Canad. Math. Bull. Vol. **28** (2), 1985

ON SOME RECENT INTERACTIONS BETWEEN MATHEMATICS AND PHYSICS

BY
RAOUL BOTT

Introduction. It gives me quite extraordinary pleasure to have been asked to deliver the Jeffrey-Williams lecture of the Canadian Mathematical Society. The reasons are manifold. First of all Canada was my home for the most formative years of my life — from 16 to 23 — and was in fact the first country willing to take me on as an adopted son. I was of course born in Budapest, but in Europe the geographical accidents of birth are not taken seriously, rather I inherited my father's status and so managed to become stateless "by induction" so to speak.

But quite apart from my debt to Canada, my debt to Lloyd Williams, after whom this lecture is named in part, is even greater. He was my first calculus teacher at McGill in 1941 and we all delighted in him from the moment he entered the class resplendent in a chalk covered academic gown. However what all of us, and especially those from foreign shores, remember most about Professor Williams was his humanity and unhesitating generosity. Without his encouragement and friendly advice I do not think that I could ever have carried out my resolve to switch from Engineering to Mathematics.

In turning now to my subject, please forgive the all encompassing title; it was, as usual, arrived at under pressure of the telephone. A more appropriate one, and certainly a more truthful one might have been: "A topologist marvels at Physics".

What there is to marvel at from the perspective of the geometer and topologist is, that the equations which the physicists after many "supple confusions" arrive at for their description of the fundamental particles, make such good sense in topology and geometry — and are indeed so inevitable that it is a scandal that the mathematicians had not studied them in their own right years ago. Of course we mathematicians have missed the boat before, and were saved from doing so most spectacularly in the equations of general relativity only by the last minute intervention of David Hilbert. So let me start my story at that point, or rather a little earlier with a lightning review of classical dynamics from the ultimate "free enterprise" point of view embodied in Hamilton's Principle of Least Action.

Received by the editors May 29, 1984.

This paper is based on the Jeffrey-Williams lecture given at the meeting of the Canadian Mathematical Society, Robson Square Media Center, Vancouver, B.C., on June 2, 1983, upon the invitation of the Research Committee of the CMS.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

1. **Hamilton's Principle**. In modern terminology the mathematical model for time-independent classical mechanics which goes back to Lagrange and Hamilton, is the following one: the configuration of a system is modeled by a manifold M of dimension n, and the forces which act on the constituents are summarized by a function $L(q, \dot{q})$ which, in the modern parlance, is defined on the "tangent bundle" TM of M. This "Lagrangian density" then determines a function $S(\mu)$ — called the action — on the space $\Omega_T(M)$, of piecewise smooth paths on M,

$$\Omega_T(M) = \{\mu | : [0, T] \rightarrow M\} \tag{1.1}$$

by the formula:

$$S(\mu) = \int_0^T L(\mu, \dot{\mu})\, \mathrm{d}t, \tag{1.2}$$

and Hamilton's Principle asserts that amongst all paths in the configuration space, joining q_1 to q_2 on M in time T the acutal motion will take place only along those paths μ, subject to $\mu(0) = q_1$, $\mu(T) = q_2$, which are extremals of S under small perturbations in μ. Indeed this principle then immediately leads to the Lagrange equations governing the motion; that is the extremal condition:

$$\frac{\delta S}{\delta \mu} = 0 \tag{1.3}$$

formally implies that:

$$\frac{\delta S}{\delta \dot{q}_i} = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\delta S}{\delta q_i}, \tag{1.4}$$

along the extremal μ, when q_i and $\dot{q}_i$ are interpreted as local coordinates on TM in the usual manner.

Now the paradigm of writing down a Lagrangian for the equations of Physics is still very much with us, and so that it stands to reason that the more fundamental the phenomenon is the more intrinsic and natural the Lagrangian governing it should be, and from this point of view the Einstein gravitational equations are nearly inevitable — once one has had the inspiration to geometrize the whole concept of gravity. Indeed in this frame work space time is modeled by a four-manifold M^4, of events, and what is sought is a Riemann metric:

$$\mathrm{d}s^2 = g_{ij}\, \mathrm{d}x^i \otimes \mathrm{d}x^j, \tag{1.5}$$

of signature $+ + + -$ on M^4, which describes gravitational phenomena in the sense that the behaviour of small test particles is to move along (time like) geodesics in this structure.

