Rend. Istit. Mat. Univ. Trieste
Volume 42 (2010), 103–128.

Gauge Theory: from Physics to Geometry[1]

Ugo Bruzzo

Abstract. *Maxwell theory may be regarded as a prototype of gauge theory and generalized to nonabelian gauge theory. We briefly sketch the history of gauge theories, from Maxwell to Yang-Mills theory, and the identification of gauge fields with connections on fibre bundles. We introduce the notion of instanton and consider the moduli spaces of such objects. Finally, we discuss some modern techniques for studying the topology of these moduli spaces.*

Keywords: Gauge Theory, Instantons, Moduli Spaces, Instanton Counting.
MS Classification 2010: 14D20, 14D21, 14J60, 32G13, 58D27, 70S15, 81T45

1. Introduction

The title of a famous paper by Eugene Wigner, "The unreasonable effectiveness of mathematics in the natural sciences" [56], has by now become almost a commonplace. Here I would like to turn it upside-down, and make it into "the unreasonable effectiveness of nature in teaching us mathematics". There have been indeed remarkable instances where physical theories have provided formidable input to mathematicians, offering the stimulus to the creation of new mathematical theories, and supplying strong evidence for highly nontrivial theorems. A striking example of this new kind of interaction between mathematics

[1] This paper is an elaboration of the contents of a talk given in a meeting in occasion of the 40th anniversary of Rendiconti dell'Istituto di Matematica dell'Università di Trieste. I thank the organizers of the meeting for their invitation, and Claudio Bartocci for a useful conversation about the development of gauge theory. I also thank Giuseppe Bruzzaniti for reading the manuscript and for helping with the pictures, and Alessandro Tanzini for suggestions. The original results I cite in this text have been obtained in collaboration with Francesco Fucito, Dimitri Markushevich, José Morales, Rubik Poghossian and Alessandro Tanzini (in different combinations). This paper was written while I was visiting the Department of Mathematics of the University of Pennsylvania for the Fall 2010 term; I thank Penn for hospitality and support, and the staff and the scientists at the Department of Mathematics for providing an enjoyable and productive atmosphere. The original research on which this paper is based was supported by PRIN "Geometria delle varietà algebriche e dei loro spazi di moduli" and the INFN project PI14 "Nonperturbative dynamics of gauge theories".

and physics is string theory, with its relation with the theory of invariants of algebraic varieties. A beautiful account of the interplay between string theory and mathematics, from the point of view of string dualities, is provided in [40].

Another major character in this story is gauge theory. If we were to pinpoint a starting point of gauge theory, we could go back to Maxwell equations. These equations crowned 200 years of experimental and theoretical research, to which one can associate such names as Nollet, Coulomb, Ampère, Arago, Ørsted, Faraday, Henry, Neumann, Maxwell.... The fields entering the Maxwell equations, the electric and magnetic fields, may be written in a suitable way as derivatives of two potentials, the scalar and the vector potential. However, these potentials are defined up a suitable combination of the derivatives of another scalar field; this is the "gauge invariance" of electromagnetism. Now, the essence of gauge theory, from the physical viewpoint, is that this gauge invariance dictates the way matter interacts via the electromagnetic fields. A first attempt to implement this idea, as a way to unify electromagnetism with gravitation, was done in 1918 by Hermann Weyl [54]. His theory was not successful, for some reasons that we cannot examine here, however it contained many ideas that found applications and were developed later on, such as the role of conformal geometry. He also introduced the term "gauge".

The first workable gauge theory after electromagnetism is Yang-Mills theory, of which we shall give some outline in the next section. The paper by Yang and Mills was published in 1954. However gauge theory entered the mathematical scene only when it was realized that a gauge field may be pictured as a connection on a fibre bundle. To my knowledge, the first paper where such a relationship was explicitly suggested is a 1958 paper by Dennis Sciama [52], even though Utiyama's paper [53] already contains the mathematics of this relationship, albeit in local, coordinate form. Precusors of this interpretation were the already mentioned paper by Weyl [54], a 1953 letter by Pauli to A. Pais [49], and others.

However, only in the late 70s the mathematics of gauge theory became a mainstream subject of study for mathematicians. A search on Mathematical Reviews will show that in the years 1977 and 1978 a huge number of papers was published on the mathematics of gauge theory, most of them related in some way to M. F. Atiyah and his collaborators. Here we shall only cite [1, 4, 6]. Afterwards, the work of S. K. Donaldson,[1] (a student of Atiyah's, and a 1986 Fields Medal recipient) showed that gauge theory is a powerful tool for the study of the geometry of four-manifolds — in particular, $SU(2)$ gauge theory.[2]

Given a (compact, oriented) four-manifold X, the moduli space of $SU(2)$

[1]See [19] and references therein.

[2]Donaldson wrote the first paper on this topic [16] when he still was a graduate student. According to the words of Michael Atiyah [2], that paper "stunned the mathematical world." Michael Atiyah himself got a Fields Medal in 1966.

instantons — a space which parametrizes connections of a particular kind, that we shall introduce in section 3 — allows one to associate with X some invariants, that are able to distinguish between different differentiable structures on X. This allowed for spectacular advances in the study of the topology and geometry of four-manifolds.

The moduli space of instantons is also at the base of some constructions that are being used to establish unexpected and highly nontrivial relations between different invariants that one can associate to geometric spaces, such as the Gromov-Witten and the Seiberg-Witten invariants. Our purpose in this paper is to give a rough sketch of the inception of gauge theories, starting from Maxwell theory and from there moving to Yang-Mills theory. We shall introduce the concept of instanton, and will briefly explain what their moduli space is. From there we shall go to the moduli spaces of framed sheaves, which provide a desingularization of the moduli space of instantons, and will show how a technique called "instanton counting" allows one to study the topology of these moduli spaces. This knowledge is important in the physical applications of this theory.

In no way this paper pretends to give a full account of the history of gauge theory,[3] or of the relations between the mathematics and the physics of gauge theory. Neither there is any claim to originality. Our only aim is to sketch a path from Maxwell theory to some modern developments of gauge theory that may highlight some points of interest and motivate further study into the subject.

2. Maxwell Equations

The Maxwell equations are a system of partial differential equations for the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$, with the electric charge density ρ and the electric current density vector $\mathbf{j}$ acting as sources.[4] In the CGS system of units they read as

$$\begin{aligned}
\operatorname{div}\mathbf{E} &= 4\pi\rho \\
\operatorname{rot}\mathbf{B} &= \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \\
\operatorname{div}\mathbf{B} &= 0 \\
\operatorname{rot}\mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}
\end{aligned}$$

[3]For a fuller account of the early developments of gauge theories the reader may wish to consult [48]. This is a collection of original papers, some translated from German, with comments and an introductory chapter by the editor L. O'Raifeartaigh.

[4]Good, classical references for the physics and mathematics of Maxwell equations, and their four-dimensional formulation, are [36, 39].

Several features of the electromagnetic field, that gave rise to highly non-trivial developments, may be drawn from these equations. The main ones that come to my mind are the following.

(i) A current is just electric charge in movement. Thus, different observers, in relative motion, will see different values for the charge and current density fields. For instance, if some observer just sees a distribution of electric charges at rest, and no electric current, another observer in relative motion will see some current, in addition to some charge. In view of Maxwell's equations, we may expect the same to be true for the electric and magnetic fields: the value of $\mathbf{E}$ and $\mathbf{B}$ will be observer-dependent, and moreover, we may expect that the transformation laws for these fields under change of observer will "mix" these fields: the value of the electric field for the observer "in motion" will depend on the values of both the electric and magnetic fields as seen by the observer "at rest", and the same for the magnetic field.

(ii) After some manipulations, from Maxwell's equations in the absence of sources (i.e., with $\mathbf{j} = \rho = 0$) one can obtain the wave equations for the electric and magnetic fields:

$$\Delta\mathbf{E} = \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}, \qquad \Delta\mathbf{B} = \frac{1}{c^2}\frac{\partial^2\mathbf{B}}{\partial t^2}\,.$$

The constant c, that appeared in Maxwell's equations, plays now the role of speed of propagation for the electromagnetic waves (also called *speed of light* since light turns out just to be a form of electromagnetic waves). The constant c can be measured in a laboratory by means of experiments in electrostatics and magnetostatics. What is striking in this state of affairs is that c appears to be the speed of light *for every observer for which the Maxwell equations hold.* If we assume — as it seems quite natural to do — that the Maxwell equation hold for any inertial observer, we have a contradiction with Galilean relativity, which would prescribe different speeds for the electromagnetic waves for different observers. This seeming contradiction is one of the roots of special relativity. According to that theory, Maxwell equations hold for any inertial observer, and the speed of light has the same value for all inertial observers. Of course, the price to be paid is that Galilean relativity should be relinquished and replaced by Einstenian relativity, with its nontrivial law of addition of velocities.

