \mathcal{D}^A \psi,\ \xi\rangle$ was assumed to be zero, it follows that

$$\int_M \langle \mathcal{D}^A \psi + \tfrac{i}{2}\vartheta_0 \bullet \varphi,\ \xi\rangle_{\mathbb{R}} > 0\,,$$

but then ξ cannot be orthogonal to the image of $(\psi, i\vartheta) \mapsto \mathcal{D}^A\psi + \frac{i}{2}\vartheta \bullet \varphi$.

Therefore, the only possibility left for a section $\xi \in \Gamma(\mathcal{W}^-)$ to be L^2-orthogonal to the image of our map is for ξ to be constantly-zero. This effectively proves that our map is onto.[33] □

Second component: varying the metric. Therefore, to ensure that $(0,0)$ is a regular value of $\mathfrak{sw}$, the problem left is to make the differential $d\mathfrak{sw}$ surjective onto the $\Gamma(\Lambda^2_+)$-factor.

There are two approaches to this: one is aesthetically-pleasing but technically harder to prove, while the other is more elementary. The first choice is to use a perturbation of the Riemannian metric:

Generic Metric Theorem. *For a generic set of Riemannian metrics on M, the obstruction spaces are all trivial, that is, $\mathcal{H}^2_{(\varphi,A)} = 0$ for all irreducible solutions (φ, A).* □

A proof of this theorem can be read in **T. Friedrich**'s ***Dirac operators in Riemannian geometry*** **[Fri00**, app A**]**. You might need some background on the space of Riemannian metrics on a manifold, see **D. Ebin**'s ***On the space of Riemannian metrics*** **[Ebi68]**.

Second component: perturbing the equations. A quite simpler approach is to perturb the equations, then deal with solutions of the perturbed equations, showing that the corresponding obstruction spaces vanish and that the resulting Seiberg–Witten invariants do not depend on this perturbation.

The perturbed Seiberg–Witten equations are

$$\begin{cases} \mathcal{D}^A\varphi = 0 \\ F_A^+ = \sigma(\varphi) + i\eta^+\,. \end{cases}$$

for some self-dual 2-form $\eta^+ \in \Gamma(\Lambda^2_+(T^*_M))$.

---

[33] Here, of course, rigorously one must use that $\Gamma(\mathcal{W}^-)$ is complete for the L^2_k-inner product used, that the image of our map is closed in this Hilbert space and thus it must admit orthogonal complements, which hence are empty only when the map is surjective, *etc.*

Dealing with all perturbations at once, we define the map

$$\begin{aligned}
&\mathcal{P}\mathfrak{sw}\colon\ \Gamma(\mathcal{W}^+)\times \mathit{Conn}(\mathcal{L})\times\Gamma(\Lambda^2_+)\ \longrightarrow\ \Gamma(\mathcal{W}^-)\times i\Gamma(\Lambda^2_+)\\
&\mathcal{P}\mathfrak{sw}(\varphi,\,A,\,\eta)=\big(\mathcal{D}^A\varphi,\ \ F^+_A-\sigma(\varphi)-i\eta^+\big)\,.
\end{aligned}$$

Its differential at an irreducible solution (φ, A, η^+), obtained as before by using the derivative at $t=0$ of a variation $t\longmapsto \mathcal{P}\mathfrak{sw}\big(\varphi+t\psi,\ \ A+it\vartheta,\ \ \eta^+ + t\alpha^+\big)$, is computed to be

$$\begin{aligned}
&d\,\mathcal{P}\mathfrak{sw}|_{(\varphi,A,\eta^+)}\colon\ \Gamma(\mathcal{W}^+)\times i\Gamma(\Lambda^1)\times\Gamma(\Lambda^2_+)\ \longrightarrow\ \Gamma(\mathcal{W}^-)\times i\Gamma(\Lambda^2_+)\\
&d\,\mathcal{P}\mathfrak{sw}|_{(\varphi,A,\eta^+)}(\psi,\,i\vartheta,\,\alpha^+)=\\
&\qquad=\big(\mathcal{D}^A\psi+\tfrac{i}{2}\vartheta\bullet\varphi,\ \ i\,d^+\vartheta-\psi\otimes\varphi^*-\varphi\otimes\psi^*-\langle\varphi,\psi\rangle\cdot id-i\alpha^+\big)\,.
\end{aligned}$$

This differential is still onto the $\Gamma(\mathcal{W}^-)$–factor, just as before, but moreover, owing to the $i\alpha^+$–term in the second component above, it is also obviously surjective onto the $\Gamma(\Lambda^2_+)$–factor. Therefore we have:

Lemma. *Assuming there are no reducible monopoles, the point* $(0,0)$ *is a regular value of the map* $\mathcal{P}\mathfrak{sw}$. □

The consequence is that the collection of all solutions (φ, A, η^+) of the perturbed Seiberg–Witten equations always make up an (infinite-dimensional) submanifold $\mathcal{P}\mathfrak{S}$ of $\Gamma(\mathcal{W}^+)\times \mathit{Conn}(L)\times\Gamma(\Lambda^2_+)$.

If we avoid reducible solutions, then the quotient of $\mathcal{P}\mathfrak{S}$ by the gauge-action of $\mathscr{G}(\mathcal{L})$ on $\Gamma(\mathcal{W}^+)\times \mathit{Conn}(L)\times\Gamma(\Lambda^2_+)$ (acting by the identity on the η^+'s) will be a smooth (infinite-dimensional) manifold $\mathcal{P}\mathfrak{M}$, *i.e.*, the moduli space of all perturbed Seiberg–Witten monopoles.

The parametrized moduli space $\mathcal{P}\mathfrak{M}$ has an obvious projection

$$\mathfrak{p}\colon \mathcal{P}\mathfrak{M}\longrightarrow\Gamma(\Lambda^2_+)\,,$$

coming from the projection of $\Gamma(\mathcal{W}^+)\times \mathit{Conn}(L)\times\Gamma(\Lambda^2_+)$ onto its third factor. Denote the fiber of $\mathfrak{p}$ over any $\eta^+\in\Gamma(\Lambda^2_+)$ by

$$\mathfrak{M}_{\eta^+}=\mathfrak{p}^{-1}[\eta^+]\,.$$

It is the moduli space of solutions to the perturbed Seiberg–Witten equations for a fixed perturbation η^+.

By the Sard–Smale theorem, the smooth map $\mathfrak{p}$ must have plenty of regular values.[34] Therefore, combining with parametrized versions of previously presented arguments, we eventually obtain:

Generic Perturbation Theorem. *For a generic set of* η^+*'s from* $\Gamma(\Lambda^2_+)$, *the moduli space* $\mathfrak{M}_{\eta^+}$ *of the* η^+*–perturbed Seiberg–Witten equations is either empty or is a smooth manifold of dimension*

$$\dim\mathfrak{M}_{\eta^+}=\tfrac{1}{4}\big(c_1(\mathfrak{s})^2-3\operatorname{sign}M-2\chi(M)\big)\,.$$

---

[34] The **Sard–Smale theorem**, in a mild version, states: *Let* $\mathfrak{f}\colon\mathfrak{X}\to\mathfrak{Y}$ *be a smooth map between infinite-dimensional (Banach) manifolds and assume that* $d\mathfrak{f}|_x$ *is Fredholm at all* $x\in\mathfrak{X}$. *Then, for most* $y\in\mathfrak{Y}$, *the preimage* $\mathfrak{f}^{-1}[y]$ *is either empty or a smooth manifold of dimension* $\dim\operatorname{Ker}d\mathfrak{f}-\dim\operatorname{Coker}d\mathfrak{f}$. It was proved in **S. Smale**'s ***An infinite dimensional version of Sard's theorem*** [**Sma65**].

Further, $\mathfrak{M}_{\eta^+}$ *is compact and orientable. Its bordism class inside* $\Gamma(\mathcal{W}^+) \times Conn(\mathcal{L}) \times \Gamma(\Lambda^2_+) \;/\; \mathcal{G}(\mathcal{L})$ *does not depend on* η^+*.* □

One can then proceed and define the Seiberg–Witten invariant $\mathcal{SW}_M(\mathfrak{s})$ as determined by the homology class of any one of the nice $\mathfrak{M}_{\eta^+}$'s for $\mathfrak{s}$, and the whole theory develops analogously to the outline from the main text.

Nonetheless, we prefer to work under the light of the generic metric theorem rather than under that of the generic perturbation theorem. That is to say, we will continue to work with unperturbed equations and unperturbed moduli spaces. (Nonetheless, a perturbation where $|\eta^+| \to \infty$ will be essential when analyzing the Seiberg–Witten equations on symplectic manifold, in the note on page 465 ahead.)

When not of simple type. Of course, when $\mathfrak{M}$ is zero-dimensional, then all one needs to do in order to obtain an invariant is to count the points of $\mathfrak{M}$, with signs. This is expected to be the case for all simply-connected 4–manifolds with $b_2^+ \geq 2$.

Nonetheless, there are plenty of examples of manifolds with $b_1^+ \leq 1$ or non-simply-connected that each have moduli spaces of dimensions as high as one pleases. It is thus worth taking a look at the definition of $\mathcal{SW}_M(\mathfrak{s})$ in case the dimension of $\mathfrak{M}$ is positive.

Ambient cohomology. If $\dim \mathfrak{M} > 0$, then we can locate the homology class $[\mathfrak{M}]$ inside $H_*(\mathfrak{B};\mathbb{Z})$ by evaluating the cohomology classes of $\mathfrak{B}$ on it. For that, of course, we need to understand the cohomology of this ambient space

$$\mathfrak{B} = \Gamma(\mathcal{W}^+) \times Conn(\mathcal{L}) \;/\; \mathcal{G}(\mathcal{L}) \,.$$

The action of $\mathcal{G}(\mathcal{L})$ has as fixed points all pairs $(0, A)$, and thus these create singularities in the quotient $\mathfrak{B}$. Removing these singularities, we are left with the space

$$\mathfrak{B}_* = \big(\Gamma(\mathcal{W}^+) \setminus 0\big) \times Conn(\mathcal{L}) \;/\; \mathcal{G}(\mathcal{L})$$

of all gauge classes of pairs (φ, A) with φ not everywhere-zero. The action of $\mathcal{G}(\mathcal{L})$ is now free, and thus $\mathfrak{B}_*$ is in fact an infinite-dimensional smooth manifold. If $b_2^+(M) \geq 1$ and reducible solutions are avoided, then the moduli space $\mathfrak{M}$ is wholly included in $\mathfrak{B}_*$.

A different way of dealing with the non-free points of $\mathcal{G}(\mathcal{L})$ is to "fix the gauge". Namely, consider the subgroup

$$\mathcal{G}^0(\mathcal{L}) = \{g \colon M \to S^1 \mid g(p) = 1\} \,,$$

where p is your favorite random point of M. We clearly have $\mathcal{G}(\mathcal{L}) = \mathcal{G}^0(\mathcal{L}) \times S^1$, with S^1 realized as constant S^1–valued functions on M.

The action of $\mathcal{G}^0(\mathcal{L})$ on $\Gamma(\mathcal{W}^+) \times Conn(\mathcal{L})$ has the advantage of *always* being free. Thus, we can define its quotient

$$\mathfrak{B}^0 = \Gamma(\mathcal{W}^+) \times Conn(\mathcal{L}) \;/\; \mathcal{G}^0(\mathcal{L}) \,,$$

and correspondingly build $\mathfrak{B}^0_*$ and $\mathfrak{M}^0$, and all these will always be smooth manifolds.

Clearly $\mathfrak{B} = \mathfrak{B}^0 \,/\, S^1$, and similarly $\mathfrak{B}_* = \mathfrak{B}^0_* \,/\, S^1$ and $\mathfrak{M} = \mathfrak{M}^0 \,/\, S^1$. Moreover, the projection

$$\mathfrak{B}^0_* \longrightarrow \mathfrak{B}_*$$

is a fiber-bundle with fiber S^1. As such, it is associated to a complex-line bundle and a corresponding Chern class

$$\mathfrak{u} \in H^2(\mathfrak{B}_*; \mathbb{Z}) .$$

These, of course, restrict to a corresponding line bundle and 2–class on $\mathfrak{M}_*$.

The class $\mathfrak{u}$ is in fact fundamental for the space $\mathfrak{B}_*$:

Theorem. *The configuration space $\mathfrak{B}_*$ is the classifying space of the gauge group: $\mathfrak{B}_* = \mathscr{B}\mathscr{G}(\mathcal{L})$. Its cohomology ring is*

$$H^*(\mathfrak{B}_*; \mathbb{Z}) = \mathbb{Z}[\mathfrak{u}] \otimes_{\mathbb{Z}} \Lambda^*\big(H^1(M;\mathbb{Z})\big) ,$$

where $\mathfrak{u}$ is the first Chern class of the bundle $\mathfrak{B}^0_* \to \mathfrak{B}_*$. □

This is shown by arguing that $\big(\Gamma(\mathcal{W}^+) \setminus 0\big) \times \mathcal{C}onn(\mathcal{L})$ is contractible, and thus its quotient by the free action of $\mathscr{G}(\mathcal{L})$ must be the classifying space[35] of $\mathscr{G}(\mathcal{L})$. Then one uses a result of R. Thom to evaluate the cohomology. For details, one can start with the third edition of **D. Husemoller**'s ***Fibre bundles*** **[Hus94**, ch 7].

Definition of the invariants. When $b_2^+(M) \geq 1$ we can avoid reducible solutions and thus make the moduli space $\mathfrak{M}$ wholly included in $\mathfrak{B}_*$. Then we can evaluate the cohomology classes of the latter on $\mathfrak{M}$.

— If the space $\mathfrak{M}$ has even dimension, which happens if and only if $b_2^+(M) + b_1(M)$ is *odd*, then, after a choice of orientation on $\mathfrak{M}$, we can define

$$\mathcal{SW}_M(\mathfrak{s}) = \int_{\mathfrak{M}} \mathfrak{u}^{\dim \mathfrak{M}/2} .$$

If $b_2^+(M) \geq 2$, then the moduli spaces for various generic metrics can be linked by a bordism inside $\mathfrak{B}_*$, and thus the invariants do not change.

If $b_2^+(M) = 1$, then the bordism might encounter a reducible solution. In this case, after removing a neighborhood of the singular point, we obtain a bordism inside $\mathfrak{B}_*$ between $\mathfrak{M}_1$ and $\mathfrak{M}_2 \cup \pm\mathbb{CP}^m$, and therefore the invariants might change by ± 1. In certain cases (such as when $c_1(\mathfrak{s})^2 > 0$), reducible solutions do not appear.

— If the space $\mathfrak{M}$ has odd dimension, which happens if and only if $b_2^+(M) + b_1(M)$ is *even*,[36] then we must define the Seiberg–Witten invariants to be trivial:

$$\mathcal{SW}_M(\mathfrak{s}) = 0 .$$

— If the manifold M is not simply-connected, then we can refine the invariants by also evaluating on $\mathfrak{M}$ the cohomology classes of $\mathfrak{B}_*$ appearing from $H^1(M;\mathbb{R})$. In this case it is best to think of $\mathcal{SW}_M(\mathfrak{s})$ as a function

$$\mathcal{SW}_M(\mathfrak{s}) \colon \ \mathbb{Z}[\mathfrak{u}] \otimes_{\mathbb{Z}} \Lambda^*\big(H^1(M;\mathbb{Z})\big) \longrightarrow \mathbb{Z} .$$

---

[35] Compare with an inserted note back on page 205, inside the end-notes of chapter 4 (classifying spaces, page 204).

[36] Notice how, if M does not admit any almost-complex structure, then the Seiberg–Witten invariants are powerless.

Note: The Seiberg–Witten proof of Donaldson's theorem

Using our better understanding of the Seiberg–Witten moduli space $\mathfrak{M}$ gained from the preceding note, we will now offer a proof of

Donaldson's Theorem. *If M is a smooth 4–manifold with negative-definite intersection form, then in fact its intersection form must be*

$$Q_M = \oplus m\,[-1]\ .$$

The presentation will take the preceding note (starting on page 439) as a requisite. Further, since reducible solutions play an essential role, we will make use of statements presented in the end-notes of the preceding chapter (connections on line bundles, page 357). This proof of Donaldson's theorem is due to P. Kronheimer, with a finishing touch from N. Elkies.

Assume that M is a smooth 4–manifold with negative-definite intersection form. In other words, $b_2^+(M) = 0$, and thus $b_2(M) = b_2^-(M)$ and $\operatorname{sign} M = -b_2(M)$.

Let $\underline{w}$ be any characteristic element of M. Then we must have $\underline{w}\cdot\underline{w} = \operatorname{sign} M$ (mod 8) and hence

$$\underline{w}\cdot\underline{w} + b_2(M) = 0 \pmod 8\ .$$

Characteristic elements $\underline{w}$ correspond to spin$^{\mathbb{C}}$ structures $\mathfrak{s}$ with $c_1(\mathfrak{s}) = \underline{w}$. The virtual dimension of the corresponding Seiberg–Witten moduli spaces is

$$\begin{aligned}\operatorname{vdim}\mathfrak{M} &= \tfrac{1}{4}\big(\underline{w}\cdot\underline{w} \;-\; 3\operatorname{sign} M \;-\; 2\chi(M)\big)\\ &= \tfrac{1}{4}\big(\underline{w}\cdot\underline{w} + b_2(M)\big) - 1\ .\end{aligned}$$

A consequence is that the dimension of the moduli space is always *odd*.

Assume there is some characteristic element $\underline{w}$ for which the virtual dimension $\operatorname{vdim}\mathfrak{M}$ is non-negative, that is to say, at least 1. Then the moduli space is either empty or a (singular) manifold of the expected dimension. Since $b_2^+ = 0$, there are always reducible solutions in $\mathfrak{M}$, which hence cannot be empty.

The space $\mathfrak{M}^0$ of Seiberg–Witten solutions modulo the action of

$$\mathscr{G}^0(\mathcal{L}) = \{g\colon M \to \mathbb{S}^1 \mid g(p) = 1\}$$

is a smooth manifold of dimension $\operatorname{vdim}\mathfrak{M} + 1$. Its dimension is even and at least 2. The group $\mathbb{S}^1$ acts on $\mathfrak{M}^0$ with fixed points at the reducible solutions, and its quotient is $\mathfrak{M}$.

Assume first that $H^1(M;\mathbb{R}) = 0$. Then there is a unique gauge class of reducible solutions. In other words, there is only one fixed point of the action of $\mathbb{S}^1$ on $\mathfrak{M}$.

For the manifold $\mathfrak{M}^0$, a discussion in terms of the complex

$$0 \longrightarrow T_{\mathscr{G}^0(\mathcal{L})}\big|_1 \xrightarrow{\;d\mathfrak{g}\;} T_{\Gamma(\mathcal{W}^+)\times Conn(\mathcal{L})}\big|_{(\varphi,A)} \xrightarrow{\;d\mathfrak{sw}\;} T_{\Gamma(\mathcal{W}^-)\times i\Gamma(\Lambda^2_+)}\big|_{(0,0)} \longrightarrow 0$$

can be undertaken, quite analogous to the one developed for $\mathfrak{M}$ in the preceding note. It leads to an identification of the tangent space of $\mathfrak{M}^0$ with the first cohomology group of this complex,

$$T_{\mathfrak{M}^0}\big|_{[\varphi,A]} = \mathcal{H}^1_{(\varphi,A)}\ .$$

The only difference now is that the tangent space $T_{\mathscr{G}^0(\mathcal{L})}|_1$ is not the full space $i\Gamma(\underline{\mathbb{R}}) = \{if \colon M \to i\,\mathbb{R}\}$, but its codimension 1 subspace of maps with $f(p) = 0$. At every reducible solution $(0, A)$, the derivatives are

$$d\mathfrak{sw}|_{(0,A)}(\psi, i\vartheta) = (\mathcal{D}^A\psi,\ \ i\,d^+\vartheta) \qquad \text{and} \qquad d\mathfrak{g}|_1(if) = (0,\ \ 2i\,df)\ ,$$

and thus the tangent space to $\mathfrak{M}^0$ at $[0, A]$ is merely

$$\mathcal{H}^1_{(0,A)} = \operatorname{Ker}\mathcal{D}^A\ ,$$

since $H^1(M; \mathbb{R})$ was assumed trivial.

We think of $T_{\mathfrak{M}^0}|_{[0,A]} = \operatorname{Ker}\mathcal{D}^A$ as an approximation to the manifold $\mathfrak{M}^0$ around the point $[0, A]$. The action of $\mathbb{S}^1$ on $\mathfrak{M}^0$, for which $[0, A]$ is a fixed point, is well approximated by the standard multiplicative action of $\mathbb{S}^1$ on the complex vector space[37] $\operatorname{Ker}\mathcal{D}^A \approx \mathbb{C}^{m+1}$, with $2(m+1) = \dim\mathfrak{M}^0$ and hence $2m = \dim\mathfrak{M} - 1$.

Therefore, the quotient $\mathfrak{M}^0/\mathbb{S}^1$ looks near $[0, A]$ like a a cone on the quotient $\mathbb{S}^{2m+1}/\mathbb{S}^1 = \mathbb{CP}^m$. Hence, a small neighborhood U in $\mathfrak{M}$ of the singularity $[0, A]$ has boundary $\partial U = \mathbb{CP}^m$, and must itself be a cone on $\mathbb{CP}^m$ $(2m = \dim\mathfrak{M} - 1)$.

One case. If $m = 0$, one should think of $\mathbb{CP}^0$ as made of a point. In this case the interior of U must be a copy of $[0, \infty)$. Since there is only one reducible solution up to gauge, this implies that $\mathfrak{M}$ is a compact manifold of dimension 1, whose boundary is made of only one point. This, of course, is impossible.

Another case. If $m \geq 1$, we reach a contradiction by arguing as follows: The bundle $\mathfrak{M}^0 \to \mathfrak{M}$ restricts to $\mathbb{CP}^m$ as the circle-bundle of its universal bundle. The universal bundle over $\mathbb{CP}^m$ is the complex-line bundle that has as fiber over each point of $\mathbb{CP}^m$ the complex-line represented by that point. Indeed, this is exactly what happens here: over each point of $\mathbb{CP}^m \subset \mathfrak{M}$ lies the circle of $\mathfrak{M}^0$ from which it is coming.

Therefore the class $\mathfrak{u}$ evaluates on ∂U as

$$\int_{\mathbb{CP}^m} \mathfrak{u}^m = \pm 1\ ,$$

since $\mathfrak{u}$, being the Chern class of $\mathfrak{M}^0 \to \mathfrak{M}$ which is universal over $\mathbb{CP}^m = \partial U$, must restrict to $\mathbb{CP}^m$ as the Poincaré-dual of $\pm[\mathbb{CP}^1] \in H_2(\mathbb{CP}^m; \mathbb{Z})$.

On the other hand, our $\mathbb{CP}^m$ bounds in $\mathfrak{B}_*$:

$$\mathbb{CP}^m = \partial(\mathfrak{M} \setminus U)\ .$$

Therefore its homology class $[\mathbb{CP}^m]$ in $H_{2m}(\mathfrak{B}_*; \mathbb{Z})$ must be trivial. Since $\mathfrak{u}$ is a global cohomology class of $\mathfrak{B}_*$ and $\mathfrak{u}^m \cdot [\mathbb{CP}^m] = \pm 1$, this is impossible.

---

[37] These statements can be made quite rigorous through the use of the so-called *Kuranishi technique.* This technique was created by **M. Kuranishi**'s ***On the locally complete families of complex analytic structures*** **[Kur62]** in the study of moduli spaces of complex varieties, and was then applied to various other moduli space settings. Its use in gauge theory originates with **M. Atiyah, N. Hitchin** and **I. Singer**'s ***Self-duality in four-dimensional Riemannian geometry*** **[AHS78]**.

We have proved that, as soon as we assume the virtual dimension of $\mathfrak{M}$ to be non-negative, contradictions ensue. Hence, for all spin$^{\mathbb{C}}$ structures on M, we must have[38] $\operatorname{vdim}\mathfrak{M} \leq -1$. That is to say,

$$\tfrac{1}{4}\big(\underline{w}\cdot\underline{w} + b_2(M)\big) - 1 \;\leq\; -1$$

for *all* characteristic elements $\underline{w}$ of M.

Theorem. *If M is a smooth manifold with negative-definite intersection form, then for every characteristic element $\underline{w}$ of M we must have*

$$\underline{w}\cdot\underline{w} \;\leq\; -b_2(M)\,.$$

Proof. The theorem was already proved above in the case when $H^1(M;\mathbb{R}) = 0$. If $H^1(M;\mathbb{R}) \neq 0$, we will reduce it to the settled case by killing H^1 through some simple surgery.

Namely, let $\{\ell_1, \ldots, \ell_m\}$ be a basis of $H_1(M;\mathbb{R})$ made of *integral classes.* Represent each of them by disjointly embedded circles in M. These circles do not bound any surface in M and (not being torsion) neither does any of their (homological) multiples.

Let S^1 be such a circle. Its neighborhood is a copy of $S^1 \times D^3$. If we remove it from M, we obtain a new manifold M° with boundary $S^1 \times S^2$, to which we can then glue a copy of $D^2 \times S^2$. Since neither S^1 nor a bunch of parallel copies of S^1 bounds any surface in M°, the added disk D^2 (or a bunch of parallel copies of it) will not cap anything that would create new homology classes. Thus, after surgery, the class of S^1 is trivial, but the 2–homology of M is unscathed.

Repeating this for all ℓ_k's results in a new manifold $\widetilde{M}$, which has $b_1(\widetilde{M}) = 0$. Moreover, since the second homology was untouched, it has $Q_{\widetilde{M}} = Q_M$, and hence the same characteristic elements as M. Then we can apply the $b_1 = 0$ part of the theorem to $\widetilde{M}$ and reach the needed conclusion. □

While the above theorem is not yet the full Donaldson theorem, it is already strong enough, for example, to exclude all *even definite* intersection forms:

Corollary. *If M is a smooth manifold with definite intersection form, then Q_M cannot be even.*

Proof. Any even form would have $\underline{w} = 0$ as characteristic element, and thus $\underline{w}\cdot\underline{w} = 0 > -b_2(M)$, which contradicts our theorem. □

For example, $E_8 \oplus E_8$ is hereby expelled from the smooth realm.

Without further ado, to obtain the full Donaldson theorem, one needs to use the following algebraic result, proved in ***A characterization of the $\mathbb{Z}^n$ lattice*** **[Elk95]**:

Lemma *(N. Elkies). Let $Q\colon Z \times Z \to \mathbb{Z}$ be a symmetric unimodular bilinear form. If Q is neither $\oplus\,[-1]$ nor $\oplus\,[+1]$, then there exists a characteristic vector $\underline{w}$ so that*

$$|\underline{w}\cdot\underline{w}| < \operatorname{rank} Q\,.$$ □

---

[38] Remember that $\operatorname{vdim}\mathfrak{M}$ is always odd.

In other words, the forms $\oplus\,[-1]$ and $\oplus\,[+1]$ are the only forms without short characteristic vectors; their shortests are $\underline{w} = \sum e_k$, with $|\underline{w}\cdot\underline{w}| = \operatorname{rank} Q$.

Since a short characteristic element in a negative-definite intersection form Q_M means $\underline{w}\cdot\underline{w} > -b_2(M)$, the full Donaldson theorem follows.

Note: Seiberg–Witten on Kähler surfaces

Seiberg–Witten theory is strongly related to complex geometry. In what follows we will explore this. We will use concepts and notations from the end-notes of the preceding chapter (complex-valued form, $\bar\partial$–operators, and holomorphic bundles, page 365) as well as results proved in a previous note (rewritings of the spinor bundle, squaring map, and Dirac operator, page 438).

Canonical spin$^{\mathbb{C}}$ structure. Assume that M is a Kähler surface and denote by ω its Kähler form $\omega(x,y) = \langle Jx, y\rangle$. The complex structure J of M induces a canonical spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J$ on M, with determinant line bundle $\mathcal{L}_{\mathfrak{s}_J} = K^*_M$, and with spinor bundles

$$\mathcal{W}^+_{\mathfrak{s}_J} = \Lambda^{0,0} \oplus \Lambda^{0,2} \qquad \text{and} \qquad \mathcal{W}^-_{\mathfrak{s}_J} = \Lambda^{0,1}\,.$$

Here $\Lambda^{0,0} = \Lambda^0 \otimes \mathbb{C} = \underline{\mathbb{C}}$, while $\Lambda^{0,1}$ denotes the half of $\Lambda^1 \otimes \mathbb{C}$ represented by complex-valued $\mathbb{C}$–anti-linear 1–forms, and $\Lambda^{0,2}$ is the subbundle of $\Lambda^2 \otimes \mathbb{C}$ containing $\mathbb{C}$–bi-anti-linear forms.

The Levi–Cività connection of the Kähler metric induces a canonical $U(1)$–connection A_0 on $K^*_M = \det_{\mathbb{C}} T^*_M = \Lambda^{0,2}$. Since A_0 is compatible with the natural holomorphic structure of K^*_M (*i.e.*, $\bar\partial_{A_0} = \bar\partial^{K^*}$), we have $F^{0,2}_{A_0} = 0$. The induced Dirac operator $\mathcal{D}^{A_0}$ is:

$$\mathcal{D}^{A_0} \colon \Gamma(\Lambda^{0,0} \oplus \Lambda^{0,2}) \longrightarrow \Gamma(\Lambda^{0,1})$$
$$\tfrac{1}{\sqrt{2}}\,\mathcal{D}^{A_0} = \bar\partial \oplus \bar\partial^*\,,$$

where $\bar\partial^* \colon \Gamma(\Lambda^{0,2}) \to \Gamma(\Lambda^{0,1})$ is the formal adjoint of $\bar\partial \colon \Gamma(\Lambda^{0,1}) \to \Gamma(\Lambda^{0,2})$, characterized by

$$\int_M \langle \bar\partial\alpha, \beta\rangle = \int_M \langle \alpha, \bar\partial^*\beta\rangle$$

for all $\alpha \in \Gamma(\Lambda^{0,1})$ and $\beta \in \Gamma(\Lambda^{0,2})$. Each operator $\bar\partial \colon \Gamma(\Lambda^{0,k}) \to \Gamma(\Lambda^{0,k+1})$ is part of the (complexified) exterior derivative $d \colon \Gamma(\Lambda^k \otimes \mathbb{C}) \to \Gamma(\Lambda^{k+1} \otimes \mathbb{C})$. Since M is a complex surface, we have $\bar\partial\,\bar\partial = 0$, and $\bar\partial f$ vanishes if and only if $f \colon M \to \mathbb{C}$ is holomorphic.

Proof that $\mathcal{D}^{A_0} = \sqrt{2}(\bar\partial + \bar\partial^*)$. *While this statement has been argued in a previous note (page 438) by using quaternions, we also present an alternative argument. The Kähler 2–form ω acts by Clifford multiplication on $\mathcal{W}^+$ and splits it into eigenbundles as $\mathcal{W}^+ = \underline{\mathbb{C}} \oplus K^*$, with ω acting on $\underline{\mathbb{C}}$ as $-2i$ and on K^* as $+2i$. Fix a section $\mathfrak{z}$ of $\underline{\mathbb{C}} \subset \mathcal{W}^+$ of constant length 1. Through Clifford multiplication, this section induces the identification*

$$\Lambda^{0,0} \oplus \Lambda^{0,2} = \mathcal{W}^+ \colon \quad f + \beta \longmapsto (f + \tfrac{1}{2}\beta) \bullet \mathfrak{z}\,.$$

The factor $\frac{1}{2}$ is there to insure that the map is an isometry. Similarly, we obtain the identification

$$\Lambda^{0,1} = \mathcal{W}^- \colon \quad \alpha \longmapsto \tfrac{1}{\sqrt{2}}\alpha \bullet \mathfrak{z}\,.$$

For any connection A on K^*, we have $\nabla^{A+i\vartheta}\mathfrak{z} = \nabla^{A}\mathfrak{z} + \frac{1}{2}i\vartheta \cdot \mathfrak{z}$, and therefore there exists a unique choice of connection A_0 such that $\nabla^{A_0}\mathfrak{z}$ is $\mathbb{C}$–orthogonal to the subbundle $\underline{\mathbb{C}}$ of $\mathcal{W}^+$, that is, $\langle \nabla^{A_0}_X \mathfrak{z}, \mathfrak{z}\rangle_{\mathbb{C}} = 0$. Using this special connection A_0, we compute:

$$\nabla^{A_0}(\omega \bullet \mathfrak{z}) = (\nabla\omega)\bullet\mathfrak{z} + \omega\bullet(\nabla^{A_0}\mathfrak{z}) \ .$$

Since we have $\omega\bullet\mathfrak{z} = -2i\mathfrak{z}$ and $\omega\bullet(\nabla^{A_0}\mathfrak{z}) = +2i\,\nabla^{A_0}\mathfrak{z}$, we obtain $\nabla^{A_0}\mathfrak{z} = \frac{i}{4}(\nabla\omega)\bullet\mathfrak{z}$. However, ω is a Kähler form, and so $\nabla\omega = 0$ and therefore

$$\nabla^{A_0}\mathfrak{z} = 0 \ .$$

Through the identification $\Lambda^{0,0}\oplus\Lambda^{0,2} = \mathcal{W}^+$, we compute

$$\nabla^{A_0}_X(f+\beta) \ = \ \nabla^{A_0}_X\big((f+\tfrac{1}{2}\beta)\bullet\mathfrak{z}\big) \ = \ \big(df(X)+\tfrac{1}{2}\nabla_X\beta\big)\bullet\mathfrak{z} \ + \ (f+\tfrac{1}{2}\beta)\bullet\nabla^{A_0}_X\mathfrak{z}$$

and the last term vanishes. The Dirac operator corresponding to A_0 is

$$\mathcal{D}^{A_0}(f+\beta) \ = \ \sum e_k\bullet\nabla^{A_0}_{e_k}(f+\beta) \ = \ \sum df(e_k)e_k\bullet\mathfrak{z} \ + \ \tfrac{1}{2}(e_k\bullet\nabla_{e_k}\beta)\bullet\mathfrak{z} \ .$$

The Clifford multiplication on forms acts as $v\bullet\alpha = v\wedge\alpha - v\lrcorner\alpha$, and thus $\sum e_k\bullet\nabla_{e_k}\beta = d\beta + d^*\beta$. Since for any 1–form γ we have $\gamma\bullet\varphi^+ = -(*\gamma)\bullet\varphi^+$ and since $d^*\beta = -*d\beta$ (as β is a self-dual 2–form), we get $(d\beta+d^*\beta)\bullet\mathfrak{z} = 2\,d^*\beta\bullet\mathfrak{z}$ and finally obtain $\mathcal{D}^{A_0}(f+\beta) = df\bullet\mathfrak{z} + d^*\beta\bullet\mathfrak{z}$. Passing this back through the identification $\Lambda^{0,1} = \mathcal{W}^-$ yields $\mathcal{D}^{A_0}(f+\beta) = \sqrt{2}(\bar\partial f + \bar\partial^*\beta)$, as claimed. □

The other spin$^{\mathbb{C}}$ structures. All other spin$^{\mathbb{C}}$ structures on M can be obtained using translations by $H^2(M;\mathbb{Z})$, as explained in section 10.2 (page 389).

Specifically, for every class ε we build a complex-line bundle L_ε such that $c_1(L_\varepsilon) = \varepsilon$. Then the determinant line bundle corresponding to the spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J+\varepsilon$ appears as $\mathcal{L}_{\mathfrak{s}_J+\varepsilon} = K^*_M \otimes L_\varepsilon^{\otimes 2}$, while its spinor bundles are

$$\mathcal{W}^+_{\mathfrak{s}_J+\varepsilon} = L_\varepsilon \oplus (L_\varepsilon\otimes\Lambda^{0,2}) \qquad \text{and} \qquad \mathcal{W}^-_{\mathfrak{s}_J+\varepsilon} = L_\varepsilon\otimes\Lambda^{0,1} \ ,$$

since $\Lambda^{0,0}\otimes L_\varepsilon$ is simply L_ε. Each term $L_\varepsilon\otimes\Lambda^{0,k}$ is understood as made of L_ε–valued $(0,k)$–forms on M.

Choose some Hermitian fiber-metric on L_ε. Any choice of a $U(1)$–connection B on L_ε combines with the special connection A_0 on K^*_M to yield a connection $A = A_0 + 2B$ on $\mathcal{L}_{\mathfrak{s}_J+\varepsilon}$. Varying $B \in \mathcal{C}onn(L_\varepsilon)$ covers all connections A from $\mathcal{C}onn(\mathcal{L}_{\mathfrak{s}_J+\varepsilon})$. (This procedure works for $\mathcal{L}_{\mathfrak{s}_J}$ itself: tensoring with the trivial bundle $L_0 = \underline{\mathbb{C}}$ and varying the connections B on $\underline{\mathbb{C}}$ covers all connections $A = A_0 + 2B$ on $\mathcal{L}_{\mathfrak{s}_J}$.)

The connection A_0+2B on $\mathcal{L}_{\mathfrak{s}_J+\varepsilon}$ induces a Dirac operator $\mathcal{D}^{A_0+2B}$. It is easy to check that we can write

$$\mathcal{D}^{A_0+2B}\colon \Gamma\big(L_\varepsilon \oplus (L_\varepsilon\otimes\Lambda^{0,2})\big) \longrightarrow \Gamma(L_\varepsilon\otimes\Lambda^{0,1})$$
$$\tfrac{1}{\sqrt{2}}\mathcal{D}^{A_0+2B} = \bar\partial_B \oplus \bar\partial_B^* \ .$$

Here, for every $\ell \in \Gamma(L_\varepsilon)$, we set $\bar\partial_B\ell$ to be the $(0,1)$–part of the L_ε–valued 1–form $d_B\ell$. Further, $\bar\partial_B^*$ denotes the the formal adjoint of $\bar\partial_B\colon \Gamma(L_\varepsilon\otimes\Lambda^{0,1}) \to \Gamma(L_\varepsilon\otimes\Lambda^{0,2})$, with the latter appearing as part of $d_B\colon \Gamma(L_\varepsilon\otimes\Lambda^1) \to \Gamma(L_\varepsilon\otimes\Lambda^2)$. We thus have

$$\int_M \langle\bar\partial_B\alpha, \beta\rangle = \int_M\langle\alpha, \bar\partial_B^*\beta\rangle$$

for all $\alpha \in \Gamma(L_\varepsilon\otimes\Lambda^{0,1})$ and $\beta\in\Gamma(L_\varepsilon\otimes\Lambda^{0,2})$.

The Seiberg–Witten equations. Choose a spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J + \varepsilon$ on M. The Seiberg–Witten equations are

$$\begin{cases} \mathcal{D}^A \varphi = 0 \\ F_A^+ = \sigma(\varphi) \, , \end{cases}$$

where φ is a self-dual spinor field from $\mathcal{W}^+_{\mathfrak{s}_J+\varepsilon}$, and A is a connection on $\mathcal{L}_{\mathfrak{s}_J+\varepsilon}$. We rewrite them while taking advantage of our special situation.

First, we write $\varphi = (\ell, \beta)$, where $\ell \in \Gamma(L_\varepsilon)$ and $\beta \in \Gamma(L_\varepsilon \otimes \Lambda^{0,2})$. Second, we use the rewriting of the squaring map obtained on page 438, so that the second equation becomes

$$F_A^{0,2} = 2\ell^*\beta \qquad \text{and} \qquad (F_A^+)^{1,1} = i(|\ell|^2 - |\beta|^2)\omega \, ,$$

where ℓ^* is the image of ℓ under the (complex anti-linear) isomorphism $L_\varepsilon \approx L_\varepsilon^*$ induced from the metric of L_ε, while $\ell^*\beta$ is the image of $\ell^* \otimes \beta$ under the evaluation map $L_\varepsilon^* \otimes (L_\varepsilon \otimes \Lambda^{0,2}) \to \Lambda^{0,2}$.

We now notice that we can write $A = A_0 + 2B$ and hence $F_A = F_{A_0} + 2F_B$, where F_B is the curvature of the connection B on L_ε. However, since A_0 defines the holomorphic structure of K^*, it follows that $F_{A_0}^{0,2} = 0$. What appears are the equations

$$\begin{cases} \bar{\partial}_B \ell + \bar{\partial}_B^* \beta = 0 \\ F_B^{0,2} = \ell^*\beta \\ (F_A^+)^{1,1} = i\left(|\ell|^2 - |\beta|^2\right)\omega \, . \end{cases}$$

In this way, a solution to the Seiberg–Witten equations is now viewed as a triplet (ℓ, β, B), with $\ell \in \Gamma(L_\varepsilon)$, $\beta \in \Gamma(L_\varepsilon \otimes \Lambda^{0,2})$ and $B \in \mathit{Conn}(L_\varepsilon)$.

Computation. Assume that (ℓ, β, B) is a solution. We apply $\bar{\partial}_B$ to the first equation above to get

$$\bar{\partial}_B \bar{\partial}_B \ell + \bar{\partial}_B \bar{\partial}_B^* \beta = 0 \, .$$

This is just

$$F_B^{0,2} \cdot \ell + \bar{\partial}_B \bar{\partial}_B^* \beta = 0 \, ,$$

and by using the second equation we obtain

$$|\ell|^2 \beta + \bar{\partial}_B \bar{\partial}_B^* \beta = 0 \, .$$

We take the inner product with β, integrate over M and use that $\bar{\partial}_B$ is adjoint to $\bar{\partial}_B^*$. The result is

$$\int_M |\ell|^2 \, |\beta|^2 + \int_M |\bar{\partial}_B^* \beta|^2 = 0 \, .$$

Therefore both integrals need to vanish, and we must have both

$$\bar{\partial}_B^* \beta \equiv 0 \qquad \text{and} \qquad |\ell| \, |\beta| \equiv 0 \, .$$

Since $|\ell| \, |\beta| \equiv 0$, it follows that $\ell^* \beta \equiv 0$, and thus that we have

$$F_B^{0,2} \equiv 0 \, ,$$

after applying the second equation. By the integrability theorem (page 369), this implies that B defines a holomorphic structure on L_ε.

On the other hand, since $\bar{\partial}_B^* \beta \equiv 0$, the first equation implies that $\bar{\partial}_B \ell \equiv 0$, which means that ℓ is a holomorphic section of L_ε, for the holomorphic structure determined by B. Dually, $* \beta \in \Gamma(\Lambda^{2,0} \otimes L_\varepsilon^*)$ is a holomorphic section in $K_M \otimes L_\varepsilon^*$ (since $K_M = \Lambda^{2,0}$).

Further, since $|\ell|\,|\beta| \equiv 0$, this means that at each point of M one of ℓ or β must vanish. Therefore there must be an open set of M on which one of them vanishes. Since both are holomorphic, this implies that either

$$\ell \equiv 0 \qquad \text{or} \qquad \beta \equiv 0 \,.$$

Therefore, either we are dealing with $(\ell, 0, B)$ and ℓ is a holomorphic section of L_ε, or we are dealing with $(0, \beta, B)$ and β is a holomorphic section of $K_M \otimes L_\varepsilon^*$, or we have $(0, 0, B)$ with $F_B^+ = 0$.

Interpretation. We are now ready to translate the Seiberg–Witten monopoles into complex-geometric terms:

One case. Assume that ℓ is nontrivial, and thus $\beta \equiv 0$. Then L_ε is a holomorphic line bundle on M, with holomorphic section ℓ. Therefore the zero set of ℓ is a $\mathbb{Z}$–linear combination of curves in M with *positive* coefficients, representing the Chern class $c_1(L_\varepsilon) = \varepsilon$. Thus, a solution of the Seiberg–Witten equations for the spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J + \varepsilon$ yields a representation of the class ε as a combination of complex curves with positive coefficients.[39]

If two solutions $(\ell', 0, B')$ and $(\ell'', 0, B'')$ are gauge-equivalent, then the induced holomorphic structures on L_ε must be isomorphic, with the isomorphism carrying the section ℓ' to ℓ''. Since a self-isomorphism of a holomorphic line bundle can only be the multiplication by constant scalar, it follows that $\ell' = c\ell''$, and their zero-set is the same. Thus a gauge-class of solutions is indeed just a positive combination of complex curves representing ε.

We are left with the third equation, namely $(F_A^+)^{1,1} = i(|\ell|^2 - |\beta|^2)\,\omega$, where $A = A_0 + 2B$ is a connection on the determinant line-bundle $\mathcal{L}_{\mathfrak{s}_J+\varepsilon} = K^* \otimes L_\varepsilon^{\otimes 2}$. Since $\beta \equiv 0$, this becomes $(F_A^+)^{1,1} = i\,|\ell|^2\,\omega$, but $c_1(\mathcal{L})$ is represented by the form $\frac{i}{2\pi}F_A$, and so it follows that $c_1(\mathcal{L}) \cdot [\omega] = \int \langle \frac{i}{2\pi} F_A,\, \omega \rangle = \int \langle \frac{i}{2\pi} (F_A^+)^{1,1},\, \omega \rangle$ and hence $c_1(\mathcal{L}) \cdot [\omega] < 0$, *i.e.*,

$$(K_M^* + 2\varepsilon) \cdot [\omega] < 0 \,.$$

The converse can be proved to be true as well: For every positive linear combination of complex curves that represents a class ε with $\varepsilon \cdot [\omega] < 0$, the spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J + \varepsilon$ admits a monopole $(\ell, 0, B)$ as above. Indeed, such a combination determines a unique holomorphic structure on L_ε and section ℓ, up to multiplication by constants. If L_ε is further endowed with a Hermitian metric, then the holomorphic structure determines uniquely a compatible connection B. The task is to find the right Hermitian metric on L_ε so that the curvature of the corresponding B satisfies the third equation. The full argument involves varying the Hermitian metric and a study of non-linear equations like $\Delta f + ae^{2f} = g$, and corresponding results from **J. Kazdan** and **F. Warner**'s ***Curvature functions for compact* 2*–manifolds*** **[KW74]**.

---

[39] In complex-geometric terms, this is an *effective divisor* of class ε.

Another case. When β is nontrivial and $\ell \equiv 0$, a similar reasoning ends up with β defining a holomorphic section of $K_M \otimes L_\varepsilon^*$ and thus representing $K_M - \varepsilon$ by a positive combination of curves, and with $c_1(\mathcal{L}) \cdot [\omega] > 0$, *i.e.*,

$$(K_M^* + 2\varepsilon) \cdot [\omega] > 0 \, .$$

Reducible case. If both ℓ and β vanish identically, then we are dealing with a reducible monopole, and the third equation implies that $(K_M^* + 2\varepsilon) \cdot [\omega] = 0$. Since no perturbations were made, these might be unavoidable (for example, when $\varepsilon = 0$ for $K3$ surfaces, since $K_{K3} = 0$).

In conclusion, we have:

Theorem *(E. Witten). Let M be a Kähler surface with $b_2^+(M) \geq 0$, and ε any class in $H^2(M; \mathbb{Z})$. Then:*

- *If $(K_M^* + 2\varepsilon) \cdot [\omega] < 0$, then the Seiberg–Witten moduli space $\mathfrak{M}_{K^*+2\varepsilon}$ coincides with the space of all representations of ε as positive-coefficient combinations of complex curves*[40] *in M.*
- *If $(K_M^* + 2\varepsilon) \cdot [\omega] > 0$, then the Seiberg–Witten moduli space $\mathfrak{M}_{K^*+2\varepsilon}$ coincides with the space of all representations of $K_M - \varepsilon$ as positive-coefficient combinations of complex curves in M.*
- *If $(K_M^* + 2\varepsilon) \cdot [\omega] = 0$, then the Seiberg–Witten moduli space $\mathfrak{M}_{K^*+2\varepsilon}$ contains only reducible solutions.* □

This result is due to **E. Witten** in the founding paper ***Monopoles and four-manifolds*** **[Wit94]**. Notice that the first two cases are linked through the Seiberg–Witten involution $\mathfrak{s} \mapsto -\mathfrak{s}$ with $\mathcal{SW}_M(-\mathfrak{s}) = \pm \mathcal{SW}_M(\mathfrak{s})$ (page 405). Indeed, if $K_M^* + 2\varepsilon' = -(K^* + 2\varepsilon'')$, then $\varepsilon' = K_M - \varepsilon''$. On the complex-geometric side, the two cases are linked by Serre duality.

Of course, one is now tempted to interpret $\mathcal{SW}_M(K^* + 2\varepsilon)$ as counting the number of representations of either ε or $K_M - \varepsilon$ as complex curves in M. The problem is that the whole argument above was made *without* any perturbations, namely, we did not perturb the metric (since we needed the Kähler structure) and we did not perturb the equations. Therefore, the above moduli spaces might not be cut transversely by the equations. In particular, the actual dimension of $\mathfrak{M}_{K^*+2\varepsilon}$ might be *bigger* than the expected dimension $K_M^* \cdot \varepsilon + \varepsilon \cdot \varepsilon$. Consequently, the above description of the moduli spaces cannot in general be used directly to evaluate the Seiberg–Witten invariants.

Nonetheless, these descriptions allow one to translate the Seiberg–Witten equations into algebraic geometric terms (Hilbert scheme of curves) and, after a careful analysis of the failure of transversality (*i.e.*, of the obstruction spaces $\mathcal{H}^2$), leads to a complete evaluation of the Seiberg–Witten invariants. This argument can be found, for example, in **R. Friedman** and **J. Morgan**'s ***Obstruction bundles, semiregularity, and Seiberg–Witten invariants*** **[FM99]** or, for a different approach (evaluating the excess intersection by using a localized Euler form), in **R. Brussee**'s ***The canonical class and the C^∞ properties of Kähler surfaces*** **[Bru96]**.

---

[40] In other words, $\mathfrak{M}$ is the space of all *effective divisors* of class ε, or the number of holomorphic sections of the bundle L_ε, up to scalar multiplication, as in $\#\mathbb{P}(\Gamma_{hol}(L_\varepsilon))$.

E. Witten himself **[Wit94]** proposed perturbing the equation $F_B^{0,2} = \ell^* \beta$ to $F_B^{0,2} = \ell^* \beta + \overline{\mu}$ for some holomorphic $(2,0)$–form $\mu \in \Gamma(\Lambda^{2,0})$. When $P_1(M) \geq 1$ (equivalently, when $b_2^+ \geq 3$), the bundle $\Lambda^{2,0} = K_M$ admits such a non-trivial holomorphic section μ. Its zeros $Zeros(\mu)$ represent the class $c_1(K_M)$ as a positive combination of complex curves in M. These perturbed Seiberg–Witten equations then lead to different moduli spaces $\mathfrak{M}^{\mu}_{K^*+2\varepsilon}$, which can be identified with the space of decompositions $Zeros(\mu) = D' + D''$ into two positive combinations of curves, so that D' represents ε. The argument can be read from **L. Nicolaescu's** ***Notes on Seiberg–Witten theory*** **[Nic00**, sec 3.2.2].

Without any of this extra technology, but by directly using the description from the above theorem, we notice that, if $\mathcal{SW}_M(K_M^* + 2\varepsilon) \neq 0$, then the above non-perturbed moduli space $\mathfrak{M}_{K^*+2\varepsilon}$ must be non-empty, and therefore certain bundles must admit holomorphic sections.

Combining this with the classification of complex surfaces directly yields restrictions on basic classes for surfaces of general type and for proper elliptic ones, as we will see in what follows. Remember from chapter 7 that a minimal Kähler surface is of general type if $K_M \cdot K_M > 0$ and $K_M \cdot [\omega] > 0$, and is proper elliptic if $K_M \cdot K_M = 0$ and $K_M \cdot [\omega] > 0$ (page 295).

Surfaces of general type.

Theorem. *If M is a Kähler surface of general type, minimal with respect to blow-downs, then its basic classes are exactly $\pm K_M^*$, with $\mathcal{SW}_M(\pm K_M^*) = \pm 1$.*

Proof. We will show that the only spin$^{\mathbb{C}}$ structures that admit monopoles are $\pm K_M^*$. Let $\mathcal{L} = K_M^* + 2\varepsilon$ be a spin$^{\mathbb{C}}$ structure on M that admits monopoles and so that the *expected* dimension of $\mathfrak{M}$ is non-negative, *i.e.*, $K_M^* \cdot \varepsilon + \varepsilon \cdot \varepsilon \geq 0$. This latter condition can be written as

$$K_M^* \cdot K_M^* \leq \mathcal{L} \cdot \mathcal{L} .$$

However, M is of general type and we must have $K_M \cdot K_M > 0$, and hence

$$\mathcal{L} \cdot \mathcal{L} > 0 .$$

Maybe after using the involution $\mathfrak{s} \mapsto -\mathfrak{s}$, we can assume that we are in the case when $(K_M^* + 2\varepsilon) \cdot [\omega] < 0$, that is, when

$$\mathcal{L} \cdot [\omega] < 0 .$$

Then a Seiberg–Witten monopole for $\mathcal{L}$ is the same with a holomorphic structure on L_ε and a non-trivial holomorphic section $\ell \in \Gamma(L_\varepsilon)$. Since ε is thus represented by a positive-coefficient combination of complex curves and since K_M is nef, it follows that $K_M \cdot \varepsilon \geq 0$. We write this as $K_M^* \cdot \varepsilon \leq 0$, then rewrite it as $K_M^* \cdot (K_M^* + 2\varepsilon - K_M^*) \leq 0$ or $K_M^* \cdot (\mathcal{L} - K_M^*) \leq 0$, and hence

$$K_M^* \cdot \mathcal{L} \leq K_M^* \cdot K_M^* .$$

Further, since M is of general type, we must have $K_M \cdot [\omega] > 0$, or

$$-K_M^* \cdot [\omega] > 0 .$$

Since on the other hand we have $\mathcal{L} \cdot [\omega] < 0$, it follows that there must exists some real $t \geq 0$ so that

$$(t\mathcal{L} - K_M^*) \cdot [\omega] = 0 .$$

Since $[\omega] \cdot [\omega] > 0$, the geometric Hodge signature theorem[41] (page 278) implies that

$$(t\mathcal{L} - K_M^*) \cdot (t\mathcal{L} - K_M^*) \leq 0 ,$$

that is, $t^2(\mathcal{L} \cdot \mathcal{L}) - 2t(K_M^* \cdot \mathcal{L}) + (K_m^* \cdot K_M^*) \leq 0$. Since $K_M^* \cdot K_M^* \leq \mathcal{L} \cdot \mathcal{L}$ (from the dimension condition) and $K_M^* \cdot \mathcal{L} \leq K_M^* \cdot K_M^*$ (from K_M^* nef and L_ε admitting sections), it follows that

$$\begin{aligned} 0 &\geq t^2(\mathcal{L} \cdot \mathcal{L}) - 2t(K_M^* \cdot \mathcal{L}) + (K_M^* \cdot K_M^*) \\ &\geq t^2(K_M^* \cdot K_M^*) - 2t(K_M^* \cdot K_M^*) + (K_M^* \cdot K_M^*) \\ &= (t^2 - 2t + 1)(K_M^* \cdot K_M^*) \\ &= (t-1)^2(K_M^* \cdot K_M^*) \geq 0 . \end{aligned}$$

Therefore the caught quantities must all be zero, t must be 1, and we have both

$$(\mathcal{L} - K_M^*) \cdot (\mathcal{L} - K_M^*) = 0 \qquad \text{and} \qquad (\mathcal{L} - K_M^*) \cdot [\omega] = 0 .$$

Using again Hodge's theorem, it follows that $\mathcal{L} - K_M^*$ is a torsion class. However, $\mathcal{L} - K_M^* = 2\varepsilon$ and ε was representable by complex curves, and therefore the only option is that $\mathcal{L} = K_M^*$.