(iii) The electric and magnetic fields can be written in terms of a scalar field ϕ (the scalar potential) and a vector field $\mathbf{A}$ (the vector potential), according to the equations

$$\mathbf{E} = -\operatorname{grad}\phi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}, \qquad \mathbf{B} = \operatorname{rot}\mathbf{A}\,. \tag{1}$$

It turns out that these potential are fixed by the electromagnetic field up to a combination of the derivatives of an indeterminate scalar field (function); let us call it ψ. If we set

$$\mathbf{A}' = \mathbf{A} + \operatorname{grad}\psi, \qquad \phi' = \phi - \frac{1}{c}\frac{\partial\psi}{\partial t} \tag{2}$$

the pairs $(\mathbf{A}, \phi)$ and $(\mathbf{A}', \phi')$ determine via the equation (1) the same electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$. This freedom may be used to "gauge" the potentials $\mathbf{A}$, ϕ in a way to simplify the treatment of some specific problem. For instance, if the potentials satisfy the condition (Lorentz gauge condition)

$$\operatorname{div}\mathbf{A} + \frac{1}{c}\frac{\partial\phi}{\partial t} = 0$$

then $\mathbf{A}$ and ϕ satisfy the inhomogeneous wave equation with sources given by the charge and current densities:

$$\Delta\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = \frac{4\pi}{c}\mathbf{j}, \qquad \Delta\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 4\pi\,\rho\,.$$

The Lorentz condition can always be met up to solving a partial differential equation: indeed, if $(\mathbf{A}, \phi)$ is any given pair of potentials, and ψ is a scalar field satisfying the inhomogeneous wave equation

$$\Delta\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = -\operatorname{div}\mathbf{A} - \frac{1}{c}\frac{\partial\phi}{\partial t}$$

then the potentials $(\mathbf{A}', \phi')$ given by the equations (2) satisfy the Lorentz gauge condition.

One outcome of this discussion is that electromagnetism should be more satisfactorily formulated in a four-dimensional setting, i.e., as a field theory on the four-dimensional Minkowski spacetime of special relativity. In this way the Maxwell equations explicitly display their invariance under the special-relativistic group of reference transformations (the Poincaré group). This invariance is not too easily detected from the three-dimensional equations we have previously written. Let us write the Maxwell equations in this way. One organizes the components of electromagnetic fields into a 4×4 matrix (the indexes μ, ν run form 0 to 3)[5]

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}$$

[5]In this part of our treatment we assume that the signature of the Minkowski metric is $(+\,-\,-\,-)$, in accordance with the usage in physics.

and from this one defines a differential 2-form (i.e., a skew-symmetric covariant two-tensor)

$$F = \tfrac{1}{2} \sum_{\mu,\nu=0,\dots,3} F_{\mu\nu}\, dx^\mu \wedge dx^\nu \,.$$

Analogously, one assembles the sources into a differential 1-form (a covariant four-vector)

$$j_\mu = (c\rho, -\mathbf{j}), \qquad j = \sum_{\mu=0}^{3} j_\mu \, dx^\mu$$

Maxwell equations may now be written as

$$dF = 0, \qquad * d * F = \tfrac{4\pi}{c}\, j$$

where d is the exterior (Cartan) differential, and $*$ denoted the Hodge dual. In component notation, these may be written as

$$\partial_\mu F_{\nu\lambda} + \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} = 0, \qquad \sum_{\nu=0}^{3} \partial^\nu F_{\nu\mu} = \tfrac{4\pi}{c} j_\mu \,.$$

The homogeneous Maxwell equations $dF = 0$ allow one to write $F = dA$ for a differential 1-form A. Again, in components this reads $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. If we set $A = (c\phi, -\mathbf{A})$, the equation $F = dA$ turns out be the four-dimensional form of equations (1). The 1-form A is called the *electromagnetic potential.*

Since $d^2 = 0$ (this is Schwarz's lemma about the symmetry of the second derivatives of a sufficiently smooth function), F is invariant under the transformation

$$A \mapsto A + d\psi \,. \tag{3}$$

Thus we recover the gauge transformations (2). The advantage of putting these equations into this new form is that it suggests a very interesting geometric interpretation.[6] Let $P \to X$ be a $U(1)$ principal bundle on a differentiable manifold X, equipped with a connection ω (we shall use the same symbol for the associated differential form on the total space P of the bundle).[7] Let $f\colon P \to P$ be a vertical automorphism of P, i.e., a diffeomorphism which maps fibres to fibres and is $U(1)$-equivariant, $f(ug) = f(u)g$ if $u \in P$ and $g \in U(1)$. The automorphism f acts on the connection by pullback

$$\omega \mapsto f^*(\omega)\,, \tag{4}$$

producing in general a new connection. The transformation (4) is called a *gauge transformation.*

[6] For an introduction to the geometry of gauge theories the reader may consult [10].

[7] For the theory of principal bundles and connections we refer the reader to [38].

The base manifold X is to be identified with spacetime, and if we want to write equations on spacetime, we need to pullback the quantities defined on P to X using a section σ of P, that is, a differentiable map $\sigma\colon X \to P$ such that $\pi \circ \sigma = \mathrm{id}_X$, where $\pi\colon P \to X$ is the projection. However such a section exists if and only if P is trivial, which in general is not the case (see the remark at the end of this section). Then we may consider *local sections* $\sigma\colon U \to P$, and set $A = \sigma^*\omega$. A section establishes an isomorphism $P_{|U} \simeq U \times U(1)$ by letting $u \mapsto (x, g)$, where $x = \pi(u)$, and g is the element of $U(1)$ such that $u = \sigma(x)g$. Under such an isomorphism, the restriction of the vertical automorphism f to $P_{|U}$ may be identified with a map $\tilde{f}\colon U \to U(1)$.

Now, if we have two sections σ, σ', and $\sigma' = \sigma \cdot \tilde{f}$, one has $A' = A + \tilde{f}^{-1} d\tilde{f}$, and, if we set $\psi = \log \tilde{f}$, we get the transformations (2). Thus, *the gauge trasformation of electromagnetism may be regarded as gauge transfomations in the sense of bundle theory.* We need to identify the electromagnetic potential with a connection on a $U(1)$ bundle: the vertical automorphisms of the bundle will reproduce the gauge transformation of electromagnetism. Moreover, the field strength $F = dA$ turns out to be the *curvature* of the connection.

If we allow the base X of the principal bundle to have nontrivial topology — so that P itself may be nontrivial — we get interesting effects. Assume for instance that a certain field configuration is time independent in some reference frame, and that the associated 3-space has the topology of $S^2 \times \mathbb{R}$. The dependence on the radial coordinate is easily separated and solved, and one is left with a $U(1)$ bundle on S^2. Such bundles are topologically classified by an integer (the first Chern class). In physics the resulting field strength is called a *Dirac monopole*, and the first Chern class is called the *charge* of the monopole.[8]

3. Yang-Mills Fields

Once electromagnetism is given this geometric interpretation, it is quite natural to argue that one can generalize it by replacing the structure group $U(1)$ with another group. In particular, one could expect nontrivial effects to arise from the choice of a nonabelian structure group. Such a generalization was indeed proposed by the physicists C. N. Yang and R. L. Mills [58] on purely physical grounds, before the intepretation of gauge fields as connections was known. In their 1954 paper, they proposed a gauge theory based on the group $SU(2)$ as a model for the so-called *isospin*. The basic idea is that the proton and the neutron are two different states of a single particle, the nucleon, which has a quantum number, the isospin, whose values correspond to the two particles. So, the observable isospin has two eigenstates, and $SU(2)$ acts on the two-dimensional complex vector space generated by these eigenstates. This

[8]More information on the Dirac monopole may be found in [44], and, from the physical viewpoint, in [36].

idea may be traced back to Heisenberg [31]; the term "isospin" was coined by Wigner [55].

Yang and Mills' idea was to promote this symmetry from a "global" to a "local" one, namely, they allowed the element of the $SU(2)$ group acting on the isospin space to depend on the spacetime position. Once this is done, the theory is no longer invariant, and to restore invariance one needs to include new fields: these are the gauge fields, which, from the physical viewpoints, are interpreted as the carriers of a physical interaction, in this case, the strong interaction (I will descrive below this mechanism in the case of electromagnetism). This theory was not entirely successful, and indeed nowadays the physics of the nucleons is explained in a completely different way by another gauge theory, called *chromodynamics*,[9] based on the group $SU(3)$ [30]. However, Yang-Mills theory has survived this drawback, and gauge theory has become the universal paradigm for the modelization of the fundamental interactions; in addition to the already mentioned chromodynamics, there is the Weinberg-Salam electroweak theory [29], a gauge theory based on the group $SU(2)\times U(1)$, which provides a unified theory of electromagnetism and the weak nuclear force. More generally, the basic structure of the Standard Model (a comprehensive theory of the fundamental interactions, excluding gravity[10]) is that of a gauge theory; and the way string theory is able to be interpreted as a unified theory of all interactions, is, at least for the electroweak and strong forces, again via gauge theory.

Let us now explain by the simplest example what the "gauge principle" is, namely, how the requirement for a global symmetry to be promoted to a local one enforces the presence of a new field, which will describe an interaction. Let us consider the Dirac equation for a spinor field ψ:

$$i\sum_{\mu=0}^{3}\gamma^\mu\,\partial_\mu\psi = m\psi\,.$$

Here γ^μ are the gamma matrices, i.e., the generators of a representation of the group $SL(2,\mathbb{C})$ on $\mathbb{C}^4$ (the group $SL(2,\mathbb{C})$ plays a role here because it is the universal covering of the Lorentz group, or to be more precise, of the proper orthocronous Lorentz group, which is the connected component of the Lorentz group containing the identity). Moreover m is the mass of the spinor field (to be identified with the electron/positron field). The Dirac equation may be derived as Euler-Lagrange equations from a variational principle associated

[9]Chromodynamics is the theory according to which heavy particles are made up by more elementary constituents, called *quarks,* which interact via the strong force; the latter is described by an $SU(3)$ gauge field, whose associated particles are called *gluons.*

[10]A good, even though somehow elementary, introduction to the Standard Model for non-specialists is given in [7]; see also [47].

with the action functional

$$S(\psi) = \frac{1}{2}\int_X \bar{\psi}\Big(i\,\textstyle\sum_\mu \gamma^\mu\,\partial_\mu - m\Big)\psi\,d^4x \quad + \quad \text{hermitian conjugate.}$$

This functional, and the Dirac equation, are invariant under the transformation

$$\psi \mapsto e^{i\alpha}\,\psi \qquad \bar{\psi} \mapsto e^{-i\alpha}\,\bar{\psi} \tag{5}$$

where α is a real constant. If α is nonconstant, i.e., it is an arbitrary function on spacetime, the action integral is no longer invariant. To make it invariant even when α is not constant, one can replace ∂_μ with $D_\mu = \partial_\mu - iA_\mu$, where A_μ is some field, and accompany the transformation rule (5) with

$$A_\mu \mapsto A_\mu + \partial_\mu \alpha\,.$$

Thus, we have rediscovered the electromagnetic gauge transformations! We may therefore interpret the field A as the electromagnetic potential, and consider an extended action integral, where (in addition to replacing ∂_μ by $D_\mu = \partial_\mu - ieA_\mu$) we include a term for the electromagnetic field. The quantity e is a “coupling constant”, to be identified with the absolute value of the electric charge of the field ψ (electron charge). The complete action now reads

$$S(\psi, A) = \int_X \left[\frac{1}{2}\bar{\psi}\Big(i\,\textstyle\sum_\mu \gamma^\mu\,D_\mu - m\Big)\psi + \frac{1}{32\pi}\textstyle\sum_{\mu\nu} F^{\mu\nu}F_{\mu\nu}\right] d^4x + \quad \text{h.c.}$$

The equations for the electron field are now

$$i\sum_{\mu=0}^{3}\gamma^\mu\,D_\mu\psi = m\psi$$

or

$$i\left(\sum_{\mu=0}^{3}\gamma^\mu\,\partial_\mu - m\right)\psi = -e\sum_{\mu=0}^{3}\gamma^\mu\,A_\mu\,\psi$$

which contains a terms that describes an interaction between ψ and A. The Euler-Lagrange equations for A read

$$\sum_\nu \partial^\nu\,F_{\nu\mu} = \frac{4\pi}{c}e\,\bar{\psi}\,\gamma_\mu\,\psi$$

i.e., we obtain the Maxwell equations with a source current term given by the electron field: indeed, the electron is a charged particle, and is the source of an electromagnetic field.

Of course, this is a fully classical description, which makes no physical sense unless it is quantized; but this is another story, i.e., quantum electrodynamics (QED). For a leisurely introduction to QED the reader may consult [21].

A similar treatment actually applies for any gauge group, for instance, for the $SU(2)$ group of Yang-Mills theory. However in that case a new phenomenon arises, due to the fact that $SU(2)$ is not abelian. The connection ω on the principal bundle is described by a differential 1-form with values in the Lie algebra of the gauge group; in the case of $SU(2)$, this is the vector space of 2×2 anti-hermitian complex matrices with zero trace, equipped with a Lie bracket given by the commutator of matrices. For this reason, the relation between the connection and its curvature is no longer linear:[11]

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]\,.$$

The field equations for the free fields (i.e., in the absence of matter) are no more linear, as it used to be in electrodynamics; this may interpreted as a self-interaction of the Yang-Mills field.

Let us consider a quite general setting. X is now a differentiable manifold, that we assume to be compact to have a finite action integral (or, if X is not compact, we assume a suitably fast decay of the fields at infinity). Moreover, we assume that a Riemmannian metric g is defined on X.[12] Let P be a principal bundle on X, with structure group a (say, compact semisimple) Lie group G. The standard action functional for a free gauge theory based on this geometric framework is

$$S(A) = -\frac{1}{2}\int_X \kappa(F, *F)\,\mathrm{vol}(g) \tag{6}$$

where A, a connection on P, is the independent variable, F is the curvature of A, while $*F$ is the Hodge dual of F, and $\mathrm{vol}(g)$ is the measure (volume form) naturally induced on X by the Riemannian metric g. Moreover, κ is the Killing-Cartan form, which is a nondegenerate bilinear form on the Lie algebra of G. The functional S can be regarded as a function on the space $\mathscr{A}$ of all connections on P (the space $\mathscr{A}$ turns out to be an infinite-dimensional affine space). Actually, the action functional (6) is gauge-invariant, in the sense that $S(A) = S(f^*(A))$ for all vertical automorphisms f of P. Therefore, denoting by $\mathscr{G}$ the group of such automorphisms, the action functional descends to a

[11] In the physical literature, the commutator term is multiplied by a dimensioned factor, which plays the role of a self-coupling constant, describing the intensity of the self-interaction of the gauge field.

[12] We assume that g is Riemannian, rather than pseudo-Riemannian. This is more convenient for the mathematical treatment, and has also a physical justification. A transition from the pseudo-Riemannian to the Riemannian signature is indeed necessary to obtain a consistent quantum treatment. In the physical theories this is achieved by a formal manipulation called the "Wick rotation" [50].

functional on the quotient space $\mathscr{B} = \mathscr{A}/\mathscr{G}$, which is called the *orbit space*. This space has in general a nontrivial topology, and is infinite-dimensional; one can do geometry on it by equipping it with a structure of Banach manifold.

The connections A at which the action functional has absolute minima are of particular interest. These are called *instantons*.[13] They may given a simple, direct geometric description: a connection A on P is an absolute minimum of the Yang-Mills functional if and only it its curvature F is *self-dual* with respect to the Hodge duality $*$ given by the Riemannian metric g, i.e., if and only if[14]

$$F = *F\,. \tag{7}$$

Being absolute minima of the action functional, from the physical viewpoint instantons represent the classical vacua of the quantum theory, and therefore play an important role in the theory of fundamental interactions. Their relevance in mathematics is the object of the next section.

4. The Instanton Moduli Space

Let $\mathscr{M} \subset \mathscr{B}$ be the subset of the orbit space $\mathscr{B}$ corresponding to gauge equivalence classes of connections whose curvature is self-dual — i.e., the moduli space $\mathscr{M}$ of instantons. The self-duality equation (7) is a nonlinear first-order PDE which is not elliptic due to the presence of the gauge freedom, i.e., an invariance under gauge transformations. However, at least locally one can fix the gauge, and the resulting equation turns out to be elliptic. Then general elliptic theory, and an application of Kuranishi's linearization technique [23, 19], imply that the space of solutions modulo gauge transformations, i.e., the space $\mathscr{M}$, may be given the structure of a smooth, finite dimensional differentiable manifold. Actually this may not work for some special, "unlucky" Riemannian metrics on X, but it does the job for a generic metric.

Let us give a precise statement. Let $\mathrm{Riem}(X)$ be the space of Riemannian structures on X. It may be given a structure of Banach manifold (see e.g. [23]), hence it is has a natural topology.

THEOREM 4.1. *[4, 23, 19] Let P be a principal G-bundle on a compact Riemannian oriented connected manifold (X, g), where G is a compact semisimple*

---

[13] For a deeper study of instantons the reader may consult [19, 23] for the mathematical theory, and [20] for the physical applications.

[14] One may consider as well *anti-self-dual connections,* namely, connections whose curvature changes sign under Hodge duality, $F = -*F$. We could call these connections *anti-instantons.* Since the Hodge $*$ operator changes sign under the reversal of the orientation, the latter operation swaps instantons with anti-instantons. The two notions are equivalent unless there is some preferred choice of orientation, as in the case of complex manifolds. We shall be vague about this distinction.

Lie group. Let $\mathscr{M}$ be the space of irreducible[15] *instanton connections on P, modulo gauge transformations.*

There is a second-category set $\mathfrak{D} \subset \mathrm{Riem}(X)$ such that, if the Riemannian metric g in X is chosen in $\mathfrak{D}$, the moduli space $\mathscr{M}$ may given the structure of a smooth differentiable manifold of dimension

$$\dim \mathscr{M} = 2\, c_2(\mathrm{Ad}(P)) - (\dim G)(1 - b_1 + b_-) \tag{8}$$

where $\mathrm{Ad}(P)$ is the adjoint bundle of P, c_2 denotes the second Chern class, b_1 is the first Betti number of X, and b_- is the dimension of the vector space of anti-self-dual harmonic 2-forms on X.