Furthermore, the holomorphic section of L_ε, since now $L_\varepsilon = \underline{\mathbb{C}}$, can only be nowhere-zero and constant, so that $\mathfrak{M}_{K_M^*}$ only contains one point. One can then verify directly that this moduli space is cut transversely, and thus that $\mathcal{SW}_M(K_M^*) = \pm 1$. (An alternative argument that $\mathcal{SW}_M(K_M^*) = \pm 1$ will be made in the next note, page 465, for symplectic manifolds in general.) □

In particular, it follows that if M' and M'' are two minimal surfaces of general type, then any diffeomorphism $f \colon M' \cong M''$ must carry the class $K_{M'}$ to $\pm K_{M''}$.

Proper elliptic surfaces.

Theorem. *If M is a proper elliptic Kähler surface, minimal with respect to blow-downs, then every Seiberg–Witten basic class of M must be a rational multiple of K_M, namely some tK_M with $t \in \mathbb{Q}$ and $|t| \leq 1$.*

Proof. We assume that M is simply-connected. As in the argument for general type surfaces, we set $\mathcal{L} = K_M^* + 2\varepsilon$ and assume that

$$\mathcal{L} \cdot [\omega] < 0 .$$

The Seiberg–Witten moduli space has *expected* dimension non-negative when $\mathcal{L} \cdot \mathcal{L} \geq K_M^* \cdot K_M^*$, but since M is elliptic we have $K_M^* \cdot K_M^* = 0$, and thus

$$\mathcal{L} \cdot \mathcal{L} \geq 0 .$$

---

[41] The Hodge signature theorem states that, if $A \cdot A > 0$ and $A \cdot B = 0$, then either $B \cdot B < 0$ or $B = 0$ or B is torsion.

If $\mathcal{SW}_M(\mathcal{L}) \neq 0$, then L_ε must admit a non-trivial holomorphic section, and therefore $K_M \cdot \varepsilon \geq 0$, which leads to $K_M^* \cdot (\mathcal{L} - K_M^*) \leq 0$. Since $K_M^* \cdot K_M^* = 0$, that means

$$K_M^* \cdot \mathcal{L} \leq 0 \,.$$

Furthermore, since M is proper elliptic, we have $K_M \cdot [\omega] > 0$, or

$$-K_M^* \cdot [\omega] > 0 \,.$$

Since $\mathcal{L} \cdot [\omega] < 0$ and $-K_M^* \cdot [\omega] > 0$, there must exist some real $t \geq 0$ so that

$$(t\mathcal{L} - K_M^*) \cdot [\omega] = 0 \,.$$

From the Hodge signature theorem, this implies that either $(t\mathcal{L} - K_M^*) \cdot (t\mathcal{L} - K_M^*) < 0$ or that $t\mathcal{L} - K_M^* = 0$.

Assume first that $(t\mathcal{L} - K_M^*) \cdot (t\mathcal{L} - K_M^*) < 0$, and so

$$t^2(\mathcal{L} \cdot \mathcal{L}) - 2t(K_M^* \cdot \mathcal{L}) < 0 \,.$$

However, $K_M^* \cdot \mathcal{L} \leq 0$ and so $-2t(K_M^* \cdot \mathcal{L}) \geq 0$. The above inequality then forces $\mathcal{L} \cdot \mathcal{L} < 0$, but the dimension requirement on the moduli space had $\mathcal{L} \cdot \mathcal{L} \geq 0$, and thus this cannot happen.

As an alternative argument, consider the function $s^2(\mathcal{L} \cdot \mathcal{L}) - 2s(K_M^* \cdot \mathcal{L})$ *in the variable* s. *The minimum of this function is* $-(K_M^* \cdot \mathcal{L})^2/(\mathcal{L} \cdot \mathcal{L})$. *Since* $K_M^* \cdot \mathcal{L} \leq 0$, *if* $\mathcal{L} \cdot \mathcal{L} < 0$, *this function never drops below* 0.

Therefore we must have $t\mathcal{L} - K_M^* = 0$. Since both $\mathcal{L}$ and K_M^* are integral classes, this coefficient t must be rational.

Now write $t\mathcal{L} - K_M^*$ as $t(K_M^* + 2\varepsilon) - K_M^* = (t-1)K_M^* + 2t\varepsilon$. Since $(t\mathcal{L} - K_M^*) \cdot [\omega] = 0$, this implies that

$$(t-1)(K_M^* \cdot [\omega]) = -2t(\varepsilon \cdot [\omega]) \,.$$

However, on one hand L_ε admits a non-trivial section and thus either $\varepsilon \cdot [\omega] > 0$ or $\varepsilon = 0$. On the other hand, $K_M^* \cdot [\omega] > 0$. Therefore we must have $t - 1 \leq 0$, that is, $0 \leq t \leq 1$. □

Remember that the canonical class of the elliptic surface $E(n)_{p,q}$ is

$$K_{E(n)_{p,q}} = (n-2)[F] + (p-1)[F_p] + (q-1)[F_q] \,,$$

where F is the generic fiber and F_p, F_q are the multiple fibers. Using the homology class $[F_{pq}] = \frac{1}{pq}[F]$, this can be written

$$K_{E(n)_{p,q}} = (npq - p - q)[F_{pq}] \,.$$

Therefore the above theorem allows as basic classes any $k[F_{pq}]$ with $|k| \leq npq - p - q$. Compare this with the actual basic classes listed in section 10.6 (page 413), where the restriction $k \equiv npq - p - q \pmod 2$ is added.

Note: Seiberg–Witten on symplectic manifolds

We will present here only the case of the canonical spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J$ with determinant line bundle K_M^* and prove the simplest result, namely that

$$\mathcal{SW}_M(\pm K_M^*) = \pm 1 .$$

As mentioned before, the proof of Taubes' general results on symplectic manifolds needs the dense 400 pages of ***Seiberg–Witten and Gromov invariants for symplectic 4–manifolds*** **[Tau00a]**. We will use notations and concepts from the endnotes of the preceding chapter (holomorphic bundles and connections, page 365), as well as from the preceding note (which started on page 457).

A simple attempt to make the argument used in the above Kähler case work in the symplectic case fails immediately. That is owing to the role now taken in many formulae by the **Nijenhuis tensor** $\mathcal{N}$. This tensor is defined by

$$\mathcal{N}(x,y) = \tfrac{1}{4}\big([Jx,Jy] - [x,y] - J[x,Jy] - J[Jx,y]\big)$$

and measures the failure of the almost-complex structure J from being integrable, *i.e.*, from corresponding to actual complex-holomorphic coordinates[42] on M. The tensor $\mathcal{N}$ appears for example in formulae like $\bar\partial\,\bar\partial f = -(\partial f)(\mathcal{N})$.

Therefore, we need a new approach for dealing with the Seiberg–Witten equations on symplectic 4–manifolds. This approach is to perturb the Seiberg–Witten equations to look like

$$\begin{cases} \mathcal{D}^A\,\varphi = 0 \\ F_A^+ - F_{A_0}^+ = \sigma(\varphi) - \rho^2\omega \,, \end{cases}$$

where A_0 is a special connection on K^*, while $\rho \in \mathbb{R}$ is a parameter that we will grow to ∞. We will show that, as ρ grows, the equations admit exactly one solution up to gauge, and therefore that $\mathcal{SW}_M(K^*) = \pm 1$.

Preparation: Almost-complex geometry. We choose an almost-complex structure J compatible with our symplectic structure ω. Since M is merely symplectic, its Nijenhuis tensor $\mathcal{N}$ does not vanish. We also choose a compatible Riemannian metric.

Complex-valued exterior forms can still be split into types. For example,

$$\Lambda^1 \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1} \,,$$

where $\Lambda^{1,0}$ is made of those complex-valued 1–forms that are complex-linear (as maps $T_M \to \mathbb{C}$), while $\Lambda^{0,1}$ contains the complex-anti-linear ones. Also,

$$\Lambda^2 \otimes \mathbb{C} = \Lambda^{2,0} \oplus \Lambda^{1,1} \oplus \Lambda^{0,2} \,,$$

---

[42] The celebrated *Newlander–Nirenberg theorem* states that an almost-complex structure J on a manifold X corresponds to an actual complex structure on X if and only if the Nijenhuis tensor $\mathcal{N}$ vanishes. A proof under the simplifying assumption that everything is real-analytic can be read from **S. Kobayashi** and **K. Nomizu**'s classic ***Foundations of differential geometry*** **[KN69, KN96**, vol II, app 8] . For the general Newlander–Nirenberg theorem, see the original paper of **A. Newlander** and **J. Nirenberg** ***Complex analytic coordinates in almost complex manifolds*** **[NN57]**, or, in monographs, **L. Hörmander**'s ***An introduction to complex analysis in several variables*** **[Hör66, Hör90]**, or **G. Folland** and **J. Kohn**'s ***The Neumann problem for the Cauchy–Riemann complex*** **[FK72]**.

where $\Lambda^{1,1}$ contains all 2–forms α with $\alpha(Jx,Jy) = \alpha(x,y)$, while $\Lambda^{0,2}$ contains complex-bi-anti-linear 2–forms.

The definitions of the $\Lambda^{p,q}$*'s cannot use* dz*'s and* $d\bar{z}$*'s, since now there are no such complex coordinates on* M*. Instead, one can start with any orthonormal local frame* $\{e_1, f_1, e_2, f_2\}$ *in* T_M *such that* $Je_1 = f_1$ *and* $Je_2 = f_2$*. This induces a dual frame* $\{e^1, f^1, e^2, f^2\}$ *in* T_M^**, which defines the complex frame* $\{\varepsilon^1 = e^1 + if^1,\ \varepsilon^2 = e^2 + if^2,\ \bar{\varepsilon}^1 = e^1 - if^1,\ \bar{\varepsilon}^2 = e^2 - if^2\}$ *in* $T_M^* \otimes \mathbb{C}$*, which is to be used as proxy for* $\{dz_1, dz_2, d\bar{z}_1, d\bar{z}_2\}$*. Then* $\Lambda^{p,q}$ *is defined as the part of* $\Lambda^{p+q}(T_M^*) \otimes \mathbb{C}$ *that is locally spanned by* p *of the* ε^k*'s and* q *of the* $\bar{\varepsilon}^k$*'s. In particular,* $\Lambda^{0,1}$ *is the complex span* $\mathbb{C}\{\bar{\varepsilon}^1, \bar{\varepsilon}^2\}$*, while* $\Lambda^{0,2}$ *is* $\mathbb{C}\{\bar{\varepsilon}^1 \wedge \bar{\varepsilon}^2\}$*.*

The first exterior derivative $d\colon \Gamma(\Lambda^0 \otimes \mathbb{C}) \to \Gamma(\Lambda^1 \otimes \mathbb{C})$ still splits as $d = \partial + \bar\partial$ with $\bar\partial\colon \Gamma(\Lambda^{0,0}) \to \Gamma(\Lambda^{0,1})$. However, on higher-degree forms the exterior derivative d is no longer exhausted by the corresponding sum $\partial + \bar\partial$. For example, for every $\alpha \in \Gamma(\Lambda^{0,1})$ we get $d\alpha = \alpha \circ \mathcal{N} + \partial\alpha + \bar\partial\alpha$, with $\alpha \circ \mathcal{N} \in \Gamma(\Lambda^{2,0})$.

Also, as mentioned above, we no longer have $\bar\partial\,\bar\partial f = 0$, but instead

$$\bar\partial\,\bar\partial f = -(\partial f) \circ \mathcal{N}\,.$$

This formula can be checked by direct computation. More generally, when dealing with sections of some bundle E endowed with a connection A, we have

$$\bar\partial_A\,\bar\partial_A \ell = F_A^{0,2} - (\partial_A \ell) \circ \mathcal{N}\,.$$

In what follows, we will also need:

Lemma *(Weitzenböck-type formula for* $\bar\partial_A$*). For every symplectic manifold* M *and every smooth complex bundle* E *on* M *endowed with a connection* A*, we have*

$$2\,\bar\partial_A^*\,\bar\partial_A \ell = d_A^* d_A \ell - i\langle F_A, \omega\rangle \cdot \ell\,.$$

Proof. We remark that for every 1–form α we have $J\alpha = *(\omega \wedge \alpha)$. Then we compute directly:

$$\begin{aligned}
2\,\bar\partial_A^*\,\bar\partial_A &= 2\,\bar\partial_A^*\,(\mathrm{proj}_{0,1} \circ d_A) = \bar\partial_A^*\,(d_A - iJd_A)\\
&= \bar\partial_A^*\,(d_A - i*(\omega \wedge d_A)) = d_A^*\,(d_A - i*(\omega \wedge d_A))\\
&= d_A^* d_A - d_A^*\,i*(\omega \wedge d_A) = d_A^* d_A - (-*d_A*)\,i*(\omega \wedge d_A)\\
&= d_A^* d_A + i*d_A**(\omega \wedge d_A) = d_A^* d_A - i*d_A(\omega \wedge d_A)\\
&= d_A^* d_A - i*(d\omega \wedge d_A + \omega \wedge d_A d_A) = d_A^* d_A - i*(\omega \wedge F_A)\\
&= d_A^* d_A - i*(\omega \wedge F_A) = d_A^* d_A - i\langle\omega, F_A\rangle\,.
\end{aligned}$$ □

This formula becomes false on merely almost-complex manifolds, where one needs to add the correction term $i\langle d^*\omega, d_A\ell\rangle$.

Rewriting and perturbing the equations. The canonical structure $\mathfrak{s}_J$ has bundles

$$\mathcal{W}^+ = \Lambda^{0,0} \oplus \Lambda^{0,2} \qquad \text{and} \qquad \mathcal{W}^- = \Lambda^{0,1}$$

and determinant line bundle $\mathcal{L} = K_M^*$. Further, one can still find a special $U(1)$–connection A_0 on K_M^* such that the induced Dirac operator is

$$\mathcal{D}^{A_0}\colon \Gamma(\Lambda^{0,0} \oplus \Lambda^{0,2}) \longrightarrow \Gamma(\Lambda^{0,1}) \qquad \tfrac{1}{\sqrt{2}}\mathcal{D}^{A_0} = \bar\partial \oplus \bar\partial^*\,.$$

The existence of such an A_0 needs that ω be symplectic.

***Proof that** $\mathcal{D}^{A_0} = \sqrt{2}(\bar\partial + \bar\partial^*)$. We follow the same thread as in the Kähler-case proof from the inserted note on page 457. Pick a unit-length section $\mathfrak{z}$ in $\mathcal{W}^+$ and choose the unique connection A_0 on K^* so that $\langle \nabla^{A_0}\mathfrak{z}, \mathfrak{z}\rangle_{\mathbb{C}} = 0$. Then, since $\omega \bullet \mathfrak{z} = -2i\mathfrak{z}$ and $\omega \bullet (\nabla^{A_0}\mathfrak{z}) = +2i\nabla^{A_0}\mathfrak{z}$, from the formula*

$$\nabla^{A_0}(\omega \bullet \mathfrak{z}) = (\nabla\omega) \bullet \mathfrak{z} + \omega \bullet (\nabla^{A_0}\mathfrak{z})$$

follows that $\nabla^{A_0}\mathfrak{z} = \frac{i}{4}(\nabla\omega) \bullet \mathfrak{z}$. Identifying $f + \beta$ from $\Lambda^{0,0} \oplus \Lambda^{0,2}$ with $(f + \frac{1}{2}\beta) \bullet \mathfrak{z}$ from $\mathcal{W}^+$, we can now compute

$$\begin{aligned} \nabla^{A_0}_X(f+\beta) \;=\; \nabla^{A_0}_X((f+\tfrac{1}{2}\beta)\bullet\mathfrak{z}) \;&=\; (df(X) + \tfrac{1}{2}\nabla_X\beta) \bullet \mathfrak{z} \;+\; (f + \tfrac{1}{2}\beta) \bullet \nabla^{A_0}\mathfrak{z} \\ &=\; (df(X) \;+\; \tfrac{1}{2}\nabla_X\beta \;+\; \tfrac{i}{4}f \cdot \nabla_X\omega \;+\; \tfrac{i}{8}\beta \bullet \nabla_X\omega) \bullet \mathfrak{z} \;. \end{aligned}$$

Since ω has constant length, $\nabla_X\omega \perp \omega$ and $\nabla_X\omega \in \Gamma(K^*) \subset \Gamma(\Lambda^2_+)$. Further, as $\beta \in \Gamma(\Lambda^{0,2})$, a quick Clifford computation in coordinates[43] *shows that $\nabla_X\omega \bullet \beta \bullet \mathfrak{z} = 0$. We are left with*

$$\nabla^{A_0}_X(f+\beta) = (df(X) \;+\; \tfrac{1}{2}\nabla_X\beta \;+\; \tfrac{i}{4}f \cdot \nabla_X\omega) \bullet \mathfrak{z} \;.$$

The corresponding Dirac operator is then

$$\begin{aligned} \mathcal{D}^{A_0}(f+\beta) \;=\; \textstyle\sum e_k \bullet \nabla^{A_0}_{e_k}(f+\beta) \;&=\; \textstyle\sum(df(e_k)e_k \;+\; \tfrac{1}{2}e_k \bullet \nabla_X\beta \;+\; \tfrac{i}{4}f \cdot e_k \bullet \nabla_X\omega) \bullet \mathfrak{z} \\ &=\; (df \;+\; d^*\beta \;+\; \tfrac{i}{2}f \cdot d^*\omega) \bullet \mathfrak{z} \;. \end{aligned}$$

However, ω is a symplectic form and thus $d^*\omega = 0$. We are left with $\mathcal{D}^{A_0}(f+\beta) = (df + d^*\beta) \bullet \mathfrak{z}$, which, after using the identification of α from $\Lambda^{0,1}$ with $\frac{1}{\sqrt{2}}\alpha \bullet \mathfrak{z}$ from $\mathcal{W}^-$, yields the formula $\mathcal{D}^{A_0}(f+\beta) = \sqrt{2}(\bar\partial f + \bar\partial^*\beta)$. □

Just as in the Kähler case, we parametrize all connections A on K^* as $A = A_0 + 2B$, with B a connection on the trivial line bundle $L_\varepsilon = \underline{\mathbb{C}}$. We then have $F_A = F_{A_0} + 2F_B$ and the equations then look like

$$\begin{cases} \bar\partial_B f \;+\; \bar\partial_B^*\beta = 0 \\ F_{A_0}^{0,2} + 2F_B^{0,2} = 2\bar f \cdot \beta \\ (F_{A_0}^+)^{1,1} + 2(F_B^+)^{1,1} = i(|f|^2 - |\beta|^2)\omega\,, \end{cases}$$

where $\varphi = (f, \beta)$ is a spinor field made of a function $f \colon M \to \mathbb{C}$ and a $(0,2)$–form $\beta \in \Gamma(\Lambda^{0,2})$, while B is a $U(1)$–connection on the trivial bundle $\underline{\mathbb{C}}$.

Further, as was seen in a previous note (page 450), the Seiberg–Witten invariants are unchanged by suitable perturbations that change the equation $F_A^+ = \sigma(\varphi)$ to $F_A^+ = \sigma(\varphi) + \eta^+$. In our case, we use the equation

$$F_A^+ = \sigma(\varphi) \;+\; F_{A_0}^+ \;-\; i\rho^2\omega$$

for some scalar ρ. The second equation above then becomes $F_B^{0,2} = \bar f \cdot \beta$, while the third is $(F_B^+)^{1,1} = \frac{i}{2}(|f|^2 - |\beta|^2 - \rho^2)\,\omega$, which we rewrite as

$$i\langle F_B, \omega\rangle = |\beta|^2 - |f|^2 + \rho^2\;.$$

In conclusion, the ρ–perturbed Seiberg–Witten equations are now:

$$\begin{cases} \bar\partial_B f \;+\; \bar\partial_B^*\beta = 0 \\ F_B^{0,2} = \bar f \cdot \beta \\ i\langle F_B, \omega\rangle = |\beta|^2 - |f|^2 + \rho^2 \end{cases}$$

---

[43] That is, one can show that, for every $\mu \in \Gamma(\Lambda^2_+)$ with $\mu \perp \omega$, every complex $\eta \in \Gamma(\Lambda^{0,2})$ and every self-dual $\varphi \in \Gamma(\mathcal{W}^+)$, one gets $\eta \bullet \mu \bullet \varphi = 0$.

and we are ready to start working on them.

Computing. We start with the aforementioned Weitzenböck-type formula

$$2\,\bar{\partial}_B^*\,\bar{\partial}_B = d_B^* d_B - i\langle F_B,\,\omega\rangle\;.$$

We apply it to $f\colon M \to \mathbb{C}$ and then use the second equation $i\langle F_B,\,\omega\rangle = |\beta|^2 - |f|^2 + \rho^2$, obtaining

$$d_B^* d_B f = 2\,\bar{\partial}_B^*\,\bar{\partial}_B f + \left(|\beta|^2 - |f|^2 + \rho^2\right)\cdot f\;.$$

We take the inner product with f and integrate over M:

$$\int_M |d_B f|^2 = 2\int_M \langle \partial_B^*\,\bar{\partial}_B f,\; f\rangle + \int_M \left(|\beta|^2 - |f|^2 + \rho^2\right)|f|^2\;.$$

Then we use the first equation $\partial_B f = -\,\bar{\partial}_B^*\,\beta$ to get:

$$\int_M |d_B f|^2 = -2\int_M \langle \beta,\;\; \bar{\partial}_B\,\bar{\partial}_B f\rangle + \int_M \left(|\beta|^2 - |f|^2 + \rho^2\right)|f|^2\;.$$

However, $\bar{\partial}_B\,\bar{\partial}_B f = F_B^{0,2}\cdot f - (\partial_B f)\circ\mathcal{N}$, and using the second equation yields:

$$\int_M |d_B f|^2 = -2\int_M \langle \beta,\;\; |f|^2\beta - (\partial_B f)\circ\mathcal{N}\rangle + \int_M \left(|\beta|^2 - |f|^2 + \rho^2\right)|f|^2\;,$$

which can then be rearranged to:

$$\int_M |d_B f|^2 = \int_M 2\langle \beta,\; (\partial_B f)\circ\mathcal{N}\rangle - |f|^2|\beta|^2 - \left(|f|^2 - \rho^2\right)^2 - \rho^2\left(|f|^2 - \rho^2\right).$$

Crucial step. We have

$$\int_M \langle F_B,\,\omega\rangle = 0\;.$$

Indeed, $\int_M \langle F_B,\,\omega\rangle = \int_M F_B \wedge *\,\omega = \int_M F_B \wedge \omega = 0$ because $[F_B] = -2\pi i c_1(\underline{\mathbb{C}})$ and thus is zero, while ω is closed and therefore $\int_M F_B \wedge \omega = [F_B]\cdot[\omega] = 0$.

The third equation states that $i\langle F_B,\,\omega\rangle = |\beta|^2 - |f|^2 + \rho^2$, and so we get

$$\int_M \left(|f|^2 - \rho^2\right) = \int_M |\beta|^2\;.$$

More computing. Picking up where we left off, we have

$$\int_M |d_B f|^2 = 2\int_M \langle \beta,\; (\partial_B f)\circ\mathcal{N}\rangle - \int_M |f|^2|\beta|^2 - \int_M \left(|f|^2 - \rho^2\right)^2 - \rho^2\int_M |\beta|^2\;.$$

Moving the terms around we get

$$\int_M \left(|d_B f|^2 + |f|^2|\beta|^2 + \left(|f|^2 - \rho^2\right)^2 + \rho^2|\beta|^2\right) = 2\int_M \langle \beta,\; (\partial_B f)\circ\mathcal{N}\rangle\;,$$

where all the terms on the left are now non-negative.

Since M is compact and $\partial_B f$ is part of $d_B f$, we see that the mysterious term $\langle \beta,\; (\partial_B f)\circ\mathcal{N}\rangle$ can be bounded above by

$$\langle \beta,\; (\partial_B f)\circ\mathcal{N}\rangle \;\leq\; C\cdot|\beta|\cdot|d_B f|$$

for some constant C that depends on $\mathcal{N}$ and the geometry of M, but on neither B nor f. Using the happy inequality $ab \leq \frac{1}{4}a^2 + b^2$, this implies that

$$\langle \beta,\; (\partial_B f)\circ\mathcal{N}\rangle \;\leq\; \tfrac{1}{4}\,|d_B f|^2 + C^2\,|\beta|^2\;.$$

Still more computing. We were saying:

$$\int_M |d_B f|^2 + |f|^2 |\beta|^2 + (|f|^2 - \rho^2)^2 + \rho^2 |\beta|^2 = 2\int_M \langle \beta, (\partial_B f) \circ \mathcal{N} \rangle .$$

Using the bound above and expressing everything in terms of L^2–norms, we get:

$$\|d_B f\|^2 + \|f\|^2 \|\beta\|^2 + \left\| |f|^2 - \rho^2 \right\|^2 + \rho^2 \|\beta\|^2 \leq \tfrac{1}{2} \|d_B f\|^2 + 2C^2 \|\beta\|^2 .$$

We can move terms around to get

$$\tfrac{1}{2} \|d_B f\|^2 + \|f\|^2 \|\beta\|^2 + \left\| |f|^2 - \rho^2 \right\|^2 \leq (2C^2 - \rho^2) \|\beta\|^2 .$$

If we push $\rho \longrightarrow \infty$ (in fact as soon as $\rho^2 > 2C^2$), we see that we must have

$$\tfrac{1}{2} \|d_B f\| = 0 \qquad \text{and} \qquad \left\| |f|^2 - \rho^2 \right\| = 0 \qquad \text{and} \qquad \|\beta\| = 0 ,$$

which implies that $|f| \equiv \rho$, and thus (after suitably rotating by a change of gauge) that f must be constant. Since $d_B f \equiv 0$, the only possible choice for the connection B on $\underline{\mathbb{C}}$ must then be the flat connection $d_B = d$.

In conclusion, for ρ big enough, the *only* Seiberg–Witten solution (f, β, B) for the canonical spin$^{\mathbb{C}}$ structure $\mathfrak{s}_J$ coincides, up to changes of gauge, with $(\rho, 0, d)$.

Conclusion. We have proved Taubes' simplest result:

Theorem (*C. Taubes*). *For every symplectic manifold M with $b_2^+ \geq 2$, we must have*

$$\mathcal{SW}_M(\pm K_M^*) = \pm 1 ,$$

where $K_M^* = c_1(\omega)$ is the anti-canonical class of M. □

This result was proved in **C. Taubes'** ***The Seiberg–Witten invariants and symplectic forms*** **[Tau94]**, by using a more sophisticated perturbation involving the Nijenhuis tensor.

Bounds on the other basic classes. Following essentially the same technique, we can also prove that all other basic classes $K^* + 2\varepsilon$ must satisfy the bounds

$$0 \leq \varepsilon \cdot [\omega] \leq -K^* \cdot [\omega] .$$

Indeed, for the other spin$^{\mathbb{C}}$ structures $K^* + 2\varepsilon$, the spinor bundles are $\mathcal{W}^+ = L_\varepsilon \oplus (\Lambda^{0,2} \otimes L_\varepsilon)$ and $\mathcal{W}^- = \Lambda^{0,1} \otimes L_\varepsilon$. Then we can write the Seiberg–Witten equations just as in the previous case,

$$\begin{cases} \bar\partial_B \ell + \bar\partial_B^* \beta = 0 \\ F_B^{0,2} = \ell^* \beta \\ i\langle F_B, \omega \rangle = |\beta|^2 - |\ell|^2 + \rho^2 \end{cases}$$

but now, instead of $f : M \longrightarrow \mathbb{C}$, we have $\ell \in \Gamma(L_\varepsilon)$; instead of $\beta \in \Gamma(\Lambda^{0,2})$, we take a L_ε–valued form $\beta \in \Gamma(\Lambda^{0,2} \otimes L_\varepsilon)$, and B is a unitary connection on L_ε.

Following essentially the same computation up to the "crucial step", we obtain

$$\int_M |d_B \ell|^2 = \int_M 2\langle \beta, (\partial_B \ell) \circ \mathcal{N} \rangle - |\ell|^2 |\beta|^2 - (|\ell|^2 - \rho^2)^2 - \rho^2(|\ell|^2 - \rho^2) .$$

At the crucial step, on one hand we have $\int\langle F_B, \omega\rangle = -2\pi i(\varepsilon\cdot[\omega])$ (since F_B is the curvature of L_ε and so $[F_B] = -2\pi i\varepsilon$) and, on the other hand, $i\langle F_B, \omega\rangle = |\beta|^2 - |\ell|^2 + \rho^2$ (from the third equation). Therefore

$$\int_M (|\ell|^2 - \rho^2) = \int_M |\beta|^2 - 2\pi(\varepsilon\cdot[\omega])\,,$$

which, plugged into our previous computation, yields

$$\int_M |d_B\ell|^2 = 2\int_M \langle\beta,\ (\partial_B\ell)\circ\mathcal{N}\rangle - \int_M |\ell|^2\,|\beta|^2 - \int_M (|\ell|^2-\rho^2)^2 - \rho^2\int_M |\beta|^2 + \rho^2\cdot 2\pi(\varepsilon\cdot[\omega])\,.$$

Proceeding with the same game plan, we get

$$\int_M \big(|d_B\ell|^2 + |\ell|^2\,|\beta|^2 + (|\ell|^2-\rho^2)^2 + \rho^2\,|\beta|^2\big) - 2\pi\rho^2(\varepsilon\cdot[\omega]) = 2\int_M \langle\beta,\ (\partial_B\ell)\circ\mathcal{N}\rangle \le \tfrac{1}{2}\,|d_B\ell|^2 + 2C^2\,|\beta|^2$$

and therefore

$$\tfrac{1}{2}\,\|d_B\ell\|^2 + \|\ell\|^2\,\|\beta\|^2 + \big\||\ell|^2-\rho^2\big\|^2 - 2\pi\rho^2(\varepsilon\cdot[\omega]) \le (2C^2-\rho^2)\,\|\beta\|^2\,.$$

As soon as $\rho^2 > 2C^2$, the term on the right is negative or zero. Therefore the sum on the left must be non-positive as well. Since the only term on the left that can be negative is $-2\pi\rho^2(\varepsilon\cdot[\omega])$, it follows that we must have

$$\varepsilon\cdot[\omega] \ge 0\,,$$

which is half of what we set out to prove.

Finally, if $\mathfrak{s}$ is a basic class, then so must $-\mathfrak{s}$. In other words, if $K^* + 2\varepsilon$ is a basic class, then so must be $-(K^*+2\varepsilon) = K^* + 2(K-\varepsilon)$. Therefore we can repeat the whole argument above with $K-\varepsilon$ instead of ε and end up with

$$(K-\varepsilon)\cdot[\omega] \ge 0\,.$$

This quickly rearranges as $K\cdot[\omega] \ge \varepsilon\cdot[\omega]$ or $-K^*\cdot[\omega] \ge \varepsilon\cdot[\omega]$. Therefore

$$0 \le \varepsilon\cdot[\omega] \le -K^*\cdot[\omega]$$

and we are done.

This latter result was proved in **C. Taubes**' ***More constraints on symplectic forms from Seiberg–Witten invariants*** **[Tau95a]**.

The analysis of the Seiberg–Witten equations on symplectic manifolds does not stop here. Essentially the same perturbation, but very delicate arguments and estimates, is used to show that a solution to the equations corresponds, in the limit $\rho \to \infty$, to holomorphic sections of a line bundle, thus making apparent the relation with the J–holomorphic curves of M. Glance through **C. Taubes**' papers, collected in the volume ***Seiberg–Witten and Gromov invariants for symplectic 4–manifolds*** **[Tau00a]**, for a feel. For a softer introduction, read **M. Hutchings** and **C. Taubes**'s ***An introduction to the Seiberg–Witten equations on symplectic manifolds*** **[HT99]**.

Note: The Gromov–Taubes invariants of symplectic 4–manifolds

The best way to state C. Taubes' general theorem on Seiberg–Witten theory and J–holomorphic curves is to relate the Seiberg–Witten invariants to certain so-called Gromov–Taubes invariants that count J–holomorphic curves in symplectic manifolds. In what follows, we will explain the setting of these invariants.

Consider a symplectic 4–manifold M, endowed with a generic compatible almost-complex structure J. We will look at the J–holomorphic curves of M.

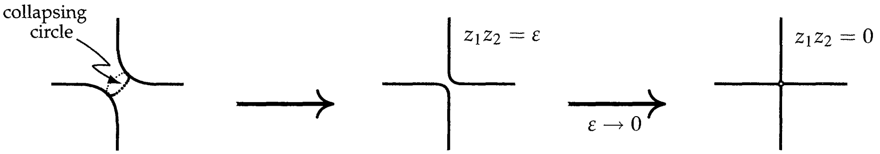
Compactness. First off, notice that a J–holomorphic curve S in a symplectic manifold is never homologically-trivial: indeed $[S] \cdot [\omega] = \int_S \omega > 0$. It is also worth noting the following fundamental result, which is a consequence of the fact that the area of a J–holomorphic curves S is exactly[44] $\int_S \omega$:

Gromov's Compactness Theorem. *Every sequence $\{f_n \colon S \to M\}$ of J–holomorphic curves has a subsequence that converges to some J–holomorphic $f \colon S^* \to M$ that might have nodal singularities (transverse double-points).* □

Here we think of S as a real surface; the map $f_n \colon S \to M$ is called J–holomorphic if the tangent bundle of $f_n[S]$ is J–invariant. Such a map is allowed to have singularities, just like a complex curve.[45] For each n, the surface S inherits a complex structure, pulled-back through[46] f_n, so that $df_n(iv) = J(df_n(v))$.

The limit-surface S^* is obtained from S by collapsing embedded circles (*à la* vanishing cycles). The nodal singularities of the limit appear in M in the same manner as the one suggested in figure 10.15: the collapse of $z_1 z_2 = \varepsilon$ to $z_1 z_2 = 0$.

The evolution of fibers in an elliptic surface are good examples to exercise one's understanding of the compactness theorem,[47] and help can be received from **R. Kirby** and **P. Melvin**'s ***The E_8–manifold, singular fibers and handlebody decompositions*** **[KM99]**.



10.15. Apparition of singularities

---

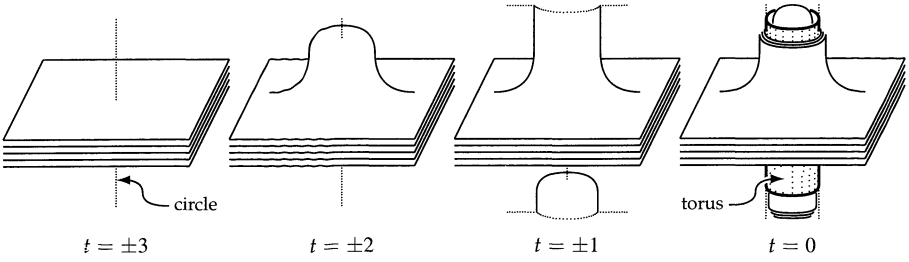
[44] One can generalize Gromov's theorem to general almost-complex manifolds by adding the requirement that the sequence f_n be bounded in area.

[45] Indeed, **D. McDuff** proved in ***Singularities of J–holomorphic curves in almost complex 4–manifolds*** **[McD92]** that the singularities that can appear have the same topology as those of a complex curve inside a complex surface.

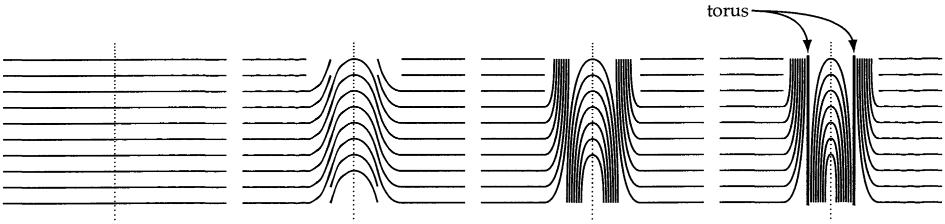
[46] Thus, the induced complex structures on S vary with n, depending on the map f_n. If S had a fixed complex structure that all f_n's were required to respect, then the nodes of the limit could only separate sphere-components ("bubbling"); in the statement above, though, for example, a torus can evolve toward a fishtail.

[47] Example: how does Gromov's theorem account for a generic torus fiber approaching a cusp fiber?

Gromov's theorem as stated in general fails for non-symplectic almost-complex structures. An example (essentially due to Y. Eliashberg) of an almost-complex structure where a sequence of J–holomorphic curves has no decent limit is suggested in figures 10.16 and 10.17 (fourth dimension in time slices, all surfaces J–holomorphic).



10.16. Bad almost-complex structure, I



10.17. Bad almost-complex structure, II

Gromov–Taubes invariants. Given a class $\alpha \in H_2(M;\mathbb{Z})$, we define

$$\mathfrak{H}_\alpha$$

to be the space of all J–holomorphic curves representing the class α in M. It is proved that, for a *generic* almost-complex structure J, the space $\mathfrak{H}_\alpha$ is in fact a *compact* smooth *oriented* manifold, with dimension

$$\dim \mathfrak{H}_\alpha = c_1(\omega) \cdot \alpha + \alpha \cdot \alpha \,.$$

(Note that "generic J" might not include "the obvious choice of J".) The strategy follows the broad outlines sketched for gauge theory in section 9.1 (page 332), and the techniques are analogous to the ones used on the Seiberg–Witten moduli space in the previous note on page 439.

From the complex adjunction formula, all the elements of $\mathfrak{H}_\alpha$ are surfaces with

$$\chi(S) = c_1(\omega) \cdot \alpha - \alpha \cdot \alpha$$

and *a priori* they might have singularities. A fundamental remark is that the curves in $\mathfrak{H}_\alpha$ can be *disconnected*, and thus this theory is at the outset distinct from Gromov–Witten theory[48] (where all curves are connected).

---

[48] Nonetheless, the Gromov–Taubes invariants are computable from the Gromov–Witten invariants, see **E. Ionel** and **T. Parker**'s ***The Gromov invariants of Ruan–Tian and Taubes*** **[IP97]**.

To reduce the dimension of $\mathfrak{H}_\alpha$ when $d = \dim \mathfrak{H}_\alpha$ is positive, we can choose $d/2$ random points in M (d is always even) and restrict to the curves from $\mathfrak{H}_\alpha$ that pass through these $d/2$ points. Denoting this space by

$$\mathfrak{H}_\alpha^0 ,$$

we have that, for a suitably generic J and generic choice of points, $\mathfrak{H}_\alpha^0$ is an oriented finite 0–dimensional manifold, and thus its points can be counted (with signs). The result of this count is the **Gromov–Taubes invariant**

$$\mathrm{Gr}(\alpha) = \#\mathfrak{H}_\alpha^0 .$$

Generically, $\mathfrak{H}_\alpha^0$ contains no singular curves. If M is Kähler, then all points of $\mathfrak{H}_\alpha^0$ have positive orientation.

In order to make this a genuine symplectic invariant, one wishes to achieve invariance under isotopies of symplectic forms. This creates problems that are solved through a very delicate and peculiar count of the tori from $\mathfrak{H}_\alpha^0$, especially the multiply-covered ones.

Relation with Seiberg–Witten theory. The full version of C. Taubes' result is:

Taubes' Theorem. *If M is symplectic and $b_2^+(M) \geq 2$, then*

$$\mathcal{SW}_M(K_M^* + 2\alpha) = \pm\, \mathrm{Gr}(\alpha) .$$

Further, if $\mathcal{SW}_M(K_M^* + 2\alpha) \neq 0$, then the dimension $d = 0$. □

Therefore, if $\mathcal{SW}_M(K_M^* + 2\alpha) \neq 0$, then no fixed points need to be chosen, and $Gr(\alpha)$ just counts all the curves representing α.

Since $\dim \mathfrak{M}_{K_M^*+2\alpha} = \dim \mathfrak{H}_\alpha$, another consequence is that all symplectic 4–manifolds are of Seiberg–Witten simple type, *i.e.*, the nontrivial Seiberg–Witten invariants occur only for almost-complex spin$^{\mathbb{C}}$ structures.

References. **C. Taubes'** theorem was first announced in ***The Seiberg–Witten and Gromov invariants*** **[Tau95b]**, while its proof appeared spread through the four heavy papers **[Tau96b, Tau99b, Tau96a, Tau99a]**, which were later gathered in the volume ***Seiberg–Witten and Gromov invariants for symplectic* 4*–manifolds*** **[Tau00a]**. The Gromov–Taubes invariant was introduced in ***Counting pseudo-holomorphic submanifolds in dimension* 4** **[Tau96a]**. See also **D. McDuff**'s ***Lectures on Gromov invariants for symplectic* 4*–manifolds*** **[McD97]**.

C. Taubes also pushed this interpretation of the Seiberg–Witten invariants beyond the symplectic world, proposing that we study symplectic structures off embedded circles[49] and count J–holomorphic curves that limit on these circles, see ***Seiberg–Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic* 2*–forms*** **[Tau99c]** and ***Seiberg–Witten invariants, self-dual harmonic* 2*–forms and the Hofer–Wysocki–Zehnder formalism*** **[Tau00b]**. Technical details are contained in ***The structure of pseudo-holomorphic subvarieties for a degenerate almost complex structure and symplectic form on*** $\mathbb{S}^1 \times \mathbb{B}^3$ **[Tau98c]**, while a recent

---

[49] "Near-symplectic" structures, in the terms of the earlier note on Lefschetz pencils, page 419.

technical paper is ***A compendium of pseudoholomorphic beasts in*** $\mathbb{R} \times (S^1 \times S^2)$ [**Tau02**]. This is still a program in progress.

M. Gromov's compactness theorem appeared in the founding paper ***Pseudoholomorphic curves in symplectic manifolds*** [**Gro85**], and discussions of this theorem and other remarkable applications of J–holomorphic curves in symplectic geometry (of general dimension) are gathered in the volume **Holomorphic curves in symplectic geometry** [**AL94**], edited by **M. Audin** and **J. Lafontaine.** See also **D. McDuff** and **D. Salamon**'s ***J–holomorphic curves and quantum cohomology*** [**MS94**], or the more recent ***J–holomorphic curves and symplectic topology*** [**MS04**]. A proof of the compactness theorem is written down in **C. Hummel**'s ***Gromov's compactness theorem for pseudo-holomorphic curves*** [**Hum97**]. For a statement of the Gromov compactness theorem generalized as much as reasonable, see **S. Ivashkovich** and **V. Shevchishin**'s ***Gromov compactness theorem for stable curves*** [**IS99**].

Note: The Bochner technique

Let X be an m–manifold, endowed with a Riemannian metric. The **Bochner technique** refers to vanishing results obtained from assumptions on curvatures. They use what are known as **Weitzenböck formulae**, namely the comparison of two Laplace operators. A **Laplace operator** on a Riemannian manifold is defined as any second-order operator with symbol $-\left|\cdot\right|^2$.

Typically one of the Laplacians is the connection Laplacian $\nabla^* \nabla$ of the Levi–Cività connection on X, while the other is the Laplacian of a Dirac-type operator, for example the familiar Hodge Laplacian $\Delta = d\,d^* + d^*d$, or the Laplacian $\mathcal{D}^*\mathcal{D}$ of a Dirac operator on spinors.

What makes the comparison useful is that the difference turns out to be a zero-order operator (instead of a first-order operator). Further, the difference can be expressed in terms of curvature contributions. The flavor of the arguments leading to vanishing results is the same as of, say, the proof of Seiberg–Witten's vanishing for positive scalar curvature (page 405): a geometric hypothesis forces an analytic conclusion.

The first applications of the Bochner technique were on harmonic functions, then later on exterior harmonic forms. For example, **S. Bochner**'s ***Curvature and Betti numbers*** [**Boc48**] proved, for every 1–form α on the Riemannian manifold X^n, the formula

$$\Delta\alpha = \nabla^* \nabla\alpha + \mathcal{R}icci(\alpha)\,,$$

where Δ is the Hodge Laplacian, ∇ is the Levi–Cività connection of X, and $\mathcal{R}icci$ is the Ricci curvature. Since every class of $H^1(X;\mathbb{R})$ can be represented by a harmonic 1–form (with $\Delta\alpha = 0$), an immediate consequence is:

Theorem. *Let X be a closed Riemannian manifold. If $\mathcal{R}icci > 0$, then the first Betti number $b_1(X)$ must vanish.* □

A similar formula and argument applied to higher-degree forms, due to **S. Gallot** and **D. Meyer**'s ***Opérateur de courbure et laplacien des formes différentielles d'une variété riemannienne*** [**GM75**], yields the following:

Theorem. *Let X^m be a Riemannian closed manifold. If the curvature operator $\mathcal{R}\colon \Lambda^2 \to \Lambda^2$ is everywhere positive-definite, then all the Betti numbers $b_1(X), \ldots, b_{m-1}(X)$ must vanish, and thus X is a rational-homology m-sphere.* □

Moving on from exterior forms to spinor fields, **A. Lichnerowicz**'s ***Laplacien sur une variété riemannienne et spineurs*** [**Lic62a**] proved that, for every spinor field φ on a spin-manifold X, we have the **Lichnerowicz formula:**

$$\mathcal{D}^* \mathcal{D}\varphi = \nabla^* \nabla\varphi + \tfrac{1}{4}\, scal \cdot \varphi\,,$$

and therefore:

Theorem. *A compact spin-manifold X with everywhere-positive scalar curvature does not admit any spinor fields with $\mathcal{D}\varphi = 0$.* □

A striking consequence was obtained in **N. Hitchin**'s ***Harmonic spinors*** [**Hit74**]:

Theorem. *In every dimension $m > 8$ with $m = 1$ or $2 \pmod 8$, there are smooth manifolds Σ^m homeomorphic to the sphere $\mathbb{S}^m$, but admitting no Riemannian metrics of positive scalar curvature.* □

Of course, all standard spheres in all dimensions admit Riemannian metrics of constant positive scalar curvature. The manifolds above must be exotic spheres.

These are but a few of the applications that the Bochner technique has found in geometry. Throughout this chapter we have seen frequent uses of the *coupled* Lichnerowicz formula

$$(\mathcal{D}^A)^* \mathcal{D}^A \varphi = (\nabla^A)^* \nabla^A\varphi + \tfrac{1}{4}\, scal \cdot \varphi + \tfrac{1}{2}F_A^+ \bullet \varphi\,.$$

In a previous note (page 466), we also used a Weitzenböck-type formula for $\bar\partial$ to investigate the Seiberg–Witten equations on symplectic manifolds.

For proofs of the above statements, the source of choice is **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** [**LM89**, sec II.8]. For more applications of the Bochner technique, try **H-h. Wu**'s monograph ***The Bochner technique in differential geometry*** [**Wu88**]. The proof of the coupled Lichnerowicz formula can also be found in **J. Morgan**'s ***The Seiberg–Witten equations and applications to the topology of smooth four-manifolds*** [**Mor96**], and, for that matter, in most comprehensive expositions of Seiberg–Witten theory.

Bibliography

General sources for the Seiberg–Witten invariants are **S.K. Donaldson**'s nice survey ***The Seiberg–Witten equations and 4–manifold topology*** [**Don96a**] (which also compares Seiberg–Witten with Donaldson theory), **J. Morgan**'s introductory book ***The Seiberg–Witten equations and applications to the topology of smooth four-manifolds*** [**Mor96**], or **L. Nicolaescu**'s textbook ***Notes on Seiberg–Witten theory*** [**Nic00**].

For a better understanding of spin$^{\mathbb{C}}$ structures, one should try to first understand spin structures, and for that the inevitable reference is **B. Lawson** and **M-L. Michelson**'s ***Spin geometry*** [**LM89**].

For a proof of the unique continuation property, one can refer to the general result of **N. Aronszajn**'s ***A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order*** [Aro57], or read from **B. Booß-Bavnbek** and **K. Wojciechowski**'s ***Elliptic boundary problems for Dirac operators*** [**BBW93**, ch 8].

The Seiberg–Witten equations were written in the physics papers of **N. Seiberg** and **E. Witten** [**SW94b, SW94a, SW94c**]. The Seiberg–Witten invariants as a tool in smooth topology were introduced in **E. Witten**'s ***Monopoles and four-manifolds*** [**Wit94**], together with their interpretation on Kähler manifolds. Most of the general properties of the Seiberg–Witten invariants were already uncovered in Witten's founding paper. A proof of the general blow-up formula appeared in **R. Fintushel** and **R. Stern**'s ***Immersed spheres in 4–manifolds and the immersed Thom conjecture*** [**FS95**], together with a direct proof of Seiberg–Witten vanishing from spheres of non-negative intersection. References for the adjunction inequality will be discussed at the end of next chapter, on page 529. A different-flavored proof for the vanishing theorem for connected sums is presented in **D. Salamon**'s ***Removable singularities and a vanishing theorem for Seiberg–Witten invariants*** [**Sal96**].

A general introduction to symplectic manifolds is **D. McDuff** and **D. Salamon**'s ***Introduction to symplectic topology*** [**MS95, MS98**]. For J–holomorphic curves, one should continue with their ***J–holomorphic curves and symplectic topology*** [**MS04**], and the collection ***Holomorphic curves in symplectic geometry*** [**AL94**], edited by **M. Audin** and **J. Lafontaine.**

The various results of **C. Taubes** appeared in the papers [**Tau94, Tau95a, Tau95b, Tau96b, Tau99b, Tau96a, Tau99a**], some of which were later gathered in the book ***Seiberg–Witten and Gromov invariants for symplectic 4–manifolds*** [**Tau00a**]. A good place to start is **M. Hutchings** and **C. Taubes**' lecture notes ***An introduction to the Seiberg–Witten equations on symplectic manifolds*** [**HT99**]. Other overviews are **C. Taubes**' short ***The geometry of the Seiberg–Witten invariants*** [**Tau98a**] and his longer [**Tau98b**] (with the same title), which also include an exposition of his work on extending the J–holomorphic interpretation beyond the symplectic world. For details of the latter, see ***Seiberg–Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic 2–forms*** [**Tau99c**], and ***Seiberg–Witten invariants, self-dual harmonic 2–forms and the Hofer–Wysocki–Zehnder formalism*** [**Tau00b**]. Some technical developments are contained in ***The structure of pseudo-holomorphic subvarieties for a degenerate almost complex structure and symplectic form on*** $S^1 \times \mathbb{B}^3$ [**Tau98c**], while a recent technical paper is ***A compendium of pseudoholomorphic beasts in*** $\mathbb{R} \times (S^1 \times S^2)$ [**Tau02**]. More papers in this direction will soon be published.

The invariants of $E(n)_{p,q}$ were computed in **R. Fintushel** and **R. Stern**'s ***Rational blowdowns of smooth 4–manifolds*** [**FS97a**]. For an different approach, see **E. Witten**'s ***Monopoles and four-manifolds*** [**Wit94**] (for $b_2^+ \geq 3$), **R. Friedman** and **J. Morgan**'s ***Obstruction bundles, semiregularity, and Seiberg–Witten invariants*** [**FM99**], and **R. Brussee** ***The canonical class and the*** C^∞ ***properties of Kähler surfaces*** [**Bru96**]. See also **R. Friedman** and **J. Morgan**'s ***Algebraic surfaces and Seiberg–Witten invariants*** [**FM97**]. Compare also with the references at the end of chapter 7 (page 299) and of chapter 8 (page 322). The computation of the invariants of elliptic surfaces can also be found in **L. Nicolaescu**'s ***Notes on Seiberg–Witten theory*** [**Nic00**, sec 3.3.2]. These also follow from general gluing formulae that will be stated in section 12.1 (page 532).

It is worth noting that, in a spirit similar to the vanishing theorem for positive scalar curvature, the interaction between the Seiberg–Witten invariant and the differential geometry of 4–manifolds is very strong, as is explored in the work of **C. LeBrun.** A survey is his ***Einstein metrics, four-manifolds, and differential topology*** [**LeB03**], and one of many important papers is ***Ricci curvature, minimal volumes, and Seiberg–Witten theory*** [**LeB01**]. Most striking consequences have been obtained on the existence of Einstein metrics on 4–manifolds.

We already mentioned the cohomotopy refinement of the Seiberg–Witten invariant due to **S. Bauer** and **M. Furuta,** which (unlike $\mathcal{SW}_M$) can be used to explore connected sums. The papers are ***A stable cohomotopy refinement of Seiberg–Witten invariants: I & II*** [**BF04**], and an exposition is ***Refined Seiberg–Witten invariants*** [**Bau03**]. See also **M. Furuta**'s survey ***Finite dimensional approximations in geometry*** [**Fur02**].