Example 4.2. The simplest nontrivial case we may consider is given by the choices $X = S^4$ (with the metric induced by the standard metric in $\mathbb{R}^5$ if we think of S^4 as the unit sphere in $\mathbb{R}^5$), and $G = SU(2)$. In this case we have $c_2(\mathrm{Ad}(P)) = 4c_2(E)$, where E is the rank 2 complex vector bundle associated with P via the natural action of $SU(2)$ on $\mathbb{C}^2$. Moreover, since S^4 has no cohomology in degree 1 and 2, we have $b_1 = b_- = 0$. If we set $k = c_2(E)$, formula (8) becomes $\dim \mathscr{M} = 8k - 3$. Instantons corresponding to various values of k can be described quite explicitly [1]. For $k = 1$ the moduli space has dimension 5, and can be identified with the open unit ball in $\mathbb{R}^5$. We shall denote this moduli space by $\mathscr{M}_1$.

Figure 1 shows the graph of the norm square of the curvature in this case, as a function of two variables on the sphere S^4. Let us imagine this as the graph of this quantity as a function of all 4 variables. (By the way the localized form of this energy density is the origin of the term "instanton", as something which is localized in time). The 4 coordinates of the center λ of the energy distribution in Figure 1, and the width ρ of the latter, defined in some conventional way, can be regarded as 5 spherical coordinates in $\mathscr{M}_1$; the four numbers in λ are angular coordinates, while the radial coordinate in $\mathscr{M}_1$ may be expressed in terms of ρ. By normalization, the height of the instanton is proportional to $1/\rho$. From this simple example we learn that the moduli space $\mathscr{M}$ is in general non compact (and indeed it is never), and that the manifold X appears as the boundary of $\mathscr{M}$. This is a general feature: the boundary of the moduli space $\mathscr{M}$ contains a component homeomorphic to X (the collar theorem, see [23]).

An important property of the moduli space $\mathscr{M}$ is that it is *orientable*. This is proved by calculating its orientation line bundle (the determinant of its tangent bundle) as the determinant of an Atiyah-Singer index bundle on $\mathscr{M}$, and checking that it is trivial [23, 19].

[15] A connection ∇ on a G-bundle P is said to be irreducible if there is no subbundle of P, with structure group a subgroup H of G, over which ∇ induces a connection by restriction. Reducible connections need to be discarded because they produce singularities in the moduli space.

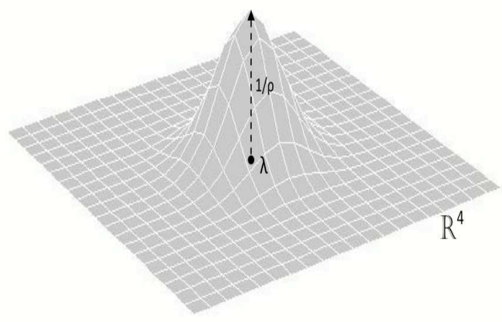

Figure 1: Square norm of the curvature for an $SU(2)$ instanton with $k = 1$.

Moreover, the moduli space carries a *universal bundle with connection* [5]. More precisely, there is a G-bundle $\mathbb{P}$ on $X \times \mathscr{M}$, with a connection $\boldsymbol{\nabla}$, enjoying the following properties: for every $m \in \mathscr{M}$, the restriction $\mathbb{P}_{|X\times\{m\}}$ is isomorphic to P, and $\boldsymbol{\nabla}_{|X\times\{m\}}$, as a connection on P, lies in the gauge equivalence class m. As a preparation for the definition of the Donaldson polynomial invariants, we may use the universal bundle $\mathbb{P}$ to define a map

$$\mu \colon H_2(X, \mathbb{Q}) \to H^2(\mathscr{M}, \mathbb{Q}) .$$

For simplicity, we only consider the case $G = SU(r)$. One defines

$$\mu(\Sigma) = c_2(\mathbb{P})\backslash\Sigma$$

where $\backslash$ is the "slant product" $H^p(X \times \mathscr{M}, \mathbb{Q}) \times H_q(X, \mathbb{Q}) \to H^{p-q}(\mathscr{M}, \mathbb{Q})$ (in our case, $p = 4$ and $q = 2$). Alternatively, by denoting p_1, p_2 the projections of $X \times \mathscr{M}$ onto its factors, we may write

$$\mu(\Sigma) = p_{2*}\left[p_1^*(PD(\Sigma)) \cup c_2(\mathbb{P})\right]$$

where p_{2*} is the Gysin morphism (push-forward) in cohomology, i.e., integration along the fibers of p_2, and PD denotes Poincaré duality. We may now define the Donaldson invariants as polynomials on the space $H_2(X, \mathbb{Q})$ by letting

$$I_d(\Sigma_1, \ldots, \Sigma_d) = \int_{\mathscr{M}} \mu(\Sigma_1) \cup \cdots \cup \mu(\Sigma_d) . \tag{9}$$

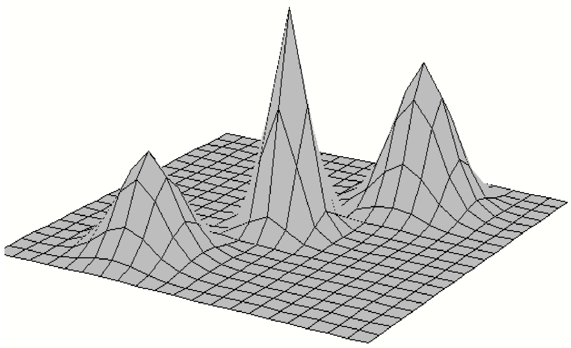

Figure 2: Square norm of the curvature for an $SU(2)$ instanton with $k = 3$.

We are assuming here that $\mathscr{M}$ is smooth, that $\dim \mathscr{M}$ is even, and set $d = \frac{1}{2}\dim \mathscr{M}$. A more important issue is the fact that, for the integral (9) to make sense, we need to compactify the moduli space $\mathscr{M}$. This is accomplished by the so-called *Uhlenbeck-Donaldson compactification.* The naive idea underlying this compactification is the following.

Let us consider the case $X = S^4$, and $G = SU(2)$. The energy density of an instanton of charge k (remember that k, the *instanton charge,* is actually the second Chern class of the bundle E, i.e., $k = c_2(E)$) is shown in Figure 2 (for $k = 3$). This is a kind of nonlinear superposition of k profiles as the one shown in Figure 1 (of course the self-duality equation are nonlinear, so that this is not a linear superposition, unless the "bumps" in Figure 2 are so far apart that the self-interaction is negligible). The $k = 1$ moduli space $\mathscr{M}_1$ (which is 5-dimensional) is compactified by letting $\rho \to 0$; this means that the "bump" in Figure 1 shrinks around its center, and its height becomes infinite. More precisely, the square norm of the curvature approaches a multiple of the Dirac delta function, concentrated at centre of the bump. The compactification boundary is diffeomorphic to S^4 (the "collar" theorem we have already mentioned). For $k > 1$, one can allow one or more bumps to shrink to zero size. So the compactification boundary is stratified, according to the number of bumps that we allow to shrink; moreover, the only information relevant to the description of the bumps that have shrunk is their position, and therefore, if we shrink m of them, we get a point in the symmetric product $\mathrm{Sym}^m(S^4)$. The $k - m$ bumps that have not been shrunk will give a point in $\mathscr{M}_{k-m}$. These configurations, corresponding to an instantons where some "bumps" have been shrunk to zero size, are called *ideal instantons*. Denoting by $\bar{\mathscr{M}}_k$ the compactified space, the resulting stratification is written as

$$\bar{\mathscr{M}}_k = \coprod_{0 \leq m \leq k} \mathscr{M}_{k-m} \times \mathrm{Sym}^m(X) \,. \tag{10}$$

The compactification is done in this way in the general case, even though it involves quite a lot of hard analysis [19, 23]; the stratification formula (10) holds true in the general case.

Donaldson's polynomial invariants are a powerful tool for the study of 4-manifolds, and therefore, also for the study of complex and algebraic manifolds of complex dimension 2. Just to give the flavour of the kind of results one can prove, we cite the following result by Donaldson [18]:

> *A non-singular, projective algebraic surface can be diffeomorphic to the connected sum of two oriented 4-manifolds only if one of them has negative-definite intersection form.*

The reader interested in this subject may consult [24]. From a physical viewpoint, it is interesting to note that Donaldson's polynomial invariants are the correlation functions of a supersymmetric topological Yang-Mills theory [57].

5. Framed Instantons and Framed Sheaves

Often one considers *framed instantons*. In the principal bundle picture, these are pairs (∇, ϕ), where ∇ is a self-dual connection on a principal bundle $P \to X$, and ϕ is a point in the fibre P_x over a fixed point $x \in X$, i.e., a "frame". Correspondingly, one restricts to consider gauge transformations that fix the frame. There are reasons for considering such pairs both in mathematics and physics. In mathematics, their moduli spaces are somehow better behaved, and have a richer mathematical structure; for instance, when $X = S^4$, and $G = SU(r)$, the resulting moduli spaces are hyperkähler [41]. The framing has a meaning also in physical theories: when the instanton moduli space represents the space of classical vacua of a quantized gauge theory, the framing has the meaning of a vacuum expectation value of some fields (technically, the scalar fields in the $N = 2$ vector multiplet).

For $X = S^4$, and $G = SU(r)$, the moduli space of framed instantons can be very nicely parametrized in terms of some linear data, called ADHM data [3, 41], from the initials of Atiyah, Drinfel'd, Hitchin and Manin. One shows that there is a one-to-one correspondence between the set of gauge equivalence classes of framed instantons of instanton charge k, and a space which is obtained by considering a space of linear data (matrices) satisfying some quadratic constraints and a nondegeneracy condition, modulo a free action of the group $U(r)$. In this way the set of gauge equivalence classes is given the structure of a smooth affine variety (over the complex numbers), of complex dimension $2rk$. Following Nakajima's notation, we shall call this moduli space $\mathscr{M}^{\mathrm{reg}}(r,k)$. This space may be constructed also by means of a hyperkähler reduction technique[16] [41], and in this way one shows that it has a hyperkähler structure. One can in a sense complete this moduli space by adding ideal instantons as in the nonframed case; in terms of ADHM data, this means to partially relax the nondegeneracy condition. However, the new moduli space that one obtains, that we denote by $\mathscr{M}_0(r,k)$, is singular. One should note that in this case $\mathscr{M}_0(r,k)$ is not compact.

In a 1993 paper [17], Donaldson showed that there is a one-to-one correspondence between "true" framed instantons on S^4, and framed bundles on $\mathbb{P}^2$, that is, holomorphic vector bundles on $\mathbb{P}^2$ with a trivialization on a fixed (projective) line. This correspondence uses a beautiful construction, called the *Atiyah-Ward correspondence,* that relates instantons on S^4 with a special class of holomorphic vector bundles on $\mathbb{P}^3$ [1, 6], and geometric invariant theory (for references about this theory we refer to Donaldson's paper [17]). $SU(r)$ instantons on S^4, with instanton charge k, correspond to rank r framed holomorphic vector bundles on $\mathbb{P}^2$, with second Chern class k. So the space $\mathscr{M}^{\mathrm{reg}}(r,k)$ is isomorphic to a moduli space $\mathscr{M}^B(r,k)$ parametrizing framed rank r vector bundles on $\mathbb{P}^2$, with second Chern class k. Now, we mentioned the fact that the "completed" moduli space $\mathscr{M}_0(r,k)$, which includes ideal instantons, is singular. We can desingularize it by the usual blowup technique, obtaining a smooth variety $\mathscr{M}(r,k)$ (the same variety can be obtained by hyperkähler reduction, by perturbing the zero-level set of the moment map). It is a very remarkable fact that $\mathscr{M}(r,k)$ is a moduli space itself, parametrizing framed *torsion-free* coherent sheaves on $\mathbb{P}^2$, with rank r and second Chern class k. The space $\mathscr{M}^B(r,k)$ sits inside $\mathscr{M}(r,k)$ as an open, dense subset, and the complement $\mathscr{M}(r,k)\setminus\mathscr{M}^B(r,k)$ is the exceptional divisor of the blowdown morphism $\pi\colon \mathscr{M}(r,k)\to\mathscr{M}_0(r,k)$. In other terms, we have a commutative diagram

$$\begin{tikzcd}
\mathscr{M}^B(r,k) \arrow[r, hook] \arrow[d, leftrightarrow, "\simeq"'] & \mathscr{M}(r,k) \arrow[d, "\pi"] \\
\mathscr{M}^{\mathrm{reg}}(r,k) \arrow[r, hook] & \mathscr{M}_0(r,k)
\end{tikzcd}$$

where the horizontal arrows are open immersions, and π is a blowdown morphism which contracts the closed subset of $\mathscr{M}(r,k)$ corresponding to framed non-locally free, torsion-free sheaves on $\mathbb{P}^2$ to the singular locus of $\mathscr{M}_0(r,k)$.

There is a kind of pattern in these correspondences. In some sense we start from $\mathbb{R}^4$; on the one hand, we compactify it by adding a point and obtaining

[16] A beautiful introduction to the ideas of the hyperkähler reduction techniques is given in [33].

S^4, and then consider framed instantons on S^4. Or, on the other hand, we choose a complex structure on $\mathbb{R}^4$ and add a projective line, obtaining $\mathbb{P}^2$, and consider on it framed holomorphic vector bundles. The two moduli spaces are isomorphic. Other instances of this pattern were studied by King and Buchdahl [37, 14]. In the first case, one starts from $\mathbb{C}^2$ blown up at the origin; adding a point we get $\overline{\mathbb{P}^2}$, i.e., $\mathbb{P}^2$ with the reversed orientation, and adding a projective line we obtain $\widehat{\mathbb{P}}^2$, that is, $\mathbb{P}^2$ blown up at a point. Framed instantons on $\overline{\mathbb{P}^2}$ correspond to framed bundles on $\widehat{\mathbb{P}}^2$. In the second case, we have framed instantons on the connected sum of n copies of $\overline{\mathbb{P}^2}$, and framed bundles on $\mathbb{P}^2$ blown up at n distinct points.

Also these moduli spaces admit ADHM descriptions. For framed bundles on $\mathbb{P}^2$ blown up at one or more points, these are given in the works of King and Buchdahl [37, 15]. An ADHM description for framed torsion-free sheaves on the multiple blowups of $\mathbb{P}^2$ has been given by A. A. Henni [32]; a similar description for framed torsion-free sheaves on Hirzebruch surfaces has been given by C. Rava [51].

A general treatment of moduli spaces of framed sheaves is given in [12]. Relying on the theory of stable framed modules as developed by Huybrechts and Lehn [34, 35], the authors of [12] study the moduli problem for torsion-free sheaves on a projective surface X, that are framed along a divisor $D \subset X$. One considers pairs $(\mathcal{E}, \phi)$, where $\mathcal{E}$ is a torsion-free sheaf on X, and ϕ is a morphism $\phi\colon \mathcal{E} \to \mathcal{F}$, where $\mathcal{F}$ is a fixed sheaf supported by D; one asks that the restriction $\phi_{|D}\colon \mathcal{E}_{|D} \to \mathcal{F}$ is an isomorphism. Under some mild conditions (one assumes D to be smooth, irreducible, big and nef,[17] and $\mathcal{F}$ to be a semistable bundle on D), one can show that a moduli space $\mathscr{M}(c)$ of framed sheaves $(\mathcal{E}, \phi)$ with invariants c exists, is a quasi-projective scheme, and is fine, that is, there is on the product $X \times \mathscr{M}(c)$ a universal framed sheaf. Here $c \in H^\bullet(X, \mathbb{Q})$ is a given set of topological invariants for the sheaf $\mathcal{E}$ (say, rank and first and second Chern class). These moduli spaces are in a sense higher rank generalizations of the Hilbert scheme of points: indeed, when we assume that the sheaves $\mathcal{E}$ have rank one, and $\mathcal{F}$ is the structure sheaf of D, the space $\mathscr{M}(c)$ turns out to be isomorphic to a Hilbert scheme of points of $X \setminus D$ (in particular, $\mathscr{M}(c) \simeq (X \setminus D)^{[n]}$ if $c = (1, 0, n)$).

Moreover, one can characterize the tangent space to the points of $\mathscr{M}(c)$:

$$T_{[(\mathcal{E},\phi)]}\mathscr{M}_X(c) \simeq \mathrm{Ext}^1(\mathcal{E}, \mathcal{E} \otimes \mathcal{O}_X(-D)), \tag{11}$$

and can compute the obstruction to the smoothness of the moduli space. For instance, if the condition $(K_X + D) \cdot D < 0$ holds, where K_X is the canonical

[17] A divisor D in a projective variety X is *nef* (which is an abbreviation for "numerically effective") if D cuts nonnegatively all curves in X, i.e., $D \cdot C \geq 0$ for all curves $C \subset X$. In terms of line bundles, the line bundle $\mathcal{O}_X(D)$ given by the linear equivalence class of D must have nonnegative degree on any curve. D is said to be *big and nef* if in addition $D^2 > 0$ (when X is a surface).

divisor of X, and $\mathcal{F}$ is taken as a trivial bundle, then the moduli space is a smooth quasi-projective variety. A typical case is when X is a rational surface, and D is a rational curve of degree 1 in it. The examples previously mentioned are all of this type.

Moduli spaces of framed sheaves are used very much in the physics literature because they provide desingularizations of moduli spaces of instantons. Very often physics papers refer to instanton moduli spaces, but really they are dealing with the moduli spaces of framed sheaves.