The Seiberg–Witten equations can also be written on 3–manifolds, and they yield torsion invariants of the 3–manifold.[50] See **G. Meng** and **C. Taubes**' ***<u>SW</u> = Milnor torsion*** [**MT96**], then **M. Hutchings** and **Y-j. Lee**'s ***Circle-valued Morse theory, Reidemeister torsion, and Seiberg–Witten invariants of* 3*–manifolds*** [**HL99b**] and ***Circle-valued Morse theory and Reidemeister torsion*** [**HL99a**], as well as **T. Mark**'s ***Torsion, TQFT, and Seiberg–Witten invariants of* 3*–manifolds*** [**Mar02**].

These torsion invariants turn out to be the Euler characteristic of a *Seiberg–Witten–Floer homology* of the 3–manifold, a complete package of which has been built by **P. Kronheimer** and **T. Mrowka,** in a volume to appear. This is relevant to 4–manifolds because every 4–dimensional cobordism between two 3–manifolds yields a morphism between the Floer homologies of its ends, and thus can be used to compute the Seiberg–Witten invariants of closed 4–manifolds by cutting these along 3–submanifolds.[51] (This type of approach proved quite fruitful for Donaldson's instanton invariants; compare also with the Ozsváth–Szabó approach below.)

The Seiberg–Witten equations are effective but mysterious. At the moment there are two geometric interpretations of the Seiberg–Witten invariants. One of them is due to C. Taubes, who generalizes his interpretation from symplectic manifolds to all 4–manifolds: on any 4–manifold there are 2–forms that are symplectic off a few embedded circles; then the Seiberg–Witten invariant counts J–holomorphic curves in the circles' complement that limit nicely near these circles; see [**Tau99c, Tau00b**].

The other interpretation is due to P. Ozsváth and Z. Szabó *via* their Floer homology of 3–manifolds and is briefly discussed below.

---

[50] The easiest way to define the Seiberg–Witten invariants of a 3–manifold N^3 is as the Seiberg–Witten invariants of the 4–manifold $N \times S^1$. Of course, there are intrinsically 3–dimensional descriptions as well.

[51] This fundamental interaction between 3– and 4–manifold theories is the most painful omission from this volume.

Ozsváth–Szabó do the Floer. Starting with a handle decomposition of a 3–manifold N^3, P. Ozsváth and Z. Szabó look at the surface S separating the 0– and 1–handles from the 2– and 3–handles. If the genus of S is g, then the decomposition of N must have g 1–handles and g 2–handles. The belt spheres of the 1–handles are a collection of disjoint embedded circles $\alpha_1, \dots, \alpha_g$ in S, while the attaching spheres of the 2–handles are a collection of disjoint embedded circles $\beta_1, \dots, \beta_g$ in S. The surface S together with the α's and the β's offer a complete recipe for the manifold N^3 and is called a **Heegaard diagram** for N.

Ozsváth and Szabó then consider the g–fold **symmetric product** $\mathrm{Sym}^g S$ of S (defined as the quotient of the g–fold product $S \times \cdots \times S$ by the action of symmetric group on g letters; a point in $\mathrm{Sym}^g S$ is a g–tuple $\{x_1, \dots, x_g\}$ of points of S, irrespective of their order; $\mathrm{Sym}^g S$ is a smooth manifold). A complex structure on S induces a complex structure on $\mathrm{Sym}^g S$, while the α's and the β's make up g–tori $\mathbb{T}_\alpha = \alpha_1 \times \cdots \alpha_g$ and $\mathbb{T}_\beta = \beta_1 \times \cdots \times \beta_g$ that are totally-real[52] in $\mathrm{Sym}^g S$. The intersection points $\mathbb{T}_\alpha \cap \mathbb{T}_\beta$ are g–tuples $\{x_1, \dots, x_g\}$ with each x_k an intersection point of α_k and β_k in S.

After this, Ozsváth and Szabó define homology theories based on chain complexes generated by these intersection points[53] of $\mathbb{T}_\alpha$ and $\mathbb{T}_\beta$, and with boundary maps determined by counting *holomorphic* Whitney disks in $\mathrm{Sym}^g S$ between these intersection points. The various homology theories—denoted by HF^∞, HF^-, HF^+, and $\widehat{HF}$—depend on a choice of spin$^{\mathbb{C}}$ structure $\mathfrak{s}$ on N^3. They call these the **Heegaard–Floer homology** (while others call them the **Ozsváth–Szabó homology**).

A fundamental remark is that a holomorphic disk $u \colon \mathbb{D}^2 \to \mathrm{Sym}^g S$ is equivalent to a holomorphic g–fold branched cover[54] $F \to \mathbb{D}^2$ together with a holomorphic map $\widetilde{u} \colon F \to S$; with care, this allows reading these holomorphic disks (and thus the homology theories) directly from the Heegaard diagram.

Any cobordism M^4 between two 3–manifolds N' and N'' induces a morphism from the homologies of N' to the homologies of N''. This eventually leads to an invariant of closed 4–manifolds (endowed with spin$^{\mathbb{C}}$ structures and with $b_2^+ \geq 2$), which is quite analogous to the Seiberg–Witten invariant. Using this invariant, Ozsváth and Szabó were able to re-prove most results analogous to statements from Seiberg–Witten theory, including Donaldson's theorem, vanishing for connected sums, non-vanishing for symplectic manifolds, an adjunction inequality, *etc.* In fact, the Seiberg–Witten invariants and the Ozsváth–Szabó 4-dimensional invariants are conjectured to coincide. A program to prove this conjecture is proposed in **Y-J. Lee**'s ***Heegaard Splittings and Seiberg–Witten monopoles*** **[Lee04]**.

An interesting feature of the Ozsváth–Szabó invariant for 4–manifolds is that, unlike Donaldson's or Seiberg–Witten's, it is not particularly helped by the 4–manifold's admitting a complex (or symplectic[55]) structure. A drawback is that

---

[52] A submanifold A inside an (almost-)complex manifold is called **totally real** if $J[T_A] \cap T_A = \varnothing$.

[53] Twisted by integral data tracking the relative position of Whitney disks in $\pi_2(\mathrm{Sym}^g S)$ (which is $\approx \mathbb{Z}$ when $g \geq 2$); also, the intersection points are grouped after spin$^{\mathbb{C}}$ structures on N.

[54] Branched covers were defined in footnote 6 on page 308.

[55] The non-vanishing property for symplectic manifolds was proved in **[OS04c]** by using Lefschetz pencils, and not directly the symplectic form.

it has not yet been proved that the only Ozsváth–Szabó basic classes of a surface of general type are $\pm K_M^*$. An advantage, though, might be that this theory will allow a much better exploration of the far-from-complex realm than was afforded by classical gauge theories. In general, the Ozsváth–Szabó invariant of a random 4–manifold is more computable than the Seiberg–Witten invariant, the former having a somewhat combinatorial nature, but the latter is more closely related to the geometries (complex, symplectic, Riemannian) of the 4–manifold.

The foundational proofs of Ozsváth–Szabó theory are very hard and are contained in **P. Ozsváth** and **Z. Szabó**'s ***Holomorphic disks and topological invariants for closed three-manifolds*** [**OS01b**] and ***Holomorphic disks and three-manifold invariants: properties and applications*** [**OS01a**]. A good place to start is the survey ***Heegaard diagrams and holomorphic disks*** [**OS04a**], followed by [**OS01b**]. For 4–manifolds, the relevant papers are ***Holomorphic triangles and invariants for smooth four-manifolds*** [**OS01c**], with further applications in ***Holomorphic triangle invariants and the topology of symplectic four-manifolds*** [**OS04c**] and in ***Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary*** [**OS03**].

It is likely that this "combinatorial" approach to gauge theory and obtaining 4–manifold invariants by slicing along 3–manifolds will prove remarkably productive. However, we have chosen not to discuss these developments at any length in this volume. Nonetheless, this is an area that might explode in the future, and the reader should keep an open eye on it.

In any case, modulo complete proofs of the equivalence conjectures, it appears that a *same* insight into 4–manifolds emerges from *three* radically different constructions: (1) Donaldson's instantons, (2) Seiberg–Witten's monopoles, and (3) the newborn Ozsváth–Szabó Heegaard combinatorics. Combining and jumping between these versions will likely be a very fertile avenue.[56] In contrast, the reader should keep in mind the fact that gauge theory, in any of its incarnations, only helps to distinguish objects, but can never show that two creatures are diffeomorphic.

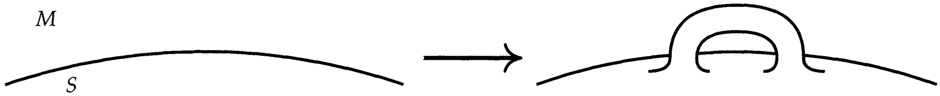
---

[56] The partially-proved equivalence between Donaldson and Seiberg–Witten has already yielded fruits: see **P. Kronheimer** and **T. Mrowka**'s ***Witten's conjecture and property P*** [**KM04**], where they prove the longstanding property-P conjecture for knots.

Chapter 11

The Minimum Genus of Embedded Surfaces

The problem of the least genus needed to represent a given homology class by embedded surfaces is the focus of this chapter. While all homology classes can be represented by smoothly embedded surfaces, the questions that arise are: How much complexity is needed? What is the minimum genus of a surface representing a given class? Can we succeed to represent it by a sphere?



11.1. Increasing genus is easy

The state of this problem before and after gauge theory offers a rather striking perspective. Before, all one dared to ask was whether a class could be represented by a sphere, and the tools were consequences of Rokhlin's theorem and various ingenious constructions.

With the advent of Seiberg–Witten theory it was shown that inside a Kähler surface the genus of a surface representing a fixed homology class is minimized by complex curves. Similarly, for symplectic manifolds the genus is minimized by J-holomorphic curves. By moving away from the complex realm, though, while one still has genus bounds involving Seiberg–Witten basic classes, it is not known when these inequalities are sharp, and the problem of determining the basic classes themselves becomes nontrivial.

We start by proving the Kervaire–Milnor generalization of Rokhlin's theorem, and show how it can be used to show that certain homology classes cannot be represented by spheres. A further generalization of Rokhlin's theorem is explained in the end-notes (page 502), and its direct proof (page 507) is in particular a complete proof of Rokhlin's theorem itself. A spin-flavored alternative proof of Rokhlin's theorem is explained in the note on page 521.

In section 11.2 (page 486) we state the Seiberg–Witten adjunction inequality for classes of positive self-intersection, which offers upper bounds on genus from basic classes; its proof is detailed in the end-notes (page 496). We continue by discussing the case of $\mathbb{CP}^2$ (the Thom conjecture), and then state the adjunction inequality for manifolds of Seiberg–Witten simple type. Since the latter can be applied to general Kähler and symplectic manifolds, it shows that complex or J–holomorphic curves minimize genus.

Finally, in section 11.3 (page 491) we take a short trip in dimension 3, where the problem of minimum genus is much better understood and intrinsically related to the theory of taut foliations.

11.1. Before gauge theory: Kervaire–Milnor

Before the advent of gauge theory, the main tool for deciding whether a homology class can be represented by a sphere was the following generalization of Rokhlin's theorem:[1]

Kervaire–Milnor Theorem. *Let M be a smooth 4–manifold. If Σ is a characteristic sphere in M, then we must have:*

$$\operatorname{sign} M - \Sigma \cdot \Sigma = 0 \pmod{16} .$$

Remember that a characteristic sphere is a sphere representing a characteristic element of M, *i.e.*, an integral lift of $w_2(T_M)$. If Σ were merely some random characteristic surface, not a sphere, then all we would know is that $\operatorname{sign} M - \Sigma \cdot \Sigma = 0 \pmod 8$.

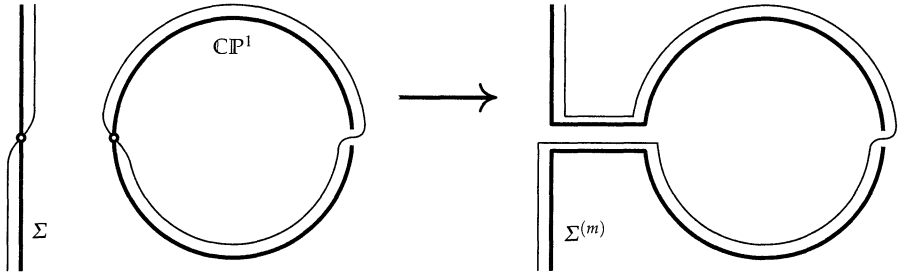
Proof. Assume given such a sphere with $\Sigma \cdot \Sigma > 0$. We write $\Sigma \cdot \Sigma = m + 1$, and connect-sum M with m copies of $\overline{\mathbb{CP}}^2$, obtaining

$$M^{(m)} = M \# m \overline{\mathbb{CP}}^2 .$$

Then we connect-sum Σ with one copy $\mathbb{CP}^1$ inside each of the m added $\overline{\mathbb{CP}}^2$'s, by using thin disjoint tubes inside $M^{(m)}$. We obtain a new *sphere* $\Sigma^{(m)}$ in $M^{(m)}$. Since each $\mathbb{CP}^1$ adds a negative self-intersection to $\Sigma^{(m)}$ (see figure 11.2), it follows that $\Sigma^{(m)}$ has self-intersection $+1$.

---

[1] Recall that Rokhlin's theorem states that, if M^4 is smooth and $w_2(T_M) = 0$, then $\operatorname{sign} M = 0 \pmod{16}$; see section 4.4 (page 170).



11.2. Joining Σ to a copy of $\mathbb{CP}^1$ for modifying its normal bundle

The boundary of a tubular neighborhood of $\Sigma^{(m)}$ in $M^{(m)}$ is a circle-bundle of Euler class $+1$ over the sphere $\Sigma^{(m)}$. It is therefore isomorphic to the Hopf bundle.[2] Hence the boundary of this tubular neighborhood must be a 3–sphere. We can then cut $\Sigma^{(m)}$, together with its surrounding neighborhood, out of $M^{(m)}$, and glue a standard 4–ball in their stead; denote the result by $M^{(m)}_{\Sigma|\mathbb{D}}$.

This means that we eliminated a characteristic surface from the manifold, and with it, the class $w_2(M)$. Therefore $w_2(M^{(m)}_{\Sigma|\mathbb{D}}) = 0$, as there is now no obstruction to the existence of a spin structure on this surgered $M^{(m)}_{\Sigma|\mathbb{D}}$. Hence Rokhlin's theorem applies, and the signature of $M^{(m)}_{\Sigma|\mathbb{D}}$ must be a multiple of 16. By keeping track of all the numerical modifications along the way, the formula above follows.

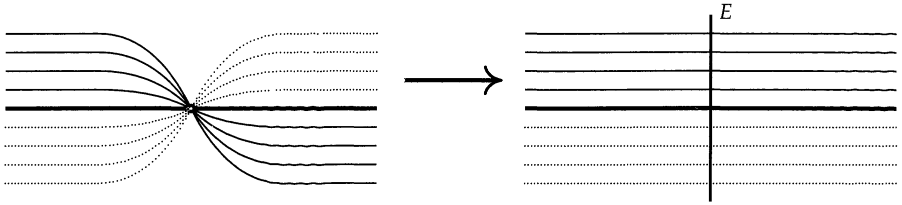
If $\Sigma \cdot \Sigma \leq 0$, we can proceed similarly, but start by connect-summing with copies of $\mathbb{CP}^2$ instead of $\overline{\mathbb{CP}}^2$, join Σ to $\overline{\mathbb{CP}}^1$'s until $\Sigma^{(m)}$ has self-intersection $+1$, and then continue as above. □

Blow it up. *As an equivalent way of describing the procedure in the $\Sigma \cdot \Sigma > 0$ case, we could have said: we blow-up M at m points that belong to Σ, and then take the proper transform of Σ. Since the blow-ups occur at m points of Σ, each of these points are removed from Σ and replaced by a copy of $\overline{\mathbb{CP}}^1$ plugged into M. The* **proper transform** *$\Sigma^{(m)}$ of Σ is simply the closure of $\Sigma \setminus \{m \text{ points}\}$ in $M^{(m)} = M \# m \overline{\mathbb{CP}}^2$. In other words, it is obtained by completing Σ at each puncture with the point of $\overline{\mathbb{CP}}^1$ that describes the direction of T_Σ at the blow-up point. Each blow-up removes one self-intersection of S, as suggested in figure 11.3 on the next page. The result is therefore a sphere $\Sigma^{(m)}$ with self-intersection $+1$.*

To see that $\Sigma^{(m)} \cdot \Sigma^{(m)} = +1$, before blowing, first perturb Σ off the blow-up points; the resulting surface Σ^0 survives the blow-ups as a surface in $M^{(m)}$ and represents there the image of Σ's homology class. The homology class of $\Sigma^{(m)}$ in $M^{(m)}$ can be written $\Sigma^{(m)} = \Sigma^0 - E_1 - \cdots - E_m$, where E_k are the

---

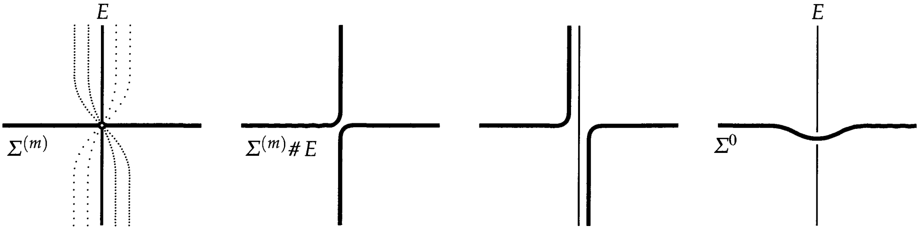
[2] The Hopf map $S^3 \to S^2$ was recalled in footnote 34 on page 129.



11.3. Ironing the normal bundle of Σ by blow-ups

$\overline{\mathbb{CP}}^1$'s from the blow-ups (the (-1)-curves). Therefore $\Sigma^{(m)} \cdot \Sigma^{(m)} = \Sigma \cdot \Sigma - m = +1$.

Further, to see that $\Sigma^{(m)} = \Sigma^0 - E_1 - \cdots - E_m$, one can think as follows: Consider the singular surface $\Sigma^{(m)} \cup E_1$ and remove the double-point in the usual fashion to obtain the surface $\Sigma^{(m)} \# E_1$. Since $E_1 \cdot E_1 = -1$ and $\Sigma^{(m)} \cdot E_1 = +1$ and the class of $\Sigma^{(m)} \# E_1$ is $\Sigma^{(m)} + E_1$, it follows that we have $(\Sigma^{(m)} \# E_1) \cdot E_1 = 0$. Therefore we can slide $\Sigma^{(m)} \# E_1$ completely off E_1. See figure 11.4. Repeating this for each added sphere E_k, we end up with the surface $\Sigma^{(m)} \# E_1 \# \cdots \# E_m$, of homology class $\Sigma^{(m)} + E_1 + \cdots + E_m$ and slid off all E_k's. Since the E_k's were the only thing added to M to make it into $M^{(m)}$, $\Sigma^{(m)} \# E_1 \# \cdots \# E_m$ can be viewed as a surface in the initial M. It can in fact be further moved until it coincides in M with Σ itself, and therefore in homology we have $\Sigma^{(m)} + E_1 + \cdots + E_m = \Sigma$.



11.4. Computing the class of the proper transform

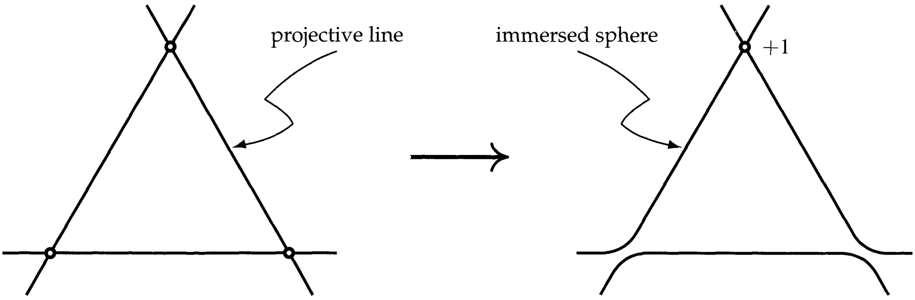
Sphere, no sphere. The Kervaire–Milnor theorem immediately imposes restrictions on which classes can be represented by spheres.

For example, the class $2\alpha + 2\bar{\alpha}$ in $S^2 \times S^2$, where $\alpha = [S^2 \times 1]$ and $\bar{\alpha} = [1 \times S^2]$, cannot be represented by any sphere. Indeed, $2\alpha + 2\bar{\alpha}$ is characteristic and its ambient $S^2 \times S^2$ has signature 0, but its self-intersection is 8.

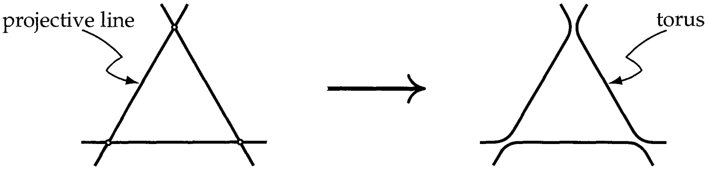
Neither can the class $3\,[\mathbb{CP}^1]$ in $\mathbb{CP}^2$ be represented by a sphere. Nonetheless, after adding a bit more room, it can:

Lemma. *The class $(3, 0)$ in $\mathbb{CP}^2 \# \mathbb{CP}^2$, i.e., the class of $3\,[\mathbb{CP}^1]$ from the first $\mathbb{CP}^2$, can be represented by a sphere.*

Proof. In the first copy of $\mathbb{CP}^2$, we represent $3\,[\mathbb{CP}^1]$ as an immersed sphere with one self-intersection point of sign $+1$. This can be achieved by taking three generic projective lines (each of them meeting another in exactly one point), and replacing two of their three intersections by annuli, as in figure 11.5. (If we were to eliminate the last intersection as well, we would end up with a torus, as sketched in figure 11.6.)



11.5. Representing $3\,[\mathbb{CP}^1]$ by an immersed sphere in $\mathbb{CP}^2$

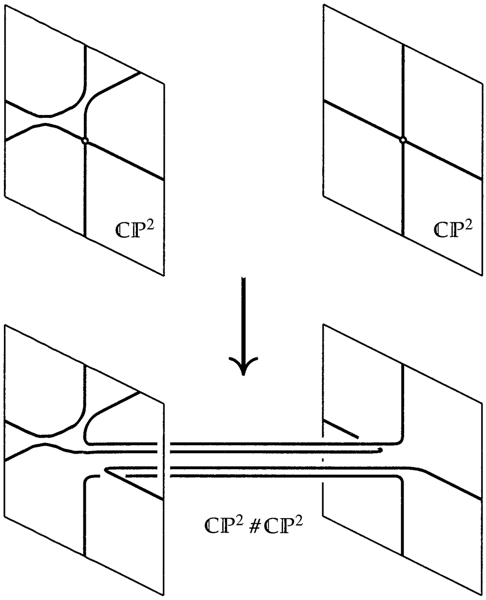


11.6. Representing $3\,[\mathbb{CP}^1]$ by an embedded torus in $\mathbb{CP}^2$

In the second copy of $\mathbb{CP}^2$, we represent the trivial class 0 by two projective lines with opposite orientations (which thus cancel homologically). These are two spheres that intersect each other at a point with negative sign.

We now connect-sum these two copies of $\mathbb{CP}^2$, but take care to do it at the double-points, *i.e.*, by cutting out from each a neighborhood of their surfaces' double-point and fitting the surfaces' leftovers to each other, as suggested in figure 11.7 on the next page. Then the double-points disappear from $\mathbb{CP}^2 \# \mathbb{CP}^2$, and the immersed sphere from the first $\mathbb{CP}^2$ joins the two spheres from the second $\mathbb{CP}^2$ to make up an embedded sphere representing the class $(3, 0)$ in $\mathbb{CP}^2 \# \mathbb{CP}^2$. □

This style of ingenious constructions and reductions were characteristic of the early attempts to deal with the sphere-representation problem, and many results of this type were thus obtained. Now that you have a taste of it, let us move on:



11.7. Eliminating double-points by connect-summing

11.2. Enter the hero: the adjunction inequality

With the advent of gauge theory, the problem of representing classes by spheres could be attacked much more effectively. Moreover, one could now tackle the more general question of the minimum genus needed to represent a class. It all started of course with Donaldson theory,[3] but the major tool came along the wings of Seiberg–Witten theory. We are talking about the adjunction inequality.[4]

Adjunction inequality for positive self-intersections

The full adjunction inequality, as stated back in section 10.4 (page 408), was proved in two stages. The first stage is the following statement:

Adjunction Inequality for Positive Self-Intersection. *Let M be simply-connected with $b_2^+(M) \geq 2$. Let S be any embedded connected surface such that*

$$S \cdot S \geq 0 ,$$

and that S is homologically nontrivial. Then, for every basic class κ of M, we must have:

$$\chi(S) + S \cdot S \leq \kappa \cdot S .$$

---

[3] Recall, for example, that our first construction of an exotic $\mathbb{R}^4$ stemmed from the impossibility of representing a certain class of $E(1) = \mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$ by a smooth sphere; see section 5.4 (page 250).

[4] Please do not confuse the Seiberg–Witten adjunction inequality with the (almost-)complex adjunction formula for J-holomorphic curves.

Notice that changing the orientation of S only changes one sign, and hence we must have $\chi(S) + S \cdot S \leq -\kappa \cdot S$ as well. Thus, we can write a sharper version of the adjunction inequality as

$$\chi(S) + S \cdot S \leq -\left|\kappa \cdot S\right| .$$

Keep in mind that, since $\chi(S) = 2 - 2\,\mathrm{genus}(S)$, the adjunction inequality by putting an upper bound on $\chi(S)$ effectively offers a *lower* bound on the genus of a surface S representing a fixed homology class, and thus measures the complexity needed to fit that class inside M. Note also that it is unknown precisely for which manifolds the inequality is sharp, *i.e.*, for which there exist surfaces that realize the minimum allowed by the adjunction inequality.

Outline of proof. Every solution (φ, A) to the Seiberg–Witten equations must satisfy the inequality

$$\left\|scal\right\| \geq 2\sqrt{2}\left\|F_A^+\right\| ,$$

where *scal* denotes the Riemannian scalar curvature of M, and we used L^2–norms. This is a direct consequence of the Lichnerowicz formula (page 393) and was already proved under the name of "integral curvature bound", on page 398. This inequality provides the needed bridge between Seiberg–Witten theory and the Riemannian geometry of M. The bridge between Riemannian geometry and the genus of S is then offered by Gauß–Bonnet.

Assume first that S has zero self-intersection. On one hand, since S has trivial normal bundle, we can embed a copy of $S \times \mathbb{S}^1 \times [0, 1]$ in M (think: a thickened normal circle-bundle of S), and then choose a Riemannian metric on M that restricts on $S \times \mathbb{S}^1 \times [0, 1]$ to the product-metric endowing the S–factor with the metric of constant scalar curvature $4\pi\chi(S)$ and area 1, assigning length 1 to the $\mathbb{S}^1$–factor, and setting the $[0, 1]$–factor to have length ρ. When stretching this cylinder by pushing $\rho \to \infty$, we get:[5]

$$\left\|scal_\rho\right\|_\rho = -\sqrt{\rho}\,4\pi\chi(S) + O(1) .$$

On the other hand, since $i\left[F_A\right] = 2\pi c_1(\mathcal{L}) = 2\pi\kappa$, a closer investigation also yields the inequality

$$\left\|F_A^+\right\|_\rho \geq \sqrt{2\rho}\,\pi(S \cdot \kappa) + O(1) .$$

---

[5] Remember that $O(f)$ is the Landau symbol that represents any quantity q that grows at most at the rate of f, *i.e.*, any $q(t)$ so that there is some constant C so that $|q(t)| < C \cdot f(t)$. Thus, q is $O(1)$ if and only if q is bounded.

Combining these three inequalities and letting the scaling parameter ρ go to infinity makes it inevitable that

$$S \cdot \kappa \geq -\chi(S) ,$$

which, after an orientation flip, is exactly the result in the $S \cdot S = 0$ case.

To finish the proof by dealing with the case $S \cdot S > 0$, a standard blow-up argument (similar to the one used in the proof of the Kervaire–Milnor theorem, but spiced-up by the blow-up formula for the Seiberg–Witten invariants) reduces the $S \cdot S > 0$ case to the above $S \cdot S = 0$ case. □

The above outline is expanded into a complete proof in the end-notes of this chapter (page 496).

The Thom conjecture

The adjunction inequality works only when $b_2^+ \geq 2$. If our 4–manifold has $b_2^+ = 0$ or $b_2^+ = 1$, then a first thing to try is to switch orientation and look at the manifold $\overline{M}$, since flipping orientations changes[6] b_2^- into b_2^+.

If this does not help, then it means that M is homeomorphic to one of

$$\mathbb{S}^4 , \qquad \text{or} \qquad \pm\,\mathbb{CP}^2 , \qquad \text{or} \qquad \mathbb{S}^2 \times \mathbb{S}^2 , \qquad \text{or} \qquad \mathbb{CP}^2 \,\#\overline{\mathbb{CP}}^2 .$$

There are no known exotic $\mathbb{S}^4$'s, $\mathbb{CP}^2$'s, $\mathbb{S}^2 \times \mathbb{S}^2$'s or $\mathbb{CP}^2 \,\#\overline{\mathbb{CP}}^2$'s (and that is in good part due to their unavoidably small b_2^+). We are thus limited to exploring their standard versions.

Spheres. Since the 4–sphere (or its possible exotic versions) has no homology, the minimum genus problem inside spheres is vacuous.

Sphere-bundles. The manifolds $\mathbb{S}^2 \times \mathbb{S}^2$ and $\mathbb{CP}^2 \,\#\overline{\mathbb{CP}}^2 = \mathbb{S}^2 \mathbin{\tilde{\times}} \mathbb{S}^2$ are well-understood:

Theorem (*D. Ruberman*). *The minimum genus of a surface representing the class $m\,\alpha + n\,\bar{\alpha}$ in $\mathbb{S}^2 \times \mathbb{S}^2$ is*

$$g_{min}(m\,\alpha + n\bar{\alpha}) = \big(|m| - 1\big)\big(|n| - 1\big) ,$$

when m and n are not zero. If $m = 0$ or $n = 0$, then the class can be represented by a sphere.

---

[6] It also changes the sign of the self-intersection of S, but we will not worry about that for now. See **T. Lawson**'s ***The minimal genus problem*** [**Law97**] for a more comprehensive survey; also see ahead the adjunction inequality for simple type (page 489) and have some faith in the simple type conjecture (page 402).

The minimum genus of a surface representing the class $m\alpha + n\bar{\alpha}$ *inside* $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2 = \mathbb{S}^2 \tilde{\times} \mathbb{S}^2$ *is*

$$g_{min}(m\alpha + n\bar{\alpha}) = \tfrac{1}{2}\big(|m| - 1\big)\big(|m| - 2\big) + \tfrac{1}{2}\,|n|\,\big(|n| - 1\big)\,,$$

when $m > n$. *If* $m < n$, *reverse the roles of* m *and* n. *If* $m = n$, *then the class can be represented by a sphere.* □

Here, as usual, α is the class $[\mathbb{S}^2 \times 1]$, while $\bar{\alpha} = [1 \times \mathbb{S}^2]$ (or, for the second case, α and $\bar{\alpha}$ are the classes of the $\mathbb{CP}^1$'s in $\mathbb{CP}^2$ and in $\overline{\mathbb{CP}}^2$).

The complex projective plane. A long-standing conjecture on genera of surfaces in $\mathbb{CP}^2$, attributed to R. Thom and proved by P. Kronheimer and T. Mrowka, is:

Thom Conjecture. *The minimum genus of a surface representing a fixed homology class in* $\mathbb{CP}^2$ *is always realized by a complex curve (with either orientation).*

Remember that, by the complex adjunction formula,[7] a complex curve of degree d in $\mathbb{CP}^2$ always has genus $\frac{1}{2}(d-1)(d-2)$. We can then rephrase: *The minimum genus of a surface representing* $d\,[\mathbb{CP}^1]$ *in* $\mathbb{CP}^2$ *is*

$$g_{min}(d) = \tfrac{1}{2}\big(|d| - 1\big)\big(|d| - 2\big)\,.$$

Idea of proof. Let S be a surface representing the class $d\,[\mathbb{CP}^1]$ and with[8] $d \geq 4$, which can be assumed to not be a sphere. Blow-up $\mathbb{CP}^2$ d^2–times at points of S, then remove singularities, ending up with a surface $S^{\#}$ of the same genus as S but with trivial normal bundle. Embed a cylinder $S^{\#} \times \mathbb{S}^1 \times [0,1]$ around $S^{\#}$ and stretch the $[0,1]$–factor. A careful study of the Seiberg–Witten equations shows that the Seiberg–Witten moduli space is non-empty (since $b_2^+(\mathbb{CP}^2) = 1$, *a priori* one expects problems with reducible solutions), and then a reasoning similar to the one in the proof of the adjunction inequality can be used. □

Adjunction inequality for simple type

Besides the restriction $b_2^+ \geq 0$ on which we commented above, the adjunction inequality as stated above is powerless in dealing with classes of *negative* self-intersection. After a delicate analysis of the Seiberg–Witten invariants, this restriction can be partly eliminated:

Adjunction Inequality for Simple Type. *Let* M *be simply-connected with* $b_2^+(M) \geq 2$. *Assume* M *is of Seiberg–Witten simple type, i.e., that the only basic classes of* M *correspond to almost-complex spin*$^{\mathbb{C}}$ *structures. Let* S *be any*

---

[7] The adjunction formula states that a complex (or J–holomorphic) curve C in M must have $\chi(C) + C \cdot C = c_1(M) \cdot C$; see section 6.3 (page 281) and section 10.1 (page 377).

[8] Since the cases $0 \leq d \leq 3$ are either obvious or dealt with by Kervaire–Milnor.

embedded connected surface in M that is not a sphere. Then for every basic class κ of M we have:

$$\chi(S) + S \cdot S \leq \kappa \cdot S\,. \qquad \square$$

This result was proved by P. Ozsváth and Z. Szabó. After flipping the orientation of S, the inequality can be sharpened to $\chi(S) + S \cdot S \leq -|\kappa \cdot S|$. Furthermore, remember that it is conjectured that all 4–manifolds that have $b_2^+ \geq 2$ and are simply-connected might be of Seiberg–Witten simple type; certainly all known ones are.

Since C. Taubes showed that all symplectic manifolds are of simple type, the above statement implies:

Corollary. *For every surface S, not a sphere, embedded in a symplectic manifold M with $b_2^+(M) \geq 2$, we must have:*

$$\chi(S) + S \cdot S \leq K_M^* \cdot S\,,$$

where $K_M^* = c_1(\omega)$ is the Chern class of the symplectic structure. □

In the particular case when M is Kähler, since any complex curve S must have $\chi(S) + S \cdot S = K_M^* \cdot S$, the following generalization of the Thom conjecture follows:

Kähler Thom Conjecture. *Inside every Kähler surface with $b_2^+ \geq 2$, non-singular complex curves always realize the minimum genus needed to represent their homology class.* □

Furthermore, just as a complex structure on a manifold distinguishes its complex submanifolds, a symplectic structure will distinguish certain submanifolds of its own. Specifically, if ω is a symplectic structure on M^4, then a surface S in M is called a **symplectic surface** of (M, ω) when we have[9]

$$\omega|_S > 0\,.$$

(In particular such an S is never homologically-trivial, since $S \cdot [\omega] \neq 0$.)

Saying "S is a symplectic surface inside (M, ω)" is equivalent to asking that there be some almost-complex structure J_S on M that is compatible with ω and makes S be J_S–holomorphic. Therefore S must satisfy the complex adjunction formula: $\chi(S) + S \cdot S = c_1(\omega) \cdot S$. We thus get a further generalization of the Thom conjecture:

Symplectic Thom Conjecture. *Inside every symplectic 4–manifold with $b_2^+ \geq 2$, symplectic surfaces always realize the minimum genus needed to represent their homology class.* □

---

[9] We write $\omega|_S > 0$ to mean $\omega(x, y) > 0$ for every pair $x, y \in T_S$ that compatibly orients S.

The above could be rephrased more succinctly as: *Inside symplectic manifolds, J–holomorphic curves always minimize genus.*

However, keep in mind that the existence of a symplectic structure is necessary. Indeed, a weaker statement like "Inside almost-complex manifolds, J–holomorphic curves minimize genus" is *false.*

> ***Keep it symplectic.*** *This is shown by an example of G. Mikhalkin*[10] *that proves that if we take the manifold $\#3\,\mathbb{CP}^2$ (which does not admit any symplectic structures*[11]*) and endow it with the almost-complex structure J of anti-canonical class*[12] *$K_J^* = (3,3,1)$, then any J–holomorphic representative of this class necessarily has genus 3; but the class $(4,0,0)$ can be represented by a torus. In light of this example, J–holomorphic curves can automatically be assumed to minimize genus only when the almost-complex structure J is tamed by a symplectic structure.*

Even with these very powerful adjunction inequalities, the minimum genus problem is far from settled. Leaving aside technical difficulties involved with actually finding the Seiberg–Witten basic classes of random far-from-complex 4–manifolds, we do not know when the adjunction inequality is *sharp:* Are there surfaces that actually realize the equality? Are there better bounds? Note that there are known examples (non-simply-connected; see below, page 494) where basic classes indeed fail to offer the best genus bound.

As is often the case with gauge theory, the farther one moves from the complex world, the less useful its tools become.

11.3. Digression: the happy case of 3–manifolds

Unlike dimension 4, in the case of dimension 3 we have a pretty good understanding of the minimum genus problem.

What is genus? On a 3–manifold, every homology class $\alpha \in H_2(N^3;\mathbb{Z})$ can be represented by an embedded surface. Such a surface though, unlike in dimension 4, often needs to be disconnected.[13] Therefore, it is not

---

[10] See **G. Mikhalkin**'s ***J–holomorphic curves in almost complex surfaces do not always minimize the genus*** **[Mik97]**.

[11] Proof: The manifold $\#3\,\mathbb{CP}^2$, being a connected sum of two manifolds with $b_2^+ \geq 1$, must have trivial Seiberg–Witten invariants and thus no basic classes. However, if $\#3\,\mathbb{CP}^2$ admitted some symplectic structure ω, then it would be necessary that $c_1(\omega)$ be a basic class. Therefore $\#3\,\mathbb{CP}^2$ does not admit any symplectic structure.

[12] Such an almost-complex structure exists owing to the existence result from section 10.1 (page 377).

[13] In a 4–manifold there is enough room to connect components with thin tubes while avoiding creation of self-intersections. Not so in dimension 3.

reasonable to talk about "minimizing genus" as such, but about maximizing the Euler–Poincaré characteristic $\chi = 2 - 2\,\mathrm{genus}$. Even so, if we admit sphere-components (which each adds 2 to χ), this might artificially bump down our evaluation of the "genus".

Thus, let us define for every surface S embedded in a 3–manifold a quantity $\overline{\chi}(S)$ as the sum, over all components S_k of S that are *not* spheres, of the negative Euler characteristic $-\chi(S_k)$:

$$\overline{\chi}(S) = \sum_{\mathrm{genus}\, S_k \geq 1} -\chi(S_k) \, .$$

Notice that minimum "genus" corresponds to maximum $\chi(S)$ but minimum $\overline{\chi}(S)$. Notice also that, since each component adds 2, more components mean a lower $\overline{\chi}(S)$. One might call $\overline{\chi}(S)$ the *complexity* of S.

For every homology class $\alpha \in H_2(N^3; \mathbb{Z})$, we define its **Thurston norm** by

$$\overline{\chi}_{min}(\alpha) = \min\{\overline{\chi}(S) \mid [S] = \alpha\} \, .$$

It detects the least complexity needed for representing α. The name of "norm" is justified by the following remarkable property:

Theorem *(W. Thurston). For every closed oriented 3–manifold N^3, the function $\overline{\chi}_{min}$ on $H_2(N; \mathbb{Z})$ satisfies the triangle inequality:*

$$\overline{\chi}_{min}(\alpha + \beta) \leq \overline{\chi}_{min}(\alpha) + \overline{\chi}_{min}(\beta) \, ,$$

and is linear on rays:

$$\overline{\chi}_{min}(n\alpha) = n\,\overline{\chi}_{min}(\alpha) \, .$$

Therefore $\overline{\chi}_{min}$ is the restriction to $H_2(N; \mathbb{Z})$ of a semi-norm on $H_2(N; \mathbb{R})$. □

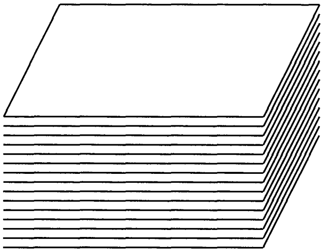
Enter foliations. Surfaces of minimum genus are strongly related to taut foliations of 3–manifolds.

A **foliation** by surfaces of a manifold N is a complete decomposition of N into surface-slices, called **leaves** of the foliations. Locally, a foliation looks like figure 11.8 on the facing page. The leaves are injectively immersed, but if they are non-compact then their ends can run and wrap around forever, and the leaf itself can be dense[14] in N.

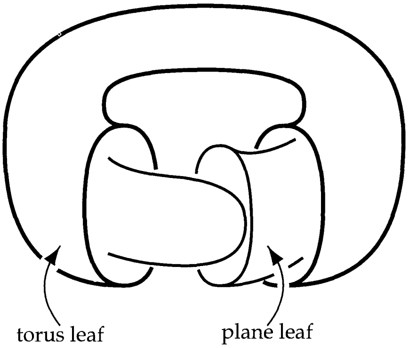
A foliation of N is called **taut** if and only if there is some Riemannian metric on N such that all leaves become *minimal surfaces* (*i.e.*, surfaces of critical area = surfaces that locally minimize area = surfaces whose mean curvature vanishes). In dimension 3, this is equivalent to the non-existence of *dead ends*, namely the absence of torus leaves that would cut N into two separated halves. A typical example of a dead end is a **Reeb component**, sketched in figure 11.9 on the next page: a solid-torus region bounded by a torus leaf and with its interior foliated by leaves that look like a paraboloid engulfing itself.

---

[14] For example, think of a line of irrational slope in the standard 2–torus.



11.8. Local model of a foliation



11.9. A Reeb component

Theorem *(W. Thurston). Assume N^3 is a closed 3-manifold that is not $\mathbb{S}^1 \times \mathbb{S}^2$. Let $\mathscr{F}$ be any taut foliation of N. Then, for every embedded surface S in N, we have:*[15]

$$\overline{\chi}(S) \geq e(T_{\mathscr{F}}) \cdot S . \qquad \square$$

One could think of this inequality as an analogue, in dimension 3, of the 4-dimensional Seiberg–Witten adjunction inequality.

A consequence of it is that, if S is the union of compact leaves of a taut foliation, then S must achieve minimum $\overline{\chi}$ in its homology class. Remarkably, the converse is also true:

Theorem *(D. Gabai). Let N^3 be an irreducible*[16] *3-manifold, and S an embedded surface of nontrivial homology class. Assume that S minimizes $\overline{\chi}$ in its homology class. Then there exists a taut foliation $\mathscr{F}$ on N such that S is a union of leaves.* □

Hence, in brief, a surface in a 3-manifold minimizes genus if and only if it is the leaf of a taut foliation.

---

[15] Here, $e(T_{\mathscr{F}}) \cdot S$ is *not* an intersection number, but the evaluation of $e(T_{\mathscr{F}}) \in H^2(N;\mathbb{Z})$ on $[S] \in H_2(N;\mathbb{Z})$. The bundle $T_{\mathscr{F}}$ is the subbundle of T_N made of the tangent planes to the leaves of $\mathscr{F}$.

[16] An n-manifold is called **irreducible** if it does not split as a connected sum of simpler manifolds (homotopy spheres do not count).

Comparison with 4–manifolds

Comparing the above results with the state of knowledge about analogous statements on 4–manifolds is a depressing affair. For 4–manifolds, the quantity similar to $\overline{\chi}$ above would be $-\big(\chi(S) + S \cdot S\big)$. In regard to its being a semi-norm, we only have the very light lemma:

Lemma. *Let S be a connected embedded surface in the 4–manifold M. Assume that either $S \cdot S > 0$, or that $S \cdot S = 0$ and S is not a sphere. Then, for every integer n, we can represent the class $n[S]$ by a connected embedded surface S_n such that*

$$\chi(S_n) + S_n \cdot S_n = n\big(\chi(S) + S \cdot S\big) .$$

> ***Proof.*** If $S \cdot S = m > 0$, then we take n push-off copies of S. Each copy will intersect each other copy exactly m times. After eliminating all these intersection points, we obtain an embedded connected surface S_n as above.
>
> If $S \cdot S = 0$, since we assumed S to have some genus, that allows us to locate, inside the normal circle-bundle of S, an n–periodic cover S_n of S (for example, S_n will twist by $2\pi/n$ while traveling over a fixed homologically-nontrivial circle of[17] S). □

Of course, this does *not* exclude that the class $n[S]$ be representable with less genus. In fact, it is not known at all whether *minimum* genus might satisfy an equality as above, or even if there are *some* 4–manifolds where that might hold true.

Finally, the only hint that there might be a relation between minimum genus surfaces and taut foliations on 4–manifolds is the following result:

Theorem (*P. Kronheimer*). *Consider the 4–manifold*

$$N^3 \times S^1 ,$$

with N a closed irreducible 3–manifold. Let $\mathscr{F}$ be a taut foliation of N with Euler class $\varepsilon = e(T_{\mathscr{F}})$. Let $\widetilde{\varepsilon} \in H^2(N \times S^1; \mathbb{Z})$ be the pull-back of $\varepsilon \in H^2(N; \mathbb{Z})$. Then for any embedded connected surface S in $N \times S^1$ we have

$$\chi(S) + S \cdot S \leq \widetilde{\varepsilon} \cdot S . \qquad \square$$

The proof uses a version of the Seiberg–Witten invariants and, in part, an argument similar to the proof of the adjunction inequality. At times the classes $\widetilde{\varepsilon}$ are *not* Seiberg–Witten basic classes (they admit monopoles, but their count cancels to zero). Thus, the adjunction inequality is *not* sharp on such (admittedly, not simply-connected) 4–manifolds.

---

[17] When S is a torus, this procedure could be suggested by the pictures used earlier when describing logarithmic transformations, back in section 8.3 (page 310).

One might conjecture that something like this theorem—taut foliations[18] on 4–manifolds yielding minimum genus bounds—might hold on more general 4–manifolds and foliations, but there is no method in sight that could plausibly be used in hunting for an answer. In fact, it is not clear whether there is any relation between general taut foliations and Seiberg–Witten monopoles. We do not know either whether a compact leaf of a taut foliation might minimize genus in a 4–manifold.

Sigh...

---

[18] A foliation of a 4–manifold by surfaces is called **taut** if its leaves are minimal surfaces for some Riemannian metric on M. Nonetheless, there is no interpretation in terms of dead ends or Reeb components, and in general the structure of codimension 2 foliations is very poorly understood. Notice that the class $\widetilde{\varepsilon}$ from Kronheimer's theorem is the Euler class of $T_{\widetilde{\mathscr{F}}}$ for the *taut* foliation $\widetilde{\mathscr{F}} = \mathscr{F} \times \mathbb{S}^1$ obtained simply by stacking copies of $\mathscr{F}$ around $\mathbb{S}^1$.

11.4. Notes

Note: Proof of the adjunction inequality

We present the proof of the adjunction inequality for surfaces of positive self-intersection. Recall its statement:

Adjunction Inequality. *Let M be simply-connected with $b_2^+(M) \geq 2$. Let S be any embedded connected surface in M such that $S \cdot S \geq 0$ and S is homologically nontrivial. Then for every basic class κ of M we have:*

$$\chi(S) + S \cdot S \leq \kappa \cdot S .$$

The result will follow from three inequalities. We begin by repeating the outline of the argument.

Outline of proof. Every solution (φ, A) to the Seiberg–Witten equations must satisfy the inequality

$$\|scal\| \geq 2\sqrt{2}\,\|F_A^+\| ,$$

where $scal$ denotes the Riemannian scalar curvature of M, and we used L^2–norms. This is a direct consequence of Lichnerowicz's formula (page 393) and was already proved in section 10.3 (page 398). This inequality provides the needed bridge between Seiberg–Witten theory and the Riemannian geometry of M. The bridge between Riemannian geometry and the genus of S will be offered by Gauß–Bonnet.

Assume first that S has zero self-intersection. On one hand, since S has trivial normal bundle, we can embed a copy of $S \times \mathbb{S}^1 \times [0,1]$ in M, then choose a Riemannian metric for M which on $S \times \mathbb{S}^1 \times [0,1]$ is the product-metric that endows the S–factor with the metric of constant scalar curvature $4\pi\chi(S)$ and area 1, assigns length 1 to the $\mathbb{S}^1$–factor, and sets the $[0,1]$–factor to have length ρ. When stretching this cylinder by pushing $\rho \to \infty$, we will evaluate

$$\|scal_\rho\|_\rho = -\sqrt{\rho}\,4\pi\chi(S) + O(1) .$$

On the other hand, since $[F_A] = -2\pi i\, c_1(\mathcal{L}) = -2\pi i \kappa$, we will also obtain the inequality

$$\|F_A^+\|_\rho \geq \sqrt{2\rho}\,\pi(S\cdot\kappa) + O(1) .$$

Combining these three inequalities and pushing the scaling parameter ρ toward infinity makes it inevitable that

$$S\cdot\kappa \geq -\chi(S) ,$$

which, after an orientation flip, is exactly the result in the $S \cdot S = 0$ case.

To finish the proof by dealing with the case $S \cdot S > 0$, a simple blow-up argument will reduce the $S \cdot S > 0$ case to the above $S \cdot S = 0$ case.

The rest of this note is taken by the expansion of this outline.

Preparation: Monopoles and scalar curvature. Start with the coupled Lichnerowicz formula

$$(\mathcal{D}^A)^* \, \mathcal{D}^A \varphi = (\nabla^A)^* \, \nabla^A \varphi + \tfrac{1}{4}\, scal \cdot \varphi + \tfrac{1}{2} F_A \bullet \varphi \, .$$

Assume (φ, A) is a Seiberg–Witten monopole, so that $\mathcal{D}^A \varphi = 0$ and $F_A^+ = \sigma(\varphi)$. Use this in the above formula, then inner-multiply with φ and integrate over M, rearrange and use the Cauchy–Schwarz inequality, use that $|\sigma(\varphi)| = \frac{1}{2\sqrt{2}} |\varphi|^2$ and obtain that, for every monopole (φ, A), we must have:

$$2\sqrt{2}\, \|F_A^+\| \leq \|scal\| \, ,$$

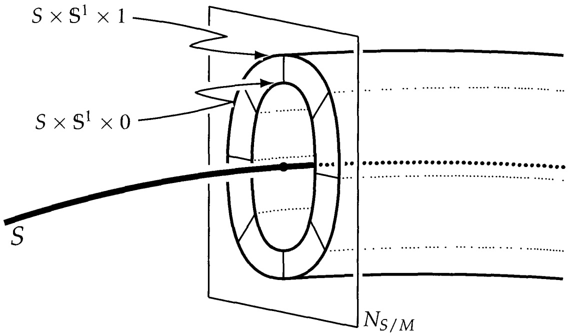
written in terms of L^2–norms.[1] The detailed proof was displayed on page 398, under the name of "integral curvature bound".

Case without self-intersection: Stretching the cylinder. Let κ be a Seiberg–Witten basic class of M. Then, for *every* metric on M, the spin$^{\mathbb{C}}$ structure corresponding to κ has *at least one* Seiberg–Witten monopole. Thus, we are free to choose our favorite metric on M.

Assume $S \cdot S = 0$. Then the normal bundle $N_{S/M}$ is trivial, and its circle bundle is a copy of $S \times \mathbb{S}^1$. Therefore we can thicken it and make appear in M an embedded cylinder

$$\mathcal{C}y\ell = S \times \mathbb{S}^1 \times [0, 1] \, ,$$

surrounding S as suggested in figure 11.10.



11.10. Cylinder $S \times \mathbb{S}^1 \times [0, 1]$ around a surface S

Next, we choose a metric on M that, when restricted to the cylinder $\mathcal{C}y\ell$, is a product metric that makes the $[0, 1]$–factor have length ρ and the $\mathbb{S}^1$–factor have length

---

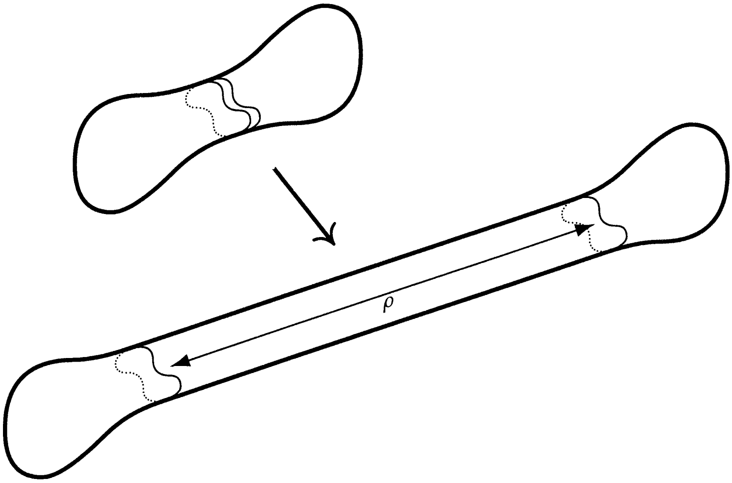
[1] Remember that the **L^2–norm** of α is $\|\alpha\|^2 = \int_M |\alpha|^2 \, vol_M$. Its value depends on the chosen Riemannian metric of M, both through the length-measurement $|\alpha|$ and the volume element vol_M.

1 and endows the S–factor with the 2–dimensional metric of constant scalar-curvature[2] $4\pi\chi(S)$ and area 1. Then, for such a choice of Riemannian metric on M, its scalar curvature *scal* is everywhere on $\mathcal{Cyl}$ equal to the constant $4\pi\chi(S)$:

$$scal\,|_{\mathcal{Cyl}} \equiv 4\pi\chi(S)\ .$$

Indeed, the scalar curvature is essentially (twice) the sum of the four sectional curvatures; owing to the product structure of the cylinder, the only non-zero sectional curvature is the one in the direction of the S–factor.