6. Instanton Counting

Moduli spaces of framed sheaves can be nicely studied when the base space X is a toric surface.[18] The toric action lifts to the moduli space of framed bundles, and can be combined with an action of the maximal torus of $GL_r(\mathbb{C})$ on the framing (we are assuming that the framing sheaf is the trivial bundle of rank r). So one has an action of the algebraic torus $(\mathbb{C}^*)^{2+r}$ on the moduli space $\mathscr{M}(c)$. Under suitable assumptions, this action has a finite number of fixed points. Then, considering the equivariant cohomology of $\mathscr{M}(c)$ with respect to this action, one cas use equivariant cohomology techniques [9] to study the geometry of these moduli spaces.

One example of such procedure is the computation of Nekrasov's partition function. This was introduced by Nekrasov [45] as the partition function of $N = 2$ topological super Yang-Mills theory. For a geometric viewpoint, it turns out that the Nekrasov partition function is the integral over the moduli space of the equivariant fundamental class. Actually, the moduli space is not compact (it is only quasi-projective) and therefore, strictly speaking, the integral is not defined. However one can formally apply the localization formula in equivariant cohomology, and the resulting expression is *by definition* Nekrasov's partition formula. This was explicitly computed in [11] for framed sheaves on $\mathbb{P}^2$, with framing provided by the trivial bundle on a line. Nakajima and Yoshioka also computed it for $\hat{\mathbb{P}}^2$, the blow-up of $\mathbb{P}^2$ at a point. A general computation for toric surfaces is given in [26]. There is a very interesting relation between the Nekrasov partition function and the Donaldson polynomials [27, 28].

These computations are done by looking at the fixed points of the toric action on the moduli space. The tangent spaces at the fixed points provide representations of the acting torus, and one can compute the characters of

[18] An n-dimensional toric variety X is an algebraic variety which contains an open dense subset over which the n-dimensional algebraic torus $(\mathbb{C}^*)^n$ acts transitively. The simplest projective example is $\mathbb{P}^n$, where the open dense subset is $\mathbb{C}^n - \{0\}$. The geometry of toric variety admits a relatively simple combinatorial description, which allows one to compute several features of the variety in a very explicit way. For an introduction to toric varieties, and the development of their theory, we refer the reader to [25, 46].

the representations. This allows one to compute the "right-hand side" of the localization formula, and therefore, to compute Nekrasov's partition function. The identification of the fixed points, and the calculations of the characters, is done with some combinatorial computations, using Young tableaux. This is what is meant (at least by mathematicians) by "instanton counting".

The same information allows one to compute the Poincaré polynomial[19] of these moduli spaces. As it was shown in [41], one can introduce a perfect Morse function on the moduli space, whose critical points coincide with the fixed points of the toric action. The index of the Morse function at the critical points can be computed in terms of the characters of the toric action.

By way of example, we show here the computation in the case of Hirzebruch surfaces[20] [13]. We denote by $\mathbb{F}_p$ the p-th Hirzebruch surface $\mathbb{F}_p = \mathbb{P}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(-p))$, which is the projective closure of the total space X_p of the line bundle $\mathcal{O}_{\mathbb{P}^1}(-p)$ on $\mathbb{P}^1$. This may be explicitly described as the divisor in $\mathbb{P}^2 \times \mathbb{P}^1$

$$\mathbb{F}_p = \{([z_0 : z_1 : z_2], [z : w] \in \mathbb{P}^2 \times \mathbb{P}^1 \mid z_1 w^p = z_2 z^p\},$$

Denoting by $f : \mathbb{F}_p \to \mathbb{P}^2$ the projection onto $\mathbb{P}^2$, we let $C_\infty = f^{-1}(l_\infty)$, where l_∞ is the "line at infinity" $z_0 = 0$. The Picard group of $\mathbb{F}_p$ is generated by C_∞ and the fibre F of the projection $\mathbb{F}_p \to \mathbb{P}^1$.

Let $\mathscr{M}^p(r,k,n)$ be the moduli space parametrizing isomorphism classes of pairs $(\mathcal{E}, \phi)$, where

- $\mathcal{E}$ is a torsion-free coherent sheaf on $\mathbb{F}_p$, whose topological invariants are the rank r, the first Chern class $c_1(\mathcal{E}) = kC$, and the discriminant
$$\Delta(\mathcal{E}) = c_2(\mathcal{E}) - \frac{r-1}{2r} c_1^2(\mathcal{E}) = n;$$
- ϕ is a framing on C_∞, i.e., an isomorphism of the restriction of $\mathcal{E}$ to C_∞ with the trivial rank r sheaf on C_∞:
$$\phi \colon \mathcal{E}_{|C_\infty} \xrightarrow{\sim} \mathcal{O}_{C_\infty}^{\oplus r}.$$

The results we have recalled in the previous section imply that the moduli space $\mathscr{M}^p(r,k,n)$, when nonempty, is a smooth quasi-projective variety of dimension $2rn$. Its tangent space at a point $[\mathcal{E}]$ is isomorphic to the vector space

---

[19] The Poincaré polynomial (or series) of a space S whose rational cohomology is finite-dimensional in all degrees is

$$P_t(S) = \sum_{n \geq 0} (\dim H^n(S, \mathbb{Q}))\, t^n .$$

[20] A useful reference about Hirzebruch surfaces is [8].

$\mathrm{Ext}^1(\mathcal{E}, \mathcal{E}(-C_\infty))$. (A nonemptiness criterium for this space is given in [51], see also [12]).

As far as the toric action is concerned, we start by noting that the two-dimensional algebraic torus $\mathbb{C}^* \times \mathbb{C}^*$ acts on $\mathbb{F}_p$ according to

$$([z_0 : z_1 : z_2], [z : w]) \xrightarrow{G_{t_1,t_2}} ([z_0 : t_1^p z_1 : t_2^p z_2], [t_1 z : t_2 w])$$

The divisors C and C_∞ are fixed under this action. Moreover, this action has four fixed points, i.e., $p_1 = ([1 : 0 : 0], [0 : 1]$ and $p_2 = ([1 : 0 : 0], [1 : 0]$ lying on the exceptional line C, and two points lying on the line at infinity C_∞. The invariance of C_∞ implies that the pullback $G^*_{t_1,t_2}$ defines an action on $\mathscr{M}^p(r, k, n)$. Moreover we have an action of the maximal torus of $Gl(r, \mathbb{C})$ on the framing. Altogether, we have an action of the torus $T = (\mathbb{C}^*)^{r+2}$ on $\mathscr{M}^p(r, k, n)$. We study now the fixed point sets for the action of T on $\mathscr{M}^p(r, k, n)$. This is basically the same statement as in [42] (see also [43] and [26]).

Proposition 6.1. *The fixed points of the action of T on $\mathscr{M}^p(r, k, n)$ are sheaves of the type*

$$\mathcal{E} = \bigoplus_{\alpha=1}^{r} \mathcal{I}_\alpha(k_\alpha C) \tag{12}$$

where $\mathcal{I}_\alpha$ is the ideal sheaf of a 0-cycle Z_α supported on $\{p_1\} \cup \{p_2\}$ and $k_1, \dots, k_r$ are integers which sum up to k. Moreover,

$$n = \ell + \frac{p}{2r}\left(r \sum_{\alpha=1}^{r} k_\alpha^2 - k^2\right) = \ell + \frac{p}{2r} \sum_{\alpha<\beta} (k_\alpha - k_\beta)^2 \tag{13}$$

where ℓ is the length of the singularity set of $\mathcal{E}$.

The exact identification of the fixed points is obtained by using some Young tableaux combinatorics [41, 45, 11]. As far as notation is concerned, $|Y|$ will denote the number of boxes in a Young tableau Y. One should attach to each fixed point an r-ple $\{Y_\alpha^{(i)}\}$ of pairs of Young tableaux (so $i = 1, 2$ and $\alpha = 1, \dots, r$). If $Z_\alpha = Z_\alpha^{(1)} \cup Z_\alpha^{(2)}$, where $Z_\alpha^{(i)}$ is supported at p_i, the Young tableau $\{Y_\alpha^{(i)}\}$ is attached to the ideal sheaf $\mathcal{I}_{Z_\alpha^{(i)}}$ as follows: choose local affine coordinates (x, y) around p_i and make a correspondence between the boxes of $\{Y_\alpha^{(i)}\}$ and monomials in x, y as shown in Figure 3. Then $\mathcal{I}_{Z_\alpha^{(i)}}$ is generated by the monomials that lie *outside* the tableau.

Now the identity (13) may be written as

$$n = \sum_\alpha \left(|Y_\alpha^1| + |Y_\alpha^2|\right) + \frac{p}{2r} \sum_{\alpha<\beta} (k_\alpha - k_\beta)^2 .$$

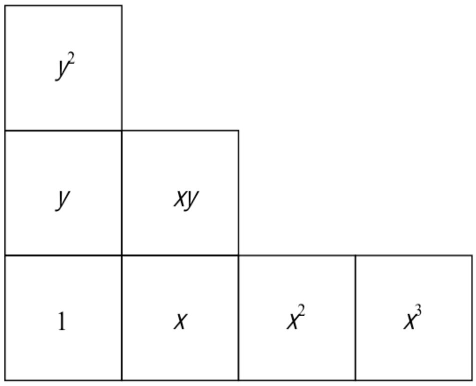

Figure 3: Labelling of the monomials generating the 0-dimensional sheaves $\mathcal{O}_X/\mathcal{I}^{(i)}_{Z_\alpha}$ at the fixed points of the toric action.

The fixed points are in a one-to-one relation with the collections of Young tableaux and strings of integers $k_1, \ldots, k_r$ satisfying this condition together with $\sum_{\alpha=1}^{r} k_\alpha = k$.

We shall show now how to determine the weight decomposition of the toric action on the tangent space to the moduli space at the fixed points, and how to use this to compute the Poincaré polynomial of the moduli spaces $\mathscr{M}^p(r,k,n)$. Actually our computations also make sense for $c_1(\mathcal{E}) = kC$ with $k = m/p$ for integer m, and $p \geq 2$. This can be justified by considering a "stacky compactification" of X_p; instead of adding the divisor C_∞, we add $\tilde{C}_\infty \simeq C_\infty/\mathbb{Z}_p$. One obtains a Deligne-Mumford stack $\mathcal{X}_p$, whose so-called coarse space may be identified with the Hirzebruch surface $\mathbb{F}_p$. Let $\widetilde{\mathscr{M}^p}(r,k,n)$ be the moduli space of torsion-free rank r sheaves $\mathcal{E}$ on $\mathcal{X}_p$, with $c_1(\tilde{\mathcal{E}}) = kC$ and discriminant n, that are framed on $\tilde{C}_\infty$ to the sheaf $\mathcal{O}^{\oplus r}_{\tilde{C}_\infty}$. The fixed points under the torus action are as in Proposition 6.1, except that in this case the k_α's have the form $k_\alpha = m_\alpha/p$, $m_\alpha \in \mathbb{Z}$.

In view of the characterization (11) and of the decomposition (12), the tangent space $T_{(\mathcal{E},\phi)}\widetilde{\mathscr{M}^p}(r,k,n)$ splits as

$$\operatorname{Ext}^1(\mathcal{E}, \mathcal{E}(-C_\infty)) = \bigoplus_{\alpha,\beta} \operatorname{Ext}^1(\mathcal{I}_\alpha(k_\alpha C), \mathcal{I}_\beta(k_\beta C - \tilde{C}_\infty)).$$

The factor $\operatorname{Ext}^1(\mathcal{I}_\alpha(k_\alpha C), \mathcal{I}_\beta(k_\beta C - \tilde{C}_\infty))$ has weight $e_\beta e_\alpha^{-1}$ under the maximal torus of $Gl(r,\mathbb{C})$. So we need only to describe the weight decomposition with respect to the remaining action of $T^2 = \mathbb{C}^* \times \mathbb{C}^*$. In this way we get

$$T_{(E,\phi)}\widetilde{\mathscr{M}}^p(r,k,n) = \sum_{\alpha,\beta=1}^{r} \Big(L_{\alpha,\beta}(t_1,t_2) + t_1^{p(k_\beta - k_\alpha)} N_{\alpha,\beta}^{\vec{Y}_1}(t_1^p, t_2/t_1) + t_2^{p(k_\beta - k_\alpha)} N_{\alpha,\beta}^{\vec{Y}_2}(t_1/t_2, t_2^p) \Big) ,$$

where

$$L_{\alpha,\beta}(t_1,t_2) = e_\beta\, e_\alpha^{-1} \sum_{\substack{i,j\geq 0, i+j-pn_{\alpha\beta}=0 \bmod p,\\ i+j\leq p(n_{\alpha\beta}-1)}} t_1^{-i} t_2^{-j}$$

for $n_{\alpha\beta} > 0$,

$$L_{\alpha,\beta}(t_1,t_2) = e_\beta\, e_\alpha^{-1} \sum_{\substack{i,j\geq 0,\ i+j+2+pn_{\alpha\beta}=0 \bmod p,\\ i+j\leq -pn_{\alpha\beta}-2}} t_1^{i+1} t_2^{j+1}$$

for $n_{\alpha\beta} \leq 0$, and

$$N_{\alpha,\beta}^{\vec{Y}}(t_1,t_2) = e_\beta e_\alpha^{-1} \times \left\{ \sum_{s\in Y_\alpha} \left(t_1^{-l_{Y_\beta}(s)} t_2^{1+a_{Y_\alpha}(s)} \right) + \sum_{s\in Y_\beta} \left(t_1^{1+l_{Y_\alpha}(s)} t_2^{-a_{Y_\beta}(s)} \right) \right\} ,$$

a well known expression for the $\mathbb{P}^2$ case, first introduced in [22]. Here $\vec{Y}$ denotes an r-ple of Young tableaux, while for a given box s in the tableau Y_α, the symbols $a_{Y_\alpha}(s)$ and $l_{Y_\alpha}(s)$ denote, respectively, the "arm" and "leg" of the box s in the tableau Y_α, that is, the number of boxes above and on the right to that box (see Figure 4).

From these data one can compute the desired Poincaré polynomial (see [13] for details).

THEOREM 6.2. *The Poincaré polynomial of* $\widetilde{\mathscr{M}}^p(r,k,n)$ *is*

$$P_t(\widetilde{\mathscr{M}}^p(r,k,n)) = \sum_{\substack{\text{fixed}\\ \text{points}}} \prod_{\alpha=1}^{r} t^{2(|Y_\alpha|-l(Y_\alpha))} \prod_{i=1}^{\infty} \frac{t^{2(m_i^{(\alpha)}+1)}-1}{t^2-1} \prod_{\alpha<\beta} t^{2(l'_{\alpha,\beta}+|Y_\alpha|+|Y_\beta|-n'_{\alpha,\beta})} .$$

Here $m_i^{(\alpha)}$ is the number of columns in Y_α whose length is i, and

$$l'_{\alpha,\beta} = \begin{cases} \frac{1}{2}[n_{\alpha\beta}]\,(p[n_{\alpha\beta}]+2-p)+p[n_{\alpha\beta}]\{n_{\alpha\beta}\} & \text{if} \quad n_{\alpha\beta}\geq 0, \\ \frac{1}{2}[n_{\beta\alpha}]\,(p[n_{\beta\alpha}]+2-p)+p[n_{\beta\alpha}]\{n_{\beta\alpha}\} - \delta_{p\{n_{\beta\alpha}\},0} & \text{otherwise.} \end{cases}$$

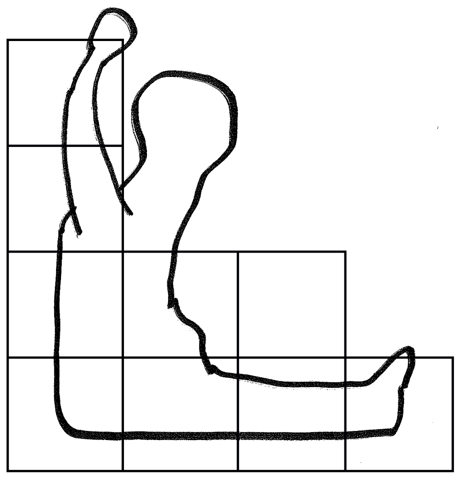

Figure 4: How to remember the meaning of "arm" and "leg" in a Young tableau.

$$n'_{\alpha,\beta} = \begin{cases} \text{number of columns of } Y_\alpha \text{ that are longer} \\ \qquad \text{than } k_\alpha - k_\beta \text{ if } k_\alpha - k_\beta \geq 0, \\ \text{number of columns of } Y_\beta \text{ that are longer} \\ \qquad \text{than } k_\beta - k_\alpha - 1 \text{ otherwise.} \end{cases}$$

Setting $t = -1$ in this formula we obtain a compact expression for the generating function of the Euler characteristics of the moduli spaces $\widetilde{\mathscr{M}}^p(r,k,n)$

$$\sum_{k,n} P_{-1}(\widetilde{\mathscr{M}}^p(r,k,n))\, q^{n+\frac{pk^2}{2r}}\, z^k = \left(\frac{\theta_3(\frac{v}{p}|\frac{\tau}{p})}{\hat{\eta}(\tau)^2} \right)^r$$

where $q = \mathrm{e}^{2\pi i\tau}$ and $z = e^{2\pi iv}$. We have used formulas for the quasi-modular functions

$$\theta_3(v|\tau) = \sum_{n\in\mathbb{Z}} q^{\frac{1}{2}n^2} \mathrm{e}^{2\pi ivn}, \qquad \hat{\eta}(\tau) = \prod_{l=1}^{\infty}(1-q^l)\,.$$

References

[1] M.F. Atiyah, *The geometry of Yang-Mills fields*, Scuola Normale Superiore, Pisa, Italy (1979).

[2] M.F. Atiyah, *On the work of Simon Donaldson*, in the *Proceedings of the International Congress of Mathematicians*, volumes 1 and 2 (Berkeley, U.S.A. (1986)), American Mathematical Society, Providence, (1987), pag. 3–6.