We now start modifying the metric of M by pushing the scaling parameter ρ to ∞, in other words, by stretching the cylinder $\mathcal{Cyl}$ as pictured in figure 11.11. Then the curvature contribution of this expanding cylinder $\mathcal{Cyl}$ will eventually overshadow anything else that might happen in the rest of M.



11.11. Stretching the cylinder

Assume that S is *not* a sphere, and hence that $4\pi\chi(S) \leq 0$. Then, in terms of the L^2–norms $\|\cdot\|_\rho$ corresponding to the ρ–scaled metrics, we have:

$$\begin{aligned}\|scal_\rho\|_\rho^2 &= \int_M (scal_\rho)^2 = \int_{\mathcal{Cyl}} (scal_\rho)^2 + C \\ &= \rho\int_S (scal)^2 + C = \rho(4\pi\chi(S))^2 + C\ ,\end{aligned}$$

where C represents the integral of $(scal)^2$ over the rest of M, which is the region where the metric is kept unchanged as ρ grows.[3] Therefore[4]

$$\|scal_\rho\|_\rho = -\sqrt{\rho}\,4\pi\chi(S) + O(1)\ ,$$

with the minus-sign appearing because we assumed $4\pi\chi(S) \leq 0$.

---

[2] Keep in mind that the scalar curvature of a surface is *twice* its Gaußian curvature.

[3] We are ignoring the small perturbation needed to keep the metric smooth while ρ changes.

[4] Remember that $O(1)$ denotes an arbitrary bounded quantity.

The stretch and the monopole. Remember that κ was a basic class. Thus, for every Riemannian metric, there must be at least one solution to the Seiberg–Witten equations. In particular, for every one of the special metrics above there is at least one monopole $(\varphi, A) = (\varphi_\rho, A_\rho)$.

Since the corresponding 2–form F_A is the curvature of a connection on $\mathcal{L}$, the class $\frac{i}{2\pi}[F_A]$ must represent the class $\kappa = c_1(\mathcal{L})$. We will first deduce that $\|F_A\| \geq \sqrt{\rho}\, 2\pi\,(S\cdot\kappa)$. This follows from the facts that

$$\tfrac{i}{2\pi}\int_S F_A = S\cdot\kappa$$

and that such an equality will still hold when integrating over every parallel copy of S inside the cylinder $\mathcal{Cyl} = S \times S^1 \times [0,1]$. Since the cylinder is stretching as $\rho \longrightarrow \infty$, the contribution to $\|F_A\|$ from the part of F_A over the cylinder overshadows all the rest. We have

$$\|F_A\|^2 = \int_M |F_A|^2 \geq \int_{\mathcal{Cyl}} |F_A|^2 = \rho\int_S |F_A|^2 \geq \rho\Big(\int_S iF_A\Big)^2 = \rho\big(2\pi(S\cdot\kappa)\big)^2$$

and therefore, as claimed,

$$\|F_A\| \geq \sqrt{\rho}\,2\pi\,(S\cdot\kappa)\,.$$

To get from $\|F_A\|$ to $\|F_A^+\|$, we recall that for every 2–form α we have

$$\|\alpha\|^2 = \|\alpha^+\|^2 + \|\alpha^-\|^2$$
$$\|\alpha^+\|^2 - \|\alpha^-\|^2 = [\alpha]\cdot[\alpha]\,.$$

The latter was proved as "lemma Z" in section 10.3 (page 399). Therefore

$$\|F_A\|^2 = 2\,\|F_A^+\|^2 - 4\pi^2\kappa\cdot\kappa\,.$$

Combining with the above inequality and cleaning up yields

$$\|F_A^+\| \geq \sqrt{\rho}\cdot\sqrt{2}\,\pi\,(S\cdot\kappa) + O(1)\,.$$

Stretching the conclusion. We have proved the following three relations:

$$\sqrt{\rho}\cdot 4\pi\,(S\cdot\kappa) + O(1) \leq 2\sqrt{2}\,\Big\|F_{A_\rho}^+\Big\|_\rho$$
$$2\sqrt{2}\,\Big\|F_{A_\rho}^+\Big\|_\rho \leq \big\|scal_\rho\big\|_\rho$$
$$\big\|scal_\rho\big\|_\rho = -\sqrt{\rho}\,4\pi\chi(S) + O(1)\,.$$

When $\rho \longrightarrow \infty$, comparing them leads to the conclusion that the only way the three can coexist peacefully is if we have

$$S\cdot\kappa \leq -\chi(S)\,,$$

which is just $\chi(S) \leq -\kappa\cdot S$. After flipping the orientation of S, we get $\chi(S) \leq \kappa\cdot S$.

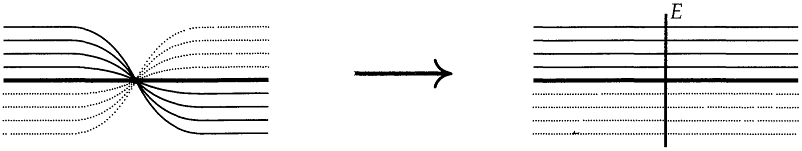
General case. For the general case, when $S \cdot S > 0$ and the normal bundle of S is *not* trivial, one can argue as follows:

Let $m = S \cdot S$. Blow M up m times at points of S, obtaining

$$M^{(m)} = M \# m \overline{\mathbb{CP}}^2 .$$

Denote by E_k the copy of $\overline{\mathbb{CP}}^1$ in the k^{th} copy of $\overline{\mathbb{CP}}^2$. In particular, we have $E_k \cdot E_k = -1$.

Just as in the proof of the Kervaire–Milnor theorem (page 483), we take the proper transform $S^{(m)}$, whose homology class is $S - E_1 - \cdots - E_m$, and therefore has $S^{(m)} \cdot \Sigma^{(m)} = 0$ and trivial normal bundle (see figure 11.12), and thus the adjunction case already proved above is applicable to it.



11.12. Straightening the normal bundle of Σ by blow-ups

Since κ is a basic class of M, from the Seiberg–Witten blow-up formula[5] it follows that $\kappa + E_1 + \cdots + E_m$ must be a basic class of $M^{(m)}$. Applying the adjunction inequality obtained above to $S^{(m)}$, we get:

$$\chi(S^{(m)}) \leq -(\kappa + E_1 + \cdots + E_m) \cdot (S - E_1 - \cdots - E_m) .$$

Since $(\kappa + E_1 + \cdots + E_m) \cdot (S - E_1 - \cdots - E_m) = \kappa \cdot S + m = \kappa \cdot S + S \cdot S$, we write:

$$\chi(S) \leq -\kappa \cdot S - S \cdot S ,$$

which we like to rearrange as

$$\chi(S) + S \cdot S \leq -\kappa \cdot S .$$

Since flipping the orientation of S only changes one sign, we can also write

$$\chi(S) + S \cdot S \leq \kappa \cdot S .$$

This concludes the proof of the adjunction inequality. □

Notice that, unlike connect-summing with $\overline{\mathbb{CP}}^2$, summing with $\mathbb{CP}^2$ destroys the Seiberg–Witten invariants.[6] Thus, if S had negative self-intersection, then we could not simply reduce to the case $S \cdot S = 0$ by increasing its self-intersection inside $M \# m \mathbb{CP}^2$.

---

[5] The blow-up formula was discussed back in section 10.4 (page 404).

[6] From Seiberg–Witten's vanishing for connect sums with $b_2^+(\mathit{terms}) \geq 1$, see section 10.4 (page 406).

Note: The Arf invariant

The Arf invariant is a $\mathbb{Z}_2$–valued invariant of a quadratic form on a $\mathbb{Z}_2$–module. It will play an essential role in the Freedman–Kirby generalization of Rokhlin's theorem that will be presented in the next note (page 502). Since the latter will be proved from scratch and thus prove Rokhlin's theorem itself, it follows that the Arf invariant underpins Rokhlin's theorem.

In fact the presence of the Arf invariant in topology is ubiquitous. One could even say that, whenever one encounters a $\mathbb{Z}_2$, there might be a corresponding Arf invariant hiding behind the scenes: in dimension 3, the Rokhlin invariant; in high-dimensions, the Kirby–Siebenmann invariant; in surgery theory, the Arf invariant holds for $(4k+2)$–manifolds the role that signatures play for $4k$–manifolds; *etc.*

Definition. Let Z be a finite-dimensional $\mathbb{Z}_2$–module. Assume that Z is endowed with a unimodular symmetric bilinear form

$$Z \times Z \longrightarrow \mathbb{Z}_2\colon \qquad (x,y) \longmapsto x \cdot y\,.$$

A most frequent case is to take $Z = H_1(S;\mathbb{Z}_2)$ for some oriented surface S, together with its intersection form. Here we are talking about the modulo 2 reduction of S's *skew-symmetric* intersection form $Q_S\colon H_1(S;\mathbb{Z}) \times H_1(S;\mathbb{Z}) \to \mathbb{Z}$, which governs the intersection of curves inside S. For brevity, we will call any abstract unimodular symmetric form, on a general $\mathbb{Z}_2$–module Z, an **intersection form** as well.

A function $q\colon Z \to \mathbb{Z}_2$ is called a **quadratic enhancement** (or **quadratic form**) for the intersection form of Z if it satisfies

$$q(x+y) = q(x) + q(y) + x \cdot y \pmod 2$$

for all $x, y \in Z$. Such an enhancement does not exist on every random $(Z, \cdot\,)$.

Pick in Z any basis $\{e_1, \dots, e_m, \bar{e}_1, \dots, \bar{e}_m\}$ such that the only non-zero intersections are $e_k \cdot \bar{e}_k = 1$. This is known as a **symplectic basis** for $(Z, \cdot\,)$ and always exists. We define the **Arf invariant** of q as

$$\operatorname{Arf}(q) = \textstyle\sum q(e_k)\, q(\bar{e}_k)\,,$$

which is a well-defined $\mathbb{Z}_2$–valued invariant of $(Z, \cdot\,, q)$.

Interpretation. Playing with the equality $q(x+y) = q(x) + q(y) + x \cdot y$, one notices that we must always have $q(0) = 0$, but also that the existence of a quadratic enhancement q on $(Z, \cdot\,)$ forces that $x \cdot x = 0$ for all $x \in Z$.

Since the intersection form of Z is unimodular, for every $x \in Z$ there exists a dual element $y \in Z$ such that $x \cdot y = 1$. Then on the pair $\{x, y\}$ the intersection form has matrix

$$\begin{bmatrix} & 1 \\ 1 & \end{bmatrix},$$

and therefore the pair $\{x, y\}$ must have an orthogonal complement in Z.

We can then keep on splitting off such hyperbolic pairs till we exhaust all of Z and split it into a sum

$$Z = H_1 \oplus \cdots \oplus H_p\,,$$

with each H_k containing four elements, $H_k = \{0,\, x,\, y,\, x+y\}$ with $x \cdot y = 1$.

Now, considering the possible behaviors of q on H_k, it turns out that either q is 1 on all three of x, y, $x+y$, or else q is 1 on one of them and 0 on the other two. Denote a pair (H_k, q) in the first case by $H^{1,1}$, and in the second by $H^{0,0}$. We have obtained a splitting

$$(Z, q) \approx \oplus m H^{1,1} \oplus n H^{0,0} .$$

Then the Arf invariant $\operatorname{Arf}(q)$ counts the number of copies of $H^{1,1}$ modulo 2:

$$\operatorname{Arf}(q) = m \pmod 2 .$$

Note that a simple change of basis from $\{x_1, y_1; x_2, y_2\}$ to $\{x_1 + y_1 + x_2,\ x_1 + y_1 + y_2,\ x_1 + x_2 + y_2,\ y_1 + x_2 + y_2\}$ establishes an isomorphism

$$H^{1,1} \oplus H^{1,1} \approx H^{0,0} \oplus H^{0,0} ,$$

so that a modulo 2 count is the very best we can hope for. In other words:

Lemma (*C. Arf*). *Two quadratic enhancements are isomorphic if and only if they have the same Arf invariant.* □

Finally, notice that evaluating the Arf invariant is merely a matter of majorities:

Voting Lemma. *We have* $\operatorname{Arf}(q) = 1$ *if and only if* q *sends more elements of* Z *to* 1 *than it sends to* 0. □

The Arf invariant was introduced in **C. Arf**'s ***Untersuchungen über quadratische Formen in Körpern der Charakteristik*** **2. *I*** [**Arf41**].

Note: The Freedman–Kirby generalization of Rokhlin's theorem

Recall that Rokhlin's theorem states: *Every smooth* 4*–manifold* M *with* $w_2(M) = 0$ *must have* $\operatorname{sign} M = 0 \pmod{16}$. Kervaire–Milnor's generalization states: *Every smooth* 4*–manifold* M *and characteristic sphere* Σ *must have* $\operatorname{sign} M - \Sigma \cdot \Sigma = 0 \pmod{16}$.

A further extension is due to **M. Freedman** and **R. Kirby**'s ***A geometric proof of Rochlin's theorem*** [**FK78**]. The restriction to spheres from Kervaire–Milnor is dropped, at the price of an extra term:

Freedman–Kirby Theorem. *Let* M *be a smooth* 4*–manifold, and* Σ *any characteristic surface in it. Then we must have:*

$$\operatorname{sign} M - \Sigma \cdot \Sigma = 8 \cdot \operatorname{Arf}(M, \Sigma) \pmod{16} .$$ □

Here $\operatorname{Arf}(M, \Sigma)$ is the Arf invariant of a suitably defined quadratic enhancement on $H_1(\Sigma; \mathbb{Z}_2)$, which turns out to depend only on the homology class of Σ.

On a merely topological 4*-manifold, this theorem does not hold (just as Rokhlin's theorem does not hold either). Nonetheless, the measure of its failure is the Kirby–Siebenmann*[7] *invariant of the manifold: we have* $\operatorname{ks}(M) = \frac{1}{8}(\operatorname{sign} M - \Sigma \cdot \Sigma) - \operatorname{Arf}(M, \Sigma)$*, for any characteristic* Σ *in* M.

In what follows we will present the motivation and the outline of proof for this generalized Rokhlin theorem. The complete proof is explained in the next note (page 507). It will, in particular, prove the original Rokhlin theorem as well.

---

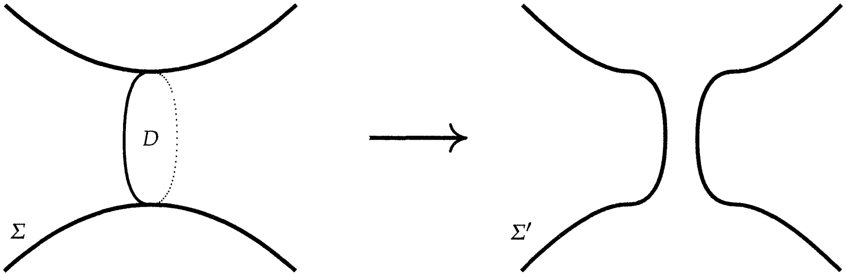
[7] For the Kirby–Siebenmann invariant, see the end-notes of chapter 4 (page 207).

Strategy. Recall van der Blij's lemma,[8] stating that, for every intersection form Q and each of its characteristic elements $\underline{w}$, we must have

$$\operatorname{sign} Q - \underline{w} \cdot \underline{w} = 0 \text{ or } 8 \pmod{16} .$$

Rokhlin's theorem states that 8 is excluded whenever $\underline{w} = 0$, while Kervaire–Milnor shows that 8 is excluded whenever $\underline{w}$ can be represented by a sphere. Therefore, a possible approach to generalizing Rokhlin's theorem is to start by representing a characteristic element $\underline{w}$ by some embedded surface Σ, and then see what obstructions we encounter while trying to modify Σ into a sphere.

Can I make you sphere? An obvious method for reducing the genus of Σ is the following: start with a circle C embedded in Σ that represents some generator in $H_1(\Sigma; \mathbb{Z})$. Assume that we are lucky and can choose a circle C for which there is a disk D embedded in M so that $\partial D = C$. Then we could cut-and-cap Σ along such a disk and thus reduce its genus, as in figure 11.13



11.13. Cutting a surface along a disk

For such a cut to actually be possible, one needs that: (1) the interior of the disk D not intersect Σ; and (2) the normal line-bundle $N_{\partial D/\Sigma}$ of C in Σ be extendable over the whole disk D as a normal line-field inside $N_{D/M}$. This second condition is needed in order to define two parallel copies of D to be used for capping the sliced-up Σ.

Algebraically, these two obstructions are reflected in two numbers: for (1) we have the intersection number

$$\Sigma \cdot D ,$$

while for (2) we can compare the normal line-field $N_{\partial D/\Sigma}$ with a trivialization of $N_{D/M}$, obtaining another integer

$$d_D .$$

Indeed, since D is a disk, there is a unique trivialization, up to homotopy, of the plane-bundle $N_{D/M}$, and therefore over the boundary ∂D we can count the twists of the line-bundle $N_{\partial D/\Sigma}$ with respect to this trivialization of $N_{D/M}$; we denote this twist number by d_D.

---

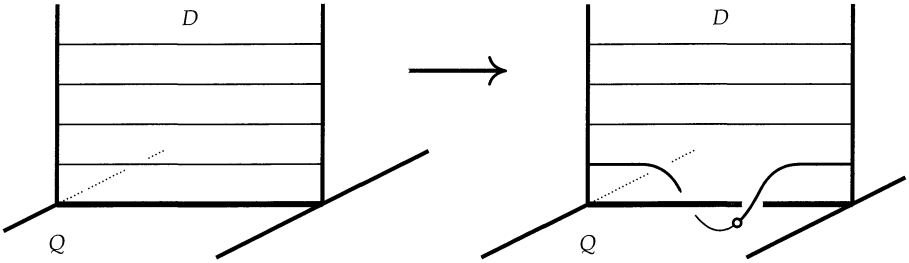
[8] Stated back in section 4.4 (page 170), proved in the end-notes of chapter 5 (page 263).

Spinning, then kissing spheres. Given such a configuration—a characteristic surface Σ, a circle C in it, and an embedded disk D meeting Σ normally along $\partial D = C$ and transversely along $\operatorname{Int} D$—we can modify it.

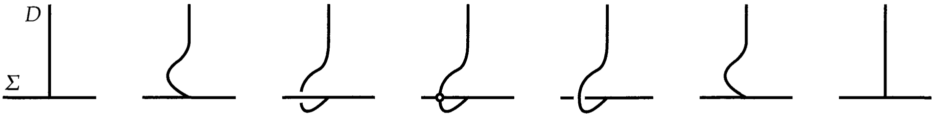
For example, we can spin the disk D around its boundary C, as suggested in figures 11.14 and 11.15. That modifies the framing of $N_{C/D}$ by 1 and creates an intersection point of D and Σ. Namely, we have:

$$d_D \longmapsto d_D \pm 1 \qquad \text{and} \qquad \Sigma \cdot D \longmapsto \Sigma \cdot D \pm 1 \, .$$

Repeating this enough times, we can arrange to have either d_D or $\Sigma \cdot D$ vanish.



11.14. Spinning a disk around its boundary



11.15. Spinning a disk around its boundary: the movie

Thus, a better invariant would seem to be the combined quantity

$$d_D - \Sigma \cdot D \, ,$$

which is unchanged by spinning.

On the other hand, assume that M contains an embedded sphere S. We can then connect D and S by using a thin tube, thus replacing D by the disk $D \# S$. Then the twist d_D changes to $d_D + S \cdot S$, while the intersections $\Sigma \cdot D$ become $\Sigma \cdot D + \Sigma \cdot S$. Hence

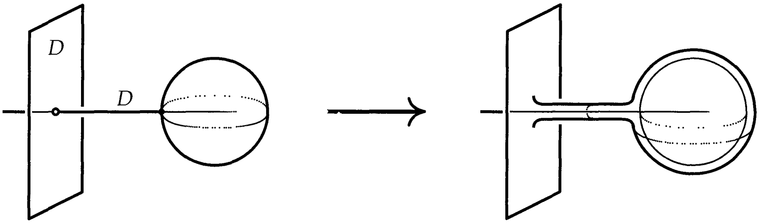
$$d_D - \Sigma \cdot D \longmapsto d_D + S \cdot S - \Sigma \cdot D - \Sigma \cdot S \, .$$

Since Σ is a characteristic surface, we must also have $\Sigma \cdot S = S \cdot S \pmod 2$, and hence we still have an unchanged quantity here, namely

$$q(C) = d_D - \Sigma \cdot D \pmod 2 \, .$$

This seems a good candidate for an obstruction to surger C away from Σ and reduce genus. Of course, there is still the issue of finding *embedded* disks D.

Stabilizations, and Mr Arf. Notice that a sum-stabilization of M (*i.e.*, connect-summing with copies of $\mathbb{S}^2 \times \mathbb{S}^2$) does not change the signature of M, and the class of Σ will still be a characteristic element. Nonetheless, stabilization certainly makes it possible to find embedded disks D for any circle C (by undoing self-intersections of some immersed disk, as recalled in figure[9] 11.16). Therefore, we can assume that for *every* generator C we can find a disk D *embedded* in some $M \# k \mathbb{S}^2 \times \mathbb{S}^2$ and thus obtain corresponding numbers $\Sigma \cdot D$ and d_D.



11.16. Eliminating an intersection by summing with a sphere

Of course, if a circle C does not have a good setting that would allow a genus-reducing cut of Σ, then we can always try our luck with another generator of $H_1(\Sigma; \mathbb{Z})$. It turns out that, up to stabilizations, whether we eventually are successful in cutting Σ all the way to a sphere depends entirely on the Arf invariant $\mathrm{Arf}(M, \Sigma)$ of the quadratic enhancement

$$q \colon H_1(\Sigma; \mathbb{Z}_2) \longrightarrow \mathbb{Z}_2$$

that assigns to each class $x \in H_1(\Sigma; \mathbb{Z}_2)$ the corresponding number

$$q(x) = d_D + \Sigma \cdot D \pmod 2$$

for any circle C representing x on Σ, and any disk D bounded by C. This q can be checked to be well-defined and quadratic with respect to the natural modulo 2 intersection form on $H_1(\Sigma; \mathbb{Z}_2)$.

One could then prove that Σ can be modified to a sphere embedded in some $M \# k\mathbb{S}^2 \times \mathbb{S}^2$ if and only if $\mathrm{Arf}(q) = 0$. In that case, since Σ is still characteristic in $M \# k \mathbb{S}^2 \times \mathbb{S}^2$ and the signature is unchanged, we could apply the Kervaire–Milnor theorem to it and deduce that we must have $\operatorname{sign} M - \Sigma \cdot \Sigma = 0 \pmod{16}$.

That would show that $\mathrm{Arf}(M, \Sigma) = 0$ implies $\operatorname{sign} M - \Sigma = 0 \pmod{16}$.

Definition of* Arf(*M*, Σ) *without stabilizations. We can define the quadratic enhancement q directly, without any use of stabilizations:

Start with a circle C embedded in a characteristic surface Σ. Then, assuming M is simply-connected, C must bound some embedded surface F (allowed to transversely-intersect Σ). Since F is homotopy-equivalent to $\mathbb{S}^1 \vee \cdots \vee \mathbb{S}^1$ and its

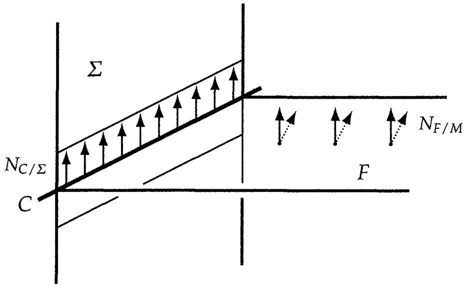
---

[9] See back in section 4.2 (page 149) for the full argument.

normal bundle $N_{F/M}$ is an orientable plane-bundle, it follows that $N_{F/M}$ is a trivial bundle. Moreover, every trivialization of $N_{F/M}$ induces the same *unique* trivialization[10] of $N_{F/M}|_C$. Thus, we always have a naturally trivialized normal bundle $N_{F/M}|_C$, and we can count the twists of the line-bundle $N_{C/\Sigma}$ inside $N_{F/M}|_C$ with respect to this trivialization. We denote the number of twists by d_F, and define

$$q(C) = d_F + \Sigma \cdot F \pmod 2 .$$

By varying the circle C over the elements of $H_1(\Sigma; \mathbb{Z}_2)$, we end up with a well-defined quadratic enhancement $q\colon H_1(\Sigma;\mathbb{Z}_2) \to \mathbb{Z}_2$, which has an associated Arf invariant $\operatorname{Arf}(M, \Sigma)$.



11.17. Comparing framings along C

Characteristic cobordisms and outline of proof. Either way, so far we only outlined an argument for showing that, if $\operatorname{Arf}(M,\Sigma) = 0$, then $\operatorname{sign} M - \Sigma \cdot \Sigma = 0 \pmod{16}$; but not that, if $\operatorname{Arf}(M,\Sigma) = 1$, then the residue must be 8.

The best and natural setting for the whole discussion is the characteristic cobordism group Ω_4^{char}, already discussed earlier in the end-notes of the preceding chapter (page 427), where it was identified with the spin$^{\mathbb{C}}$ cobordism group and fully evaluated. This setting, in fact, allows us to prove the Freedman–Kirby theorem directly, without relying on Rokhlin's theorem (under the guise of Kervaire–Milnor), and thus in particular does prove Rokhlin's theorem itself.

Recall that the **characteristic cobordism group**

$$\Omega_4^{char}$$

is the cobordism group[11] generated by pairs (M, Σ) where M is a smooth 4–manifold and Σ is a characteristic surface inside M. Further, $M \setminus \Sigma$ is endowed with a spin structure that does not extend across Σ. Two such pairs (M', Σ') and (M'', Σ'') are considered cobordant when there is a cobordism W^5 between M' and M'', containing an unoriented 3–manifold Y^3 that is itself a cobordism between Σ' and Σ''. Further, it is required that Y^3 be dual to $w_2(W)$, and that $W \setminus Y$ be endowed with a spin structure not extending across Y and inducing on each of $M' \setminus \Sigma'$ and $M'' \setminus \Sigma''$ their chosen spin structures.

---

[10] Any two trivializations of $N_{F/M}$ differ by a map $g\colon F \to SO(2)$. Since ∂F is homologically-trivial in F, and since $SO(2) = S^1 = K(\mathbb{Z}, 1)$ and so $H^1(F;\mathbb{Z}) = [F, SO(2)]$, a simple argument shows that $g|_{\partial F}$ must always be homotopic to the trivial map $g|_{\partial F} \equiv 1$.

[11] Compare with other cobordism groups in the end-notes of chapter 4 (page 227).

It was proved in the end-notes of the preceding chapter (page 427) that we have an isomorphism

$$\Omega_4^{char} \approx \mathbb{Z} \oplus \mathbb{Z} \qquad \textit{via} \qquad (M, \Sigma) \longmapsto \left(\operatorname{sign} M,\ \tfrac{1}{8}(\Sigma \cdot \Sigma - \operatorname{sign} M)\right) ,$$

and that Ω_4^{char} thus has generators $\left(\mathbb{CP}^2, \mathbb{CP}^1\right)$ and $\left(\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2,\ \# 3\, \mathbb{CP}^1 \# \overline{\mathbb{CP}}^1\right)$.

To prove the Freedman–Kirby theorem, one shows that the Arf invariant of a characteristic surface Σ inside some M is invariant under cobordisms. In other words, that the Arf invariant induces a well-defined morphism

$$\operatorname{Arf} \colon \Omega_4^{char} \longrightarrow \mathbb{Z}_2 .$$

It is then checked directly on the generators of Ω_4^{char} that

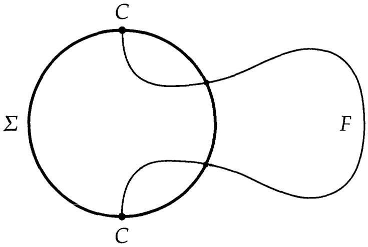
$$\operatorname{Arf}(\Sigma) = \tfrac{1}{8}(\operatorname{sign} M - \Sigma \cdot \Sigma) \pmod 2 .$$

Then this relation must extend to the whole group Ω_4^{char}, and hence finishes the proof. The full details are presented in the following note.

Note: Proof of Freedman–Kirby's and Rokhlin's theorems

In the preceding note (starting on page 502) we have provided motivation and heuristics for the proof to follow. Other requisites for fully absorbing this note is the earlier note defining the Arf invariant of an algebraic quadratic enhancement (page 501), as well as the end-note of the preceding chapter (page 427) where the characteristic cobordism group Ω_4^{char} was shown to be $\mathbb{Z} \oplus \mathbb{Z}$.

Definition of the quadratic enhancement. Let Σ be a characteristic surface in M. Let C be any family of circles embedded in Σ, and let F be any oriented surface bounded by C. Arrange that F meets Σ normally along $C = \partial F$ and transversely otherwise, see figure 11.18. At times we will call F a **membrane** for the circles C.



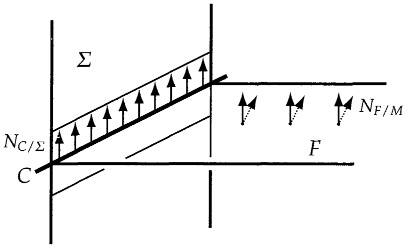
11.18. Characteristic surface, 1–cycle, and membrane

The normal bundle $N_{C/\Sigma}$ of the circles C in Σ is a trivial real-line bundle. Since F is an orientable surface with boundary, it is homotopically-equivalent to $S^1 \vee \cdots \vee S^1$, and thus its (orientable) normal bundle $N_{F/M}$ in M is a trivial real-plane bundle. The line bundle $N_{C/\Sigma}$ is a subbundle of the plane bundle $N_{F/M}|_C$, and we can count how many times $N_{C/\Sigma}$ twists with respect to some trivialization of $N_{F/M}$.

Specifically, choose a trivializing frame f_σ of $N_{C/\Sigma}$ (hence f_σ is tangent to Σ) and add a complementary section[12] f_m (normal to both F and Σ) to obtain the trivialization $fr_\Sigma = \{f_\sigma, f_m\}$ of $N_{F/M}|_C$ induced by Σ. On the other hand, choose a trivialization fr_F of the whole $N_{F/M}$ over F. Compare the two induced framings fr_Σ and $fr_F|_C$ of $N_{F/M}|_C$. Since $\pi_1 SO(2) = \mathbb{Z}$, the result of the comparison over each circle is an integer (which does not depend on the chosen trivialization of $N_{F/M}$). We reduce these integers modulo 2 and add them over all circles of C, obtaining an element

$$d_C = d_C(fr_\Sigma, fr_F) \in \mathbb{Z}_2 .$$

In other words, d_C is the modulo 2 count of the twists of $N_{C/\Sigma}$ with respect to a trivialization of $N_{F/M}$; that is to say, d_C is the number of twists of Σ along C with respect to a trivialization of $N_{F/M}$.



11.19. Comparing framings along C

That d_C is independent of the various choices can be seen by thinking of d_C as the obstruction to extending the trivialization fr_Σ of $N_{F/M}|_C$ across the whole membrane F. In other words, we have

$$d_C(fr_\Sigma, fr_F) = w_2(N_{F/M}, fr_\Sigma)[F, \partial] ,$$

where $w_2(N_{F/M}, fr_\Sigma) \in H^2(F, \partial F; \mathbb{Z}_2)$ is the Stiefel–Whitney class of $N_{F/M}$ *relative* to the framing fr_Σ on $\partial F = C$, while $[F, \partial] \in H_2(F, \partial F; \mathbb{Z}_2)$ is the $\mathbb{Z}_2$-orientation cycle of F. Of course, here we think of w_2 in terms of obstruction theory.[13] Notice that, if we actually define d_C by the above formula, then there is no need for the membrane F to be oriented or even orientable.

Another numerical data that can be extracted from F is the number of intersections (modulo 2) of Int F and Σ; denote the latter simply by

$$\Sigma \cdot F \in \mathbb{Z}_2 .$$

We add these two to obtain

$$q(F) = d_C(fr_\Sigma, fr_F) + \Sigma \cdot F .$$

As we will see, this q depends only on C and not on F.

Furthermore, we will prove that q is invariant under appropriate cobordisms of the configuration (M, Σ, C), and hence in particular invariant under cobordisms

---

[12] The reason for these awkward notations is that the frame $\{f_\sigma, f_m\}$ will soon be part of a 5-frame $\{f_\sigma, f_m, f_f, f_c, f_v\}$, where simple numerical indices would be harder to follow. We will set f_f tangent to F, f_c tangent to C, and f_v normal to M, which left us to denote by f_m the one that is tangent to M.

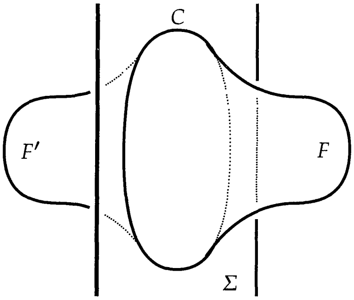
[13] For an overview of obstruction theory, see the end-notes of chapter 4 (page 197).

that merely change the C's. Therefore, q in fact descends to a well-defined map $q\colon H_1(\Sigma;\mathbb{Z}_2) \to \mathbb{Z}_2$, which is a quadratic enhancement of the intersection form of Σ and thus has an Arf invariant, denoted by $\operatorname{Arf}(M,\Sigma)$.

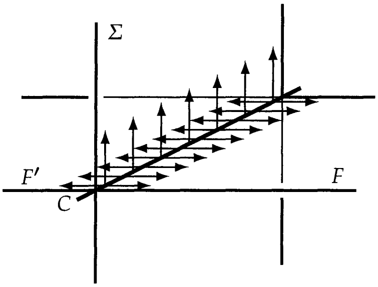
Independence from membrane. As a warm-up, we start by proving that q depends only on C and not on F, even though this is also a consequence of the cobordism invariance that we will prove afterwards.

Lemma. *The number $q(F)$ depends only on the family of circles $C = \partial F$*

> ***Proof.*** As in figure 11.20, take another surface F' bounded by the same circles C. Spin F' around C until the inner-pointing normals of C in F and in F' are everywhere-opposite along C, as in figure 11.21, and thus $F \cup F'$ makes up a closed immersed surface in M. Notice that such a 2π-spin of F' around C changes $d_C(fr_\Sigma, fr_{F'})$ by 1 but also creates an intersection point between $\operatorname{Int} F'$ and Σ, and therefore $q(F')$ is unchanged by this spinning. Denote by $B = F \cup F'$ the immersed closed surface formed from F and F'.



11.20. Two membranes with the same boundary



11.21. Fit F and F' into an immersed surface B

Trivialize $N_{F/M}$ with a framing fr_F, then trivialize $N_{F'/M}$ with a framing $fr_{F'}$. These two induce framings along C, which can be compared modulo 2 yielding what we denote by $d_C(fr_F, fr_{F'}) \in \mathbb{Z}_2$. We can also use these framings to push B off itself and compute its self-intersection modulo 2. A part of the self-intersections is counted by $d_C(fr_F, fr_{F'})$, the other part is caught by $F \cdot F'$. Since any intersection point of F and F' is a double point of B, it creates two intersection points between B and its push-off, as in figure 11.22. Thus, the contribution of $F \cdot F'$ is invisible modulo 2. We are thus left with

$$B \cdot B = d_C(fr_F, fr_{F'}) \ .$$

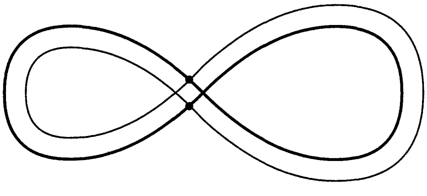
The latter can be split into the sum of the comparisons of each of fr_F and $fr_{F'}$ with the framing fr_Σ induced by the position of Σ in $N_{B/M}$:

$$B \cdot B = d_C(fr_\Sigma, fr_F) + d_C(fr_\Sigma, fr_{F'}) \pmod 2 \ .$$

On the other hand, Σ is a characteristic surface in M, and hence

$$B \cdot B = \Sigma \cdot B = \Sigma \cdot F + \Sigma \cdot F' \pmod 2 \ .$$

It follows that $q(F) = q(F') \pmod 2$. □



11.22. Double-points and self-intersections

Invariance under cobordisms. Consider now two configurations (M', Σ', C') and (M'', Σ'', C''), made of a 4–manifold, a characteristic surface embedded in it, and a family of circles embedded in the surface. We will call two such configurations cobordant if there exists a configuration (W, Y, A), made of an oriented 5–manifold W^5, an embedded unoriented 3–manifold Y^3 that is Poincaré-dual to $w_2(T_W)$, and an unoriented surface A embedded in Y, such that $\partial W = \overline{M'} \cup M''$, $\partial Y = \Sigma' \cup \Sigma''$, and $\partial A = C' \cup C''$. See figure[14] 11.23 on the facing page. We prove that two cobordant configurations lead to the same value of q:

Lemma. *If (M', Σ', C') and (M'', Σ'', C'') are cobordant, then $q(C') = q(C'')$.*

Proof. The proof is a refinement of the previous proof, with the added unpleasantness of having to cross the bridge A between C' and C''.

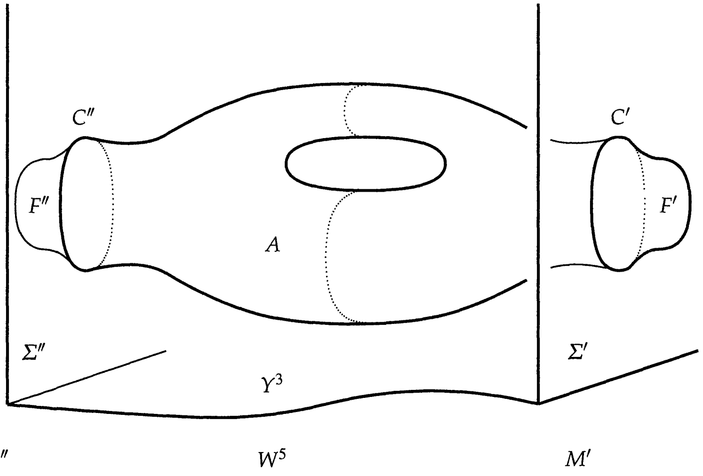
Choose a membrane F' in M' with $\partial F' = C'$ and normal to Σ', and similarly F'' in M''. We want to show that

$$d_{C'}(fr_{\Sigma'}, fr_{F'}) + \Sigma' \cdot F' \equiv d_{C''}(fr_{\Sigma''}, fr_{F''}) + \Sigma'' \cdot F'' \pmod 2 \ .$$

We build the closed surface $B = F' \cup A \cup F''$. The manifold Y^3 determines a class in $H_3(W, \partial W; \mathbb{Z}_2)$, dual to $w_2(T_W) \in H^2(W; \mathbb{Z}_2)$; the surface B determines a class in $H_2(W; \mathbb{Z}_2)$. We can thus couple Y and B and compute their intersection pairing $Y \cdot B$. The strategy of the proof is to compute $Y \cdot B$ in two ways, once using geometric intersections, and once by evaluating $w_2(T_W)$ on B. The argument stretches on through page 514.

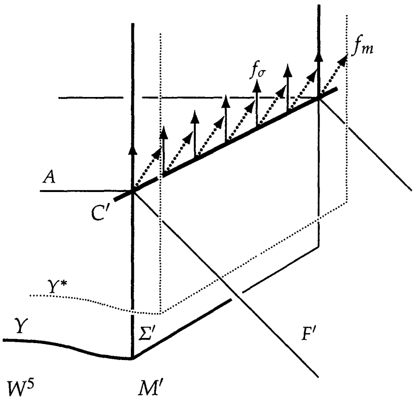
---

[14] In figure 11.23, notice how A does not smoothly continue F' or F''. Indeed, F' is normal to Σ' in M', while Y starts near M' as Σ' crossed with its normal direction into W. Therefore, F' is normal to Y at C', and thus normal to A. Similarly for F'' and A.



11.23. (W, Y, A) is a cobordism between (M', Σ', C') and (M'', Σ'', C'')

Evaluate by intersections. On one hand, we push Y off itself and count intersections with B. This push-off must be done by using a vector field normal to Y, but also tangent to ∂W and normal to F' and F''. Indeed, we need the pushed-off Y^* to still be a nicely embedded 3–manifold with $\partial Y^* \subset \partial W$, and we want its boundary to be transverse to F' and F''. Such a vector field can be obtained by extending across Y the vector field f_m on C', which was normal to both Σ' and F'; similarly along C''. See figure 11.24. (We have already encountered f_m as the second component of the framing $fr_{\Sigma'} = \{f_\sigma, f_m\}$ of $N_{F'/M'}|_{C'}$.)



11.24. Pushing Y off itself, parallel to ∂W

After pushing off, we can count the intersections with each of F', F'', and A:

$$Y \cdot B = \Sigma' \cdot F' + Y \cdot A + \Sigma'' \cdot F'' .$$

The intersection number $Y \cdot A$ is an intersection in the 5–manifold W, while $\Sigma' \cdot F'$ and $\Sigma'' \cdot F''$ are intersections inside the 4–manifolds M' and M''.

The middle term $Y \cdot A$ can be interpreted as the obstruction to extending f_m across A as a nowhere-zero vector field normal to Y. We complete f_m to a framing $\{f_m, f_f\}$ of $N_{Y/W}|_{C'}$ by adding the inner-pointing normals f_f of C' in F', see figure 11.25. (Indeed, since $N_{Y/W}|_{M'} = N_{\Sigma'/M'}$, and both f_m and f_f are normal to Σ', they are normal to Y as well.) Similarly along C''. Thus, we have

$$Y \cdot A \;=\; w_2\big(N_{Y/W},\ \{f_m, f_f\}\big)[A, \partial] \quad (\text{mod } 2)\ .$$

Therefore

$$Y \cdot B \;=\; \Sigma' \cdot F' \;+\; \Sigma'' \cdot F'' \;+\; w_2\big(N_{Y/W},\ \{f_m, f_f\}\big)[A, \partial]\ .$$

Evaluate by Stiefel–Whitney. On the other hand, since Y is dual to $w_2(T_W)$, we evaluate

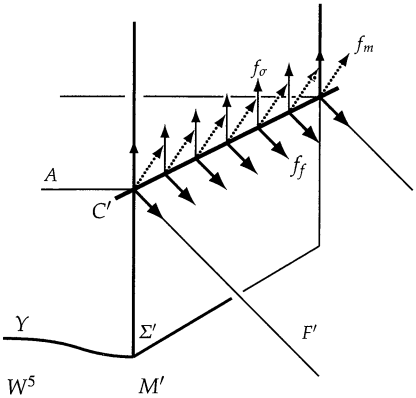
$$\begin{aligned} Y \cdot B \;&=\; w_2(T_W)[B] \quad (\text{mod } 2) \\ &=\; w_2(T_W|_B) \;=\; w_2(T_B \oplus N_{B/W}) \\ &=\; w_2(T_B) \;+\; w_2(N_{B/W}) \;+\; w_1(T_B) \cdot w_1(N_{B/W})\ . \end{aligned}$$

However, W is oriented, so $w_1(T_W|_B) = 0$, and hence $w_1(T_B) + w_1(N_{B/W}) = 0$. Therefore, we have:

$$\begin{aligned} Y \cdot B \;&=\; w_2(T_B) \;+\; w_2(N_{B/W}) \;-\; \big(w_1(T_B)\big)^2 \\ &=\; w_2(T_B) \;+\; w_2(N_{B/W}) \;-\; w_2(T_B)\ , \end{aligned}$$

after using Wu's formula for surfaces. So:

$$Y \cdot B \;=\; w_2(N_{B/W})\ .$$



11.25. Framing the bridge

Take the vector field f_f normal to C' in F'; it is normal to Σ' and thus normal to Y. Then the frame $\{f_\sigma, f_m, f_f\}$ trivializes $N_{A/W}|_{C'}$ (with f_σ spanning $N_{A/Y}$ and $\{f_m, f_f\}$ spanning $N_{Y/W}$), as suggested in figure 11.25. Similarly along C''. Using this framing of $N_{B/W}$ along the seams $C' \cup C''$, we split $w_2(N_{B/W})$

into relative classes following the splitting $B = F' \cup F'' \cup A$:

$$\begin{aligned} Y \cdot B &= w_2\big(N_{F'/M'} \cup N_{F''/M''} \cup N_{A/W}\big)[B] \\ &= w_2\big(N_{F'/M'},\ \{f_\sigma, f_m\}\big)[F', \partial] \ +\ w_2\big(N_{F''/M''},\ \{f_\sigma, f_m\}\big)[F'', \partial] \\ &\qquad\qquad +\ w_2\big(N_{A/W},\ \{f_\sigma, f_m, f_f\}\big)[A, \partial] \\ &= d_C(fr_{\Sigma'}, fr_{F'}) \ +\ d_{C''}(fr_{\Sigma''}, fr_{F''}) \ +\ w_2\big(N_{A/W},\ \{f_\sigma, f_m, f_f\}\big)[A, \partial]\,. \end{aligned}$$

We will show that the last term evaluates to $Y \cdot A$.

Middle intersections. Split the 3–plane bundle $N_{A/W}$ into the line bundle $N_{A/Y}$ and the 2–plane bundle $N_{Y/W}$:

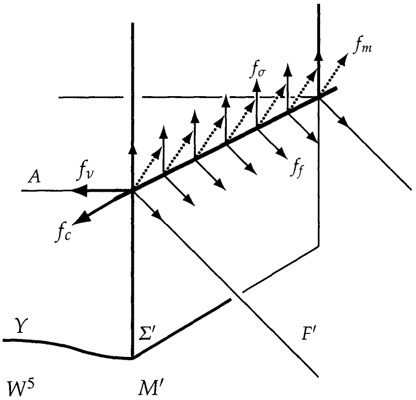
$$N_{A/W} = N_{A/Y} \oplus N_{Y/W}\ .$$

Along the seam C', the bundle $N_{A/Y}$ coincides with $N_{C'/\Sigma'}$, and thus is framed by f_σ. Similarly along C''. The bundle $N_{Y/W}$ is framed along $C' \cup C''$ by $\{f_m, f_f\}$. We thus write:

$$\begin{aligned} & w_2\big(N_{A/W},\ \{f_\sigma, f_m, f_f\}\big) \\ &\quad = w_2\big(N_{A/Y} \oplus N_{Y/W},\ f_\sigma \oplus \{f_m, f_f\}\big) \\ &\quad = w_1(N_{A/Y},\ f_\sigma) \cdot w_1\big(N_{Y/W},\ \{f_m, f_f\}\big) \ +\ w_2\big(N_{Y/W},\ \{f_m, f_f\}\big)\ . \end{aligned}$$

We argue that the product-term vanishes.

The framing $\{f_\sigma, f_m, f_f\}$ of $N_{A/W}|_{C'}$ can be completed by using a section f_c of $T_{C'}$ and the normal field f_v of C' in A, to a framing $\{f_\sigma, f_m, f_f, f_c, f_v\}$ of $T_W|_{C'}$, as in figure 11.26. Similarly over C''.



11.26. Framing the circles

We have:

$$\begin{aligned} & w_1\big(T_W,\ \{f_v, f_c, f_\sigma, f_m, f_f\}\big) = 0 \\ & w_1\big(T_A \oplus N_{A/Y} \oplus N_{Y/W},\ \ \{f_v, f_c\} \oplus f_\sigma \oplus \{f_m, f_f\}\big) = 0 \\ & w_1(T_A,\ \{f_v, f_c\}) + w_1(N_{A/Y}, f_\sigma) + w_1\big(N_{Y/W},\ \{f_m, f_f\}\big) = 0 \\ & w_1(T_A,\ \{f_v, f_c\}) = w_1(N_{A/Y}, f_\sigma) + w_1\big(N_{Y/W},\ \{f_m, f_f\}\big) \quad (\mathrm{mod}\ 2)\ . \end{aligned}$$

Applying Wu's formula, we have that $w_1(T_A, \{f_v, f_c\}) \cdot \alpha = \alpha \cdot \alpha$ for all $\alpha \in H^1(A, \partial A; \mathbb{Z}_2)$. For $\alpha = w_1(N_{A/Y}, f_\sigma)$, this becomes:

$$\begin{aligned} w_1(T_A, \{f_v, f_c\}) \cdot w_1(N_{A/Y}, f_\sigma) &= \big(w_1(N_{A/Y}, f_\sigma)\big)^2 \\ \big(w_1(N_{A/Y}, f_\sigma) + w_1(N_{Y/W}, \{f_m, f_f\})\big) \cdot w_1(N_{A/Y}, f_\sigma) &= \big(w_1(N_{A/Y}, f_\sigma)\big)^2 \\ \big(w_1(N_{A/Y}, f_\sigma)\big)^2 + w_1(N_{Y/W}, \{f_m, f_f\}) \cdot w_1(N_{A/Y}, f_\sigma) &= \big(w_1(N_{A/Y}, f_\sigma)\big)^2 \\ w_1(N_{Y/W}, \{f_m, f_f\}) \cdot w_1(N_{A/Y}, f_\sigma) &= 0\,. \end{aligned}$$

Tracking backwards, we have:

$$\begin{aligned} & w_2(N_{A/W}, \{f_\sigma, f_m, f_f\}) \\ &\quad = w_1(N_{A/Y}, f_\sigma) \cdot w_1(N_{Y/W}, \{f_m, f_f\}) + w_2(N_{Y/W}, \{f_m, f_f\}) \\ &\quad = w_2(N_{Y/W}, \{f_m, f_f\})\,. \end{aligned}$$

Recall that f_m was the vector field normal to Y that we used to push Y off itself and compute the intersection $Y \cdot A$. Then the obstruction to extending $\{f_m, f_f\}$ over A is that intersection number, and hence

$$w_2(N_{A/W}, \{f_\sigma, f_m, f_f\})[A, \partial] = Y \cdot A\,.$$

Tracking back some more, we have

$$\begin{aligned} Y \cdot B &= d_{C'}(fr_{\Sigma'}, fr_{F'}) + d_{C''}(fr_{\Sigma''}, fr_{F''}) + w_2(N_{A/W}, \{f_\sigma, f_m, f_f\})[A, \partial] \\ &= d_{C'}(fr_{\Sigma'}, fr_{F'}) + d_{C''}(fr_{\Sigma''}, fr_{F''}) + Y \cdot A\,, \end{aligned}$$

and the same intersection $Y \cdot B$ was also evaluated at the beginning as

$$Y \cdot B = \Sigma' \cdot F' + \Sigma'' \cdot F'' + Y \cdot A\,.$$

It follows that

$$d_{C'}(fr_{\Sigma'}, fr_{F'}) + \Sigma \cdot F = d_C(fr_\Sigma, fr_{F''}) + \Sigma \cdot F'' \pmod 2$$

or, in other words, that $q(C') = q(C'')$, which is what we set out to show. □

We have a quadratic. An immediate consequence of the invariance of q under cobordisms is

Corollary. *The formula $q(C) = d_C(fr_\Sigma, fr_F) + \Sigma \cdot F$ describes a well-defined function $q \colon H_1(\Sigma; \mathbb{Z}_2) \to \mathbb{Z}_2$.* □

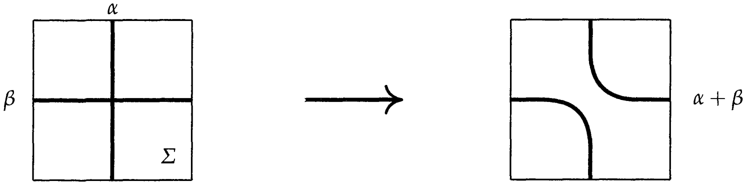
When M is not simply-connected, not all elements of $H_1(\Sigma; \mathbb{Z}_2)$ bound membranes in M. Nonetheless, we can modify M by a cobordism in the usual way, by adding 2–handles to kill $H_1(M; \mathbb{Z})$ without touching Σ, and then every C in Σ will bound some surface F as required. Invariance under cobordisms ensures that the result does not depend on the particular surgery.

Lemma. *The function $q \colon H_1(\Sigma; \mathbb{Z}) \to \mathbb{Z}_2$ is a quadratic enhancement of the intersection form of Σ.*

Proof. Assume that α and β are two 1–cycles from $H_1(\Sigma; \mathbb{Z}_2)$ and assume that they are represented by two circles meeting at only one point. Then a circle representing $\alpha + \beta$ is obtained simply by eliminating the double point, as in figure 11.27. We wish to show that

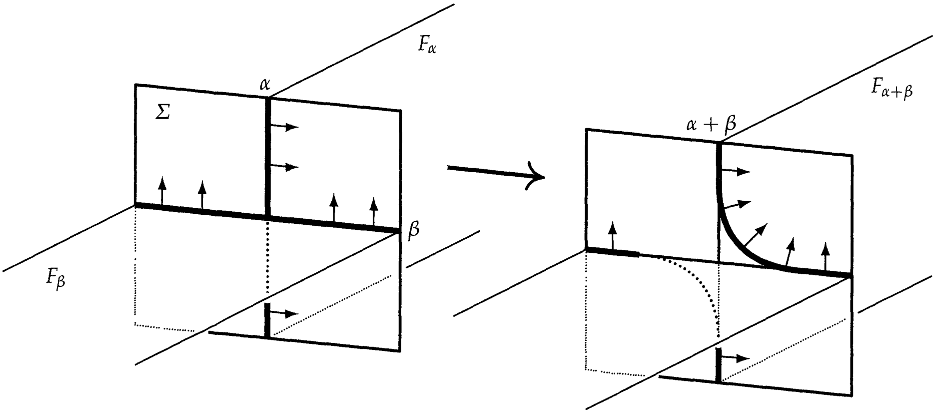
$$q(\alpha + \beta) = q(\alpha) + q(\beta) + \alpha \cdot \beta \pmod 2 ,$$

which in this case has $\alpha \cdot \beta = 1$.



11.27. Adding two 1–cycles on Σ

Let F_α and F_β be the corresponding membranes for α and β. Figures 11.28 and 11.29 explain how to obtain a membrane $F_{\alpha+\beta}$ for the sum, essentially by gluing a twisted ribbon to unite F_α and F_β, then pushing the interior of the ribbon off Σ in the fourth dimension.



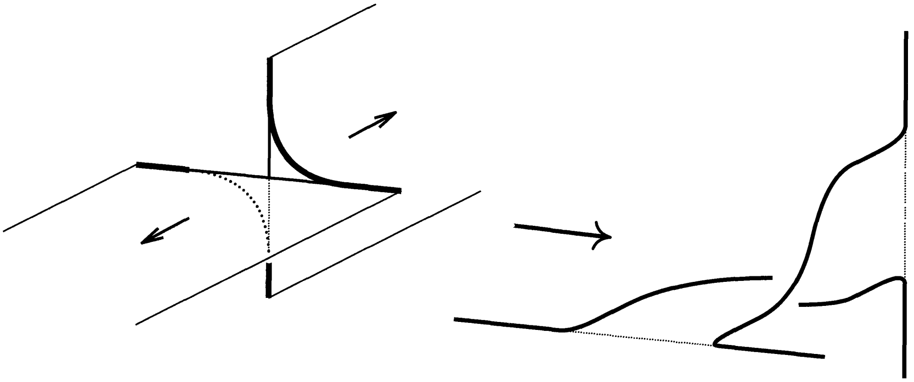
11.28. Building the membrane for the sum

To compute $q(\alpha + \beta)$, on one hand we notice that $F_{\alpha+\beta}$ intersects Σ only where F_α and F_β did, but not along the added ribbon. Thus,

$$\Sigma \cdot F_{\alpha+\beta} = \Sigma \cdot F_\alpha + \Sigma \cdot F_\beta .$$

On the other hand, in order to compute the obstruction term $d_{\alpha+\beta}$, we extend the normal framing of α and of β to a framing of $\alpha + \beta$ in Σ, as suggested back in figure 11.28. The obstruction $d_{\alpha+\beta}$ to extending this framing as a normal field across all $F_{\alpha+\beta}$ is the obstruction d_α to extend across F_α, together with the obstruction d_β to extend across F_β, together, finally, with the obstruction to extend across the ribbon.

The obstruction to extend the normal framing over the ribbon is the self-intersection of the ribbon; that is to say, we push the ribbon off itself so that its

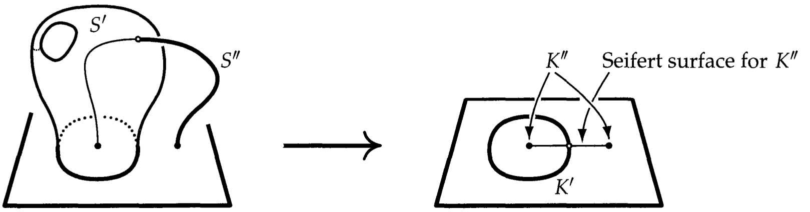


11.29. Understanding the sum-membrane

boundary moves according to its normal framing, and we count intersection points modulo 2. We remember, from an argument made during the proof of Whitehead's theorem (page 148) and recalled in figure 11.30, that the intersection numbers of two surfaces in dimension 4, bounding knots in a 3–dimensional slice, are given by the linking numbers of those knots. Thus, it is enough to focus on the boundary of our ribbon, push it off using the framing to get another loop, then notice that the two circles link once in the 3–dimensional slice at which we are looking. See figure 11.31 on the facing page. Therefore, the ribbon and its push-off will have to meet once, and hence the obstruction over the ribbon is 1. We have

$$d_{\alpha+\beta} = d_\alpha + d_\beta + 1\,.$$

Therefore, $q(\alpha+\beta) = q(\alpha) + q(\beta) + \alpha\cdot\beta$, as we wished. The other cases are argued in a similar style. □



11.30. The linking numbers of the K's are the intersection numbers of the S's

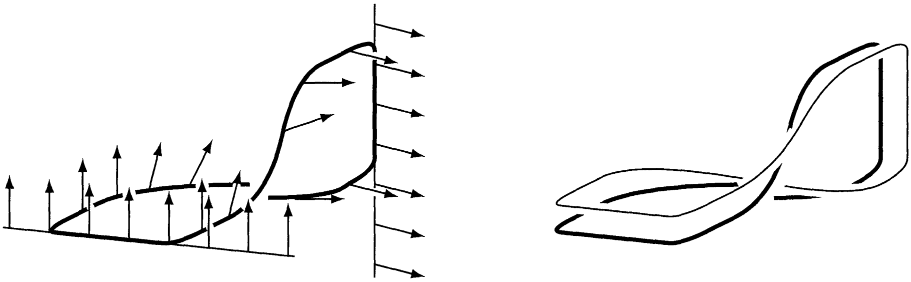
The Arf invariant. Since now we know q to be a quadratic enhancement, of course we look at its complete invariant, namely its Arf invariant. Thus, for any 4–manifold M and embedded characteristic surface Σ, we denote by

$$\mathrm{Arf}(M,\Sigma) \in \mathbb{Z}_2$$

the Arf invariant of the corresponding q.

Lemma. *The Arf invariant* $\mathrm{Arf}(q)$ *of* q *depends only on the characteristic cobordism class* (M,Σ). *Therefore, we have a well-defined additive morphism*

$$\mathrm{Arf}\colon \Omega_4^{char} \longrightarrow \mathbb{Z}_2\,.$$



11.31. Measuring the new obstruction

Proof. Let (M', Σ') and (M'', Σ'') be cobordant in Ω_4^{char} through (W, Y). Consider the inclusion $i \colon \Sigma' \cup \Sigma'' \subset Y$ of Y's boundary and denote its 1–homology kernel by

$$K = \operatorname{Ker}\bigl(i_* \colon H_1(\Sigma' \cup \Sigma''; \mathbb{Z}_2) \to H_1(Y; \mathbb{Z}_2)\bigr) .$$

Following a modulo 2 version of the argument used for proving that signatures of boundaries must vanish (page 120), we have[15]

$$\dim K = \tfrac{1}{2} \dim H_1(\Sigma' \cup \Sigma''; \mathbb{Z}_2) .$$

Consider $q|_{\Sigma' \cup \Sigma''} = q|_{\Sigma'} \oplus q|_{\Sigma''}$, defined on $H_1(\Sigma' \cup \Sigma''; \mathbb{Z}_2)$; clearly

$$\operatorname{Arf}(q|_{\Sigma' \cup \Sigma''}) = \operatorname{Arf}(q|_{\Sigma'}) + \operatorname{Arf}(q|_{\Sigma''}) .$$

We will now argue that $q|_{\Sigma' \cup \Sigma''}$ vanishes on all elements of K. Indeed, let $C' \cup C''$ represent a class in K and let A be a surface in Y such that $\partial A = C' \cup C''$. Now, instead of interpreting (W, Y, A) as a cobordism from (M', Σ', C') to (M'', Σ'', C''), view it instead as a cobordism from $(M' \cup \overline{M''},\ \Sigma' \cup \Sigma'',\ C' \cup C'')$ to $(\varnothing, \varnothing, \varnothing)$. Since we proved q to be invariant by cobordisms, it follows that

$$q(C' \cup C'') = q(\varnothing) = 0 .$$

Therefore q vanishes on all K.

Then q cannot be 1 on a majority of elements of $H_1(\Sigma' \cup \Sigma''; \mathbb{Z}_2)$, and hence $\operatorname{Arf}(q|_{\Sigma' \cup \Sigma''}) = 0$ (from the "voting lemma", page 502). By additivity, it follows that $\operatorname{Arf}(q|_{\Sigma'}) = \operatorname{Arf}(q|_{\Sigma''})$, better written as $\operatorname{Arf}(M', \Sigma') = \operatorname{Arf}(M'', \Sigma'')$. □

Corollary. *The Arf invariant* $\operatorname{Arf}(M, \Sigma)$ *depends only on the homology class of* Σ*; thus, we can define* $\operatorname{Arf}(M, \underline{w})$ *for any characteristic element* $\underline{w}$ *of* M. □

---

[15] Review: Argue first that, if c', c'' both bound some a', a'', then $c' \cdot c'' = 0$, since their intersection points are linked by the arcs making up $a' \cap a''$. Then, if c does not bound, then it admits a dual $a \in H_2(Y, \partial Y; \mathbb{Z}_2)$ so that $c \cdot a = 1$ in Y; then $c^* = \partial a$ in ∂Y will have $c \cdot c^* = 1$ and, since it bounds, $c^* \cdot c^* = 0$. Therefore, c and c^* span a unimodular subspace for the modulo 2 intersection form of ∂Y; split it off and proceed. In the end, we obtain a splitting of $H_1(\partial Y; \mathbb{Z}_2)$ into c's that do not bound and c^*'s that do. Hence $\dim K = \frac{1}{2} \dim H_1(\partial Y; \mathbb{Z}_2)$.

Characteristic cobordisms. First, two fundamental examples:

Lemma. *We have* $\mathrm{Arf}(\mathbb{CP}^2,\ \mathbb{CP}^1) = 0$ *and* $\mathrm{Arf}(\mathbb{CP}^2,\ \#3\,\mathbb{CP}^1) = 1$.

> ***Proof.*** Since $\mathbb{CP}^1$ is a sphere and thus has no 1–homology, the first case follows trivially.
>
> For the second case, notice that the class $\#3\,\mathbb{CP}^1$ can be represented by a torus. One way to see that its Arf invariant is 1 is as follows: Since the Arf invariant is a cobordism invariant, we have the additivity
>
> $$\mathrm{Arf}(M' \# M'',\ \Sigma' \# \Sigma'') = \mathrm{Arf}(M', \Sigma') + \mathrm{Arf}(M'', \Sigma'') .$$
>
> Since $\mathrm{Arf}(\mathbb{CP}^2,\ \mathbb{CP}^1) = 0$, it follows that
>
> $$\mathrm{Arf}(\mathbb{CP}^2,\ \#3\,\mathbb{CP}^1) = \mathrm{Arf}(\mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2,\ \ \#3\,\mathbb{CP}^1 \# 9\,\overline{\mathbb{CP}}^1) .$$
>
> The latter is simply $\mathrm{Arf}(E(1),\ T)$, where $E(1)$ is the rational elliptic surface discussed in section 8.1 (page 302) and T is a regular torus fiber.[16] If we choose the fibration so that T sits near a cusp fiber, then that implies that two circle generators of $H_1(T; \mathbb{Z}_2)$ are vanishing cycles, *i.e.*, they bound disks D embedded in $M \setminus T$ with self-intersection -1. However, this self-intersection modulo 2 is what we denoted by $d_C(fr_T, fr_D)$, *i.e.*, the obstruction to extending $N_{C/T}$ across D. Therefore q assigns 1 to both generators, and hence $\mathrm{Arf}(q) = 1$. □

Now remember the characteristic cobordism group Ω_4^{char}, with

$$\Omega_4^{char} \approx \mathbb{Z} \oplus \mathbb{Z} \qquad \textit{via} \qquad (M, \Sigma) \longmapsto \big(\operatorname{sign} M,\ \tfrac{1}{8}(\Sigma \cdot \Sigma - \operatorname{sign} M)\big) .$$

This was proved in the end-notes of the preceding chapter (page 427). Then the characteristic cobordism group can be viewed as generated by

$$(\mathbb{CP}^2, \mathbb{CP}^1) \equiv (1,0) \qquad \text{and} \qquad (\mathbb{CP}^2,\ \#3\,\mathbb{CP}^1) \equiv (1,1) .$$

Both functions

$$(M, \Sigma) \longmapsto \tfrac{1}{8}(\Sigma \cdot \Sigma - \operatorname{sign} M) \qquad \text{and} \qquad (M, \Sigma) \longmapsto \mathrm{Arf}(M, \Sigma)$$

are additive group morphisms $\Omega_4^{char} \to \mathbb{Z}_2$, and they coincide on the generators. Therefore, they must coincide over the whole Ω_4^{char}, and we have:

Corollary (*Freedman–Kirby Theorem*). *For all characteristic surfaces Σ embedded in any smooth 4–manifold M, we have*

$$\mathrm{Arf}(M, \Sigma) = \tfrac{1}{8}(\Sigma \cdot \Sigma - \operatorname{sign} M) \pmod 2 . \qquad \square$$

Corollary (*Kervaire–Milnor Theorem*). *If M is smooth and contains a characteristic sphere Σ, then*

$$\Sigma \cdot \Sigma = \operatorname{sign} M \pmod{16} . \qquad \square$$

Corollary (*Rokhlin's Theorem*). *If M is smooth and has $w_2(T_M) = 0$, then*

$$\operatorname{sign} M = 0 \pmod{16} . \qquad \square$$

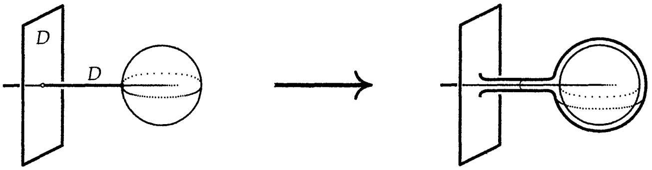
---

[16] Indeed, recall how $E(1)$ was obtained: we picked a cubic family, containing tori (some singular) meeting in nine points, blew up those nine points, and obtained an elliptic fibration; its fiber is $\#3\,\mathbb{CP}^1$ blown-up nine times, *i.e.*, summed with one $\overline{\mathbb{CP}}^1$ from each of the nine $\overline{\mathbb{CP}}^2$'s.

Stable genus bounds. Getting back to the main topic of this chapter—minimum genus of surfaces—a consequence of the above discussion is a solution of the genus problem, up to stabilizations:

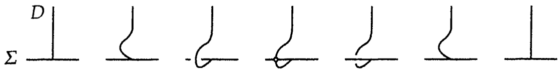
Theorem *(M. Freedman and R. Kirby). Assume M is simply-connected, and let $\underline{w}$ be a characteristic element of M. If $\operatorname{Arf}(M, \underline{w}) = 0$, then there is an integer k so that $\underline{w}$ can be represented by an embedded sphere in $M \# k S^2 \times S^2$. If $\operatorname{Arf}(M, \underline{w}) = 1$, then $\underline{w}$ can be represented by an embedded torus.*

> ***Proof.*** Recall from section 4.2 (page 149) that the main advantage of stabilizing is that the extra spheres from $S^2 \times S^2$ allow us to undo intersections and self-intersections.
>
> Let Σ be a characteristic surface in M and assume that $\operatorname{Arf}(M, \Sigma) = 0$. Then there must be a subspace K of $H_1(\Sigma; \mathbb{Z}_2)$ so that q vanishes on it, the intersection pairing vanishes as well, and K covers half of $H_1(\Sigma; \mathbb{Z}_2)$, *i.e.*, $\dim K = \frac{1}{2} \dim H_1(\Sigma; \mathbb{Z}_2)$.
>
> If we could find nice disks for the generators of K and cut-and-cap Σ along them (as explained in the preceding note, on page 502), then Σ would become a sphere.
>
> Let C be an element of K. Since M is simply-connected, there is an immersed disk D in M that is bounded by C. By connect summing with enough $S^2 \times S^2$'s, we can eliminate all self-intersections of D, as was explained in section 4.2 (page 149) and is recalled in figure 11.32. We end up with a disk D, embedded in $M \# k S^2 \times S^2$. Further, Σ is untouched and still characteristic.



11.32. Eliminating an intersection by summing with a sphere

> We have $q(C) = 0$, meaning that $d_C + \Sigma \cdot D = 0 \pmod 2$. Spin D around C (as explained on page 504 and recalled in figure 11.33) till $\Sigma \cdot D = 0$ as an *integer.* This means that the intersection points of Σ and D appear in pairs with opposite signs. By adding even more copies of $S^2 \times S^2$, we can use the Whitney trick to eliminate all common points of Σ and $\operatorname{Int} D$.



11.33. Spinning a membrane around its boundary

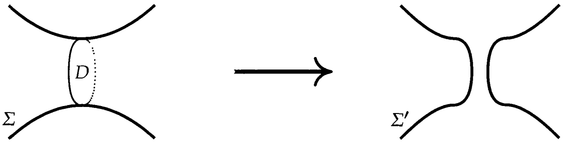
> Thus we are left with $d_C = 0 \pmod 2$, and so $d_C(fr_\Sigma, fr_D) = 2m$ for some integer[17] m. Connect sum with m more copies of $S^2 \times S^2$. Join D with either the

---

[17] Here we think of d_C as the integral relative Euler class, instead of the modulo 2 relative Stiefel–Whitney class.

diagonal or the anti-diagonal sphere S from each new copy of $\mathbb{S}^2 \times \mathbb{S}^2$. Homologically, $S = \mathbb{S}^2 \times 1 \pm 1 \times \mathbb{S}^2$, and hence $S \cdot S = \pm 2$. Therefore, substituting D by $D \# S$ will change d_C to $d_C + 2$ or $d_C - 2$.

We eventually obtain an embedded disk D that only touches Σ along its boundary C, and whose normal bundle can be trivialized by extending the normal framing of C in Σ. We can then cut Σ open along C and cap it to a surface of lesser genus, as recalled in figure 11.34.



11.34. Cutting a surface along a disk

Repeating this procedure for whoever is left in K eventually surgers Σ to an embedded sphere in $M \# (huge)\, \mathbb{S}^2 \times \mathbb{S}^2$. □

For non-characteristic elements, we have a sharper result:

Theorem *(C.T.C. Wall). Assume M is simply-connected and Q_M is indefinite. Let $\alpha \in H_2(M;\mathbb{Z})$ be an indivisible class that is not characteristic. Then α can be represented by a sphere embedded in $M \# \mathbb{S}^2 \times \mathbb{S}^2$.*

This result follows from Wall's theorem on diffeomorphisms (page 153), by moving spheres from $\mathbb{S}^2 \times \mathbb{S}^2$ (or from[18] $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$) around and using the realization of automorphisms of $H_2(M \# \mathbb{S}^2 \times \mathbb{S}^2; \mathbb{Z})$ by self-diffeomorphisms. It was proved in **C.T.C. Wall**'s ***Diffeomorphisms of 4–manifolds*** **[Wal64b]**.

References. Our exposition follows **M. Freedman** and **R. Kirby**'s original paper ***A geometric proof of Rochlin's theorem*** **[FK78]**. Other proofs are **L. Guillou** and **A. Marin**'s ***Une extension d'un théorème de Rohlin sur la signature*** **[GM86c]**, and **Y. Matsumoto**'s ***An elementary proof of Rochlin's signature theorem and its extension by Guillou and Marin*** **[Mat86]**, both inside the volume ***À la recherche de la topologie perdue*** **[GM86a]**. Both the latter papers extend the result to *unoriented* characteristic surfaces, which leads to a modulo 4 version. The origin of all these geometric proofs of Rokhlin's theorem involving the Arf invariant can be traced to lectures of **A. Casson** from around 1975.

A proof of Rokhlin's theorem, streamlined through the use of spin structures, can be found in **R. Kirby**'s ***The topology of 4–manifolds*** **[Kir89**, ch XI**]** and will be explained in the following note.

For more references and historical comments on Rokhlin's theorem, see the references back on page 235, at the end of chapter 4.

---

[18] Recall that, when Q_M is odd, $M \# \mathbb{S}^2 \times \mathbb{S}^2 \cong M \# \mathbb{S}^2 \tilde{\times} \mathbb{S}^2$. Spheres from $\mathbb{S}^2 \tilde{\times} \mathbb{S}^2$ are needed to realize classes of odd self-intersection.

Note: Alternative proof of Rokhlin's theorem

In what follows we explain an alternative geometric proof of Rokhlin's theorem, taken from **R. Kirby**'s ***The topology of 4–manifolds*** [Kir89, ch XI]. Instead of the acrobatics with characteristic classes from the preceding note, we will play with spin structures, both partial spin structures on 4–manifolds and spin structures on lower-dimensional manifolds.

The characteristic cobordism group still has a crucial role, but we will also invoke lower-dimensional spin cobordism groups. Thus, maybe a glance back at the end-notes of chapter 4 (cobordism groups, page 227) and a quick visit with the end-notes of the preceding chapter (characteristic cobordism group, page 427) are recommended.[19]

We start by discussing spin structures on 1– and 2–dimensional manifolds:

Spin structures and spin cobordism in low dimensions. Recall that a spin structure on a manifold X^n is in general defined as a trivialization of T_X over the 1–skeleton of X that can be extended across the 2–skeleton of X. Furthermore, two spin manifolds X' and X'' are said to be spin-cobordant if there is some manifold Y^{n+1} with $\partial Y = \overline{X'} \cup X''$ and endowed with a spin structure that induces[20] the chosen spin structures on X' and X''. This leads to the spin cobordism group Ω_n^{Spin}, as was mentioned already in the end-notes of chapter 4 (page 229).

Since $\pi_1 SO(n) = \mathbb{Z}_2$ only for n at least 3, defining spin structures on lower-dimensional manifolds needs a bit of care, by first raising the fiber-dimension of T_X by stabilization.

Specifically, for a surface S a spin structure on S is a trivialization of $T_S \oplus \underline{\mathbb{R}}$ over the 1–skeleton of S that extends across all S. For a 1–dimensional manifold C, a spin structure is a trivialization of $T_C \oplus \underline{\mathbb{R}}^2$.

Spin structure on circles. On a circle C, there are exactly two[21] spin structures, as pictured in figure[22] 11.35 on the following page. One of them appears from seeing the circle as bounding a disk D, with the trivialization of $T_C \oplus \underline{\mathbb{R}}$ induced from the natural trivialization of T_D. The other one is often called the **Lie-group spin structure,** as it can be obtained by translating a frame using the multiplication of S^1. Let us denote the bounding spin structure by $\mathfrak{s}_\partial$ and the Lie-group one by $\mathfrak{s}_{Lie}$.

Since two circles with the Lie-group spin structure spin-bound the cylinder $C \times [0, 1]$, while a lone Lie-group circle does not spin-bound anything, it follows that the spin cobordism group in dimension 1 is exactly $\mathbb{Z}_2$:

Lemma. *We have* $\Omega_1^{Spin} = \mathbb{Z}_2$, *with generator* $(S^1, \mathfrak{s}_{Lie})$. □

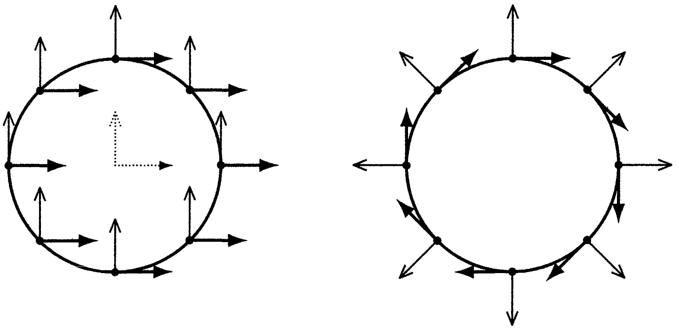
---

[19] If you read the preceding note and survived, then there is no need for that.

[20] A spin structure on Y induces a spin structure on its boundary in a manner analogous with the way an orientation on Y induces an orientation on its boundary: arrange that the trivialization of T_Y over the 1–skeleton of ∂Y fits the decomposition $T_Y|_{\partial Y} = T_{\partial Y} \oplus N_{\partial Y/Y}$; then this defines a trivialization of $T_{\partial Y}$ over the 1–skeleton.

[21] "Two", just as in $\pi_1 SO(3) = \mathbb{Z}_2$, of course.

[22] The third vector of the trivializing frame of $T_C \oplus \underline{\mathbb{R}}^2$ should be imagined perpendicular to the page.



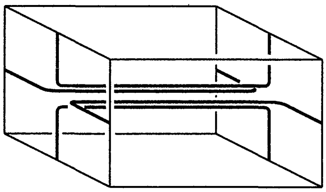
bounding spin structure non-bounding spin structure

11.35. Spin structures on a circle

Spin structures on surfaces. Since every orientable surface is a connected-sum of tori, we first investigate the case of spin structures on a torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$.

Recall that $H^1(X^n; \mathbb{Z}_2)$ acts transitively on the spin structures of[23] X^n. Since $H^1(\mathbb{T}^2; \mathbb{Z}_2) = \oplus 2\,\mathbb{Z}_2$, we expect to find exactly four distinct spin structures on $\mathbb{T}^2$. Indeed, these can be built immediately as products of spin structures on circles. Specifically, we have $(\mathbb{S}^1, \mathfrak{s}_{Lie}) \times (\mathbb{S}^1, \mathfrak{s}_{Lie})$, which does not spin-bound any 3–manifold, as well as $(\mathbb{S}^1, \mathfrak{s}_{Lie}) \times (\mathbb{S}^1, \mathfrak{s}_\partial)$, $(\mathbb{S}^1, \mathfrak{s}_\partial) \times (\mathbb{S}^1, \mathfrak{s}_{Lie})$ and $(\mathbb{S}^1, \mathfrak{s}_\partial) \times (\mathbb{S}^1, \mathfrak{s}_\partial)$, which spin-bound, respectively, $\mathbb{S}^1 \times \mathbb{D}^2$, $\mathbb{D}^2 \times \mathbb{S}^1$, and both. Since $\mathfrak{s}_{Lie} \times \mathfrak{s}_{Lie}$ can be seen as obtained from a frame translated around $\mathbb{T}^2$ using the Lie group structure of $\mathbb{T}^2$, we will call it the **Lie-group spin structure** on $\mathbb{T}^2$ and denote it by $\mathfrak{s}_{Lie}$ as well.

Moving on to general surfaces $S = \#m\,\mathbb{T}^2$, first we notice that a spin structure on S can be built from spin structures on its torus-terms. If all the torus-terms are endowed with bounding spin structures, then their connected-sum S will spin-bound a suitable boundary-sum of $\mathbb{S}^1 \times \mathbb{D}^2$'s and $\mathbb{D}^2 \times \mathbb{S}^1$'s. At the other extreme, if $S = \#2k\,\mathbb{T}^2$ is a sum of an even number of tori, each endowed with the non-bounding spin structure, then S is nonetheless the spin-boundary of the 3–manifold $\big((\#k\,\mathbb{T}^2) \setminus \mathit{disk}\big) \times [0,1]$. (This is in tune with the general fact that $X \# \overline{X}$ bounds $(X \setminus \mathit{ball}) \times [0,1]$, as suggested in figure 11.36.)



11.36. $X \# \overline{X}$ is the boundary of $(X \setminus \mathit{ball}) \times [0,1]$

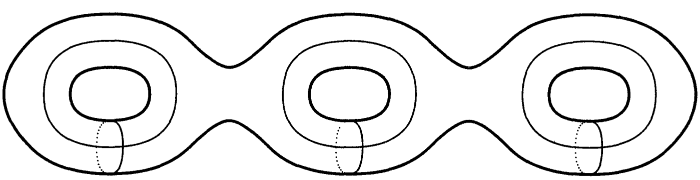
Such reasoning leads us to:

---

[23] This was mentioned, for example, in the inserted note on page 164, and of course follows neatly from obstruction theory, *etc.*

Lemma. *We have* $\Omega_2^{Spin} = \mathbb{Z}_2$*, with generator* $(\mathbb{T}^2, \mathfrak{s}_{Lie})$. □

Finally, notice that a spin structure on a surface S is completely determined by the spin structures induced on a family of circle-generators of $H_1(S;\mathbb{Z})$. Indeed, let $C_1, \dots, C_m, \overline{C}_1, \dots, \overline{C}_m$ be a family of circles embedded in S so that their only intersections are one for each pair C_k and $\overline{C}_k$, and so that they generate $H_1(S;\mathbb{Z})$, as in figure 11.37. Such a choice of circles corresponds to a splitting of S as a connected sum of tori $C_k \times \overline{C}_k$. If each of the circles is endowed with a spin structure, then so will each $C_k \times \overline{C}_k$ and thus S itself. (This can also be deduced from the isomorphism $H^1(S;\mathbb{Z}_2) \approx H^1(\cup C_k \cup \overline{C}_k;\, \mathbb{Z}_2)$, induced by inclusion.)



11.37. Circles on a surface

Spin cobordism of surfaces and the Arf invariant. The isomorphism $\Omega_2^{Spin} = \mathbb{Z}_2$ can be realized by means of the Arf invariant[24] of a suitable quadratic form on $H_1(S;\mathbb{Z}_2)$.

Specifically, let S be a random surface endowed with a spin structure $\mathfrak{s}$, *i.e.*, with a trivialization of $T_S \oplus \underline{\mathbb{R}}$. Then every circle C embedded in S, as it has trivial normal bundle, inherits a spin structure $\mathfrak{s}|_C$. Define

$$q(C) = \big[C, \mathfrak{s}|_C\big] \quad \text{in} \quad \Omega_1^{Spin} ,$$

in other words, set $q(C) = 0$ if C inherits the bounding spin structure and set $q(C) = 1$ if C inherits the Lie-group spin structure. It turns out that this q descends to a well-defined map $q\colon H_1(S;\mathbb{Z}_2) \to \mathbb{Z}_2$ that is a quadratic enhancement of the intersection form of S, and therefore q has an Arf invariant.

Lemma. *The map* $\Omega_2^{Spin} \longrightarrow \mathbb{Z}_2\colon\ (S,\mathfrak{s}) \longmapsto \operatorname{Arf}(q)$ *is an isomorphism.* □

> ***Sketch of proof.*** On one hand, if $\operatorname{Arf}(q) = 0$, then there exists a symplectic basis of[25] $H_1(S;\mathbb{Z}_2)$ corresponding to a decomposition[26] $\big(H_1(S;\mathbb{Z}_2), q\big) = \oplus\, m\, H^{0,0}$. With some care, this decomposition can be realized as a connected-sum splitting $S = \#m\,\mathbb{T}^2$, with each $\mathbb{T}^2$ corresponding to an $H^{0,0}$-term, and thus endowed with a bounding spin structure. On the other hand, if $\operatorname{Arf}(q) = 1$, then we have $\big(H_1(S;\mathbb{Z}_2), q\big) = \oplus\, m\, H^{0,0} \oplus H^{1,1}$, and the last term can be made to correspond to a torus-term with the Lie-group spin structure. □

---

[24] The Arf invariant was explained in the note on page 501.

[25] A symplectic basis is a basis $e_1, \dots, e_m, \bar{e}_1, \dots, \bar{e}_m$ with the only non-zero intersections $e_k \cdot \bar{e}_k = 1$.

[26] We use the notation from the earlier note on the Arf invariant (page 501), where $H^{0,0}$ denoted the $\mathbb{Z}_2$-module $\{0, x, y, x+y\}$ with $x \cdot y = 1$ and $q(x) = 0$ and $q(y) = 0$.

Spin structures on the* 3*–torus. All 3–manifolds have trivial tangent bundle, and thus admit spin structures. Furthermore, it is known that

$$\Omega_3^{SO} = 0 \qquad \text{and} \qquad \Omega_3^{Spin} = 0\,.$$

We will not prove these statements here, but merely focus on the simple case of 3–tori.

The 3–torus $\mathbb{T}^3 = \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$ has $H^1(\mathbb{T}^3;\mathbb{Z}_2) = \oplus 3\,\mathbb{Z}_2$ and thus must admit eight distinct spin structures. These can all be realized, just as for the 2–torus, by products of spin structures on its circle-factors. In particular, seven of these spin structures will contain at least one $(\mathbb{S}^1, \mathfrak{s}_\partial)$–factor and thus spin-bound some product like $\mathbb{D}^2 \times \mathbb{S}^1 \times \mathbb{S}^1$.

The eighth spin structure on $\mathbb{T}^3$ comes from three Lie-group $\mathbb{S}^1$'s and will be called the **Lie-group spin structure** of $\mathbb{T}^3$ and denoted by $\mathfrak{s}_{Lie}$. Nonetheless, $(\mathbb{T}^3, \mathfrak{s}_{Lie})$ does spin-bound a 4–manifold, namely the complement of a generic fiber in the rational elliptic fibration $E(1)$:

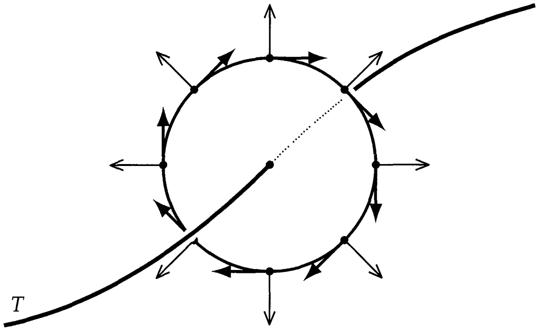
Lemma. *The spin* 3*–torus* $(\mathbb{T}^3, \mathfrak{s}_{Lie})$ *spin-bounds.*

> ***Proof.*** Let $E(1) = \mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2$ be the rational elliptic surface. Let T be a generic torus-fiber near a cusp fiber. Its normal bundle is trivial, and thus its normal circle-bundle $\mathbb{S}N_{T/E(1)}$ is isomorphic to $T \times \mathbb{S}^1$, a 3–torus. Cut a tubular neighborhood of T out of $E(1)$ and denote the resulting 4–manifold by $M = E(1) \setminus \mathbb{D}N_{T/E(1)}$. The boundary of M is the 3–torus $\mathbb{S}N_{T/E(1)}$.
>
> The fiber T has homology class $3e_0 + e_1 + \cdots + e_9$ (a cubic torus in $\mathbb{CP}^2$ blown-up nine times) and therefore is a characteristic surface in $E(1)$. As T represents the obstruction w_2 to the existence of a spin structure on $E(1)$, this implies that $E(1) \setminus T$ (and thus M) admits a (unique)[27] spin structure that does not extend across T. We will argue that this spin structure of M induces on its boundary 3–torus $\mathbb{S}N$ the Lie-group spin structure.
>
> > ***Begging the question.*** *Assume that the spin structure on* M *induced on* $\mathbb{S}N = \partial M$ *is a spin structure other than the Lie-group one. This would mean that the* 3*–torus* $\mathbb{S}N$ *spin-bounds some* $\mathbb{D}^2 \times \mathbb{S}^1 \times \mathbb{S}^1$*, which could then be glued to* M *in order to obtain a closed manifold, endowed with a natural spin structure from its nicely-glued pieces. However, that would create a closed spin* 4*–manifold with signature* 8*, which is forbidden by Rokhlin's theorem. Therefore the spin structure on* $\mathbb{S}N$ *can only be the Lie-group one. Neat—if only our eventual goal were not to actually prove Rokhlin's theorem.*
>
> The actual argument will split $\mathbb{S}N$ into a product of three circles with the Lie-group spin structure. The first such circle is easy to locate:
>
> *The normal circle-fibers.* Since the outside spin structure does not extend across T, it must induce the Lie-group spin structure on each circle-fiber[28] of $\mathbb{S}N \to T$, as in figure 11.38 on the next page. Thus, one $\mathbb{S}^1$–factor of the 3–torus $\mathbb{S}N$ has the Lie-group spin structure. We will show that two other circle-factors of $\mathbb{S}N$ must also inherit the Lie-group spin structure from $E(1) \setminus T$.

---

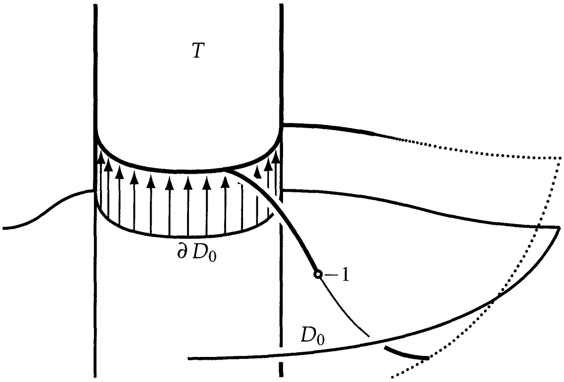
[27] Unique, since $E(1) \setminus T$ is simply-connected, and thus $H^1 = 0$.

[28] This is a general fact about characteristic surfaces and was mentioned in the end-notes of chapter 4 (spin structures, page 179).



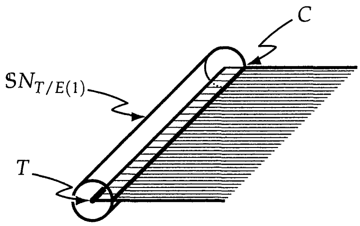
11.38. Outside spin structure, not extending across T

Vanishing cycles. Since $E(1)$ contains a cusp-fiber near T, it follows that two generating circles of $H_1(T;\mathbb{Z})$ are vanishing cycles, *i.e.*, each bounds a disk embedded in $E(1) \setminus T$ with self-intersection -1, as suggested in figure 11.39. We will show that the (-1)–twist around these disks has to be counterbalanced by a twist in the spin structures of their boundary-circles, which thus have to inherit the Lie-group spin structure. To rigorously set this up, we will split T_M above such a disk into its normal and tangent part; in the normal part will live the (-1)–twist of the disk, while in the tangent part will appear the spin-twisting.



11.39. Vanishing cycle

Take such a vanishing cycle C_0 in T, bounding a disk D_0 embedded in $E(1) \setminus T$ with self-intersection -1. We move everything into $M = E(1) \setminus \mathbb{D}N_{T/E(1)}$. Namely, $D_0 \cap M$ is a smaller disk D, bounding a circle $C = D_0 \cap \partial M$ in the 3–torus $\mathbb{S}N$, as pictured in figure 11.40 on the next page. The disk D still has self-intersection -1. This self-intersection of D is detected as follows: take a nowhere-zero vector field normal to C in T (seen as a factor in $\mathbb{S}N = T \times S^1$), extend it as a section of $N_{D/M}$, then push D following this normal section and count the intersections between D and its pushed copy, as already suggested in figure 11.39.



11.40. Disk, and smaller disk

Translation. This is equivalent to the following: Choose some random reference trivialization $N_{D/M} \approx D \times \mathbb{R}^2$. On the boundary-circle C, the normal plane-bundle $N_{D/M}|_C$ already has a natural trivialization induced from the line-bundle $N_{C/T}$, namely as $N_{D/M}|_C \approx N_{C/T} \oplus T_{\mathbb{S}^1}|_C$ (where $\mathbb{S}^1$ denotes the circle-fiber of the bundle $\mathbb{S}N$). The latter trivialization, when viewed through the reference trivialization of all $N_{D/M}$, determines a loop ℓ' in $SO(2)$. If $[\ell']$ were 0 in $\pi_1 SO(2) = \mathbb{Z}$, then the boundary-trivialization could be extended across D. In general, the integer $[\ell']$ counts the self-intersections of D. Therefore, in our case, $[\ell'] = -1$.

The rest of $T_M|_D$. On the tangent side of things, the bundle T_D has a natural trivialization as $T_D \approx D \times \mathbb{R}^2$. On the boundary C, another trivialization of $T_D|_C$ appears from the spin structure of C. Namely, a spin structure on C means a trivialization of $T_C \oplus \underline{\mathbb{R}}^2$, or a trivialization of $T_C \oplus \underline{\mathbb{R}}$ considered only up to 4π–twists. (Indeed, the natural morphism $\pi_1 SO(2) \to \pi_1 SO(3)$ merely forgets 4π–twists and projects $\mathbb{Z}$ onto $\mathbb{Z}_2$.) Identifying $T_C \oplus \underline{\mathbb{R}}$ with $T_C \oplus N_{C/D} = T_D|_C$, we obtain a trivialization of $T_D|_C$, determined by the spin structure of C only up to 4π–twists. When compared with the natural trivialization of the whole T_D, the spin-induced trivialization of $T_D|_C$ determines a loop ℓ'' in $SO(2)$. The parity of $[\ell'']$ in $\pi_1 SO(2) = \mathbb{Z}$ detects whether the spin structure of C matches the trivialization of T_D. In other words, $[\ell''] = 0 \pmod 2$ if C has the bounding spin structure, while $[\ell''] = 1 \pmod 2$ if C has the Lie-group spin structure.

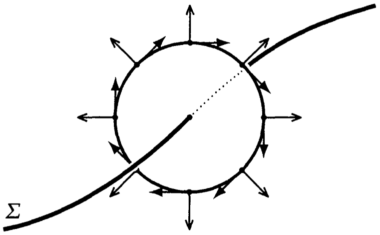
Assembly. Think now of D as part of the 2–skeleton of M and of C as part of its 1–skeleton. The spin structure of M offers a trivialization of $T_M|_C$ that can be extended across D. This means that, after picking a reference trivialization of $T_M|_D \approx D \times \mathbb{R}^4$, the trivialization of $T_M|_C$ must determine a loop ℓ in $SO(4)$ that is null-homotopic. However, $T_M|_C = T_C \oplus N_{C/D} \oplus N_{C/T} \oplus T_{\mathbb{S}^1}$, while $T_M|_D = T_D \oplus N_{D/M}$. As we have seen, on one hand, the matching of $N_{C/T} \oplus T_{\mathbb{S}^1}$ with $N_{D/M}$ is measured by the loop ℓ' in $SO(2)$, whose class is $[\ell'] = -1$, while on the other hand, the matching of $T_C \oplus N_{C/D}$ (trivialized by the spin structure on C) with T_D is measured by the loop ℓ'' in $SO(2)$. The overall matching of $T_M|_C$ with $T_M|_D$ (*i.e.*, extendability) is measured by the loop $\ell = \ell' \oplus \ell''$ in $SO(2) \oplus SO(2) \subset SO(4)$. Since $[\ell] = 0$ and $[\ell'] = -1$, it follows that $[\ell'']$ must be odd. However, this is equivalent with the spin structure of C being non-bounding.

Therefore, if C_0 is a circle of T that bounds a disk D_0 with self-intersection -1, then the circle $C = D_0 \cap \mathbb{S}N$ must inherit from $E(1) \setminus T$ the Lie-group spin structure. Since T contains two such independent vanishing cycles, this implies (together with the Lie-group spin structure on the normal circle-fibers of $\mathbb{S}N$) that the 3-torus $\mathbb{S}N$ inherits the Lie-group spin structure. □

Inducing spin structures on characteristic surfaces. Let M be some 4-manifold and let Σ be a characteristic surface in M. Assume $M \setminus \Sigma$ is endowed with a spin structure that does not extend across Σ. The normal-circle bundle $\mathbb{S}N_{\Sigma/M}$ inherits a spin structure from $M \setminus \Sigma$. We will show that this allows us to endow Σ itself with a canonical spin structure, depending only on the chosen spin structure of $M \setminus \Sigma$. In other words, we will describe a well-defined map

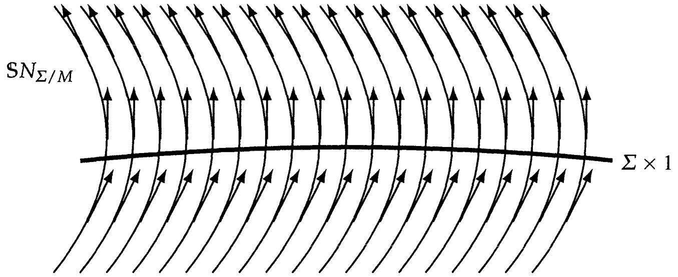
$$\eta\colon \{\text{non-extendable spin structures on } M \setminus \Sigma\} \longrightarrow \{\text{spin structures on } \Sigma\} .$$

The fundamental fact is that, since the spin structure on $M \setminus \Sigma$ does not extend across Σ, each circle-fiber of the circle-normal bundle $\mathbb{S}N_{\Sigma/M}$ must inherit the Lie-group spin structure.



11.41. Spin structure on $M \setminus \Sigma$, not extending across Σ

Assume first that the normal bundle of Σ is trivial. Then $\mathbb{S}N_{\Sigma/M} = \Sigma \times \mathbb{S}^1$. As a 3-submanifold of $M \setminus \Sigma$ (with trivial normal bundle), $\mathbb{S}N_{\Sigma/M}$ inherits a spin structure $\mathfrak{s}_{M\setminus\Sigma}$, that is to say, a trivialization of its tangent bundle. Since each circle-fiber has the Lie-group spin structure, it follows that the spin-induced trivialization of $T_{\mathbb{S}N}$ can be arranged to fit the splitting $T_\Sigma \times T_{\mathbb{S}^1}$ and thus offer a trivialization of T_Σ, as sketched in figure 11.42. In other words, there is a spin structure on Σ fitting in the formula $\left(\Sigma \times \mathbb{S}^1, \mathfrak{s}_{M\setminus\Sigma}\right) = \Sigma \times \left(\mathbb{S}^1, \mathfrak{s}_{Lie}\right)$.



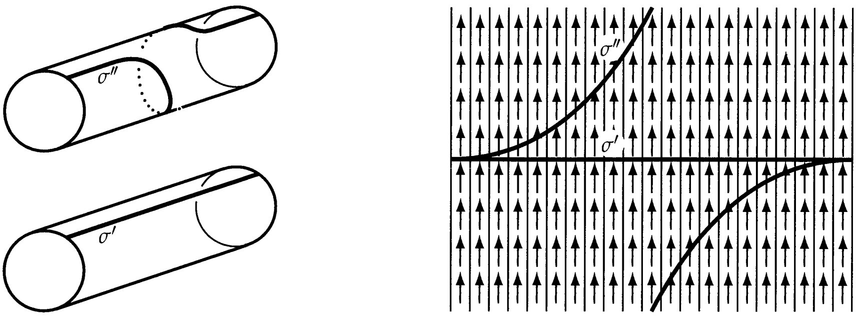
11.42. If $\mathbb{S}N_{\Sigma/M}$ were trivial

In general, when Σ's normal bundle is not trivial, such an argument does not work. Nonetheless, one can still think of $\mathbb{S}N_{\Sigma/M}$ as a twisted product $\Sigma \tilde{\times} \mathbb{S}^1$. While this bundle does not admit any global sections, a reasoning as above could still be applied locally. Moreover, it turns out that we do not need to fit all of Σ inside $\mathbb{S}N_{\Sigma/M}$ to endow it with a spin structure, but only enough of it to cover all of $H_1(\Sigma; \mathbb{Z}_2)$.

Specifically, choose a family of circles $C_1, \dots, C_k, \overline{C}_1, \dots, \overline{C}_k$ embedded in Σ, with $C_k \cdot \overline{C}_k = 1$ the only non-zero intersections, so that the circles generate $H_1(\Sigma; \mathbb{Z}_2)$. Take a thin neighborhood $U \subset \Sigma$ of their union. Notice that, if we endow U with a spin structure, then it automatically extends as a spin structure across the whole Σ. Furthermore, this extension is unique.

Since $N_{\Sigma/M}$ is a plane-bundle over a surface, a generic section in $N_{\Sigma/M}$ has only isolated zeros. Thus, we can always choose a section that is non-zero along U and think of it as a section σ in $\mathbb{S}N_{\Sigma/M}|_U$. The 3–manifold $\mathbb{S}N_{\Sigma/M}$ has a spin structure, inherited from $M \setminus \Sigma$. Its submanifold $\sigma[U]$ has a normal line-bundle, with fibers along the circle fibers of $\mathbb{S}N$, and thus trivialized. Therefore $\sigma[U]$ inherits a spin structure from $\mathbb{S}N$. This can be pulled back to a spin structure on U, which then extends to a spin structure of Σ.

This spin structure on Σ does not depend on the choice of section σ. Indeed, along each circle C_k (or $\overline{C}_k$), two choices of sections σ' and σ'' can only differ by twists around the normal circle-fiber of $\mathbb{S}N$, as in figure 11.43. Since these circle-fibers have the Lie-group spin structure, the induced spin structures on $\sigma'[U]$ and $\sigma''[U]$ are in fact equivalent.



11.43. Two sections in $\mathbb{S}N_{\Sigma/M}$

As an example, consider again the elliptic surface $E(1)$ and its generic torus fiber T. Since $E(1) \setminus T$ is simply-connected, its spin structure is unique. Moreover, as we have seen earlier (page 524), the normal circle-bundle $\mathbb{S}N_{T/E(1)}$ is a 3–torus that inherits the Lie-group spin structure. Since $\mathbb{S}N_{T/E(1)} = T \times \mathbb{S}^1$, this implies that the induced spin structure on T must be the Lie-group one as well. In brief:

$$\eta\big(E(1), T\big) = \big(\mathbb{T}^2, \mathfrak{s}_{Lie}\big) .$$

Finally, the proof. We have all the ingredients for proving Rokhlin's theorem.

Consider the characteristic cobordism group Ω_4^{char}, generated by pairs (M, Σ) with M a 4–manifold, Σ a characteristic surface, and $M \setminus \Sigma$ endowed with a spin structure that does not extend across Σ. Two such pairs (M', Σ') and (M'', Σ'') are considered cobordant if there is a 5–manifold W^5 with $\partial W = \overline{M'} \cup M''$, with a 3–submanifold Y^3 dual to $w_2(W)$ and such that $\partial Y = \overline{\Sigma'} \cup \Sigma''$, and together with a spin structure on $W \setminus Y$ that restricts to the chosen spin structures of $M' \setminus \Sigma'$ and $M'' \setminus \Sigma''$.

As we have seen above, the spin structures on $M' \setminus \Sigma'$ and $M'' \setminus \Sigma''$ induce unique spin structures on Σ' and Σ''. However, arguing in a similar manner as above, one can show that the spin structure on $W \setminus Y$ induces a spin structure on Y. Furthermore, this Y now establishes a spin-cobordism between Σ' and Σ''. In other words, there is a well-defined morphism

$$\eta \colon \Omega_4^{char} \longrightarrow \Omega_2^{Spin} .$$

On one hand, we have seen earlier that $\Omega_2^{Spin} = \mathbb{Z}_2$. On the other hand, we have proved in the end-notes of the preceding chapter (page 427) that we have

$$\Omega_4^{char} \approx \mathbb{Z} \oplus \mathbb{Z} \qquad via \qquad (M, \Sigma) \longmapsto \left(\operatorname{sign} M,\ \tfrac{1}{8}(\Sigma \cdot \Sigma - \operatorname{sign} M)\right) .$$

We will argue that the morphism $\eta \colon \Omega_4^{char} \to \Omega_2^{Spin}$ is just the modulo 2 reduction of the second component of the above isomorphism, *i.e.*,

$$\eta(M, \Sigma) = \tfrac{1}{8}(\Sigma \cdot \Sigma - \operatorname{sign} M) \pmod 2 .$$

All we need to do is check on the generators of Ω_4^{char}. These are $(\mathbb{CP}^2, \mathbb{CP}^1)$ and $(\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2,\ \# 3\,\mathbb{CP}^1 \# \overline{\mathbb{CP}}^1)$, or, more useful for us, $(\mathbb{CP}^2, \mathbb{CP}^1)$ and $(E(1), T)$. For $(\mathbb{CP}^2, \mathbb{CP}^1)$, since $\mathbb{CP}^1$ is a sphere, the only spin structure it can inherit is the trivial one, and thus $\eta(\mathbb{CP}^2, \mathbb{CP}^1) = 0$, fitting the claimed formula. For $(E(1), T)$, as we have seen, $\eta(E(1), T) = [\mathbb{T}^2, \mathfrak{s}_{Lie}]$ and thus $\eta(E(1), T) = 1$, also fitting the formula. Therefore, the above formula, being true on generators, extends to the whole cobordism group Ω_4^{char}.

Immediate consequences are both Rokhlin's theorem and the Kervaire–Milnor generalization. Indeed, when $w_2(M) = 0$, we can take $\Sigma = \varnothing$ and $\operatorname{sign} M = 0 \pmod{16}$ follows. If $w_2(M)$ can be represented by a sphere Σ, then $\eta(M, \Sigma) = 0$ and $\Sigma \cdot \Sigma - \operatorname{sign} M = 0 \pmod{16}$ follows.

Furthermore, by combining the above formula for $\eta \colon \Omega_4^{char} \to \Omega_2^{Spin}$ with the isomorphism Arf: $\Omega_2^{Spin} \approx \mathbb{Z}_2$ (page 523), a slightly different version of Freedman–Kirby's theorem appears. It is instructive to compare both the techniques of the proofs and the actual conclusions of this note and of the Freedman–Kirby theorem/proof from the previous notes.

Bibliography

A nice survey of the state of affairs of the genus problem before gauge theory can be found in **Terry Lawson**'s ***Smooth embeddings of* 2*–spheres in* 4*–manifolds*** [**Law92**]. The same author also surveyed the post-Seiberg–Witten situation in ***The***

minimal genus problem [**Law97**]. The Kervaire–Milnor theorem appeared in ***On* 2*–spheres in* 4*–manifolds*** [**KM61**], together with applications to the genus problem.

The major gauge-theoretic crack at the genus problem came from **P. Kronheimer** and **T. Mrowka**'s ***Embedded surfaces and the structure of Donaldson's polynomial invariants*** [**KM95**], where they proved using Donaldson theory that the adjunction inequality (for $S \cdot S \geq 0$) holds for many Kähler surfaces (with $b_2^+ \geq 2$). Using Seiberg–Witten theory, this was then extended to all Kähler surfaces and surfaces of non-negative intersection in **P. Kronheimer** and **T. Mrowka**'s ***The genus of embedded surfaces in the projective plane*** [**KM94a**] (Thom conjecture, for $\mathbb{CP}^2$) and **J. Morgan, Z. Szabó** and **C. Taubes**' ***A product formula for the Seiberg–Witten invariants and the generalized Thom conjecture*** [**MST96**] (for $b_2^+ \geq 2$, using gluing techniques for Seiberg–Witten). The proof of the adjunction inequality as presented here is due to **P. Kronheimer** and **T. Mrowka;** it is an extension of their argument for the Thom conjecture from [**KM94a**] and can be found for example in **P. Kronheimer**'s expository ***Embedded surfaces and gauge theory in three and four dimensions*** [**Kro98**]. In light of C. Taubes's results on symplectic manifolds, all these arguments can be used to prove that symplectic surfaces of non-negative intersection minimize genus.

The restriction to non-negative self-intersections was first attacked in **R. Fintushel** and **R. Stern**'s ***Immersed spheres in* 4*–manifolds and the immersed Thom conjecture*** [**FS95**], where they discuss immersed spheres.[29] Finally, the restriction on self-intersections fell in **P. Ozsváth** and **Z. Szabó**'s ***The symplectic Thom conjecture*** [**OS00b**]. A more general adjunction for 4–manifolds not of simple type is obtained in their ***Higher type adjunction inequalities in Seiberg–Witten theory*** [**OS00a**]. The genus of surfaces in $\mathbb{S}^2 \times \mathbb{S}^2$ and $\mathbb{CP}^2 \# n \overline{\mathbb{CP}}^2$ is discussed in **D. Ruberman**'s ***The minimal genus of an embedded surface of non-negative square in a rational surface*** [**Rub96**].