[3] M.F. Atiyah, N.J. Hitchin, V.G. Drinfel'd and Yu. I. Manin, *Construction of instantons*, Phys. Lett. **A 65** (1978), 185–187.
[4] M.F. Atiyah, N.J. Hitchin and I.M. Singer, *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. London Ser. A **362** (1978), 425–461.
[5] M.F. Atiyah and I.M. Singer, *Dirac operators coupled to vector potentials*, Proc. Nat. Acad. Sci. U.S.A. **81** (1984), 2597–2600.
[6] M.F. Atiyah and R.S. Ward, *Instantons and algebraic geometry*, Commun. Math. Phys. **55** (1977), 117–124.
[7] R.M. Barnett, H. Mühry and H.R. Quinn, *The charm of strange quarks*, Springer, Berlin (2000).
[8] W.P. Barth, K. Hulek, C.A. M. Peters, and A. Van de Ven, *Compact complex surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete volume 4, Springer, Berlin (2004).
[9] N. Berline, E. Getzler and M. Vergne, *Heat kernels and Dirac operators*, Grundlehren der Mathematischen Wissenschaften volume 298, Springer, Berlin (1992).
[10] D. Bleecker, *Gauge theory and variational principles*, Addison-Wesley Publishing Co., Reading (1981).
[11] U. Bruzzo, F. Fucito, J.F. Morales and A. Tanzini, *Multi-instanton calculus and equivariant cohomology*, JHEP 05(2003)054. 24 (electronic).
[12] U. Bruzzo and D. Markushevich, *Moduli of framed sheaves on projective surfaces*, arXiv:0906.1436, to appear in Doc. Math.
[13] U. Bruzzo, R. Poghossian and A. Tanzini, *Poincaré polynomial of moduli spaces of framed sheaves on (stacky) Hirzebruch surfaces*, arXiv:0909.1458.
[14] N. Buchdahl, *Instantons on* $n\overline{\mathbb{CP}}_2$, J. Diff. Geom. **37** (1993), 669–687.
[15] N. Buchdahl, *Blowups and gauge fields*, Pacific J. Math. **196** (2000), 69—111.
[16] S.K. Donaldson, *Self-dual connections and the topology of smooth 4-manifolds*, Bull. Amer. Math. Soc. **8** (1983), 81–83.
[17] S.K. Donaldson, *Instantons and geometric invariant theory*, Comm. Math. Phys. **93** (1984), 453–460.
[18] S.K. Donaldson, *Polynomial invariants for smooth four-manifolds*, Topology **29** (1990), 257–315.
[19] S.K. Donaldson and P.B. Kronheimer, *The geometry of four-manifolds*, Oxford University Press, Oxford (1990).
[20] N. Dorey, T.J. Hollowood, V.V. Khoze and M.P. Mattis, *The calculus of many instantons*, Phys. Rept. **371** (2002), 231–459.
[21] R. Feynman, *QED: The strange theory of light and matter*, Princeton University Press, Princeton (1985).
[22] R. Flume and R. Poghossian, *An algorithm for the microscopic evaluation of the coefficients of the Seiberg-Witten prepotential*, Int. J. Mod. Phys. **A 18** (2003), 2541–2563.
[23] D.S. Freed and K.K. Uhlenbeck, *Instantons and four-manifolds*, 2^{nd} edition. Springer, Berlin (1991).
[24] R. Friedman and J.W. Morgan, *Smooth four-manifolds and complex surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete volume 27, Springer, Berlin (1994).

[25] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies volume 131, Princeton University Press, Princeton (1993).
[26] E. Gasparim and C.-C.M. Liu, *The Nekrasov conjecture for toric surfaces*, Comm. Math. Phys. **293** (2010), 661–700.
[27] L. Göttsche, H. Nakajima and K. Yoshioka, *Instanton counting and Donaldson invariants*, J. Diff. Geom. **80** (2008), 343–390.
[28] L. Göttsche, H. Nakajima and K. Yoshioka, *K-theoretic Donaldson invariants via instanton counting*, Pure Appl. Math. Q. **5** (2009), 1029–1111.
[29] W. Greiner and B. Müller, *Gauge theory of weak interactions*, Springer, Berlin (2000).
[30] W. Greiner and A. Schäfer, *Quantum chromodynamics*, Springer, Berline (1994).
[31] W. Heisenberg, *Über den Bau der Atomkerne*, Z. Phys. **77** (1932), 1–11; translated in D.M. Brink, *Nuclear forces*, Pergamon Press, Oxford (1965), pag. 145–154.
[32] A.A. Henni, *Monads for torsion-free sheaves on multi-blowups of the projective plane*, arXiv:0903.3190.
[33] N.J. Hitchin, A. Karlhede, U. Lindström and M. Roček, *Hyper-Kähler metrics and supersymmetry*, Comm. Math. Phys. **108** (1987), 535–589.
[34] D. Huybrechts and M. Lehn, *Framed modules and their moduli*, Int. J. Math. **6** (1995), 297–324.
[35] D. Huybrechts and M. Lehn, *Stable pairs on curves and surfaces*, J. Alg. Geom.**4** (1995), 67–104.
[36] J.D. Jackson, *Classical electrodynamics*, 2$^{\mathrm{nd}}$ edition, John Wiley & Sons Inc., New York (1975).
[37] A. King, *Instantons and holomorphic bundles on the blown-up plane*, Ph.D. thesis, Oxford University, Oxford (1989).
[38] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, volume I*, John Wiley & Sons Inc., New York (1996).
[39] L.D. Landau and E.M. Lifshitz, *The classical theory of fields*, Course of Theoretical Physics volume 2, Pergamon Press, Oxford (1962).
[40] M. Mariño, *La teoria delle stringhe,* in C. Bartocci and P. Odifreddi, *La matematica, volume 4: "Pensare il mondo"*, Einaudi, Torino (2010), pages 641–686.
[41] H. Nakajima, *Lectures on Hilbert schemes of points on surfaces*, University Lecture Series volume 18, American Mathematical Society, Providence (1999).
[42] H. Nakajima and K. Yoshioka, *Instanton counting on blowup. I. 4-dimensional pure gauge theory*, Inv. Math. **162** (2005), 313–355.
[43] H. Nakajima and K. Yoshioka, *Lectures on instanton counting*, in *Algebraic structures and moduli spaces*, CRM Proc. Lecture Notes volume 38, American Mathematical Society, Providence (2004), pages 31–101.
[44] C. Nash and S. Sen, *Topology and geometry for physicists*, Academic Press, London (1983).
[45] N.A. Nekrasov, *Seiberg-Witten prepotential from instanton counting*, Adv. Theor. Math. Phys. **7** (2003), 831–864.
[46] T. Oda, *Convex bodies and algebraic geometry. An introduction to the theory of*

toric varieties, Ergebnisse der Mathematik und ihrer Grenzgebiete volume 15, Springer, Berlin (1988).
[47] R. Oerter, *The theory of almost everything: the Standard Model, the unsung triumph of modern physics*, Plume, New York (2006).
[48] L. O'Raifeartaigh, *The dawning of gauge theory*, Princeton University Press, Princeton (1997).
[49] W. Pauli, *Meson-nucleon interaction and differential geometry*, letter to A. Pais (1953), printed in [48], pages 171–181.
[50] P. Ramond, *Field theory: a modern primer*, Westview Press, Boulder (1997).
[51] C.L.S. Rava, *ADHM data for framed sheaves on Hirzebruch surfaces*, Ph.D. thesis, SISSA, Trieste, Italy (2010).
[52] D.W. Sciama, *On a geometrical theory of the electromagnetic field*, Nuovo Cim. **8** (1958), 417–431.
[53] R. Utiyama, *Invariant theoretical interpretation of interaction*, Phys. Rev. **101** (1956), 1597–1607; also in [48], pages 213–239.
[54] H. Weyl, *Gravitation und Elektrizität*, Sitzungsber. Kön. Preuss. Akad. Wiss. Berlin **465** (1918), 465–480 (in German, translated into English in [48], 24–37).
[55] E. Wigner, *On the consequences of the symmetry of the nuclear Hamiltonian on the spectroscopy of nuclei*, Physical Review **51** (1937), 106–119.
[56] E. Wigner, *The unreasonable effectiveness of mathematics in the natural sciences*, Comm. Pure Appl. Math. **13** (1960), 1–14.
[57] E. Witten, *Topological quantum field theory*, Comm. Math. Phys. **117** (1988), 353–386.
[58] C.N. Yang and R.L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Physical Rev. **96** (1954), 191–195 (also in [48], 186–196).

Author's address:

Ugo Bruzzo[21]
Department of Mathematics
University of Pennsylvania, David Rittenhouse Laboratory
209 S 33rd Street, Philadelphia, PA 19104, U.S.A.
and
Istituto Nazionale di Fisica Nucleare, Sezione di Trieste
via A. Valerio 2, 34127 Trieste, Italia
E-mail: bruzzo@sissa.it

Received October 20, 2010
Revised November 2, 2010

[21] On leave of absence from Scuola Internazionale Superiore di Studi Avanzati, Trieste.