For 3–manifolds, the results quoted can be found in **W. Thurston**'s ***A norm for the homology of* 3*–manifolds*** [**Thu86**] and **D. Gabai**'s ***Foliations and the topology of* 3*–manifolds*** [**Gab83**]. See also **A. Candel** and **L. Conlon**'s ***Foliations. II*** [**CC03**, part 3]. **P. Kronheimer**'s result on foliations on 4–manifolds appeared in ***Minimal genus in*** $\mathbb{S}^1 \times M^3$ [**Kro99**]. A nice discussion of the the problem of embedded surfaces in both 3– and 4–manifolds in light of Seiberg–Witten theory can be found in **P. Kronheimer**'s survey ***Embedded surfaces and gauge theory in three and four dimensions*** [**Kro98**].

Note also that Ozsváth–Szabó's Heegaard–Floer homology[30] yields genus bounds on 3–manifold, as shown in **P. Ozsváth** and **Z. Szabó**'s ***Holomorphic disks and genus bounds*** [**OS04b**], and was used to re-prove the Thom conjecture in ***Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary*** [**OS03**]. An adjunction inequality holds for the 4–manifold invariant arising from this theory, see ***Holomorphic triangles and invariants for smooth four-manifolds*** [**OS01c**].

---

[29] An interesting consequence of their work there is the following statement: *If M contains an immersed sphere S with m* positive *double-points and $[S] \cdot [S] = 2m - 2$, then M must be of Seiberg–Witten simple type.* Note also that the same paper proves the Seiberg–Witten blow-up formula.

[30] See the references on page 478 at the end of the preceding chapter.

Chapter 12

Wildness Unleashed: The Fintushel–Stern Surgery

We present a remarkable procedure for modifying the smooth structure of a 4–manifold without altering the underlying topological 4–manifold. This easily produces many new infinite families of homeomorphic but non-diffeomorphic 4–manifolds.

Roughly, the procedure removes a neighborhood $\mathbb{T}^2 \times \mathbb{D}^2$ of a nice torus in M and replaces it with a knotted version. The knotted version is obtained from the complement of a knot K in $\mathbb{S}^3$. While M and the resulting M_K are homeomorphic, they are rarely diffeomorphic. Indeed, the Seiberg–Witten invariants of M_K and M differ exactly by the Alexander polynomial of K.

The chapter starts by quoting results about the Seiberg–Witten invariants of manifolds obtained by gluing along 3–tori. Some of these statements were used by R. Fintushel and R. Stern in the proof of their result; others were not, but they nonetheless are of independent interest. In section 12.2 (page 539) we review the Alexander polynomial of a knot, while in section 12.3 (page 541) we describe the surgery procedure outlined above and the Seiberg–Witten invariants of the result. Section 12.4 (page 545) contains some quick applications, such as the construction of infinitely-many exotic $K3$'s that do not admit any symplectic structures.

12.1. Gluing results in Seiberg–Witten theory

In this section we review gluing results for the Seiberg–Witten invariants of manifolds built by attachment along boundary 3–tori. These boundaries usually appear after cutting out neighborhoods of embedded 2–tori.

Preparation

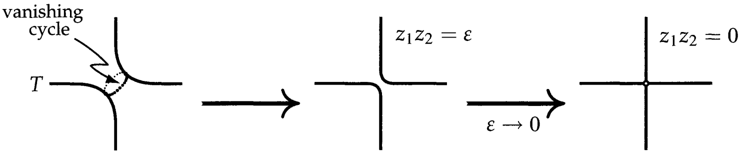
We present the type of 2–tori most relevant for the surgeries of this chapter. Then we rewrite the Seiberg–Witten invariants as a Laurent polynomial, which will help us to write the gluing formulae succinctly and comfortably.

Nice tori. Let M be a simply-connected 4–manifold. Let T be a torus embedded in M, homologically nontrivial, and with zero self-intersection:

$$T \cdot T = 0 .$$

Notice that, owing to the adjunction inequality, T must be orthogonal to all basic classes[1] of M.

Furthermore, we assume that $H_1(T;\mathbb{Z})$ is generated by two embedded circles, intersecting each other exactly once, so that each bounds a disk of self-intersection -1 in M (they are *vanishing cycles*, as in figure 12.1).



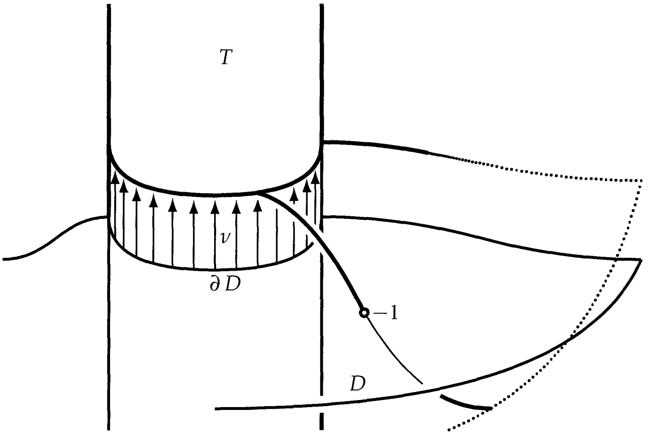
12.1. A vanishing cycle

> *This concept of "disk of self-intersection -1" needs a bit of care: Consider a disk D in M, with ∂D embedded as a circle in T and* Int D *included in the complement of T. Pick a random orientation of D. This induces an orientation of ∂D inside the oriented torus T, and thus ∂D has a preferred nowhere-vanishing normal vector field v inside T. Extend v across the whole disk D as a normal vector field and use it to push D off itself. Then count with signs the intersection points of D with this pushed-off copy to get its self-intersection number. See figure 12.2 on the facing page. The result does not depend on the choices made along the way.*

In what follows, an embedded torus T will be called **near-cusp embedded** (or ***c*–embedded**) if a couple of generators of $H_1(T;\mathbb{Z})$ are vanishing cycles, *i.e.*, if they bound such disks of self-intersection -1. Notice that every near-cusp embedded torus automatically has zero self-intersection.

---

[1] We have $\chi(T) + T \cdot T \leq -|\kappa \cdot T|$ and hence $\kappa \cdot T = 0$. Notice that this does not exclude T itself from being a basic class.



12.2. Self-intersection of a disk

The name of "near-cusp embedded" is justified as follows: A torus T is near-cusp embedded if and only if a neighborhood of T in M looks like a neighborhood U of a generic torus fiber inside some elliptic fibration, so that U contains a *cusp fiber* and so that T corresponds to a regular fiber.

> *Indeed, take a thickened* 2*–torus* $\mathbb{T}^2 \times \mathbb{D}^2$*, choose two circles* C'*,* C'' *in* $\mathbb{T}^2$ *that generate* $H_1(\mathbb{T}^2; \mathbb{Z})$*, then attach to each of them a thickened disk* (2*–handle*) $\mathbb{D}^2 \times \mathbb{D}^2$*, by identifying the attaching circle* $\mathbb{S}^1 \times 0$ *with* $C' \times 1$ *or* $C'' \times 1$ *respectively, and gluing the thickening* $0 \times \mathbb{D}^2$ *with a twist of* -1 *along the circle. In short, attach to each of* $C' \times 1$ *and* $C'' \times 1$ *a* 2*–handle with framing* -1*. The resulting* 4*–manifold with boundary is diffeomorphic to a neighborhood of a cusp fiber inside an elliptic fibration, and* $\mathbb{T}^2 \times 0$ *corresponds to a regular fiber.*

In what follows, we will glue 4–manifolds along near-cusp embedded tori, and thus generalize the fiber-sum and logarithmic transformations that were defined earlier for elliptic surfaces.[2]

The Seiberg–Witten series. To more concisely express the results of this chapter, we need to introduce a few notations: First, for any class $\alpha \in H^2(M; \mathbb{Z})$, we introduce its formal exponential e^α. We then have

$$e^\alpha \cdot e^\beta = e^{\alpha+\beta} .$$

This formal rule is in fact the whole point of introducing these exponentials. We now formally rewrite the Seiberg–Witten invariant as a combination of such exponentials:

$$\mathcal{SW}_M = \sum_{\kappa \text{ basic}} \mathcal{SW}_M(\kappa) \cdot e^\kappa .$$

---

[2] See back, section 8.2 (page 306) and section 8.3 (page 310).

For example, for the $K3$ surface, since 0 is its only basic class and the value of the invariant is 1, we write $\mathcal{SW}_{K3} = 1 \cdot e^0 = 1$.

> ***Formalities.*** *An algebraically more satisfying method for dealing with the above convention is the following: Consider $H^2(M;\mathbb{Z})$ as an Abelian group (with addition), but write its group operation multiplicatively. In particular, the trivial homology class will be denoted by 1, not 0. Then build the group-ring $\mathbb{Z}\big[H^2(M;\mathbb{Z})\big]$. Its elements are of the form $\sum m_k \alpha_k$, with addition defined on the coefficients m_k, and multiplication set as $(m_i\alpha_i)\cdot(m_j\alpha_j) = m_im_j(\alpha_i+\alpha_j)$. Then one writes simply $\mathcal{SW}_M = \sum \mathcal{SW}_M(\alpha)\,\alpha$, as an element of $\mathbb{Z}\big[H^2(M;\mathbb{Z})\big]$. While on one hand this makes disappear cumbersome exponentials and the troubling word "formal series", it leads to conflicts with the usual notations.[3] In what follows, we will stick with the exponentials.[4]*

Generalized fiber sums

Assume that we have two smooth 4–manifolds M' and M'', each containing a homologically-nontrivial near-cusp embedded torus, T' and T''. We can then choose small tubular neighborhoods around these tori and remove them from M' and M''; for convenience we denote the results by $M' \setminus T'$ and $M'' \setminus T''$ and call them directly "the complements of T' and T'' in M' and M''".

Since both T' and T'' have zero self-intersection, their normal bundles are trivial, and thus the boundaries of their complements $M' \setminus T'$ and $M'' \setminus T''$ are two copies of the 3–torus $\mathbb{T}^3 = \mathbb{T}^2 \times \mathbb{S}^1$. Any identification of the torus T' with T'' induces an (orientation-reversing) identification of $\partial(M' \setminus T')$ with[5] $\partial(M'' \setminus T'')$. Then $M' \setminus T'$ and $M'' \setminus T''$ can be glued together in the usual fashion, as in figure 12.3 on the next page. Denote the result by

$$M' \ \#_{T'=T''} M'' .$$

This is a generalization of the *fiber sum* used to build the elliptic surfaces $E(n)$ in section 8.2 (page 306).

Theorem *(J. Morgan and T. Mrowka and Z. Szabó). Let M' and M'' be two 4–manifolds with $b_2^+ \geq 2$ and assume that each contains a homologically-nontrivial near-cusp embedded torus, T' and T''. Then:*

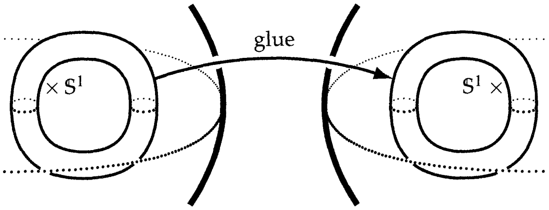
$$\mathcal{SW}_{(M' \#_{T'=T''} M'')} = \mathcal{SW}_{M'} \cdot \mathcal{SW}_{M''} \cdot \big(e^{T'} - e^{-T'}\big) \cdot \big(e^{T''} - e^{-T''}\big) . \qquad \square$$

---

[3] For example, $\alpha_i \cdot \alpha_j$ is now supposed to mean $\alpha_i + \alpha_j$ and not $Q_M(\alpha_i, \alpha_j)$ as before.

[4] Of course, there are middle ways as well, denoting by t_α the element $\alpha \in H^2(M;\mathbb{Z})$ when understood as an element of the multiplicatively-written Abelian group, with $t_\alpha \cdot t_\beta = t_{\alpha+\beta}$. Pick your own favorites.

[5] Simply multiply the chosen identification $T' \cong T''$ with a reflection (complex conjugation) on $\mathbb{S}^1$ to get the required orientation-reversing diffeomorphism $T' \times \mathbb{S}^1 \cong \overline{T'' \times \mathbb{S}^1}$.



12.3. Building $M' \#_{T'=T''} M''$

Example. *Since* $E(4) = E(2) \#_{fiber} E(2)$, *we have that* $\mathcal{SW}_{E(4)} = (e^F - e^{-F})^2 = e^{2F} - 2e^0 + e^{-2F}$, *and thus* $E(4)$ *has basic classes* $\pm 2F$ *and* 0, *with values* 1 *and* -2 *respectively.*

Since the rational elliptic surface $E(1) = \mathbb{CP}^2 \# 9 \overline{\mathbb{CP}}^2$ has $b_2^+ = 1$, generalized fiber sums with it need to be treated as a separate case:

Theorem (*J. Morgan and T. Mrowka and Z. Szabó*). *Let* M *be a* 4*–manifold with* $b_2^+(M) \geq 2$ *and assume that* M *contains a homologically-nontrivial near-cusp embedded torus* T. *By gluing* $M \setminus T$ *to the complement of some regular fiber* F *in* $E(1)$, *we have:*

$$\mathcal{SW}_{(M \#_{T=F} E(1))} = \mathcal{SW}_M \cdot \left(e^T - e^{-T}\right) . \qquad \square$$

Example. *Thus, knowing that the K3 surface has* $\mathcal{SW}_{K3} = 1$ *and that all* $E(n) = K3 \#_{fiber} (n-2) E(1)$, *one immediately obtains the formula*

$$\mathcal{SW}_{E(n)} = \left(e^F - e^{-F}\right)^{n-2}$$

for all $n \geq 2$. *This is a concise way of rewriting the results stated in section 10.6 (page 413).*

Finally, assume that a single 4–manifold M contains two nontrivial disjoint near-cusp embedded tori T' and T''. Then their complement $M \setminus T' \cup T''$ can be closed-up by identifying the 3–torus surrounding T' with the 3–torus surrounding T''. Denote the resulting self-sum by

$$\#_{T'=T''} M .$$

The corresponding gluing formula is:

Theorem (*J. Morgan and T. Mrowka and Z. Szabó*). *Let* M *be a* 4*–manifold with* $b_2^+(M) \geq 2$ *and assume that it contains two disjoint homologically-nontrivial near-cusp embedded tori* T' *and* T''. *Then:*

$$\mathcal{SW}_{(\#_{T'=T''} M)} = \mathcal{SW}_M \,|_{T'=T''} \cdot \left(e^{T'} - e^{-T'}\right)^2 . \qquad \square$$

Generalized logarithmic transformations

In what follows, we will cut a neighborhood of a torus out of a 4–manifold, then glue it back after twisting by a diffeomorphism of the boundaries. (One could think of this procedure as a 4–dimensional analogue of Dehn surgery on 3–manifolds.)

Let M be a closed 4–manifold and let T be any torus embedded in M, with zero self-intersection. Consider a tubular neighborhood of T in M (which is a copy of $T \times \mathbb{D}^2$) and cut it out of M. Denote the resulting 4–manifold by $M \setminus T$; its boundary is a copy of $\mathbb{T}^3 = \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$.

Any orientation-reversing self-diffeomorphism of $\mathbb{T}^3$ can now be used to glue $T \times \mathbb{D}^2$ back into $M \setminus T$. The 3–torus is an especially nice space, and every self-diffeomorphism of $\mathbb{T}^3$ is isotopic to a *linear* self-diffeomorphism. That is to say, the self-diffeomorphisms of $\mathbb{T}^3$ are classified, up to isotopy, by $\mathrm{Aut}_{\mathbb{Z}}(H_1(\mathbb{T}^3;\mathbb{Z}))$. If we choose a basis in $H_1(\mathbb{T}^3;\mathbb{Z})$, then all orientation-reversing self-diffeomorphisms of $\mathbb{T}^3$ are represented by integral 3×3 matrices with determinant -1.

Let $\{\mu, \alpha', \alpha''\}$ be a basis in $H_1(\mathbb{T}^3;\mathbb{Z})$, with $\mathbb{T}^3 = T \times \mathbb{S}^1$ viewed as the boundary of $T \times \mathbb{D}^2$, and so that μ is the class of *point* $\times\, \mathbb{S}^1$, while α' and α'' come from some fixed splitting $T = \mathbb{S}^1 \times \mathbb{S}^1$.

Let $\varphi \colon \mathbb{T}^3 \cong \mathbb{T}^3$ be any orientation-reversing self-diffeomorphism. Then in homology we have

$$\varphi_*(\mu) = p\mu + m\alpha' + n\alpha''$$

for some integers p, m, n. Just as μ, the element $\varphi_*(\mu)$ must be *indivisible* in $H_1(\mathbb{T}^3;\mathbb{Z})$, *i.e.*, p, m, n have no common divisor. Furthermore, the values of $\varphi_*(\alpha')$ and $\varphi_*(\alpha'')$ have no influence on the result of attaching $\mathbb{D}^2 \times T$ to $M \setminus T$. Indeed, if two orientation-reversing self-diffeomorphisms φ and ψ of $T \times \mathbb{S}^1$ have $\varphi_*(\mu) = \psi_*(\mu)$, then they can be linked through a self-diffeomorphism of the whole $T \times \mathbb{D}^2$. Conversely, for any three integers p, m, n with no common divisors there exist orientation-reversing self-diffeomorphisms $\varphi \colon \mathbb{T}^3 \cong \mathbb{T}^3$ with $\varphi_*(\mu) = p\mu + m\alpha' + n\alpha''$.

In conclusion, given any three integers p, m, n with no common divisor, there is a *unique* smooth 4–manifold

$$M_{(p,m,n)} = M \setminus T \cup_{\varphi} T \times \mathbb{D}^2 ,$$

obtained by attaching $T \times \mathbb{D}^2$ to $M \setminus T$ by using some orientation-reversing self-diffeomorphism φ of $\partial(T \times \mathbb{D}^2)$ with $\varphi_*(\mu) = p\mu + m\alpha' + n\alpha''$. Of course, we have $M_{(1,0,0)} = M$.

If E is an elliptic surface and T is a regular torus fiber in E, then the above procedure turns out to smoothly coincide with the logarithmic transformation of multiplicity p. Indeed, we have:

Theorem (*R. Gompf*). *Let E be any elliptic surface containing a cusp fiber and let T be a regular fiber of E. Then, for every p, m, n with no common divisor, we have*

$$E_p \cong E_{(p,m,n)} ,$$

where E_p denotes the logarithmic transformation of multiplicity p. □

In light of this, we will call $M_{(p,m,n)}$ the **generalized logarithmic transformation** of M along T, no matter whether M is elliptic or not, or whether T is anything more than merely an embedded torus with zero self-intersection (not necessarily near-cusp embedded).

Analogous to the case of a multiple fiber appearing from a logarithmic transformation, the homology class of the initial T from M survives as a class in $M_{(p,m,n)}$, and there this class is p times the homology class represented by the copy of T that is plugged into $M_{(p,m,n)}$ packaged in $T \times \mathbb{D}^2$. Thus, when talking homology, we will be careful and denote by T only the class of the initial torus T from M, while the class of the new torus plugged in $M_{(p,m,n)}$ will be denoted by T_p; hence we write

$$T = p \cdot T_p$$

in $H^2(M_{(p,m,n)}; \mathbb{Z})$.

For the Seiberg–Witten invariants, we have:

Theorem (*R. Fintushel and R. Stern*). *Let M be 4–manifold and T a near-cusp embedded torus, homologically nontrivial, non-torsion, and with zero self-intersection. Then we have:*

$$\mathcal{SW}_{M_{(0,m,n)}} = 0 ,$$

$$\mathcal{SW}_{M_{(p,m,n)}} \cdot \left(e^{T_p} - e^{-T_p}\right) = \mathcal{SW}_M \cdot \left(e^{T} - e^{-T}\right) .$$

Writing everything in terms of T_p,

$$\mathcal{SW}_{M_{(p,m,n)}} = \mathcal{SW}_M \cdot \left(e^{(p-1)T_p} + e^{(p-2)T_p} + \cdots + e^{-(p-1)T_p}\right) .$$ □

Example. In the particular case of the elliptic surfaces $E(n)$ with $n \geq 2$, it follows that performing one logarithmic transformation yields

$$\mathcal{SW}_{E(n)_p} = \left(e^F - e^{-F}\right)^{n-2} \cdot \left(e^{(p-1)F_p} + \cdots + e^{-(p-1)F_p}\right) ,$$

where F_p is the class of the multiple fiber being created and $F = pF_p$ is the class of the regular fiber.

Performing two logarithmic transformations of coprime multiplicities gets

$$\mathcal{SW}_{E(n)_{p,q}} = \left(e^F - e^{-F}\right)^{n-2} \cdot \left(e^{(p-1)F_p} + \cdots\right) \cdot \left(e^{(q-1)F_q} + \cdots\right) ,$$

where F_p and F_q are the classes of the multiple fibers and $F = pF_p = qF_q$ is the class of the regular fiber of $E(n)$. (Remember that there also exists a class F_{pq} so that $F = pqF_{pq}$, and the above can be expressed fully in terms of F_{pq}.)

These are merely a rewriting of the results stated in section 10.6 (page 413), which led to the infinite families of homeomorphic but non-diffeomorphic elliptic surfaces from section 8.4 (page 314). Of course, these results were obtained easier and earlier than the general gluing results above by using complex geometry and the special shape of the Seiberg–Witten invariants on Kähler surfaces.

When T is *not* near-cusp embedded, the invariants of $M_{(p,m,n)}$ will no longer depend only on[6] p, but will involve the values of m and n as well. The corresponding gluing result is:

Theorem (*J. Morgan and T. Mrowka and Z. Szabó*). *Assume that M has $b_2^+(M) \geq 3$ and let T be any torus embedded in M with zero self-intersection. On one hand, if κ is a characteristic element of $M_{(p,m,n)}$ that restricts nontrivially to the plugged-in $T \times \mathbb{D}^2$, i.e., if $\kappa \cdot T_p \neq 0$, then we have*

$$\mathcal{SW}_{M_{(p,m,n)}}(\kappa) = 0 .$$

On the other hand, let c be any characteristic element of $M \setminus T$ and denote by $\{c\}_{(p,m,n)}$ the set of all characteristic elements κ of $M_{(p,m,n)}$ that restrict to c on $M \setminus T$, i.e., $\{c\}_{(p,m,n)} = \{\kappa \text{ characteristic in } M_{(p,m,n)},\ \kappa|_{M\setminus T} = c\}$. We have

$$\sum_{\kappa\in\{c\}_{(p,m,n)}} \mathcal{SW}_{M_{(p,m,n)}}(\kappa) =$$

$$p \cdot \sum_{\kappa\in\{c\}_{(1,0,0)}} \mathcal{SW}_{M_{(1,0,0)}}(\kappa) + m \cdot \sum_{\kappa\in\{c\}_{(0,1,0)}} \mathcal{SW}_{M_{(0,1,0)}}(\kappa) + n \cdot \sum_{\kappa\in\{c\}_{(0,0,1)}} \mathcal{SW}_{M_{(0,0,1)}}(\kappa) .$$

Further, if the class of $p\mu + m\alpha' + n\alpha''$ from $H_1(\partial(T \times \mathbb{D}^2); \mathbb{Z})$ is indivisible in $H_1(M_{(p,m,n)}; \mathbb{Z})$, then the sum on the left reduces to at most one term. □

For simplicity, in the statement above we assumed that M has no 2–torsion.

A frequently-used particular case of the theorem is when T is null-homologous in $M = M_{(1,0,0)}$. In this case, all characteristic elements κ of M have $\kappa \cdot T = 0$, and thus can be identified with unique characteristic elements in each of the $M_{(p,m,n)}$'s. Further, the classes of the copies of T plugged back into $M_{(0,1,0)}$ and $M_{(0,0,1)}$ are nontrivial and are the only classes not inherited from $M \setminus T$. Therefore, for every characteristic element κ of M, all the characteristic elements of $M_{(0,1,0)}$ or $M_{(0,0,1)}$ that restrict on $M \setminus T$ to κ can be written as $\kappa + 2tT_0$.

---

[6] Keep in mind that, even though the Seiberg–Witten invariants depend only on p, nobody knows whether $M_{(p,m',n')}$ and $M_{(p,m'',n'')}$ are diffeomorphic or not; we merely know that Seiberg–Witten does not distinguish them.

Corollary. *Assume further that T is null-homologous. Then we have*

$$\begin{aligned} \mathcal{SW}_{M_{(p,m,n)}}(\kappa) &= p \cdot \mathcal{SW}_M(\kappa) \\ &\quad + m \cdot \sum_{t \in \mathbb{Z}} \mathcal{SW}_{M_{(0,1,0)}}(\kappa + 2tT_0) + n \cdot \sum_{t \in \mathbb{Z}} \mathcal{SW}_{M_{(0,0,1)}}(\kappa + 2tT_0) , \end{aligned}$$

where T_0 denotes the class of the torus T plugged along with $T \times \mathbb{D}^2$ into $M_{(0,1,0)}$ and $M_{(0,0,1)}$ respectively. □

Of course, since there are only finitely-many basic classes, each of the sums above necessarily has only finitely-many non-zero terms.

12.2. Review: the Alexander polynomial of a knot

Remember now that the **Alexander polynomial**

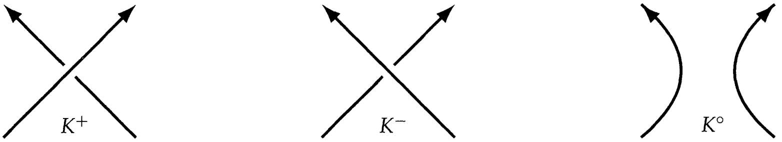
$$\Delta_K \in \mathbb{Z}[t, t^{-1}]$$

of an *oriented* knot K is the Laurent polynomial that can be defined through the following *skein relation*:

$$\Delta_{K^+}(t) = \Delta_{K^-}(t) + (t^{1/2} - t^{-1/2}) \cdot \Delta_{K^\circ}$$

for any three knots K^+, K^-, K° whose diagrams differ only in a neighborhood that looks like figure 12.4. We normalize the values of Δ by requiring that on the unknot $\bigcirc$ it yields

$$\Delta_{\bigcirc}(t) = 1 .$$



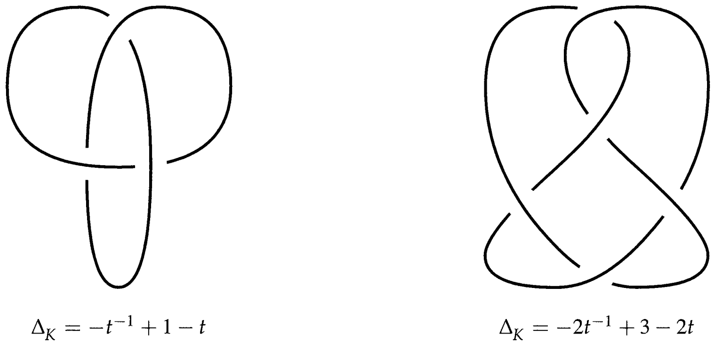
12.4. Crossings for the skein relation

Since changing crossings eventually allows one to untie any knot, the above conditions completely determine the Alexander polynomial of a knot.[7] Of particular help in computations is the property that the Alexander polynomial is zero on any split link (*i.e.*, any link whose components can be drawn apart).

Keep in mind that there are plenty of knot invariants which are quite finer than the Alexander polynomial. For example, the Jones polynomial distinguishes between a knot and its mirror image, while the Alexander polynomial does not.

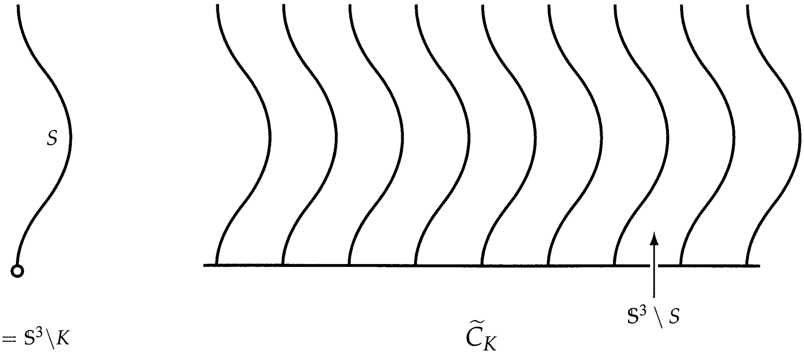
---

[7] The hard part is to prove that Δ_K is well-defined.



12.5. Two knots, two polynomials

Cyclic-cover definition. *An equivalent definition of the Alexander polynomial of K uses the infinite-cyclic cover of its complement $C_K = S^3 \setminus K$. Indeed, we have $H_1(C_K; \mathbb{Z}) = \mathbb{Z}$ and thus $\pi_1 C_K \,/\, [\pi_1 C_K, \pi_1 C_K] = \mathbb{Z}$, leading to a cover map $\widetilde{C}_K \to C_K$ whose group of deck transformations is infinite-cyclic. The cover space $\widetilde{C}_K$ can be built starting with some choice of Seifert surface S for K in S^3, as suggested in figure 12.6.*



12.6. Building the infinite-cyclic cover of a knot complement

Denote by t the generator of the deck transformations of $\widetilde{C}_K \to C_K$. Encoding the action of the deck transformations on the homology of $\widetilde{C}_K$, we consider $H_1(\widetilde{C}_K; \mathbb{Z})$ as a module over $\mathbb{Z}[t, t^{-1}]$. Then $H_1(\widetilde{C}_K; \mathbb{Z})$ is the quotient of $\mathbb{Z}[t, t^{-1}]$ by an ideal, and that ideal must in fact be a principal ideal, i.e., generated by a single element $\Delta_K(t)$:

$$H_1(\widetilde{C}_K; \mathbb{Z}) = \mathbb{Z}[t, t^{-1}] \,/\, \langle \Delta_K \rangle \,,$$

where $\langle \Delta_K \rangle = \mathbb{Z}[t, t^{-1}] \cdot \Delta_K(t)$ is the ideal generated by the polynomial Δ_K. Notice that there are several possible choices of polynomials for such a generating $\Delta_K \in \mathbb{Z}[t, t^{-1}]$, differing by multiplication by some t^m. We choose Δ_K to be the symmetric *representative, i.e., the one for which the coefficients of each t^k and t^{-k} are the same, and we call it the Alexander polynomial of K.*

While talking about the knot complement, it is worth mentioning the following remarkable result:

Theorem (*C. Gordon and J. Luecke*). *The knot complement determines the knot completely. In other words, if two knots K' and K'' in $\mathbb{S}^3$ have diffeomorphic complements $\mathbb{S}^3 \setminus K' \cong \mathbb{S}^3 \setminus K''$, then the two knots are equivalent.*[8] □

12.3. The knot surgery

Finally, we are ready to describe the Fintushel–Stern surgery:

Carve the torus out. Let M be a simply-connected smooth 4–manifold containing a near-cusp embedded torus T, that is to say, an embedded torus T that is homologically nontrivial, of zero self-intersection, and with a basis of $H_1(T;\mathbb{Z})$ represented by two circles each bounding a disk with self-intersection -1. Assume further that the complement $M \setminus T$ is *simply-connected.*

Take a small tubular neighborhood of T and cut it out. This means removing a thickened torus $T \times \mathbb{D}^2$ from M. For simplicity, denote by

$$M \setminus T$$

the resulting 4–manifold; its boundary is a 3–torus.

In what follows, we will use the complement of a knot in $\mathbb{S}^3$ to build a *homological* copy of $T \times \mathbb{D}^2$. The purpose is that, when we cut $T \times \mathbb{D}^2$ out of M and replace it with this homological copy, the homology of the resulting 4–manifold be the same as that of M, with the same intersection form, and hence that the two 4–manifolds be homeomorphic. However, they will not be diffeomorphic.

Build the cork. Choose your favorite knot K in $\mathbb{S}^3$, take a small tubular neighborhood of it and remove it to get the *knot complement,* which for convenience we denote by

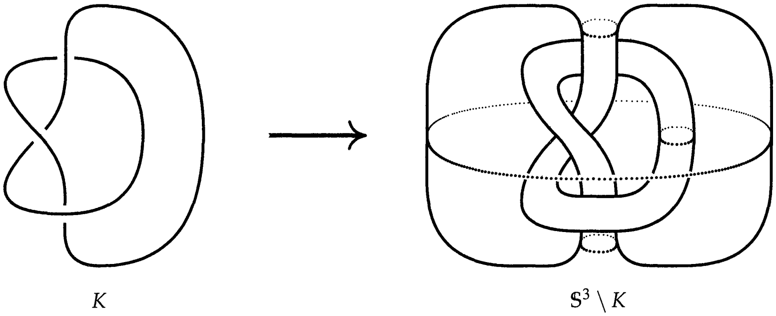
$$\mathbb{S}^3 \setminus K\,.$$

This is a 3–manifold having as boundary a 2–torus, see figure 12.7 on the following page.

It is not hard to see that *homologically* the knot complement is indistinguishable from a solid torus $\mathbb{S}^1 \times \mathbb{D}^2$, as in figure 12.8 on the next page.

---

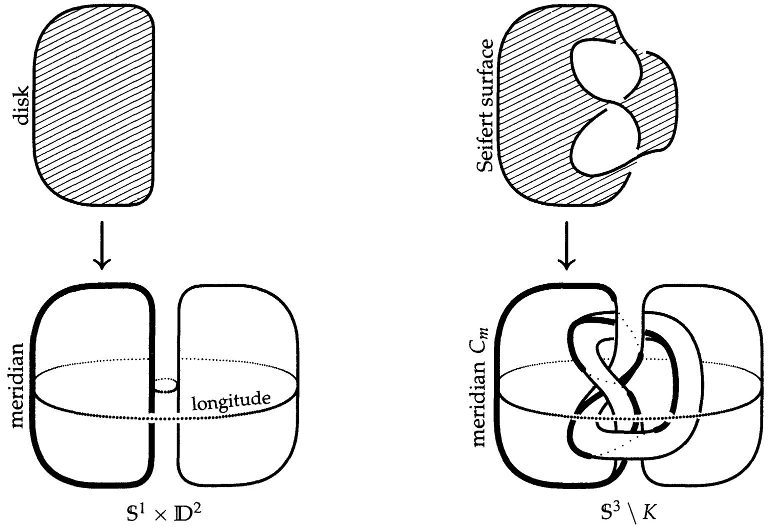
[8] Two knots are called **equivalent** if there is an (orientation-preserving) diffeomorphism $\mathbb{S}^3 \to \mathbb{S}^3$ that takes K' to K''.



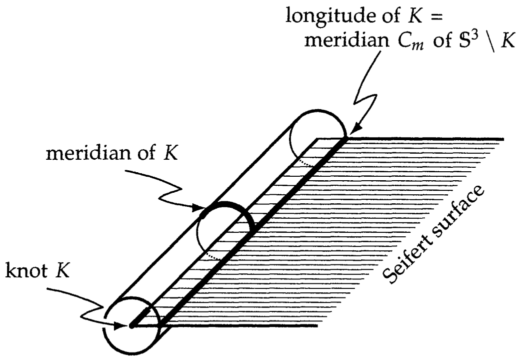
12.7. Knot, and knot complement

Indeed, the knot K bounds a Seifert surface in $\mathbb{S}^3$, which draws a circle C_m on the torus boundary $\partial(\mathbb{S}^3 \setminus K)$ of $\mathbb{S}^3 \setminus K$, as in figure 12.9 on the facing page. The circle C_m is a longitude *of the tubular neighborhood of K. Nonetheless, for the homology of the knot complement $\mathbb{S}^3 \setminus K$, it plays the same role that the* meridian $1 \times \mathbb{S}^1$ *plays for the solid torus $\mathbb{S}^1 \times \mathbb{D}^2$. Since we are focused on the complement, we will call C_m a* **meridian** *of $\mathbb{S}^3 \setminus K$. (In short, a longitude of K is a meridian of $\mathbb{S}^3 \setminus K$.) The homology class of a meridian C_m of $\mathbb{S}^3 \setminus K$ is uniquely determined, up to sign. Indeed, the class of C_m generates the kernel of the natural map $H_1(\partial(\mathbb{S}^3 \setminus K); \mathbb{Z}) \to H_1(\mathbb{S}^3 \setminus K; \mathbb{Z})$ induced by inclusion.*

Similarly, the meridian C_ℓ of the neighborhood of K represents the generator of $H_1(\mathbb{S}^3 \setminus K; \mathbb{Z}) = \mathbb{Z}$, and homologically plays the same role for $\mathbb{S}^3 \setminus K$ that the longitude $\mathbb{S}^1 \times 1$ plays for the solid torus $\mathbb{S}^1 \times \mathbb{D}^2$. Thus, we call C_ℓ the **longitude** *of the knot complement. (In short, a meridian of K is a longitude of $\mathbb{S}^3 \setminus K$, and vice-versa.)*



12.8. Homologically, a knot complement is a solid torus



12.9. A longitude bounds in the knot complement

Since the knot complement $\mathbb{S}^3 \setminus K$ is homologically identical to the solid torus $\mathbb{S}^1 \times \mathbb{D}^2$, it follows that the 4–manifold

$$(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$$

will be homologically indistinguishable from a thickened torus $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{D}^2$. In other words, $(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$ is a homological copy of $T \times \mathbb{D}^2$ (where T is the nice torus of M that was expelled earlier). Further, both $(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$ and $T \times \mathbb{D}^2$ have boundary a 3–torus. We identify these boundaries in a manner that respects the homological identification of $(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$ with $T \times \mathbb{D}^2$.

> *That is to say, we choose an identification of $\partial\big((\mathbb{S}^3 \setminus K) \times \mathbb{S}^1\big)$ with $\partial(T \times \mathbb{D}^2)$ such that the meridian $C_m \times 1$ of $(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$ be sent to the meridian $1 \times \mathbb{S}^1$ of $T \times \mathbb{D}^2$. (The longitude $C_\ell \times 1$ of $(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$ can be sent to any generator of $H_1(\mathbb{T}^2; \mathbb{Z})$ in $\mathbb{T}^2 \times \mathbb{S}^1$; there is no canonical choice here.)*

Plug it in. Using such an identification, we glue $(\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$ to $M \setminus T$, as in figure 12.10 on the next page. Since homologically this is the same as gluing $T \times \mathbb{D}^2$ back into $M \setminus T$ and rebuilding M, it follows that the resulting 4–manifold

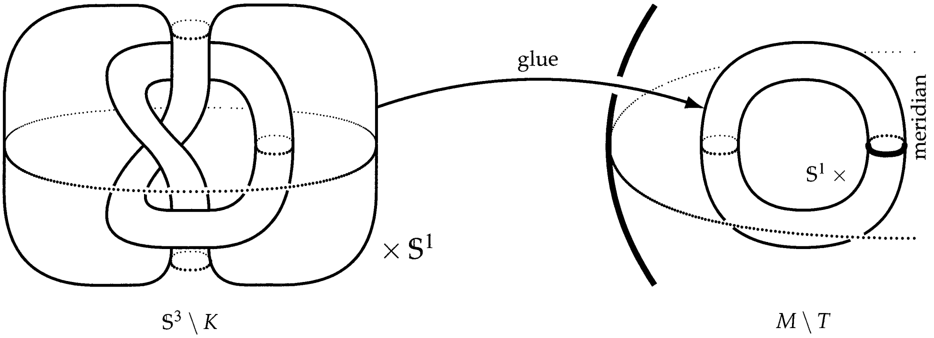
$$M_K \;=\; M \setminus T \;\;\cup_{(1\times\mathbb{S}^1) = (C_m \times 1)}\; (\mathbb{S}^3 \setminus K) \times \mathbb{S}^1$$

has the same homology as M, and in particular the same intersection form.

Since we assumed that $M \setminus T$ was simply-connected, the resulting M_K will be simply-connected as well, and therefore it follows from Freedman's classification that the two 4-manifolds are in fact homeomorphic:

$$M_K \;\simeq\; M\,.$$

Since the construction of M_K depends upon several other choices besides picking the knot K, in the sequel M_K will denote *any* manifold obtained



12.10. Knot surgery

after such a surgery. (It is possible that all such are diffeomorphic, but nothing is known about that.)

The result. The gluing statements presented earlier allow one to compare the Seiberg–Witten invariants of M_K and M and show that in general M_K and M are *not* diffeomorphic. Indeed, the Seiberg–Witten invariants of M and M_K differ exactly by the Alexander polynomial Δ_K of the knot K:

Fintushel–Stern Theorem. *Let M be a 4–manifold with $b_2^+(M) \geq 2$. Assume that it contains a homologically-nontrivial near-cusp embedded torus T. Perform surgery along T using some knot K and denote by M_K the resulting 4–manifold. Then:*

$$\mathcal{SW}_{M_K} = \mathcal{SW}_M \cdot \Delta_K ,$$

where the Alexander polynomial $\Delta_K(t)$ is here evaluated on $t = e^{2T}$.

One can rewrite the above result explicitly as:

$$\sum_{\kappa \text{ basic in } M_K} \mathcal{SW}_{M_K}(\kappa) \cdot e^{\kappa} = \Big(\sum_{\kappa \text{ basic in } M} \mathcal{SW}_M(\kappa) \cdot e^{\kappa} \Big) \cdot \Delta_K(e^{2T}) .$$

> ***Idea of the proof.*** By using the earlier Seiberg–Witten gluing formulae and various ingenious constructions, one obtains a skein-like relation for the Seiberg–Witten invariants of knot-surgered manifolds. Since this skein relation reveals itself to be the same with the one defining the Alexander polynomial, the result follows. Indeed, when $t = e^{2T}$, the factor $(e^T - e^{-T})$ from the gluing results corresponds to $(t^{1/2} - t^{-1/2})$ for the Alexander polynomial. □

Looking closer at the proof of this result, one can notice that in fact the knot surgery is interpreted as a series of generalized logarithmic transformations on null-homologous tori. Indeed, the unknotting of K is done by a series of crossing changes, each achieved through a simple surgery on a surrounding torus.

Finally, it should come as no surprise that M_K is fragile: just one stabilization is enough to undo it: $M_K \# S^2 \times S^2 \cong M \# S^2 \times S^2$.

12.4. Applications

The technique above can be applied to any smooth 4–manifold that contains a suitable torus. As an example, consider the $K3$ surface:

Exotic K3's. The $K3$ surface has $\mathcal{SW}_{K3} = 1$. Then, picking for T above a generic torus fiber F of $K3$ and for K some random knot, we get

$$\mathcal{SW}_{K3_K} = \Delta_K\big(e^{2F}\big) .$$

Since it is known that *every* symmetric Laurent polynomial $a_n t^{-n} + \cdots + a_n t^n$ with coefficient sum $a_n + \cdots + a_0 + \cdots + a_n = \pm 1$ can be realized as the Alexander polynomial of some knot, we conclude that:

Corollary. *There are infinitely-many 4–manifolds homeomorphic but not diffeomorphic to K3.* □

Of course, something like this was already known from performing logarithmic transformations on $E(2)$, which yielded infinitely-many complex surfaces homeomorphic but not diffeomorphic[9] to $K3$. What is essentially new here is that most of these $K3_K$'s are not even *symplectic:*

Non-symplectic manifolds. No matter what M is, if it contains a suitable torus, then the result M_K of a random knot surgery will most likely *not* admit any symplectic structures:

Corollary. *If Δ_K is* not *monic, then M_K cannot admit any symplectic structures.*

> ***Proof.*** This happens because M_K has no class in $H^2(M;\mathbb{Z})$ that might play the role of the anti-canonical class K^*_ω. If M_K were symplectic, then $\mathcal{SW}_{M_K}(K^*_\omega) = \pm 1$. However, since Δ_K is not monic, there is no class κ for which we could have $\mathcal{SW}_{M_K}(\kappa) = \pm 1$. □

Thus, beyond the close-to-complex realm lie truly vast fields of unexplored non-symplectic 4–manifolds, of which we know close to nothing.

Knots and mazes. Remember that the manifold M_K depends on certain choices made during the surgery. A first mystery is whether the various possible versions of M_K are diffeomorphic or not; the Seiberg–Witten invariants are blind here.

---

[9] See section 8.4 (page 314).

In fact, using certain distinct knots with the same Alexander polynomial and a more involved construction, Fintushel and Stern built (non-simply-connected) symplectic 4–manifolds that have the same Seiberg–Witten invariants, but are non-diffeomorphic (they were distinguished by the Seiberg–Witten invariants of their cyclic covers).

Moreover, in light of the Fintushel–Stern surgery theorem, it is clear that 4–dimensional topology sees some knot theory. The question is whether it sees only the Alexander polynomial, or if it also detects differences between knots to which the Alexander polynomial is oblivious. While, as we mentioned, the complement of a knot determines completely the knot, it is not clear how much of this information survives the surgery procedure.

For example, a proposed conjecture is:

Fintushel–Stern Conjecture (*open*)*.* *Two 4–manifolds*

$$K3_{K'} \qquad \textit{and} \qquad K3_{K''}$$

are diffeomorphic if and only if K' and K'' are equivalent knots.

As stated, the conjecture is not true, since we have

Lemma (*S. Akbulut*)*. If $-K$ denotes the mirror-image of the knot K, then we have a diffeomorphism*

$$M_K \cong M_{-K} . \qquad \square$$

Of course, it is still possible that this is the only reduction, and that two non-equivalent knots yield diffeomorphic M_K's if and only if they are mirror-images of each other. If true, this would mean that 4–manifold topology includes essentially the whole complexity of knot theory.

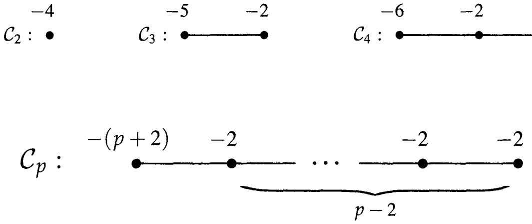
12.5. Notes

Note: Rational blow-downs

In this note we present an interesting construction, due to R. Fintushel and R. Stern, called *rational blow-down*. A usual blow-down replaces the neighborhood of a (-1)–sphere by a 4–ball. A rational blow-down replaces a neighborhood of a certain configuration of spheres by a rational-homology 4–ball.

The relevance of this construction is that, on one hand, any logarithmic transformation can be described as a sequence of blow-ups followed by a rational blow-down. On the other hand, the change of the Seiberg–Witten invariants under a rational blow-down is particularly amenable. Finally, rational blow-downs can be used to reduce the homology of a 4–manifold, and recently rational blow-downs and their generalizations have been used to build exotic 4–manifolds with small homology, for example exotic $\mathbb{CP}^2 \# 5\,\overline{\mathbb{CP}}^2$'s.

Ingredients. Take $p-1$ copies of S^2 and build on them disk-bundles of Euler classes $-(p+2), -2, \dots, -2$, then plumb these according to the diagram in figure 12.11. The result is a simply-connected 4–manifold C_p, whose boundary is the lens space[1] $L(p^2, p-1)$. In particular, $\pi_1(\partial C_p) = \mathbb{Z}_{p^2}$.



12.11. Plumbing diagram for C_p

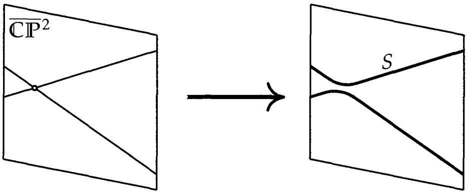
Lemma. *The 4–manifold C_p can be embedded in $\#(p-1)\overline{\mathbb{CP}}^2$.*

Proof. All we need to do is locate a configuration of $p-1$ embedded spheres in $\#(p-1)\overline{\mathbb{CP}}^2$ that intersect according to the diagram above; then a neighborhood of those spheres will be a copy of C_p.

Start with $\overline{\mathbb{CP}}^2$ and denote by e_1 the class of $\overline{\mathbb{CP}}^1$. Take two projective lines in $\overline{\mathbb{CP}}^2$ and eliminate their single intersection point in the usual fashion, as suggested in figure 12.12 on the next page. The result is a sphere S of homology class $2e_1$, with self-intersection -4. Therefore, a neighborhood of S is the manifold C_2.

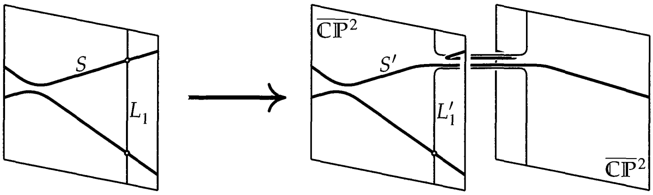
---

[1] The **lens space** $L(m,n)$ is the closed 3–manifold obtained as the quotient of $S^3 \subset \mathbb{C}^2$ by the action of $\mathbb{Z}_m$ induced by $(z_1, z_2) \mapsto (e^{2\pi i/m} z_1,\ e^{n\,2\pi i/m} z_2)$. Its fundamental group is $\mathbb{Z}_m$.



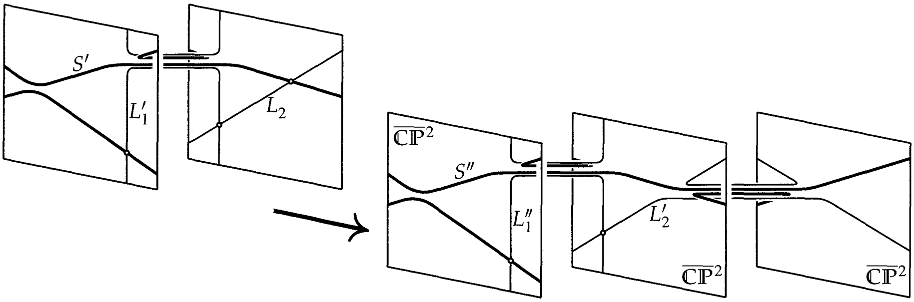
12.12. Embedding C_2 in $\overline{\mathbb{CP}}^2$

Take now another projective line L_1 in $\overline{\mathbb{CP}}^2$, meeting S in two points, as in figure 12.13. Orient L_1 so that both intersection points are positive (thus, $[L_1] = -e_1$). Blow-up at one of the two intersection points and denote by e_2 the class of the added exceptional curve. The result in $(\overline{\mathbb{CP}}^2)' = \overline{\mathbb{CP}}^2 \# \overline{\mathbb{CP}}^2$ is one sphere S' of class $2e_1 - e_2$ and another one L_1' of class $-e_1 - e_2$. These spheres have self-intersections -5 and -2 and meet in one positive point. Therefore, a neighborhood of $S' \cup L_1'$ is a copy of C_3.



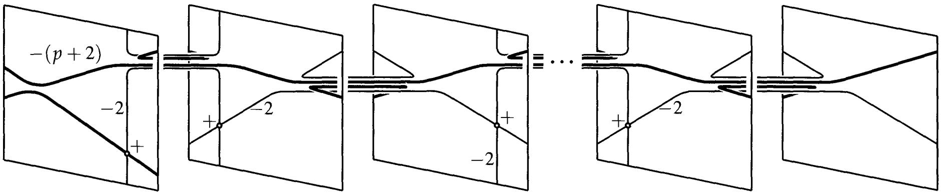
12.13. Embedding C_3 in $\# 2\overline{\mathbb{CP}}^2$

The exceptional sphere of the blown-up $\overline{\mathbb{CP}}^2$ can be represented by a sphere L_2 of class $-e_2$, meeting each of S' and L_1' in one positive point, as sketched in figure 12.14. If we blow-up the intersection point of L_2 with S', the result in $\# 3\overline{\mathbb{CP}}^2$ is a configuration of three spheres: a sphere S'' of class $2e_1 - e_2 - e_3$, a sphere L_1'' of class $-e_1 - e_2$, and a sphere L_2' of class $-e_2 - e_3$. A neighborhood of $S'' \cup L_1'' \cup L_2'$ is a copy of C_4.



12.14. Embedding C_4 in $\# 3\overline{\mathbb{CP}}^2$

This procedure can be continued in the obvious manner, yielding embeddings of C_p in $\#(p-1)\overline{\mathbb{CP}}^2$, as a neighborhood of spheres of classes $2e_1 - e_2 - \cdots - e_{p-1}$, $-e_2 - e_3$, $-e_3 - e_4$, $\ldots$, $-e_{p-2} - e_{p-1}$. See Figure 12.15. □



12.15. Embedding $\mathcal{C}_p$ in $\#(p-1)\overline{\mathbb{CP}}^2$

Now notice that, instead of embedding $\mathcal{C}_p$ in $\#(p-1)\overline{\mathbb{CP}}^2$, we can reverse orientations and embed $\overline{\mathcal{C}}_p$ inside $\#(p-1)\mathbb{CP}^2$. Then the complement $\#(p-1)\mathbb{CP}^2 \setminus \overline{\mathcal{C}}_p$ is a 4–manifold $\mathcal{B}_p$, with boundary $L(p^2, p-1)$. The inclusion $\partial \mathcal{B}_p \subset \mathcal{B}_p$ induces a surjective morphism $\pi_1(\partial \mathcal{B}_p) \to \pi_1(\mathcal{B}_p)$ and we have $\pi_1(\mathcal{B}_p) = \mathbb{Z}_p$. Nonetheless, since the homology of $\overline{\mathcal{C}}_p$ exhausts all of $H_2(\#(p-1)\mathbb{CP}^2;\ \mathbb{Q})$, it follows that $\mathcal{B}_p$ has $H_*(\mathcal{B}_p;\ \mathbb{Q}) = 0$. In conclusion,

Lemma. *The manifold C_p has the same boundary as a rational-homology 4–ball $\mathcal{B}_p$.* □

Construction of a rational blow-down. If $\mathcal{C}_p$ is embedded inside some 4–manifold M, then we could cut it out of M and replace it by a copy of $\mathcal{B}_p$. Specifically, if M contains a configuration of embedded 2–spheres as prescribed in the plumbing diagram of $\mathcal{C}_p$, then a neighborhood of this configuration in M must be a copy of $\mathcal{C}_p$. Then M can be split as

$$M = M^\circ \cup_\partial \mathcal{C}_p .$$

By replacing $\mathcal{C}_p$ by $\mathcal{B}_p$ we obtain the new manifold

$$M_{(p)} = M^\circ \cup_\partial \mathcal{B}_p ,$$

which is called the **rational blow-down**[2] of $\mathcal{C}_p$ from M.

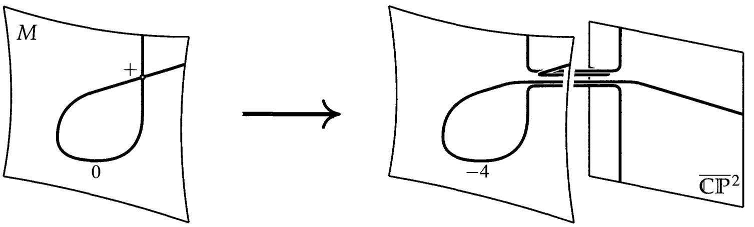
Notice that, if both M and M° are simply-connected, then so is $M_{(p)}$. The homology $H_2(M_{(p)};\ \mathbb{Z})$ can be identified with the Q_M–orthogonal complement of $H_2(\mathcal{C}_p; \mathbb{Z})$ in $H_2(M; \mathbb{Z})$; in other words, with the complement of the classes represented by the spheres in M used to embed $\mathcal{C}_p$. Since moving from M to $M_{(p)}$ eliminates classes of negative self-intersection, it follows that $b_2^+ M = b_2^+ M_{(p)}$ while the signature has increased.

Logarithmic transformations. Imagine that some 4–manifold M contains an immersed sphere S with only one double-point of positive sign. Assume further that S has zero self-intersection in homology, $[S] \cdot [S] = 0$. The typical example is when M is an elliptic surface and S is a fishtail fiber.

In this situation, the double-point of S can be blown-up, creating a sphere S' embedded in $M \# \overline{\mathbb{CP}}^2$ and of homology class $[S'] = [S] - 2e_1$, as in figure 12.16 on the following page. This is a sphere of self-intersection -4 and thus its neighborhood is a copy of $\mathcal{C}_2$.

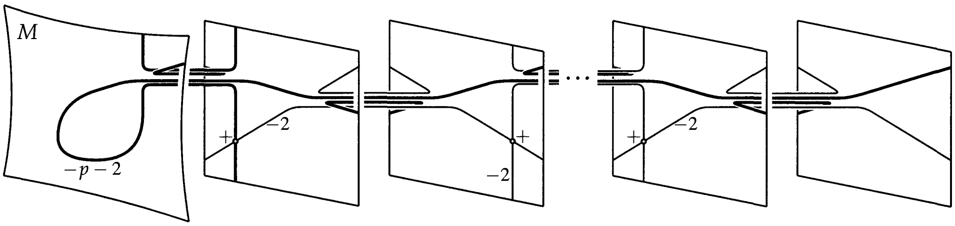
---

[2] The resulting $M_{(p)}$ does not depend on the particular choice of gluing used to fit M° to $\mathcal{B}_p$ along their common boundary $L(p^2, p-1)$, because one can prove that any self-diffeomorphism of $L(p^2, p-1)$ extends to a self-diffeomorphism of $\mathcal{B}_p$.



12.16. C_2 in M' from an immersed sphere in M

One can settle with rationally blowing-down C_2, or one can continue with more blow-ups. Then, just as in our construction of the embedding of C_p in $\#(p-1)\overline{\mathbb{CP}}^2$, one obtains an embedding of C_p in $M\,\#\,(p-1)\,\overline{\mathbb{CP}}^2$, with spheres of classes $[S]-2e_1-e_2-\cdots-e_{p-1}$, e_1-e_2, e_2-e_3, … , $e_{p-2}-e_{p-1}$. See figure 12.17.



12.17. Building C_p on an immersed sphere

In the case when M is an elliptic surface and S is a fishtail fiber, the result of such a sequence of blow-ups followed by a rational blow-down is in fact diffeomorphic to a logarithmic transformation:

Theorem *(R. Fintushel and R. Stern). If E is an elliptic surface containing a cusp fiber and C_p is built inside $E\,\#(p-1)\,\overline{\mathbb{CP}}^2$ by starting with a fishtail fiber S of E, then the result of the rational blow-down of C_p is diffeomorphic to the result of a logarithmic transformation of multiplicity p, performed on a regular torus fiber near S:*

$$\big(E\,\#(p-1)\,\overline{\mathbb{CP}}^2\big)_{(p)} \cong E_p\,.$$

Idea of proof. One uses Kirby calculus. The cusp fiber is present to ensure that there is no ambiguity for the logarithmic transformation. See **R. Gompf** and **A. Stpsicz**'s ***4–Manifolds and Kirby calculus*** [GS99, sec 8.5]. □

Seiberg–Witten invariants. Consider a 4–manifold M containing a suitable configuration of spheres so that C_p embeds in M. Then M splits as $M = M^\circ \cup_\partial C_p$ and leads to the corresponding rational blow-down $M_{(p)} = M^\circ \cup_\partial \mathcal{B}_p$. The Seiberg–Witten invariants of M and $M_{(p)}$ fit rather nicely:

Theorem *(R. Fintushel and R. Stern). Assume that both M and $M_{(p)}$ are simply-connected. For every characteristic element κ of $M_{(p)}$ there exists a characteristic element $\overline{\kappa}$ of M so that $\kappa|_{M^\circ} = \overline{\kappa}|_{M^\circ}$ and $\kappa\cdot\kappa - \overline{\kappa}\cdot\overline{\kappa} = p-1$. Then we have*

$$\mathcal{SW}_{M_{(p)}}(\kappa) = \mathcal{SW}_M(\overline{\kappa})\,.$$

Idea of proof. First off, one can identify $H_2(M_{(p)};\mathbb{Z})$ with the complement of the spheres from $\mathcal{C}_p$ in $H_2(M;\mathbb{Z})$. The requirement that $\kappa\cdot\kappa-\bar{\kappa}\cdot\bar{\kappa}=p-1$ merely ensures that the Seiberg–Witten moduli spaces corresponding to κ on $M_{(p)}$ and to $\bar{\kappa}$ on M have the same (expected) dimension, $\dim\mathfrak{M}_\kappa=\dim\mathfrak{M}_{\bar{\kappa}}$.

For the actual formula, the starting point is the fact that $\mathcal{B}_p$ admits a metric of positive scalar curvature. Then the argument flows similarly to the proof of the usual Seiberg–Witten blow-up formula (page 407), by stretching the connecting neck $\partial\times(-\varepsilon,\varepsilon)$ between M° and $\mathcal{B}_p$. □

Notice that the above result is interesting only when $\bar{\kappa}\cdot\bar{\kappa}\geq 3\operatorname{sign}M+2\chi(M)$, since otherwise the expected dimension of $\mathfrak{M}_{\bar{\kappa}}$ is negative and then automatically $\mathcal{SW}_M(\bar{\kappa})=0$ and $\mathcal{SW}_{M_{(p)}}(\kappa)=0$.

Using this theorem, one can evaluate the Seiberg–Witten invariants of 4–manifolds obtained as rational blow-downs from better-understood 4–manifolds. In particular, the Seiberg–Witten invariants of elliptic surfaces can be evaluated in this fashion.

References. The rational blow-down procedure and its effect on gauge theory invariants was first explored in **R. Fintushel** and **R. Stern**'s ***Rational blowdowns of smooth 4–manifolds*** **[FS97a]**, where the invariants of $E(n)_{p,q}$ were thus evaluated. A nice exposition can be read from **R. Gompf** and **A. Stpsicz**'s ***4–Manifolds and Kirby calculus*** [**GS99**, sec 8.5].

The rational blow-down construction can be extended to more general configurations of spheres. Namely, one can consider the 4–manifold $\mathcal{C}_{p,q}$ obtained by plumbing a string of disk-bundles with Euler classes specified by the coefficients of the continued fraction expansion of $-p^2/pq-1$, with $p>q$ and coprime. The boundary of $\mathcal{C}_{p,q}$ is the lens space $L(p^2,pq-1)$, which was known since **A. Casson** and **J. Harer**'s ***Some homology lens spaces which bound rational homology balls*** **[CH81]** to bound a rational-homology 4–ball $\mathcal{B}_{p,q}$. Therefore such configurations of spheres can be used for a **generalized rational blow-down**. This generalization was studied in **M. Symington**'s ***Generalized symplectic rational blowdowns*** **[Sym01]** and **J. Park**'s ***Seiberg–Witten invariants of generalised rational blow-downs*** **[Par97]**. Recently, generalized rational blow-downs have been used to build new exotic 4–manifolds, for example exotic $\mathbb{CP}^2\,\#5\,\overline{\mathbb{CP}^2}$'s.

Bibliography

The gluing formulae for the Seiberg–Witten invariants appeared in **J. Morgan, T. Mrowka** and **Z. Szabó**'s ***Product formulae along*** $\mathbb{T}^3$ ***for Seiberg–Witten invariants*** **[MMS97]**, with the case of near-cusp logarithmic transformations obtained in **R. Fintushel** and **R. Stern**'s ***Rational blowdowns of smooth 4–manifolds*** **[FS97a]**. The fact that generalized logarithmic transformations on elliptic surfaces result in manifolds diffeomorphic to the complex logarithmic transformation was proved by **R. Gompf**'s ***Nuclei of elliptic surfaces*** **[Gom91a]**.

Other, more general gluing results were published in **C. Taubes**' ***The Seiberg–Witten invariants and 4–manifolds with essential tori*** **[Tau01]**, and **D. Park**'s version

A gluing formula for the Seiberg–Witten invariant along $\mathbb{T}^3$ **[Par02b]**, the former with somewhat more emphasis on the analytic side, while the latter, on the geometric side. Results about gluing along surfaces other than tori can be found in **J. Morgan, Z. Szabó** and **C. Taubes'** ***A product formula for the Seiberg–Witten invariants and the generalized Thom conjecture*** **[MST96]**.

One of the techniques use for proving such gluing results is the study of the Seiberg–Witten equations on 4–manifolds with cylindrical ends. A different approach is to study the Seiberg–Witten–Floer homology of the 3–manifold along which you glue, see for example **V. Muñoz** and **B. Wang's** ***Seiberg–Witten–Floer homology of a surface times a circle*** **[MW99]**. The setting of more general results about gluing along 3–manifolds inevitably involves this Seiberg–Witten–Floer homology.

Gluing result have been used for proving the adjunction inequality for positive self-intersection in **J. Morgan, Z. Szabó** and **C. Taubes'** ***A product formula for the Seiberg–Witten invariants and the generalized Thom conjecture*** **[MST96]**, or for building irreducible non-symplectic manifold in **Z. Szabó's** ***Simply-connected irreducible 4–manifolds with no symplectic structures*** **[Sza98]**.

The Alexander polynomial was discovered in **J. Alexander's** ***Topological invariants of knots and links*** **[Ale28]**. For this and other classic topics in knot theory, one should look at **D. Rolfsen's** beautifully geometric ***Knots and links*** **[Rol76, Rol90, Rol03]**. The same author offers a surgery interpretation of the Alexander polynomial in **[Rol75]**. A recent monograph on knot theory, including descriptions of invariants finer than the Alexander polynomial, is **R. Lickorish's** ***An introduction to knot theory*** **[Lic97]**. The fact that knots are determined by their complements was proved in **C. Gordon** and **J. Luecke's** **[GL89a, GL89b]**.

The knot surgery results have appeared in **R. Fintushel** and **R. Stern's** ***Knots, links, and 4–manifolds*** **[FS98b]**. The link between the Seiberg–Witten invariants and the Alexander polynomial first became apparent in the case of 3–manifold invariants, in **G. Meng** and **C. Taubes'** ***$\underline{SW}$ = Milnor torsion*** **[MT96]**. A handle description (Kirby calculus) of the Fintushel–Stern surgery can be found in **S. Akbulut's** ***A fake cusp and a fishtail*** **[Akb99]**. This is used in ***Variations on Fintushel–Stern knot surgery on 4–manifolds*** **[Akb02]** to show that $M_K \cong M_{-K}$.

A construction similar to knot surgery ("rim surgery") can be used to change the smooth embedding of a surface in a 4–manifold, see **R. Fintushel** and **R. Stern's** ***Surfaces in 4–manifolds*** **[FS97b]**. Examples of 4–manifolds with the same Seiberg–Witten invariants appeared in their ***Nondiffeomorphic symplectic 4–manifolds with the same Seiberg–Witten invariants*** **[FS99]**. In ***Families of simply-connected 4–manifolds with the same Seiberg–Witten invariants*** **[FS02]** are constructed infinite families of homeomorphic 4–manifolds that have the same Seiberg–Witten invariants, but one might expect them to be non-diffeomorphic (there are no tools to make such a decision). A quick survey of their work related to knot surgery is ***Constructions of smooth 4–manifolds*** **[FS98a]**. Examples of non-symplectic 4–manifolds with exactly one Seiberg–Witten basic class were built in ***Nonsymplectic 4–manifolds with one basic class*** **[FS00]** (the previous examples of non-symplectics had many basic classes).

Vast geographies. At first, there was the conjecture that all simply-connected 4–manifolds might split into connected sums of complex surfaces, with either orientation. This fell with the construction of irreducible[3] non-complex examples in **R. Gompf** and **T. Mrowka**'s ***Irreducible 4–manifolds need not be complex*** **[GM93]**. Then, there was hope that all simply-connected 4–manifolds might decompose into sums of symplectic manifolds, with either orientations. This was negated in **Z. Szabó**'s ***Simply-connected irreducible 4–manifolds with no symplectic structures*** **[Sza98]**. Today, we have no conjecture about what the building blocks of all 4–manifolds might be.

Certainly before any conjectures can even be attempted, we need to construct enough examples, and hope that eventually some patterns will become apparent. Thus, some energy has been devoted in recent years toward this goal, by using mainly fiber sums, logarithmic transformations, (generalized) rational blowdowns[4] and Fintushel–Stern surgeries, often in the direction of building exotic manifolds with smaller-and-smaller homology. A starting reference is **R. Stern**'s optimistically-titled ***Will we ever classify simply-connected smooth 4–manifolds?*** **[Ste05]**. In what follows, we gather a rather random sampling of such constructive results:

On one hand, for every even form

$$Q = \oplus \pm 2m\, E_8 \oplus n\, H$$

that satisfies the 3/2–conjecture (*i.e.,*[5] $\chi \geq \frac{3}{2}\,|\mathrm{sign}|$) and also has $b_2^+ = n$ odd and sign $= 8m$ non-zero, **D. Park** and **Z. Szabó**'s ***The geography problem for irreducible spin four-manifolds*** **[PS00]** built simply-connected irreducible 4–manifolds realizing that form; for each such Q, they had both symplectic (for the orientation with sign < 0) and non-symplectic examples. (Symplectic examples for the orientation with sign > 0 were constructed in **[Par02d]**.) The case of even forms with sign $= 0$, that is to say, of manifolds homeomorphic to $\#m\,\mathbb{S}^2 \times \mathbb{S}^2$, was attacked by **J. Park**'s ***The geography of Spin symplectic 4–manifolds*** **[Par02d]**, which proved that there must be some (undetermined) m_0 such that for all odd $m \geq m_0$ the friendly manifold $\#m\,\mathbb{S}^2 \times \mathbb{S}^2$ must admit infinitely-many smooth structures.

On the other hand, on the side of odd intersection forms

$$Q = \oplus\, m\,[+1] \oplus n\,[-1]\ ,$$

we have seen exotic smooth structures on $\#3\,\mathbb{CP}^2 \# 19\,\overline{\mathbb{CP}}^2$ and $\mathbb{CP}^2 \# 9\,\overline{\mathbb{CP}}^2$ from logarithmic transformations on $E(2)$ and $E(1)$. Then **J. Park**'s ***Exotic smooth structures on 4–manifolds*** **[Par02c]** showed that each form Q with $n \geq m+7$ and $m = b_2^+$ odd ≥ 3 is realized by infinitely-many smooth manifolds; all these are irreducible when $2\chi + 3\,\mathrm{sign} \geq 0$. For the study of the positive-signature cases, Lefschetz fibrations were used in **A. Stispicz**'s ***Simply connected symplectic 4–manifolds with positive signature*** **[Sti99]**, and combined with Fintushel–Stern surgery in **J. Park**'s ***Exotic smooth structures on 4–manifolds. II*** **[Par03a]**.

---

[3] Remember that a manifold M^4 is called **irreducible** if it does not split into a connected sum of two manifold, none of which is homeomorphic to $\mathbb{S}^4$, *i.e.*, each term splits off some homology from M.

[4] See the preceding note, on page 547.

[5] Where χ is viewed as determined by the intersection form through $\chi = 2 + b_2^+ + b_2^-$.

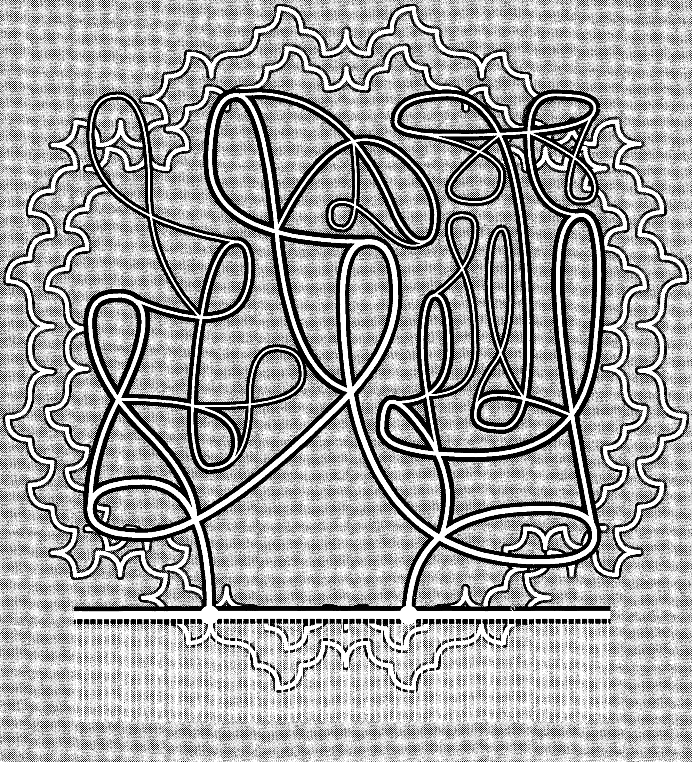
For manifolds homeomorphic to $\#3\,\mathbb{CP}^2\,\#n\,\overline{\mathbb{CP}}^2$, we have **D. Park**'s ***Constructing infinitely many smooth structures on*** $\mathbf{3\,\mathbb{CP}^2\,\#\,n\,\overline{\mathbb{CP}}^2}$ **[Par02a]**; his structures are both symplectic and non-symplectic, are built for all $n \geq 10$, and are irreducible when $10 \leq n \leq 13$. Recently, infinitely-many smooth structures on $\#3\,\mathbb{CP}^2\,\#\,9\,\overline{\mathbb{CP}}^2$ were built in **A. Stipsicz** and **Z. Szabó**'s ***Small exotic 4–manifolds with*** $\boldsymbol{b_2^+ = 3}$ **[SS05]**.

For manifolds homeomorphic to $\mathbb{CP}^2\,\#n\,\overline{\mathbb{CP}}^2$, we had **D. Kotschick**'s ***On manifolds homeomorphic to*** $\mathbf{\mathbb{CP}^2\,\#\,8\,\overline{\mathbb{CP}}^2}$ **[Kot89]**, which proved that the Barlow surface[6] is an exotic $\mathbb{CP}^2\,\#\,8\,\overline{\mathbb{CP}}^2$. More recently, **J. Park**'s ***Simply connected symplectic 4–manifolds with*** $\boldsymbol{b_2^+ = 1}$ ***and*** $\boldsymbol{c_1^2 = 2}$ **[Par03b]** used rational blow-downs to build an exotic $\mathbb{CP}^2\,\#7\,\overline{\mathbb{CP}}^2$ and inspired a whole series of new constructions. It was first followed by **A. Stipsicz** and **Z. Szabó**'s ***An exotic smooth structure on*** $\mathbf{\mathbb{CP}^2\,\#\,6\,\overline{\mathbb{CP}}^2}$ **[SS04]**. Then **R. Fintushel** and **R. Stern**'s ***Double node neighborhoods and families of simply connected 4–manifolds with*** $\boldsymbol{b^+ = 1}$ **[FS04]** built infinitely-many distinct (non-symplectic) smooth structures on $\mathbb{CP}^2\,\#n\,\overline{\mathbb{CP}}^2$ for $n = 6, 7, 8$. Shortly afterwards, **J. Park, A. Stipsicz** and **Z. Szabó**'s ***Exotic smooth structures on*** $\mathbf{\mathbb{CP}^2\,\#\,5\,\overline{\mathbb{CP}}^2}$ **[PSS04]** dealt with the $n = 5$ case. How low can you go?

New results in this direction seem to appear all the time, and thus the above paragraphs might become obsolete even before this volume reaches bookstores...

---

[6] The Barlow surface is a simply-connected complex surface of general type, built in **R. Barlow**'s ***A simply connected surface of general type with*** $\boldsymbol{p_g = 0}$ **[Bar85]**.



Epilogue

Under the light of the various examples seen in this book, it seems reasonable to conjecture that, if a topological 4–manifold admits a smooth structure at all, then it might admit infinitely many.

While gauge theory was how the door was opened on those vast unexplored realms, it might not be how these will be charted. We have seen that there are whole realms where the Seiberg–Witten invariants cannot help us. For example, the theory is blind on 4–manifolds that admit metrics of positive scalar curvature, on homology 4–spheres (which in particular leaves the *smooth* 4–dimensional Poincaré conjecture with no solution in sight), on all manifolds with b_2^+ even, and in general on 4–manifolds that are far from complex.

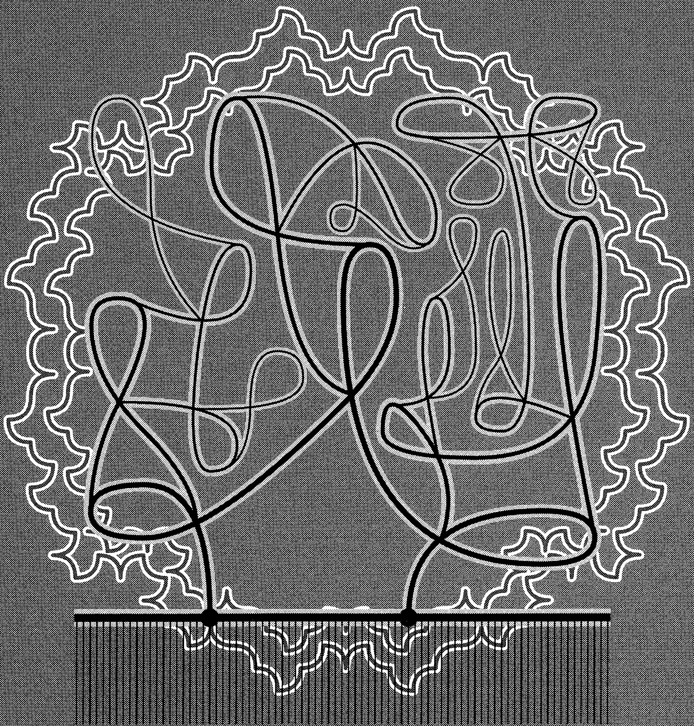
More, gauge theory offers only negative results (as in "two manifolds are *not* diffeomorphic"). Indeed, the field of 4–manifolds lacks enough techniques for obtaining affirmative results (as in "two manifolds *are* diffeomorphic"). Looking back, the only affirmative results we encountered came either from *ad hoc* constructions, from Kirby calculus, or from complex geometry. The field also lacks techniques for building enough examples, which might one day be organized into any sort of classification scheme. We are lost in an ever-growing jungle.

Hence the final conclusion of this volume can only be that

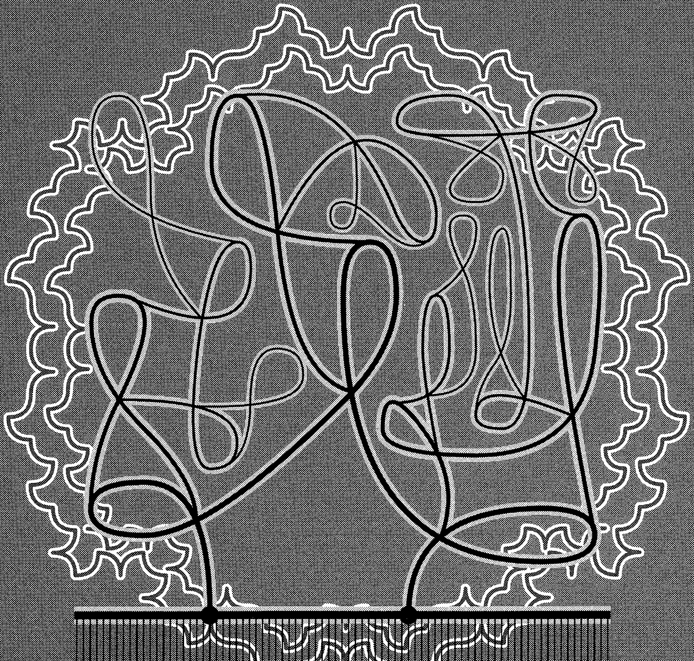
We know that we don't know.

This only makes it all the more exciting...

January 2005



List of Figures and Tables



Bibliography

[ADK04] Denis Auroux, Simon K. Donaldson, and Ludmil Katzarkov, *Singular Lefschetz pencils*, 2004. `arXiv:math.DG/0410332`.

[AHS78] Michael F. Atiyah, Nigel J. Hitchin, and Isadore M. Singer, *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. London Ser. A **362** (1978), no. 1711, 425–461. MR 80d:53023.

[Akb91] Selman Akbulut, *A fake compact contractible* 4*-manifold*, J. Differential Geom. **33** (1991), no. 2, 335–356. MR 92b:57025.

[Akb96] ______, *Lectures on Seiberg–Witten invariants*, Turkish J. Math. **20** (1996), no. 1, 95–118. MR 97g:57038. `arXiv:alg-geom/9510012`.

[Akb99] ______, *A fake cusp and a fishtail*, Proceedings of 6th Gökova Geometry-Topology Conference, vol. 23, Turkish J. Math., no. 1, 1999, pp. 19–31. MR 2000g:57035. `arXiv:math.GT/9904058`.

[Akb02] ______, *Variations on Fintushel–Stern knot surgery on* 4*-manifolds*, Turkish J. Math. **26** (2002), no. 1, 81–92. MR 2002m:57039. `arXiv:math.GT/0201156`.

[AL94] Michèle Audin and Jacques Lafontaine (eds.), *Holomorphic curves in symplectic geometry*, Progress in Mathematics, vol. 117, Birkhäuser, Basel, 1994. MR 95i:58005.

[Ale28] J. W. Alexander, *Topological invariants of knots and links*, Trans. American Mathematical Society **30** (1928), no. 2, 275–306. MR 1 501 429.

[AM90] Selman Akbulut and John D. McCarthy, *Casson's invariant for oriented homology* 3*-spheres*, Mathematical Notes, vol. 36, Princeton University Press, Princeton, NJ, 1990. MR 90k:57017.

[Arf41] Cahit Arf, *Untersuchungen über quadratische Formen in Körpern der Charakteristik 2. I*, J. Reine Angew. Math. **183** (1941), 148–167. MR 4,237f.

[Aro57] N. Aronszajn, *A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order*, J. Math. Pures Appl. (9) **36** (1957), 235–249. MR 19,1056c.

[Ati67] Michael F. Atiyah, *K-theory*, W. A. Benjamin, Inc., New York-Amsterdam, 1967. MR 36 #7130.

[Ati89] ______, *K-theory*, second ed., Advanced Book Classics, Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, 1989. MR 90m:18011.

[Aur97] Denis Auroux, *Asymptotically holomorphic families of symplectic submanifolds*, Geom. Funct. Anal. **7** (1997), no. 6, 971–995. MR MR1487750 (99b:57069).

[Bar63] Dennis Barden, *The structure of manifolds*, Ph.D. thesis, Cambridge University, 1963.

[Bar85] Rebecca Barlow, *A simply connected surface of general type with* $p_g = 0$, Invent. Math. **79** (1985), no. 2, 293–301. MR 87a:14033.

[Bas68] Hyman Bass, *Algebraic K-theory*, W. A. Benjamin, New York–Amsterdam, 1968. MR 40 #2736.

[Bau92] Stefan Bauer, *Some nonreduced moduli of bundles and Donaldson invariants for Dolgachev surfaces*, J. Reine Angew. Math. **424** (1992), 149–180. MR MR1144164 (93e:14043).

[Bau03] ———, *Refined Seiberg–Witten invariants*, 2003. `arXiv:math.GT/0312523`.

[BBB81] Lionel Bérard-Bergery and Marcel Berger (eds.), *Géométrie riemannienne en dimension* 4, Textes Mathématiques, vol. 3, CEDIC, Paris, 1981, Papers from the Arthur Besse seminar held at the Université de Paris VII, Paris, 1978/1979. MR 86i:53020.

[BBW93] Bernhelm Booß-Bavnbek and Krzysztof P. Wojciechowski, *Elliptic boundary problems for Dirac operators*, Birkhäuser Boston, Boston, MA, 1993. MR 94h:58168.

[Bea78] Arnaud Beauville, *Surfaces algébriques complexes*, Astérisque, vol. 54, Société Mathématique de France, Paris, 1978. MR 58 #5686.

[Bea83] ———, *Complex algebraic surfaces*, London Mathematical Society Lecture Note Series, vol. 68, Cambridge University Press, Cambridge, 1983. MR 85a:14024.

[Bea96] ———, *Complex algebraic surfaces*, second ed., London Mathematical Society Student Texts, vol. 34, Cambridge University Press, Cambridge, 1996, Translated from the 1978 French original. MR 97e:14045.

[Ber98] Marcel Berger, *Riemannian geometry during the second half of the twentieth century*, Jahresber. Deutsch. Math.-Verein. **100** (1998), no. 2, 45–208. MR 99h:53001.

[Ber00] ———, *Riemannian geometry during the second half of the twentieth century*, University Lecture Series, vol. 17, American Mathematical Society, Providence, RI, 2000, Reprint of the 1998 original. MR 2000h:53002.

[Ber03] ———, *A panoramic view of Riemannian geometry*, Springer-Verlag, Berlin, 2003. MR 2004h:53001.

[Bes87] Arthur L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 10, Springer, Berlin, 1987. MR 88f:53087.

[BF04] Stefan Bauer and Mikio Furuta, *A stable cohomotopy refinement of Seiberg–Witten invariants: I & II*, Invent. Math. **155** (2004), no. 1, 1–19 & 21–40. `arXiv:math.DG/0204340` & `arXiv:math.DG/0204267`.

[BG96] Žarko Bižaca and Robert E. Gompf, *Elliptic surfaces and some simple exotic* $\mathbb{R}^4$*'s*, J. Differential Geom. **43** (1996), no. 3, 458–504. MR 97j:57044.

[BHPVdV04] W. Barth, K. Hulek, C. Peters, and A. Van de Ven, *Compact complex surfaces*, second enlarged ed., Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 4, Springer, Berlin, 2004.

[Bin59] R. H. Bing, *An alternative proof that* 3*-manifolds can be triangulated*, Ann. of Math. (2) **69** (1959), 37–65. MR 20 #7269.

[Biž94] Žarko Bižaca, *A handle decomposition of an exotic* $\mathbb{R}^4$, J. Differential Geom. **39** (1994), no. 3, 491–508. MR 96a:57045.

[Biž95] ———, *An explicit family of exotic Casson handles*, Proc. American Mathematical Society **123** (1995), no. 4, 1297–1302. MR 95e:57034.

[Boc48] Salomon Bochner, *Curvature and Betti numbers*, Ann. of Math. (2) **49** (1948), 379–390. MR 9,618d.

[Boh00] Christian Bohr, *Embedded surfaces and almost complex structures*, Proc. American Mathematical Society **128** (2000), no. 7, 2147–2154. MR 2000m:57051. `arXiv:math.DG/9812056`.

[Bor99] Jorge Luis Borges, *John Wilkins' analytical language*, Selected non-fictions (Eliot Weinberger, ed.), Viking, New York, NY, 1999, translated by Eliot Weinberger, pp. 229–232.

[BPVdV84] W. Barth, C. Peters, and A. Van de Ven, *Compact complex surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 4, Springer, Berlin, 1984. MR 86c:32026.

[Bri66] Egbert Brieskorn, *Beispiele zur Differentialtopologie von Singularitäten*, Invent. Math. **2** (1966), 1–14. MR 34 #6788.

[Bro60] Morton Brown, *A proof of the generalized Schoenflies theorem*, Bull. Amer. Math. Soc. **66** (1960), 74–76. MR 22 #8470b.

[Bro72] William Browder, *Surgery on simply-connected manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 65, Springer, New York, 1972. MR 50 #11272.

[Bru96] Rogier Brussee, *The canonical class and the C^∞ properties of Kähler surfaces*, New York J. Math. **2** (1996), 103–146 (electronic). MR 98a:57041. `arXiv:alg-geom/9503004`.

[BT82] Raoul Bott and Loring Tu, *Differential forms in algebraic topology*, Graduate Texts in Mathematics, vol. 82, Springer, New York, 1982. MR 83i:57016.

[Cai61] Stewart S. Cairns, *The manifold smoothing problem*, Bull. Amer. Math. Soc. **67** (1961), 237–238. MR 23 #A1383.

[Cas86] Andrew J. Casson, *Three lectures on new-infinite constructions in 4–dimensional manifolds*, À la recherche de la topologie perdue [**GM86a**], Progr. Math., vol. 62, Birkhäuser Boston, Boston, MA, 1986, pp. 201–244. MR 900 253.

[CC03] Alberto Candel and Lawrence Conlon, *Foliations. II*, Graduate Studies in Mathematics, vol. 60, American Mathematical Society, Providence, RI, 2003. MR 2004e:57034.

[Cer64] Jean Cerf, *La nullité de $\pi_0(\mathrm{Diff}\, S^3)$: Théorèmes de fibration des espaces de plongements. Applications; 1. Position du problème; 2. Espaces fonctionnels liés aux décompositions d'une sphére plongée dans $\mathbb{R}^3$; 3. Construction d'une section pour le revêtement $\mathcal{R}$.*, Séminaire Henri Cartan, 1962/63, Exp. 8, 9-10, 20, & 21, Secrétariat mathématique, Paris, 1964, p. 13. MR 33 #6641.

[Cer68a] ———, *Sur les difféomorphismes de la sphère de dimension trois ($\Gamma_4 = 0$)*, Lecture Notes in Mathematics, No. 53, Springer, Berlin, 1968. MR 37 #4824.

[Čer68b] A. V. Černavskiĭ, *Local contractibility of the group of homeomorphisms of a manifold.*, Dokl. Akad. Nauk SSSR **182** (1968), 510–513. MR 38 #5241.

[Čer69] ———, *Local contractibility of the group of homeomorphisms of a manifold.*, Mat. Sb. (N.S.) **79 (121)** (1969), 307–356. MR 41 #4554.

[Cer70] Jean Cerf, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. (1970), no. 39, 5–173. MR 45 #1176.

[CFHS96] C. L. Curtis, M. H. Freedman, W-C. Hsiang, and R. Stong, *A decomposition theorem for h–cobordant smooth simply-connected compact 4–manifolds*, Invent. Math. **123** (1996), no. 2, 343–348. MR 97e:57020.

[CH81] Andrew J. Casson and John L. Harer, *Some homology lens spaces which bound rational homology balls*, Pacific J. Math. **96** (1981), no. 1, 23–36. MR 83h:57013.

[Cha74] T. A. Chapman, *Topological invariance of Whitehead torsion*, Amer. J. Math. **96** (1974), 488–497. MR 52 #11931.

[Cha93] Isaac Chavel, *Riemannian geometry—a modern introduction*, Cambridge Tracts in Mathematics, vol. 108, Cambridge University Press, Cambridge, 1993. MR 95j:53001.

[Coh73] Marshall M. Cohen, *A course in simple-homotopy theory*, Graduate Texts in Mathematics, vol. 10, Springer, New York, 1973. MR 50 #14762.

[dC92] Manfredo Perdigão do Carmo, *Riemannian geometry*, Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1992, Transl from second Portuguese ed. MR 92i:53001.

[Deh10] Max Dehn, *Über die Topologie des dreidimensionalen Raume*, Math. Ann. **69** (1910), 137–168.

[Dim92] Alexandru Dimca, *Singularities and topology of hypersurfaces*, Universitext, Springer-Verlag, New York, 1992. MR 94b:32058.

[DK90] Simon K. Donaldson and Peter B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 1990. MR 92a:57036.

[DK01] James F. Davis and Paul Kirk, *Lecture notes in algebraic topology*, Graduate Studies in Mathematics, vol. 35, American Mathematical Society, Providence, RI, 2001. MR 2002f:55001.

[DMF92] Stefano De Michelis and Michael H. Freedman, *Uncountably many exotic* $\mathbb{R}^4$*'s in standard* 4*-space*, J. Differential Geom. **35** (1992), no. 1, 219–254. MR 93d:57036.

[Dol80] Albrecht Dold, *Lectures on algebraic topology*, second ed., Grundlehren der Mathematischen Wissenschaften, vol. 200, Springer-Verlag, Berlin, 1980. MR 82c:55001.

[Dol95] ______, *Lectures on algebraic topology*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Reprint of the 1972 edition. MR 96c:55001.

[Don83] Simon K. Donaldson, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. **18** (1983), no. 2, 279–315. MR 85c:57015.

[Don86] ______, *Connections, cohomology and the intersection forms of* 4*-manifolds*, J. Differential Geom. **24** (1986), no. 3, 275–341. MR 88g:57033.

[Don87] ______, *Irrationality and the* h*-cobordism conjecture*, J. Differential Geom. **26** (1987), no. 1, 141–168. MR 88j:57035.

[Don90] ______, *Polynomial invariants for smooth four-manifolds*, Topology **29** (1990), no. 3, 257–315. MR 92a:57035.

[Don96a] ______, *The Seiberg–Witten equations and* 4*-manifold topology*, Bull. Amer. Math. Soc. (N.S.) **33** (1996), no. 1, 45–70. MR 96k:57033.

[Don96b] ______, *Symplectic submanifolds and almost-complex geometry*, J. Differential Geom. **44** (1996), no. 4, 666–705. MR 98h:53045.

[Don98] ______, *Lefschetz fibrations in symplectic geometry*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), Doc. Math., 1998, Extra Vol. II, pp. 309–314 (electronic). MR 99i:57044.

[Don99] ______, *Lefschetz pencils on symplectic manifolds*, J. Differential Geom. **53** (1999), no. 2, 205–236. MR 2002g:53154.

[DW59] Albrecht Dold and Hassler Whitney, *Classification of oriented sphere bundles over a* 4*-complex*, Ann. of Math. (2) **69** (1959), 667–677. MR 23 #A659.

[Ebe90] Wolfgang Ebeling, *An example of two homeomorphic, nondiffeomorphic complete intersection surfaces*, Invent. Math. **99** (1990), no. 3, 651–654. MR 91c:57038.

[Ebi68] David G. Ebin, *On the space of Riemannian metrics*, Bull. Amer. Math. Soc. **74** (1968), 1001–1003. MR 37 #6965.

[EK62] James Eells, Jr. and Nicolaas Kuiper, *An invariant for certain smooth manifolds*, Ann. Mat. Pura Appl. (4) **60** (1962), 93–110. MR 27 #6280.

[Elk95] Noam D. Elkies, *A characterization of the* $\mathbb{Z}^n$ *lattice*, Math. Res. Lett. **2** (1995), no. 3, 321–326. MR 96h:11064. `arXiv:math.NT/9906019`.

[Enr49] Federigo Enriques, *Le superficie algebriche*, Nicola Zanichelli, Bologna, 1949. MR 11,202b.

[FK72] G. B. Folland and J. J. Kohn, *The Neumann problem for the Cauchy–Riemann complex*, Annals of Mathematics Studies, vol. 75, Princeton University Press, Princeton, NJ, 1972. MR 57 #1573.

[FK78] Michael H. Freedman and Robion Kirby, *A geometric proof of Rochlin's theorem*, Algebraic and geometric topology (Stanford, CA, 1976), Part 2, Proc. Sympos. Pure Math., vol. 32, American Mathematical Society, Providence, RI, 1978, pp. 85–97. MR 80f:57015.

[FKM01] Mikio Furuta, Y. Kametani, and H. Matsue, *Spin* 4*-manifolds with signature* $= -32$, Math. Res. Lett. **8** (2001), no. 3, 293–301. MR 2003e:57043.

[FL89] Michael H. Freedman and Feng Luo, *Selected applications of geometry to low-dimensional topology*, University Lecture Series, vol. 1, American Mathematical Society, Providence, RI, 1989. MR 91k:57002.

[FL98a] Paul M. N. Feehan and Thomas G. Leness, *$PU(2)$ monopoles and relations between four-manifold invariants*, Topology Appl. **88** (1998), no. 1-2, 111–145, Symplectic, contact and low-dimensional topology (Athens, GA, 1996). MR 2000d:57049. `arXiv:dg-ga/9709022`.

[FL98b] ———, *$PU(2)$ monopoles. I. Regularity, Uhlenbeck compactness, and transversality*, J. Differential Geom. **49** (1998), no. 2, 265–410. MR 2000e:57052. `arXiv:dg-ga/9710032`.

[FL01a] ———, *$PU(2)$ monopoles and links of top-level Seiberg–Witten moduli spaces*, J. Reine Angew. Math. **538** (2001), 57–133. MR 2002f:57067. `arXiv:math.DG/0007190`.

[FL01b] ———, *$PU(2)$ monopoles. II. Top-level Seiberg–Witten moduli spaces and Witten's conjecture in low degrees*, J. Reine Angew. Math. **538** (2001), 135–212. MR 2002f:57068. `arXiv:dg-ga/9712005`.

[FL02a] ———, *A general $SO(3)$-monopole cobordism formula relating Donaldson and Seiberg–Witten invariants*, 2002. `arXiv:math.DG/0203047`.

[FL02b] ———, *$SO(3)$ monopoles, level-one Seiberg–Witten moduli spaces, and Witten's conjecture in low degrees*, Proceedings of the 1999 Georgia Topology Conference (Athens, GA), vol. 124, Topology Appl., no. 2, 2002, pp. 221–326. MR 2003i:57049.

[FL03] ———, *On Donaldson and Seiberg–Witten invariants*, Topology and geometry of manifolds (Athens, GA, 2001), Proc. Sympos. Pure Math., vol. 71, American Mathematical Society, Providence, RI, 2003, pp. 237–248. MR 2 024 636. `arXiv:math.DG/0106221`.

[Flo88] Andreas Floer, *An instanton-invariant for 3-manifolds*, Comm. Math. Phys. **118** (1988), no. 2, 215–240. MR MR956166 (89k:57028).

[FM66] Ralph H. Fox and John W. Milnor, *Singularities of 2-spheres in 4-space and cobordism of knots*, Osaka J. Math. **3** (1966), 257–267. MR 35 #2273.

[FM88] Robert Friedman and John W. Morgan, *On the diffeomorphism types of certain algebraic surfaces. I & II*, J. Differential Geom. **27** (1988), no. 3, 297–369 & 371–398. MR 89d:57046 & 89d:57047.

[FM89] ———, *Complex versus differentiable classification of algebraic surfaces*, Proceedings of the 1987 Georgia Topology Conference (Athens, GA, 1987), vol. 32, Topology Appl., no. 2, 1989, pp. 135–139. MR 90f:57039.

[FM94a] ———, *Smooth four-manifolds and complex surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 27, Springer, Berlin, 1994. MR 95m:57046.

[FM94b] ———, *Smooth four-manifolds and complex surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 27, Springer-Verlag, Berlin, 1994. MR MR1288304 (95m:57046).

[FM97] ———, *Algebraic surfaces and Seiberg–Witten invariants*, J. Algebraic Geom. **6** (1997), no. 3, 445–479. MR 99b:32045. `arXiv:alg-geom/9502026`.

[FM99] ———, *Obstruction bundles, semiregularity, and Seiberg–Witten invariants*, Comm. Anal. Geom. **7** (1999), no. 3, 451–495. MR 2000e:14074. `arXiv:alg-geom/9509007`.

[FQ90] Michael H. Freedman and Frank Quinn, *Topology of 4-manifolds*, Princeton Mathematical Series, vol. 39, Princeton University Press, Princeton, NJ, 1990. MR 94b:57021.

[FQ95] Robert Friedman and Zhenbo Qin, *On complex surfaces diffeomorphic to rational surfaces*, Invent. Math. **120** (1995), no. 1, 81–117. MR MR1323983 (96d:57032).

[FR79] Roger Fenn and Colin Rourke, *On Kirby's calculus of links*, Topology **18** (1979), no. 1, 1–15. MR 80c:57005.

[Fre79] Michael H. Freedman, *A fake $S^3 \times \mathbb{R}$*, Ann. of Math. (2) **110** (1979), no. 1, 177–201. MR MR541336 (80j:57029).

[Fre82] ———, *The topology of four-dimensional manifolds*, J. Differential Geom. **17** (1982), no. 3, 357–453. MR 84b:57006.

[Fre84] ———, *The disk theorem for four-dimensional manifolds*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983) (Warsaw), PWN, 1984, pp. 647–663. MR MR804721 (86m:57016).

[Fri95] Robert Friedman, *Vector bundles and $SO(3)$-invariants for elliptic surfaces*, J. American Mathematical Society **8** (1995), no. 1, 29–139. MR 95k:14052.

[Fri97] ———, *Donaldson and Seiberg–Witten invariants of algebraic surfaces*, Algebraic geometry (Santa Cruz, CA, 1995), Proc. Sympos. Pure Math., vol. 62, American Mathematical Society, Providence, RI, 1997, pp. 85–100. MR 99b:14032. `arXiv:alg-geom/9605006`.

[Fri00] Thomas Friedrich, *Dirac operators in Riemannian geometry*, American Mathematical Society, Providence, RI, 2000, Translated from the 1997 German original. MR 1 777 332.

[FS95] Ronald Fintushel and Ronald J. Stern, *Immersed spheres in 4-manifolds and the immersed Thom conjecture*, Turkish J. Math. **19** (1995), no. 2, 145–157. MR 96j:57036.

[FS97a] ———, *Rational blowdowns of smooth 4-manifolds*, J. Differential Geom. **46** (1997), no. 2, 181–235. MR 98j:57047. `arXiv:alg-geom/9505018`.

[FS97b] ———, *Surfaces in 4-manifolds*, Math. Res. Lett. **4** (1997), no. 6, 907–914. MR 98k:57047. `arXiv:dg-ga/9705009`.

[FS98a] ———, *Constructions of smooth 4-manifolds*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), Doc. Math., 1998, Extra Vol. II, pp. 443–452 (electronic). MR 99g:57033. `arXiv:math.GT/9907178`.

[FS98b] ———, *Knots; links, and 4-manifolds*, Invent. Math. **134** (1998), no. 2, 363–400. MR 99j:57033. `arXiv:dg-ga/9612014`.

[FS99] ———, *Nondiffeomorphic symplectic 4-manifolds with the same Seiberg–Witten invariants*, Proceedings of the Kirbyfest (Berkeley, CA, 1998) (Coventry), Geom. Topol. Monogr., vol. 2, Geom. Topol. Publ., 1999, pp. 103–111 (electronic). MR 2000j:57065. `arXiv:math.SG/9811019`.

[FS00] ———, *Nonsymplectic 4-manifolds with one basic class*, Pacific J. Math. **194** (2000), no. 2, 325–333. MR 2001g:57060. `arXiv:math.GT/9907179`.

[FS02] ———, *Families of simply-connected 4-manifolds with the same Seiberg–Witten invariants*, 2002. `arXiv:math.GT/0210206`.

[FS04] ———, *Double node neighborhoods and families of simply connected 4-manifolds with $b^+ = 1$*, 2004. `arXiv:math.GT/0412126`.

[FT86] Michael H. Freedman and Laurence R. Taylor, *A universal smoothing of four-space*, J. Differential Geom. **24** (1986), no. 1, 69–78. MR 88a:57044.

[FU84] Daniel S. Freed and Karen K. Uhlenbeck, *Instantons and four-manifolds*, Mathematical Sciences Research Institute Publications, vol. 1, Springer, New York, 1984. MR 86c:57031.

[Fur01] Mikio Furuta, *Monopole equation and the $^{11}/_{8}$-conjecture*, Math. Res. Lett. **8** (2001), no. 3, 279–291. MR 2003e:57042.

[Fur02] ———, *Finite dimensional approximations in geometry*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002) (Beijing), Higher Ed. Press, 2002, pp. 395–403. MR MR1957050 (2004a:57038).

[Gab83] David Gabai, *Foliations and the topology of 3-manifolds*, J. Differential Geom. **18** (1983), no. 3, 445–503. MR 86a:57009.

[GH78] Phillip Griffiths and Joseph Harris, *Principles of algebraic geometry*, Pure and Applied Mathematics, Wiley-Interscience [John Wiley & Sons], New York, 1978. MR 80b:14001.

[GH94] ———, *Principles of algebraic geometry*, Wiley Classics Library, John Wiley & Sons, New York, 1994, Reprint of the 1978 original. MR 95d:14001.

[GL89a] C. McA. Gordon and J. Luecke, *Knots are determined by their complements*, Bull. Amer. Math. Soc. (N.S.) **20** (1989), no. 1, 83–87. MR 90a:57006b.

[GL89b] ———, *Knots are determined by their complements*, J. Amer. Math. Soc. **2** (1989), no. 2, 371–415. MR 90a:57006a.

[Glu62] Herman Gluck, *The embedding of two-spheres in the four-sphere*, Trans. Amer. Math. Soc. **104** (1962), 308–333. MR MR0146807 (26 #4327).

[GM75] S. Gallot and D. Meyer, *Opérateur de courbure et laplacien des formes différentielles d'une variété riemannienne*, J. Math. Pures Appl. (9) **54** (1975), no. 3, 259–284. MR 56 #13128.

[GM86a] Lucien Guillou and Alexis Marin (eds.), *À la recherche de la topologie perdue. I. Du côté de chez Rohlin; II. Le côté de Casson*, Progress in Mathematics, vol. 62, Birkhäuser Boston, Boston, MA, 1986. MR 88k:57002.

[GM86b] ______, *Commentaires sur les quatre articles précédents de V. A. Rohlin*, À la recherche de la topologie perdue [**GM86a**], Progr. Math., vol. 62, Birkhäuser Boston, Boston, MA, 1986, pp. 25–95. MR 900 246.

[GM86c] ______, *Une extension d'un théorème de Rohlin sur la signature*, À la recherche de la topologie perdue [**GM86a**], Progr. Math., vol. 62, Birkhäuser Boston, Boston, MA, 1986, pp. 97–118. MR 900 247.

[GM93] Robert E. Gompf and Tomasz S. Mrowka, *Irreducible 4-manifolds need not be complex*, Ann. of Math. (2) **138** (1993), no. 1, 61–111. MR 95c:57053.

[God58] Roger Godement, *Topologie algébrique et théorie des faisceaux*, Actualités Sci. Ind. No. 1252. Publ. Math. University Strasbourg, vol. 13, Hermann, Paris, 1958. MR 21 #1583.

[God73] ______, *Topologie algébrique et théorie des faisceaux*, Hermann, Paris, 1973, Troisième édition revue et corrigée. MR 49 #9831.

[Gom83] Robert E. Gompf, *Three exotic* $\mathbb{R}^4$*'s and other anomalies*, J. Differential Geom. **18** (1983), no. 2, 317–328. MR 85b:57038.

[Gom85] ______, *An infinite set of exotic* $\mathbb{R}^4$*'s*, J. Differential Geom. **21** (1985), no. 2, 283–300. MR 87h:57030.

[Gom88] ______, *On sums of algebraic surfaces*, Invent. Math. **94** (1988), no. 1, 171–174. MR 89j:57028.

[Gom91a] ______, *Nuclei of elliptic surfaces*, Topology **30** (1991), no. 3, 479–511. MR 92f:57042.

[Gom91b] ______, *Sums of elliptic surfaces*, J. Differential Geom. **34** (1991), no. 1, 93–114. MR 92f:57044.

[Gom93] ______, *An exotic menagerie*, J. Differential Geom. **37** (1993), no. 1, 199–223. MR 93k:57041.

[Gom97] ______, *Spin*$^{\mathbb{C}}$*-structures and homotopy equivalences*, Geom. Topol. **1** (1997), 41–50 (electronic). MR 98i:57048. `arXiv:math.GT/9705218`.

[GP74] Victor Guillemin and Alan Pollack, *Differential topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974. MR 50 #1276.

[Gro85] M. Gromov, *Pseudoholomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), no. 2, 307–347. MR 87j:53053.

[GS99] Robert E. Gompf and András I. Stipsicz, *4-Manifolds and Kirby calculus*, American Mathematical Society, Providence, RI, 1999. MR 2000h:57038.

[Hat02] Allen Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002. MR 2002k:55001. `http://www.math.cornell.edu/~hatcher/AT/ATpage.html`.

[Hem76] John Hempel, *3-Manifolds*, Ann. of Math. Studies, vol. 86, Princeton University Press, Princeton, NJ, 1976. MR 54 #3702.

[Her81] Ralph J. Herbert, *Multiple points of immersed manifolds*, Mem. American Mathematical Society **34** (1981), no. 250, xiv+60. MR 84c:57022.

[HH58] Friedrich Hirzebruch and Heinz Hopf, *Felder von Flächenelementen in 4-dimensionalen Mannigfaltigkeiten*, Math. Ann. **136** (1958), 156–172. MR 20 #7272.

[Hir53] Friedrich Hirzebruch, *On Steenrod's reduced powers, the index of inertia, and the Todd genus*, Proc. Nat. Acad. Sci. U. S. A. **39** (1953), 951–956. MR 16,159f.

[Hir56] ______, *Neue topologische Methoden in der algebraischen Geometrie*, Ergebnisse der Mathematik und ihrer Grenzgebiete (N.F.), vol. 9, Springer, Berlin, 1956. MR 18,509b.

[Hir63] Morris W. Hirsch, *Obstruction theories for smoothing manifolds and maps*, Bull. Amer. Math. Soc. **69** (1963), 352–356. MR 26 #6980.

[Hir66] Friedrich Hirzebruch, *Topological methods in algebraic geometry*, Die Grundlehren der Mathematischen Wissenschaften, vol. 131, Springer, New York, 1966. MR 34 #2573.

[Hir76] Morris W. Hirsch, *Differential topology*, Graduate Texts in Mathematics, vol. 33, Springer, New York, 1976. MR 56 #6669.

[Hir94] ______, *Differential topology*, Springer, New York, 1994, Corrected reprint of the 1976 original. MR 96c:57001.

[Hir95] Friedrich Hirzebruch, *Topological methods in algebraic geometry*, Classics in Mathematics, Springer, Berlin, 1995, Reprint of the 1978 edition. MR 96c:57002.

[Hit74] Nigel J. Hitchin, *Harmonic spinors*, Advances in Math. **14** (1974), 1–55. MR 50 #11332.

[HKK86] John Harer, Arnold Kas, and Robion Kirby, *Handlebody decompositions of complex surfaces*, Mem. American Mathematical Society **62** (1986), no. 350, iv+102. MR 88e:57030.

[HL99a] Michael Hutchings and Yi-Jen Lee, *Circle-valued Morse theory and Reidemeister torsion*, Geom. Topol. **3** (1999), 369–396 (electronic). MR 2000h:57063. `arXiv:dg-ga/9706012`.

[HL99b] ______, *Circle-valued Morse theory, Reidemeister torsion, and Seiberg–Witten invariants of* 3*-manifolds*, Topology **38** (1999), no. 4, 861–888. MR 2000i:57055. `arXiv:dg-ga/9612004`.

[HM68] Friedrich Hirzebruch and K. H. Mayer, *$O(n)$-Mannigfaltigkeiten, exotische Sphären und Singularitäten*, Lecture Notes in Mathematics, No. 57, Springer-Verlag, Berlin, 1968. MR 37 #4825.

[HM74] Morris W. Hirsch and Barry Mazur, *Smoothings of piecewise linear manifolds*, Annals of Mathematics Studies, vol. 80, Princeton University Press, Princeton, NJ, 1974. MR 54 #3711.

[HNK71] Friedrich Hirzebruch, W. D. Neumann, and S. S. Koh, *Differentiable manifolds and quadratic forms*, vol. 4, Marcel Dekker Inc., New York, 1971. MR 49 #6250.

[Hör66] Lars Hörmander, *An introduction to complex analysis in several variables*, D. Van Nostrand Co., Princeton, NJ–Toronto, Ont.–London, 1966. MR 34 #2933.

[Hör90] ______, *An introduction to complex analysis in several variables*, third ed., North-Holland Mathematical Library, vol. 7, North-Holland Publishing Co., Amsterdam, 1990. MR 91a:32001.

[HS69] W-C. Hsiang and J. L. Shaneson, *Fake tori, the annulus conjecture, and the conjectures of Kirby*, Proc. Nat. Acad. Sci. U.S.A. **62** (1969), 687–691. MR 42 #5267.

[HT99] Michael Hutchings and Clifford H. Taubes, *An introduction to the Seiberg–Witten equations on symplectic manifolds*, Symplectic geometry and topology (Park City, UT, 1997), IAS/Park City Math. Ser., vol. 7, American Mathematical Society, Providence, RI, 1999, pp. 103–142. MR 2000j:57066.

[Hum97] Christoph Hummel, *Gromov's compactness theorem for pseudo-holomorphic curves*, Progress in Mathematics, vol. 151, Birkhäuser, Basel, 1997. MR 98k:58032.

[Hus66] Dale Husemoller, *Fibre bundles*, McGraw-Hill Book Co., New York, 1966. MR 37 #4821.

[Hus94] ______, *Fibre bundles*, third ed., Springer, New York, 1994. MR 94k:55001.

[IP97] Eleny-Nicoleta Ionel and Thomas H. Parker, *The Gromov invariants of Ruan–Tian and Taubes*, Math. Res. Lett. **4** (1997), no. 4, 521–532. MR 98i:58038. `arXiv:alg-geom/9702008`.

[IS99] S. Ivashkovich and V. Shevchishin, *Gromov compactness theorem for stable curves*, 1999. `arXiv:math.DG/9903047`.

[Jos95] Jürgen Jost, *Riemannian geometry and geometric analysis*, Universitext, Springer, Berlin, 1995. MR 96g:53049.

[Jos02] ______, *Riemannian geometry and geometric analysis*, third ed., Universitext, Springer, Berlin, 2002. MR 2002i:53001.

[Kap01] Michael Kapovich, *Hyperbolic manifolds and discrete groups*, Progress in Mathematics, vol. 183, Birkhäuser Boston Inc., Boston, MA, 2001. MR 2002m:57018.

[Kel55] John L. Kelley, *General topology*, D. Van Nostrand Company, Toronto–New York–London, 1955. MR 16,1136c.

[Kel75] ______, *General topology*, Graduate Texts in Mathematics, vol. 27, Springer, New York, 1975, Reprint of the 1955 edition. MR 51 #6681.

[Ker60] Michel A. Kervaire, *A manifold which does not admit any differentiable structure*, Comment. Math. Helv. **34** (1960), 257–270. MR 25 #2608.

[Kir69] Robion Kirby, *Stable homeomorphisms and the annulus conjecture*, Ann. of Math. (2) **89** (1969), 575–582, also reprinted in *Foundational essays...* [**KS77**]. MR 39 #3499.

[Kir78] ———, *A calculus for framed links in* S^3, Invent. Math. **45** (1978), no. 1, 35–56. MR 57 #7605.

[Kir89] ———, *The topology of* 4*-manifolds*, Springer, Berlin, 1989. MR 90j:57012.

[Kir96] ———, *Akbulut's corks and* h*-cobordisms of smooth, simply connected* 4*-manifolds*, Turkish J. Math. **20** (1996), no. 1, 85–93. MR 97j:57055. `arXiv:math.GT/9712231`.

[Kis64] J. M. Kister, *Microbundles are fibre bundles*, Ann. of Math. (2) **80** (1964), 190–199. MR 31 #5216.

[KM60] Michel A. Kervaire and John Milnor, *Bernoulli numbers, homotopy groups, and a theorem of Rohlin*, Proc. Internat. Congress Math. 1958, Cambridge University Press, New York, 1960, pp. 454–458. MR 22 #12531.

[KM61] ———, *On* 2*-spheres in* 4*-manifolds*, Proc. Nat. Acad. Sci. U.S.A. **47** (1961), 1651–1657. MR 24 #A2968.

[KM63] ———, *Groups of homotopy spheres. I*, Ann. of Math. (2) **77** (1963), 504–537. MR 26 #5584.

[KM91] Robion Kirby and Paul Melvin, *The* 3*-manifold invariants of Witten and Reshetikhin–Turaev for* $\mathfrak{sl}(2,\mathbb{C})$, Invent. Math. **105** (1991), no. 3, 473–545. MR 92e:57011.

[KM94a] Peter B. Kronheimer and Tomasz S. Mrowka, *The genus of embedded surfaces in the projective plane*, Math. Res. Lett. **1** (1994), no. 6, 797–808. MR 96a:57073.

[KM94b] ———, *Recurrence relations and asymptotics for four-manifold invariants*, Bull. Amer. Math. Soc. (N.S.) **30** (1994), no. 2, 215–221. MR 94k:57046.

[KM95] ———, *Embedded surfaces and the structure of Donaldson's polynomial invariants*, J. Differential Geom. **41** (1995), no. 3, 573–734. MR 96e:57019.

[KM99] Robion Kirby and Paul Melvin, *The* E_8*-manifold, singular fibers and handlebody decompositions*, Proceedings of the Kirbyfest (Berkeley, CA, 1998) (Coventry), Geom. Topol. Monogr., vol. 2, Geom. Topol. Publ., 1999, pp. 233–258 (electronic). MR 2000j:57051. `arXiv:math.GT/9911253`.

[KM04] Peter B. Kronheimer and Tomasz S. Mrowka, *Witten's conjecture and property* P, Geom. Topol. **8** (2004), 295–310 (electronic). MR 2 023 280. `arXiv:math.GT/0311489`.

[KN69] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry. Vol I & II*, Interscience Publishers John Wiley & Sons, New York–London–Sydney, 1963 & 1969. MR 27 #2945 & 38 #6501.

[KN96] ———, *Foundations of differential geometry. Vol. I & II*, Wiley Classics Library, John Wiley & Sons, New York, 1996, Reprint of the 1969 original. MR 97c:53001.

[Kod63] Kunihiko Kodaira, *On compact complex analytic surfaces. I–III*, Ann. of Math. (2) **71–78** (1960–1963).

[Kod68] ———, *On the structure of compact complex analytic surfaces. I–IV*, Amer. J. Math. **86–90** (1964–1968).

[Kos93] Antoni A. Kosinski, *Differential manifolds*, Pure and Applied Mathematics, vol. 138, Academic Press Inc., Boston, MA, 1993. MR 95b:57001.

[Kot89] Dieter Kotschick, *On manifolds homeomorphic to* $\mathbb{CP}^2 \# 8\,\overline{\mathbb{CP}}^2$, Invent. Math. **95** (1989), no. 3, 591–600. MR 90a:57047.

[Kro98] Peter B. Kronheimer, *Embedded surfaces and gauge theory in three and four dimensions*, Surveys in differential geometry, Vol. III (Cambridge, MA, 1996), International Press, Boston, MA, 1998, pp. 243–298. MR 2000a:57086.

[Kro99] ———, *Minimal genus in* $S^1 \times M^3$, Invent. Math. **135** (1999), no. 1, 45–61. MR 2000c:57071.

[KS69] Robion Kirby and Lawrence C. Siebenmann, *On the triangulation of manifolds and the Hauptvermutung*, Bull. Amer. Math. Soc. **75** (1969), 742–749, also reprinted in *Foundational essays...* **[KS77]**. MR 39 #3500.

[KS77] Robion Kirby and Laurence C. Siebenmann, *Foundational essays on topological manifolds, smoothings, and triangulations*, Annals of Mathematics Studies, vol. 88, Princeton University Press, Princeton, NJ, 1977. MR 58 #31082.

[KS79] Robion Kirby and Martin G. Scharlemann, *Eight faces of the Poincaré homology* 3*–sphere*, Geometric topology (Proc. Georgia Topology Conf., Athens, GA, 1977), Academic Press, New York, 1979, pp. 113–146. MR MR537730 (80k:57042).

[Kur62] Masatake Kuranishi, *On the locally complete families of complex analytic structures*, Ann. of Math. (2) **75** (1962), 536–577. MR MR0141139 (25 #4550).

[KW74] Jerry L. Kazdan and F. W. Warner, *Curvature functions for compact* 2*–manifolds*, Ann. of Math. (2) **99** (1974), 14–47. MR 49 #7949.

[Lam81] Klaus Lamotke, *The topology of complex projective varieties after S. Lefschetz*, Topology **20** (1981), no. 1, 15–51. MR MR592569 (81m:14019).

[Lau73] François Laudenbach, *Sur les* 2*–sphères d'une variété de dimension* 3, Ann. of Math. (2) **97** (1973), 57–81. MR MR0314054 (47 #2606).

[Law85] H. Blaine Lawson, Jr., *The theory of gauge fields in four dimensions*, CBMS Regional Conference Series in Mathematics, vol. 58, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1985. MR 87d:58044.

[Law92] Terry Lawson, *Smooth embeddings of* 2*–spheres in* 4*–manifolds*, Exposition. Math. **10** (1992), no. 4, 289–309. MR 93m:57041.

[Law97] ———, *The minimal genus problem*, Exposition. Math. **15** (1997), no. 5, 385–431. MR 98g:57055.

[LeB01] Claude LeBrun, *Ricci curvature, minimal volumes, and Seiberg–Witten theory*, Invent. Math. **145** (2001), no. 2, 279–316. MR 2002h:53061. `arXiv:math.DG/0003068`.

[LeB03] ———, *Einstein metrics, four-manifolds, and differential topology*, Surveys in differential geometry, Vol. VIII (Boston, MA, 2002), Surv. Differ. Geom., VIII, Int. Press, Somerville, MA, 2003, pp. 235–255. MR 2 039 991. `arXiv:math.DG/0404359`.

[Lee04] Yi-Jen Lee, *Heegaard Splittings and Seiberg–Witten monopoles*, 2004. `arXiv:math.GT/0409536`.

[Lef24] Solomon Lefschetz, *L'analysis situs and et la géometrie algébrique*, Gauthier–Villars, Paris, 1924.

[Les96] Christine Lescop, *Global surgery formula for the Casson–Walker invariant*, Annals of Mathematics Studies, vol. 140, Princeton University Press, Princeton, NJ, 1996. MR MR1372947 (97c:57017).

[Lic62a] André Lichnerowicz, *Laplacien sur une variété riemannienne et spineurs*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. (8) **33** (1962), 187–191. MR 27 #5203.

[Lic62b] W. B. Raymond Lickorish, *A representation of orientable combinatorial* 3*–manifolds*, Ann. of Math. (2) **76** (1962), 531–540. MR 27 #1929.

[Lic97] ———, *An introduction to knot theory*, Graduate Texts in Mathematics, vol. 175, Springer-Verlag, New York, 1997. MR 98f:57015.

[Lis94] Paolo Lisca, *On the Donaldson polynomials of elliptic surfaces*, Math. Ann. **299** (1994), no. 4, 629–639. MR 95j:57036.

[LM89] H. Blaine Lawson, Jr. and Marie-Louise Michelsohn, *Spin geometry*, Princeton University Press, Princeton, NJ, 1989. MR 91g:53001.

[LP72] François Laudenbach and Valentin Poénaru, *A note on* 4*–dimensional handlebodies*, Bull. Soc. Math. France **100** (1972), 337–344. MR MR0317343 (47 #5890).

[LR65] Richard Lashof and M. Rothenberg, *Microbundles and smoothing*, Topology **3** (1965), 357–388. MR 31 #752.

[LT84] Richard Lashof and Laurence Taylor, *Smoothing theory and Freedman's work on four-manifolds,* Algebraic topology (Aarhus, 1982), Lecture Notes in Math., vol. 1051, Springer, Berlin, 1984, pp. 271–292. MR 86b:57009.

[Man79] Richard Mandelbaum, *Decomposing analytic surfaces,* Geometric topology (Proc. Georgia Topology Conf., Athens, GA, 1977), Academic Press, New York, 1979, pp. 147–217. MR 80h:57013.

[Man80] ______, *Four-dimensional topology: an introduction,* Bull. Amer. Math. Soc. (N.S.) **2** (1980), no. 1, 1–159. MR 81j:57001.

[Man85] ______, *Lefschetz fibrations of Riemann surfaces and decompositions of complex elliptic surfaces,* Combinatorial methods in topology and algebraic geometry (Rochester, N.Y., 1982), Contemp. Math., vol. 44, Amer. Math. Soc., Providence, RI, 1985, pp. 291–310. MR 87f:32050.

[Mar02] Thomas Mark, *Torsion, TQFT, and Seiberg–Witten invariants of 3–manifolds,* Geom. Topol. **6** (2002), 27–58 (electronic). MR 2003c:57032. `arXiv:math.DG/9912147`.

[Mat82] Yukio Matsumoto, *On the bounding genus of homology 3-spheres,* J. Fac. Sci. Univ. Tokyo Sect. IA Math. **29** (1982), no. 2, 287–318. MR 84g:57010.

[Mat86] ______, *An elementary proof of Rochlin's signature theorem and its extension by Guillou and Marin,* À la recherche de la topologie perdue **[GM86a]**, Progr. Math., vol. 62, Birkhäuser Boston, Boston, MA, 1986, pp. 119–139. MR 900 248.

[Mat96] R. Matveyev, *A decomposition of smooth simply-connected h–cobordant 4–manifolds,* J. Differential Geom. **44** (1996), no. 3, 571–582. MR 98a:57033. `arXiv:dg-ga/9505001`.

[Mat02] Yukio Matsumoto, *An introduction to Morse theory,* Translations of Mathematical Monographs, vol. 208, American Mathematical Society, Providence, RI, 2002, Translated from the 1997 Japanese original, Iwanami Series in Modern Mathematics. MR 2002m:57043.

[Maz59] Barry Mazur, *On embeddings of spheres,* Bull. Amer. Math. Soc. **65** (1959), 59–65. MR MR0117693 (22 #8469).

[Maz61] ______, *On embeddings of spheres,* Acta Math. **105** (1961), 1–17. MR MR0125570 (23 #A2870).

[Maz63] ______, *Relative neighborhoods and the theorems of Smale,* Ann. of Math. (2) **77** (1963), 232–249. MR 27 #773.

[McD92] Dusa McDuff, *Singularities of J–holomorphic curves in almost complex 4–manifolds,* J. Geom. Anal. **2** (1992), no. 3, 249–266. MR MR1164604 (93g:58032).

[McD97] ______, *Lectures on Gromov invariants for symplectic 4–manifolds,* Gauge theory and symplectic geometry (Montreal, PQ, 1995), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 488, Kluwer Acad. Publ., Dordrecht, 1997. MR 98j:58018. `arXiv:dg-ga/9606005`.

[Mey84] A. Meyer, *Über die Auflösung der Gleichung* $ax^2 + by^2 + cz^2 + du^2 + ev^2 = 0$ *in ganzen Zahlen,* Vierteljahrs. Naturfors. Gesells. Zürich **29** (1884), 220–222.

[MH73] John Milnor and Dale Husemoller, *Symmetric bilinear forms,* Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 73, Springer, New York, 1973. MR 58 #22129.

[Mik97] G. Mikhalkin, *J–holomorphic curves in almost complex surfaces do not always minimize the genus,* Proc. American Mathematical Society **125** (1997), no. 6, 1831–1833. MR 97g:57047.

[Mil56a] John Milnor, *Construction of universal bundles. II,* Ann. of Math. (2) **63** (1956), 430–436. MR 17,1120a.

[Mil56b] ______, *On manifolds homeomorphic to the 7–sphere,* Ann. of Math. (2) **64** (1956), 399–405. MR 18,498d.

[Mil58a] ______, *Differentiable manifolds which are homotopy spheres,* mimeographed notes, 1958.

[Mil58b] ______, *On simply connected 4–manifolds,* Symposium internacional de topología algebraica, Universidad Nacional Autónoma de México and UNESCO, Mexico City, 1958, pp. 122–128. MR 21 #2240.

[Mil59] ______, *Differentiable structures on spheres,* Amer. J. Math. **81** (1959), 962–972. MR 22 #990.

[Mil61] ______, *Two complexes which are homeomorphic but combinatorially distinct,* Ann. of Math. (2) **74** (1961), 575–590. MR MR0133127 (24 #A2961).

[Mil63a] ______, *Morse theory*, Annals of Mathematics Studies, vol. 51, Princeton University Press, Princeton, NJ, 1963. MR 29 #634.

[Mil63b] ______, *Spin structures on manifolds*, Enseignement Math. (2) **9** (1963), 198–203. MR 28 #622.

[Mil63c] ______, *Topological manifolds and smooth manifolds*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), Inst. Mittag-Leffler, Djursholm, 1963, pp. 132–138. MR 28 #4553a.

[Mil64] ______, *Microbundles. I*, Topology **3** (1964), no. suppl. 1, 53–80. MR 28 #4553b.

[Mil65a] ______, *Lectures on the h-cobordism theorem*, Princeton University Press, Princeton, NJ, 1965. MR 32 #8352.

[Mil65b] ______, *Topology from the differentiable viewpoint*, The University Press of Virginia, Charlottesville, VA, 1965. MR 37 #2239.

[Mil66] ______, *Whitehead torsion*, Bull. Amer. Math. Soc. **72** (1966), 358–426. MR 33 #4922.

[Mil68] ______, *Singular points of complex hypersurfaces*, Annals of Mathematics Studies, vol. 61, Princeton University Press, Princeton, NJ, 1968. MR 39 #969.

[Mil71] ______, *Introduction to algebraic K-theory*, Annals of Mathematics Studies, vol. 72, Princeton University Press, Princeton, NJ, 1971. MR 50 #2304.

[Mil97] ______, *Topology from the differentiable viewpoint*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997, Revised reprint of the 1965 original. MR 98h:57051.

[MM76] Richard Mandelbaum and Boris Moishezon, *On the topological structure of non-singular algebraic surfaces in* $\mathbb{CP}^3$, Topology **15** (1976), no. 1, 23–40. MR 53 #9251.

[MM80] ______, *On the topology of simply connected algebraic surfaces*, Trans. Amer. Math. Soc. **260** (1980), no. 1, 195–222. MR 81g:14018.

[MM93] John W. Morgan and Tomasz S. Mrowka, *On the diffeomorphism classification of regular elliptic surfaces*, Internat. Math. Res. Notices (1993), no. 6, 183–184. MR 94g:57035.

[MMS97] John W. Morgan, Tomasz S. Mrowka, and Zoltán Szabó, *Product formulas along* $\mathbb{T}^3$ *for Seiberg-Witten invariants*, Math. Res. Lett. **4** (1997), no. 6, 915–929. MR 99f:57039.

[MO93] John W. Morgan and Kieran G. O'Grady, *Differential topology of complex surfaces*, Lecture Notes in Mathematics, vol. 1545, Springer-Verlag, Berlin, 1993, Elliptic surfaces with $p_g = 1$: smooth classification, With the collaboration of Millie Niss. MR MR1312610 (96e:14041).

[Moi54] Edwin E. Moise, *Affine structures in 3-manifolds. I–VIII*, Ann. of Math. (2) **54–59** (1951–1954).

[Moi77a] ______, *Geometric topology in dimensions* 2 *and* 3, Graduate Texts in Mathematics, vol. 47, Springer, New York, 1977. MR 58 #7631.

[Moi77b] Boris Moishezon, *Complex surfaces and connected sums of complex projective planes*, Lecture Notes in Mathematics, vol. 603, Springer, Berlin, 1977. MR 58 #10931.

[Mor31] Marston Morse, *The critical points of a function of n variables*, Trans. American Mathematical Society **33** (1931), no. 1, 72–91. MR 1 501 576.

[Mor96] John W. Morgan, *The Seiberg–Witten equations and applications to the topology of smooth four-manifolds*, Princeton University Press, Princeton, NJ, 1996. MR 97d:57042.

[MS74] John Milnor and James D. Stasheff, *Characteristic classes*, Annals of Mathematics Studies, vol. 76, Princeton University Press, Princeton, NJ, 1974. MR 55 #13428.

[MS94] Dusa McDuff and Dietmar Salamon, *J-holomorphic curves and quantum cohomology*, University Lecture Series, vol. 6, American Mathematical Society, Providence, RI, 1994. MR 95g:58026.

[MS95] ______, *Introduction to symplectic topology*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1995. MR 97b:58062.

[MS98] ______, *Introduction to symplectic topology*, second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1998. MR 2000g:53098.

[MS04] ———, *J-holomorphic curves and symplectic topology*, American Mathematical Society Colloquium Publications, vol. 52, American Mathematical Society, Providence, RI, 2004. MR MR2045629 (2004m:53154).

[MST96] John W. Morgan, Zoltán Szabó, and Clifford H. Taubes, *A product formula for the Seiberg–Witten invariants and the generalized Thom conjecture*, J. Differential Geom. **44** (1996), no. 4, 706–788. MR 97m:57052.

[MT96] Guowu Meng and Clifford H. Taubes, *$\underline{SW}$ = Milnor torsion*, Math. Res. Lett. **3** (1996), no. 5, 661–674. MR 98j:57049.

[Mum76] David Mumford, *Algebraic geometry. I. Complex projective varieties*, Die Grundlehren der Mathematischen Wissenschaften, vol. 221, Springer, Berlin, 1976. MR 56 #11992.

[Mum95] ———, *Algebraic geometry. I. Complex projective varieties*, Classics in Mathematics, Springer, Berlin, 1995, Reprint of the 1976 edition. MR 96d:14001.

[Mun59] James Munkres, *Obstructions to the smoothing of piecewise-differentiable homeomorphisms*, Bull. Amer. Math. Soc. **65** (1959), 332–334. MR 22 #3005.

[Mun60a] ———, *Differentiable isotopies on the* 2*-sphere*, Michigan Math. J. **7** (1960), 193–197. MR 26 #1899.

[Mun60b] ———, *Obstructions to the smoothing of piecewise-differentiable homeomorphisms*, Ann. of Math. (2) **72** (1960), 521–554. MR 22 #12534.

[Mun64] ———, *Obstructions to imposing differentiable structures*, Illinois J. Math. **8** (1964), 361–376. MR 31 #5209.

[Mun65] ———, *Higher obstructions to smoothing*, Topology **4** (1965), 27–45. MR 31 #758.

[Mun67] ———, *Concordance of differentiable structures—two approaches*, Michigan Math. J. **14** (1967), 183–191. MR 35 #3671.

[MW99] Vicente Muñoz and Bai-Ling Wang, *Seiberg–Witten–Floer homology of a surface times a circle*, 1999. `arXiv:math.DG/9905050`.

[Nic00] Liviu I. Nicolaescu, *Notes on Seiberg–Witten theory*, Graduate Studies in Mathematics, vol. 28, American Mathematical Society, Providence, RI, 2000. MR 2001k:57037.

[Nie73] Hans-Volker Niemeier, *Definite quadratische Formen der Dimension* 24 *und Diskriminate* 1, J. Number Theory **5** (1973), 142–178. MR 47 #4931.

[NN57] A. Newlander and L. Nirenberg, *Complex analytic coordinates in almost complex manifolds*, Ann. of Math. (2) **65** (1957), 391–404. MR 19,577a.

[OS00a] Peter Ozsváth and Zoltán Szabó, *Higher type adjunction inequalities in Seiberg–Witten theory*, J. Differential Geom. **55** (2000), no. 3, 385–440. MR 2002j:57061. `arXiv:math.DG/0005268`.

[OS00b] ———, *The symplectic Thom conjecture*, Ann. of Math. (2) **151** (2000), no. 1, 93–124. MR 2001a:57049. `arXiv:math.DG/9811087`.

[OS01a] ———, *Holomorphic disks and three-manifold invariants: properties and applications*, 2001, to appear in Ann. of Math. `arXiv:math.SG/0105202`.

[OS01b] ———, *Holomorphic disks and topological invariants for closed three-manifolds*, 2001, to appear in Ann. of Math. `arXiv:math.SG/0101206`.

[OS01c] ———, *Holomorphic triangles and invariants for smooth four-manifolds*, 2001. `arXiv:math.SG/0110169`.

[OS03] ———, *Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary*, Adv. Math. **173** (2003), no. 2, 179–261. MR 2003m:57066. `arXiv:math.SG/0110170`.

[OS04a] ———, *Heegaard diagrams and holomorphic disks*, 2004. `arXiv:math.GT/0403029`.

[OS04b] ———, *Holomorphic disks and genus bounds*, Geom. Topol. **8** (2004), 311–334 (electronic). MR 2 023 281. `arXiv:math.GT/0311496`.

[OS04c] ———, *Holomorphic triangle invariants and the topology of symplectic four-manifolds*, Duke Math. J. **121** (2004), no. 1, 1–34. MR 2004m:57059. `arXiv:math.SG/0201049`.

[OT95a] Christian Okonek and Andrei Teleman, *The coupled Seiberg–Witten equations, vortices, and moduli spaces of stable pairs*, Internat. J. Math. **6** (1995), no. 6, 893–910. MR MR1354000 (97a:32012). `arXiv:alg-geom/9505012`.

[OT95b] ———, *Les invariants de Seiberg–Witten et la conjecture de van de Ven*, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), no. 4, 457–461. MR MR1351097 (96m:57049). `arXiv:alg-geom/9505011`.

[OVdV86] C. Okonek and A. Van de Ven, *Stable bundles and differentiable structures on certain elliptic surfaces*, Invent. Math. **86** (1986), no. 2, 357–370. MR 88a:32027.

[Pap57a] C. D. Papakyriakopoulos, *On Dehn's lemma and the asphericity of knots*, Ann. of Math. (2) **66** (1957), 1–26. MR 19,761a.

[Pap57b] ———, *On solid tori*, Proc. London Math. Soc. (3) **7** (1957), 281–299. MR 19,441d.

[Par97] Jongil Park, *Seiberg–Witten invariants of generalised rational blow-downs*, Bull. Austral. Math. Soc. **56** (1997), no. 3, 363–384. MR 99b:57067.

[Par02a] B. Doug Park, *Constructing infinitely many smooth structures on* $3\,\mathbb{CP}^2 \# n\,\overline{\mathbb{CP}}^2$, Math. Ann. **322** (2002), no. 2, 267–278. MR 2003f:57061.

[Par02b] ———, *A gluing formula for the Seiberg–Witten invariant along* $\mathbb{T}^3$, Michigan Math. J. **50** (2002), no. 3, 593–611. MR 2003i:57051.

[Par02c] Jongil Park, *Exotic smooth structures on* 4*–manifolds*, Forum Math. **14** (2002), no. 6, 915–929. MR 2003i:57046. `arXiv:math.GT/0008161`.

[Par02d] ———, *The geography of Spin symplectic* 4*–manifolds*, Math. Z. **240** (2002), no. 2, 405–421. MR 2003c:57030.

[Par03a] ———, *Exotic smooth structures on* 4*–manifolds. II*, Topology Appl. **132** (2003), no. 2, 195–202. MR 2004d:57033.

[Par03b] ———, *Simply connected symplectic* 4*–manifolds with* $b_2^+ = 1$ *and* $c_1^2 = 2$, 2003. `arXiv:math.GT/0311395`.

[Per81] Ulf Persson, *Chern invariants of surfaces of general type*, Compositio Math. **43** (1981), no. 1, 3–58. MR 83b:14012.

[Per02] Grisha Perelman, *The entropy formula for the Ricci flow and its geometric applications*, 2002. `arXiv:math.DG/0211159`.

[Per03a] ———, *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, 2003. `arXiv:math.DG/0307245`.

[Per03b] ———, *Ricci flow with surgery on three-manifolds*, 2003. `arXiv:math.DG/0303109`.

[Pet98] Peter Petersen, *Riemannian geometry*, Graduate Texts in Mathematics, vol. 171, Springer-Verlag, New York, 1998. MR 98m:53001.

[Pet99] ———, *Aspects of global Riemannian geometry*, Bull. Amer. Math. Soc. (N.S.) **36** (1999), no. 3, 297–344. MR 2000e:53037.

[Pid94] Victor Ya. Pidstrigach, *Some glueing formulas for spin polynomials and a proof of the Van de Ven conjecture*, Izv. Ross. Akad. Nauk Ser. Mat. **58** (1994), no. 6, 105–122. MR MR1317211 (96m:57050).

[Pon49a] L. S. Pontryagin, *The homotopy group* $\pi_{n+1}(K^n)$ $(n \geq 2)$ *of dimension* $n+1$ *of a connected finite polyhedron* K^n *of arbitrary dimension, whose fundamental group and Betti groups of dimensions* $2,\ldots,n-1$ *are trivial*, Doklady Akad. Nauk SSSR (N.S.) **65** (1949), 797–800. MR 11,122b.

[Pon49b] ———, *On the classification of four-dimensional manifolds*, Uspehi Matem. Nauk (N.S.) **4** (1949), no. 4(32), 157–158. MR 11,194g.

[Pon50] ———, *Homotopy classification of the mappings of an* $(n+2)$*–dimensional sphere on an* n*–dimensional one*, Doklady Akad. Nauk SSSR (N.S.) **70** (1950), 957–959. MR 13,57b.

[Pon55] ———, *Gladkie mnogoobraziya i ih primeneniya v teorii gomotopiĭ*, Trudy Mat. Inst. im. Steklov. no. 45, Izdat. Akad. Nauk SSSR, Moscow, 1955. MR 17,181c.

[Pon59] ———, *Smooth manifolds and their applications in homotopy theory*, American Mathematical Society Translations, Ser. 2, Vol. 11, American Mathematical Society, Providence, RI, 1959, pp. 1–114. MR 22 #5980.

[PPX96] Ulf Persson, Chris Peters, and Gang Xiao, *Geography of spin surfaces*, Topology **35** (1996), no. 4, 845–862. MR 98h:14046.

[PS00] B. Doug Park and Zoltán Szabó, *The geography problem for irreducible spin four-manifolds*, Trans. Amer. Math. Soc. **352** (2000), no. 8, 3639–3650. MR 2000m:57037.

[PSS04] Jongil Park, András I. Stipsicz, and Zoltán Szabó, *Exotic smooth structures on* $\mathbb{CP}^2 \# 5 \overline{\mathbb{CP}}^2$, 2004. `arXiv:math.GT/0412216`.

[PT92] Victor Ya. Pidstrigach and Andrei N. Tyurin, *The smooth structure invariants of an algebraic surface defined by the Dirac operator*, Izv. Ross. Akad. Nauk Ser. Mat. **56** (1992), no. 2, 279–371. MR MR1180377 (93m:14036).

[PT95] ———, *Localisation of the Donaldson's invariants along Seiberg–Witten classes*, 1995. `arXiv: dg-ga/9507004`.

[Qui73] Daniel Quillen, *Higher algebraic K-theory. I*, Algebraic K-theory, I: Higher K-theories (Proc. Conf., Battelle Memorial Inst., Seattle, WA, 1972), Springer, Berlin, 1973, pp. 85–147. Lecture Notes in Math., Vol. 341. MR 49 #2895.

[Qui82] Frank Quinn, *Ends of maps. III. Dimensions* 4 *and* 5, J. Differential Geom. **17** (1982), no. 3, 503–521. MR 84j:57012.

[Qui84] ———, *Smooth structures on* 4*-manifolds*, Four-manifold theory (Durham, NH, 1982), Contemp. Math., vol. 35, American Mathematical Society, Providence, RI, 1984, pp. 473–479. MR 86f:57015.

[Qui86] ———, *Isotopy of* 4*-manifolds*, J. Differential Geom. **24** (1986), no. 3, 343–372. MR 88f:57020.

[Ran02] Andrew Ranicki, *Algebraic and geometric surgery*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, Oxford, 2002. MR 2 061 749.

[Rei97] Miles Reid, *Chapters on algebraic surfaces*, Complex algebraic geometry (Park City, UT, 1993), IAS/Park City Math. Ser., vol. 3, American Mathematical Society, Providence, RI, 1997, pp. 3–159. MR 98d:14049. `arXiv:alg-geom/9602006`.

[Rok52] V. A. Rokhlin, *New results in the theory of four-dimensional manifolds*, Doklady Akad. Nauk SSSR (N.S.) **84** (1952), 221–224. MR 14,573b.

[Rok72] ———, *Proof of a conjecture of Gudkov*, Funkcional. Anal. i Priložen. **6** (1972), no. 2, 62–64. MR 45 #5131.

[Rok86] ———, *Quatre articles de V. A. Rohlin*, À la recherche de la topologie perdue **[GM86a]**, Progr. Math., vol. 62, Birkhäuser Boston, Boston, MA, 1986, Translated from the Russian by S. Ochanine, pp. 3–23. MR 900 245.

[Rol75] Dale Rolfsen, *A surgical view of Alexander's polynomial*, Geometric topology (Proc. Conf., Park City, UT, 1974), Lecture Notes in Mathematics, vol. 438, Springer, Berlin, 1975, pp. 415–423. MR 52 #11892.

[Rol76] ———, *Knots and links*, Mathematics Lecture Series, vol. 7, Publish or Perish, Berkeley, CA, 1976. MR 58 #24236.

[Rol90] ———, *Knots and links*, Mathematics Lecture Series, vol. 7, Publish or Perish, Houston, TX, 1990, Corrected reprint of the 1976 original. MR 95c:57018.

[Rol03] ———, *Knots and links*, AMS Chelsea Publishing, American Mathematical Society, Providence, RI, 2003, Corrected reprint of the 1976 original.

[Ros94] Jonathan Rosenberg, *Algebraic K-theory and its applications*, Graduate Texts in Mathematics, vol. 147, Springer, New York, 1994. MR 95e:19001.

[RS72] C. P. Rourke and B. J. Sanderson, *Introduction to piecewise-linear topology*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 69, Springer, New York, 1972. MR 50 #3236.

[RT91] N. Reshetikhin and V. G. Turaev, *Invariants of* 3*-manifolds via link polynomials and quantum groups*, Invent. Math. **103** (1991), no. 3, 547–597. MR 92b:57024.

[Rub96] Daniel Ruberman, *The minimal genus of an embedded surface of non-negative square in a rational surface*, Turkish J. Math. **20** (1996), no. 1, 129–133. MR 97k:57036.

[Rud01] Yuli B. Rudyak, *Piecewise linear structures on topological manifolds*, 2001. `arXiv:math.AT/0105047`.

[Sak96] Takashi Sakai, *Riemannian geometry*, Translations of Mathematical Monographs, vol. 149, American Mathematical Society, Providence, RI, 1996, Transl from 1992 Japanese original. MR 97f:53001.

[Sal96] Dietmar Salamon, *Removable singularities and a vanishing theorem for Seiberg–Witten invariants*, Turkish J. Math. **20** (1996), no. 1, 61–73. MR 97f:57035.

[Sar42] Arthur Sard, *The measure of the critical values of differentiable maps*, Bull. Amer. Math. Soc. **48** (1942), 883–890. MR 4,153c.

[Sav02] Nikolai Saveliev, *Invariants for homology 3–spheres*, Encyclopaedia of Mathematical Sciences, vol. 140: Low-Dimensional Topology, I, Springer-Verlag, Berlin, 2002. MR 2004c:57026.

[Sch93] Matthias Schwarz, *Morse homology*, Progress in Mathematics, vol. 111, Birkhäuser, Basel, 1993. MR 95a:58022.

[Sco83] Peter Scott, *The geometries of 3–manifolds*, Bull. London Math. Soc. **15** (1983), no. 5, 401–487. MR 84m:57009.

[Sco03] Alexandru Scorpan, *Existence of foliations on 4–manifolds*, Algebr. Geom. Topol. **3** (2003), 1225–1256 (electronic). `arXiv:math.GT/0302318`.

[Sei35] Herbert Seifert, *Über das Geschlecht von Knoten*, Math. Ann. **110** (1935), 571–592.

[Ser51] Jean-Pierre Serre, *Homologie singulière des espaces fibrés. III. Applications homotopiques*, C. R. Acad. Sci. Paris **232** (1951), 142–144. MR 12,521a.

[Ser56] ———, *Géométrie algébrique et géométrie analytique*, Ann. Inst. Fourier, Grenoble **6** (1955–1956), 1–42. MR 18,511a.

[Ser62] ———, *Formes bilinéaires symétriques entières à discriminant* ± 1, Séminaire Henri Cartan, 1961/62, Exp. 14-15, Secrétariat mathématique, Paris, 1961/1962, p. 16. MR 28 #3443.

[Ser70] ———, *Cours d'arithmétique*, Le Mathématicien, vol. 2, Presses Universitaires de France, Paris, 1970. MR 41 #138.

[Ser73] ———, *A course in arithmetic*, Graduate Texts in Mathematics, vol. 7, Springer-Verlag, New York, 1973, Translated from the French. MR 49 #8956.

[Ser77] ———, *Cours d'arithmétique*, second corrected ed., Le Mathématicien, vol. 2, Presses Universitaires de France, Paris, 1977. MR 58 #16473.

[Sha74] Igor R. Shafarevich, *Basic algebraic geometry*, Die Grundlehren der mathematischen Wissenschaften, vol. 213, Springer, New York, 1974. MR 51 #3163.

[Sha94] ———, *Basic algebraic geometry. I: Varieties in projective space; & II: Schemes and complex manifolds*, second ed., Springer, Berlin, 1994. MR 95m:14001 & 95m:14002.

[Sma59] Stephen Smale, *Diffeomorphisms of the 2–sphere*, Proc. Amer. Math. Soc. **10** (1959), 621–626. MR 22 #3004.

[Sma61] ———, *Generalized Poincaré's conjecture in dimensions greater than four*, Ann. of Math. (2) **74** (1961), 391–406. MR 25 #580.

[Sma62] ———, *On the structure of manifolds*, Amer. J. Math. **84** (1962), 387–399. MR 27 #2991.

[Sma65] ———, *An infinite dimensional version of Sard's theorem*, Amer. J. Math. **87** (1965), 861–866. MR 32 #3067.

[Spa66] Edwin H. Spanier, *Algebraic topology*, McGraw-Hill Book Co., New York, 1966. MR 35 #1007.

[Spa81] ———, *Algebraic topology*, Springer, New York, 1981, Corrected reprint. MR 83i:55001.

[SS94] András I. Stipsicz and Zoltán Szabó, *The smooth classification of elliptic surfaces with* $b^+ > 1$, Duke Math. J. **75** (1994), no. 1, 1–50. MR 96d:57033.

[SS04] ———, *An exotic smooth structure on* $\mathbb{CP}^2 \# 6\overline{\mathbb{CP}}^2$, 2004. `arXiv:math.GT/0411258`.

[SS05] ______, *Small exotic 4–manifolds with $b_2^+ = 3$*, 2005. `arXiv:math.GT/0501273`.

[Sta60a] John R. Stallings, *On the loop theorem*, Ann. of Math. (2) **72** (1960), 12–19. MR 22 #12526.

[Sta60b] ______, *Polyhedral homotopy-spheres*, Bull. Amer. Math. Soc. **66** (1960), 485–488. MR 23 #A2214.

[Sta65] ______, *On infinite processes leading to differentiability in the complement of a point*, Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), Princeton University Press, Princeton, NJ, 1965, pp. 245–254. MR 31 #5213.

[Sta67] ______, *Lectures on polyhedral topology*, Tata Institute of Fundamental Research Lectures on Mathematics, vol. 43, Tata Institute of Fundamental Research, Bombay, 1967. MR 38 #6605.

[Sta71] ______, *Group theory and three-dimensional manifolds*, Yale University Press, New Haven, CT, 1971. MR 54 #3705.

[Ste51] Norman Steenrod, *The topology of fibre bundles*, Princeton Mathematical Series, vol. 14, Princeton University Press, Princeton, NJ, 1951. MR 12,522b.

[Ste99] ______, *The topology of fibre bundles*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1999, Reprint of the 1957 edition. MR 2000a:55001.

[Ste05] Ronald J. Stern, *Will we ever classify simply-connected smooth 4–manifolds?*, 2005. `arXiv:math.GT/0502164`.

[Sti99] András I. Stipsicz, *Simply connected symplectic 4–manifolds with positive signature*, Turkish J. Math. **23** (1999), no. 1, 145–150, Proceedings of 6th Gökova Geometry-Topology Conference. MR 2001e:57029.

[Sto68] Robert E. Stong, *Notes on cobordism theory*, Mathematical notes, Princeton University Press, Princeton, NJ, 1968. MR 40 #2108.

[SW94a] N. Seiberg and E. Witten, *Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory*, Nuclear Phys. B **426** (1994), no. 1, 19–52. MR 95m:81202a.

[SW94b] ______, *Erratum: "Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory"*, Nuclear Phys. B **430** (1994), no. 2, 485–486. MR 95m:81202b.

[SW94c] ______, *Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD*, Nuclear Phys. B **431** (1994), no. 3, 484–550. MR 95m:81203.

[Sym01] Margaret Symington, *Generalized symplectic rational blowdowns*, Algebr. Geom. Topol. **1** (2001), 503–518 (electronic). MR 2002g:57056. `arXiv:math.SG/0108075`.

[Sza98] Zoltán Szabó, *Simply-connected irreducible 4–manifolds with no symplectic structures*, Invent. Math. **132** (1998), no. 3, 457–466. MR 99f:57033.

[Tau82] Clifford H. Taubes, *Self-dual Yang–Mills connections on non-self-dual 4–manifolds*, J. Differential Geom. **17** (1982), no. 1, 139–170. MR 83i:53055.

[Tau87] ______, *Gauge theory on asymptotically periodic 4–manifolds*, J. Differential Geom. **25** (1987), no. 3, 363–430. MR 88g:58176.

[Tau90] ______, *Casson's invariant and gauge theory*, J. Differential Geom. **31** (1990), no. 2, 547–599. MR MR1037415 (91m:57025).

[Tau94] ______, *The Seiberg–Witten invariants and symplectic forms*, Math. Res. Lett. **1** (1994), no. 6, 809–822. MR 95j:57039.

[Tau95a] ______, *More constraints on symplectic forms from Seiberg–Witten invariants*, Math. Res. Lett. **2** (1995), no. 1, 9–13. MR 96a:57075.

[Tau95b] ______, *The Seiberg–Witten and Gromov invariants*, Math. Res. Lett. **2** (1995), no. 2, 221–238. MR 96a:57076.

[Tau96a] ______, *Counting pseudo-holomorphic submanifolds in dimension 4*, J. Differential Geom. **44** (1996), no. 4, 818–893. MR 97k:58029.

[Tau96b] ______, *SW $\Rightarrow$ Gr: from the Seiberg–Witten equations to pseudo-holomorphic curves*, J. American Mathematical Society **9** (1996), no. 3, 845–918. MR 97a:57033.

[Tau98a] ———, *The geometry of the Seiberg–Witten invariants,* Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), no. Extra Vol. II, 1998, pp. 493–504 (electronic). MR 99j:57035.

[Tau98b] ———, *The geometry of the Seiberg–Witten invariants,* Surveys in differential geometry, Vol. III (Cambridge, MA, 1996), Int. Press, Boston, MA, 1998, pp. 299–339. MR 2000j:57067.

[Tau98c] ———, *The structure of pseudo-holomorphic subvarieties for a degenerate almost complex structure and symplectic form on* $\mathbb{S}^1 \times \mathbb{B}^3$, Geom. Topol. **2** (1998), 221–332 (electronic). MR 99m:57029.

[Tau99a] ———, *Gr = SW: counting curves and connections,* J. Differential Geom. **52** (1999), no. 3, 453–609. MR 2002i:53119.

[Tau99b] ———, *Gr* $\Rightarrow$ *SW: from pseudo-holomorphic curves to Seiberg–Witten solutions,* J. Differential Geom. **51** (1999), no. 2, 203–334. MR 2000i:53123.

[Tau99c] ———, *Seiberg–Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic* 2*-forms,* Geom. Topol. **3** (1999), 167–210 (electronic). MR 1 697 181. `arXiv:math.SG/9907199`.

[Tau00a] ———, *Seiberg–Witten and Gromov invariants for symplectic* 4*-manifolds,* International Press, Somerville, MA, 2000. MR 1 798 809.

[Tau00b] ———, *Seiberg–Witten invariants, self-dual harmonic* 2*-forms and the Hofer–Wysocki–Zehnder formalism,* Surveys in differential geometry, Surv. Differ. Geom., VII, International Press, Somerville, MA, 2000, pp. 625–672. MR 2003k:53119.

[Tau01] ———, *The Seiberg-Witten invariants and* 4*-manifolds with essential tori,* Geom. Topol. **5** (2001), 441–519 (electronic). MR 2002d:57025.

[Tau02] ———, *A compendium of pseudoholomorphic beasts in* $\mathbb{R} \times (\mathbb{S}^1 \times \mathbb{S}^2)$, Geom. Topol. **6** (2002), 657–814 (electronic). MR 1 943 381.

[Tho53a] René Thom, *Sous-variétés et classes d'homologie des variétés différentiables. I. Le théorème général. II. Résultats et applications,* C. R. Acad. Sci. Paris **236** (1953), 453–454 & 573–575. MR 14,1005.

[Tho53b] ———, *Variétés différentiables cobordantes,* C. R. Acad. Sci. Paris **236** (1953), 1733–1735. MR 14,1112b.

[Tho54] ———, *Quelques propriétés globales des variétés différentiables,* Comment. Math. Helv. **28** (1954), 17–86. MR 15,890a.

[Tho60] ———, *Des variétés triangulées aux variétés différentiables,* Proc. Internat. Congress Math. 1958, Cambridge Univ. Press, New York, 1960, pp. 248–255. MR 22 #12536.

[Thu76] William P. Thurston, *Some simple examples of symplectic manifolds,* Proc. Amer. Math. Soc. **55** (1976), no. 2, 467–468. MR 53 #6578.

[Thu86] ———, *A norm for the homology of* 3*-manifolds,* Mem. American Mathematical Society **59** (1986), no. 339, i–vi and 99–130. MR 88h:57014.

[Thu97] ———, *Three-dimensional geometry and topology. Vol. 1,* Princeton Mathematical Series, vol. 35, Princeton University Press, Princeton, NJ, 1997, Edited by Silvio Levy. MR 97m:57016.

[Uhl82a] Karen K. Uhlenbeck, *Connections with* L^p *bounds on curvature,* Comm. Math. Phys. **83** (1982), no. 1, 31–42. MR 83e:53035.

[Uhl82b] ———, *Removable singularities in Yang–Mills fields,* Comm. Math. Phys. **83** (1982), no. 1, 11–29. MR 83e:53034.

[vdB59] F. van der Blij, *An invariant of quadratic forms mod* 8, Nederl. Akad. Wetensch. Proc. Ser. A 62 = Indag. Math. **21** (1959), 291–293. MR 21 #7183.

[Wal60] Andrew H. Wallace, *Modifications and cobounding manifolds,* Canad. J. Math. **12** (1960), 503–528. MR 23 #A2887.

[Wal62] C. T. C. Wall, *On the orthogonal groups of unimodular quadratic forms,* Math. Ann. **147** (1962), 328–338. MR 25 #2009.

[Wal64a] ———, *On the orthogonal groups of unimodular quadratic forms. II*, J. Reine Angew. Math. **213** (1963/1964), 122–136. MR 27 #5732.

[Wal64b] ———, *Diffeomorphisms of* 4*-manifolds*, J. London Math. Soc. **39** (1964), 131–140. MR 29 #626.

[Wal64c] ———, *On simply-connected* 4*-manifolds*, J. London Math. Soc. **39** (1964), 141–149. MR 29 #627.

[Wal69a] ———, *Non-additivity of the signature*, Invent. Math. **7** (1969), 269–274. MR MR0246311 (39 #7615).

[Wal69b] ———, *On homotopy tori and the annulus theorem*, Bull. London Math. Soc. **1** (1969), 95–97. MR 39 #3498.

[Wal70] ———, *Surgery on compact manifolds*, London Mathematical Society Monographs, vol. 1, Academic Press, London, 1970. MR 55 #4217.

[Wal92] Kevin Walker, *An extension of Casson's invariant*, Annals of Mathematics Studies, vol. 126, Princeton University Press, Princeton, NJ, 1992. MR 93e:57031.

[Wal99] C. T. C. Wall, *Surgery on compact manifolds*, second ed., Mathematical Surveys and Monographs, vol. 69, American Mathematical Society, Providence, RI, 1999. MR 2000a:57089.

[Whi44] Hassler Whitney, *The self-intersections of a smooth n-manifold in $2n$-space*, Ann. of Math. (2) **45** (1944), 220–246. MR 5,273g.

[Whi49a] J. H. C. Whitehead, *Combinatorial homotopy. I & II*, Bull. Amer. Math. Soc. **55** (1949), 213–245 & 453–496. MR MR0030759 (11,48c) & MR0030760 (11,48c).

[Whi49b] ———, *On simply connected,* 4*-dimensional polyhedra*, Comment. Math. Helv. **22** (1949), 48–92. MR 10,559d.

[Whi50] ———, *Simple homotopy types*, Amer. J. Math. **72** (1950), 1–57. MR 11,735c.

[Whi61] ———, *Manifolds with transverse fields in euclidean space*, Ann. of Math. (2) **73** (1961), 154–212. MR 23 #A2225.

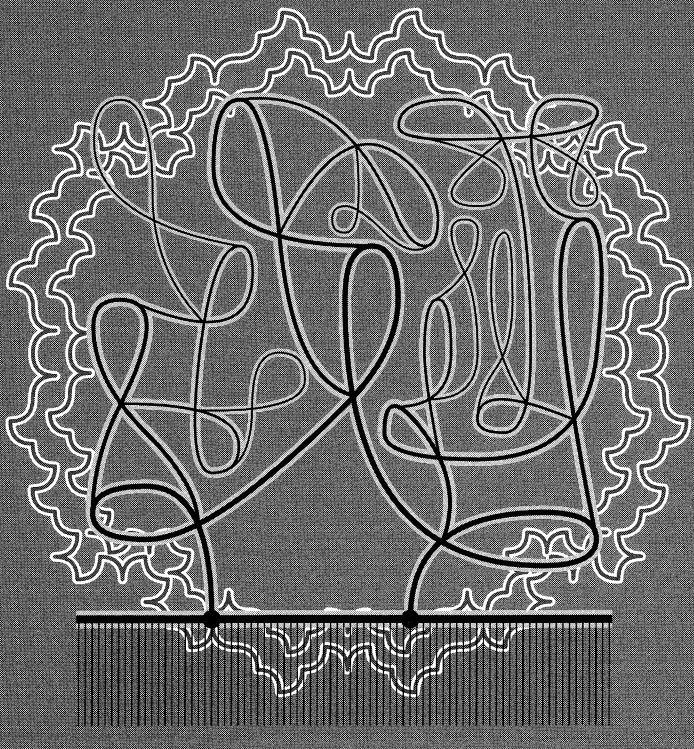
[Wit89] Edward Witten, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. **121** (1989), no. 3, 351–399. MR 90h:57009.

[Wit94] ———, *Monopoles and four-manifolds*, Math. Res. Lett. **1** (1994), no. 6, 769–796. MR 96d:57035. `arXiv:hep-th/9411102`.

[Wu52] Wen-tsun Wu, *Sur les classes caractéristiques des structures fibrées sphériques*, Publ. Inst. Math. University Strasbourg, vol. 11, Hermann & Cie, Paris, 1952, pp. 5–89, 155–156. MR 14,1112d.

[Wu88] Hung-hsi Wu, *The Bochner technique in differential geometry*, Math. Rep. **3** (1988), no. 2, i–xii and 289–538. MR 91h:58031.

[Zee61] E. C. Zeeman, *The generalised Poincaré conjecture*, Bull. Amer. Math. Soc. **67** (1961), 270. MR 23 #A2215.



Index

Page numbers in **bold** refer to the definition, construction, or in other way the main treatment of the item. Items typed in ***italic bold*** are theorems or conjectures. Page numbers followed by *fn* refer to content in a footnote. Notations involving a Greek letter are alphabetized under the corresponding Roman letter, *i.e.*, Δ under D, Λ under L, etc. Similarly, $\mathbb{CP}$ is placed as CP would, *etc.*.