If one applies the "Lagrangian philosophy" to this problem in free space one is led to seek a natural or intrinsic Lagrangian density depending on the metric $\mathrm{d}s^2$, and for those familiar with the "pure mathematics" of the calculus on manifolds the Einstein-Hilbert Lagrangian is then certainly the leading candidate.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

Recall first that in extending the calculus from $\mathbf{R}^n$ to manifolds — which are after all only the shapes that arise by gluing $\mathbf{R}^n$'s smoothly together — one must take care to distill out of the many derivatives with which the calculus provides us those combinations which are in fact independent of their local descriptions. And of course at the present stage of the art we understand these phenomena rather well. The smooth functions – $F(M)$ – on M, or 0-forms $\Omega^0(M)$, as we topologists call them, make perfectly good sense on M, and so do their total derivatives, but these are already "twisted functions" or "tensors of type (1, 0)" on M. The topologists call these the 1-forms "$\Omega^1(M)$" on M.

In local coordinates a 1-form is written as $\omega = \Sigma\, a_i\, \mathrm{d}x^i$ by the mathematicians and simply as $\{a_i\}$ by the physicists, and the functions a_i are the local representatives of the form ω. Thus in local coordinates the total derivative of a function f is given by

$$\mathrm{d}f = \sum \frac{\delta f}{\delta x^i}\, \mathrm{d}x^i. \tag{1.6}$$

Combined with the fundamental constructions of linear algebra (i.e. the dual space and tensor product) these 1-forms then generate the so called tensor-algebra $T^{p,q}(M)$ of tensors fields of type (p, q) over M, so that, for instance, a Riemann structure $\mathrm{d}s^2$ on M is technically a tensor-field in $T^{2,0}(M)$ — usually denoted by $g = \{g_{ij}\}$. Now the primary tensorial invariants of such a $\{g_{ij}\}$ is the "Riemann curvature tensor" R^i_{jkl}, out of which one generates an invariant function of the $g_{ij}: R(g) = R^i_{jij}$ by contraction. This "scalar curvature" of g combined with the "natural volume" $\mathrm{vol}(g)$ determined by g then collaborates to yield the simplest possible candidate for a natural Lagrangian density depending on $\mathrm{d}s^2$, with corresponding action $S(g)$ given by the integral:

$$S(g) = \int_M R(g)\, \mathrm{vol}(g). \tag{1.7}$$

And indeed the formal consequences of the variational equations

$$\frac{\delta S}{\delta g} = 0, \tag{1.8}$$

are precisely the Einstein equations for the gravitational potential $\{g_{ij}\}$ in free space:

$$R_{ij} - \frac{1}{2} g_{ij} R = 0 \qquad R_{ij} = R^k_{ijk}. \tag{1.9}$$

Let us next explore how the equations of electromagnetism fit into this geometric and topological frame work, for, properly understood, their generalization to our ultimate goal, the Yang-Mills equations, is again — alas belatedly — immediate.

2. **The Maxwell equations**. In free space the electromagnetic field has components $\boldsymbol{E} = \{E_x, E_y, E_z\}$ and $\boldsymbol{B} = \{B_x, B_y, B_z\}$ in terms of which the Maxwell equations take the form:

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$(2.1)\qquad \begin{array}{ll} \nabla \times \boldsymbol{E} = \dfrac{\delta \boldsymbol{B}}{\delta t} & \nabla \times \boldsymbol{B} = -\dfrac{\delta \boldsymbol{E}}{\delta t} \\[2ex] \operatorname{div} \boldsymbol{E} = 0 & \operatorname{div} \boldsymbol{B} = 0. \end{array}$$

This sort of no-nonsense description has many virtues but it completely obscures the geometro-topological aspects of these equations. These become apparent only when one combines the E and B fields into the *alternating* tensor: $F = \{F_{\mu\nu}\}$ in space time $\mathbf{R}^4$, by setting

$$(2.2)\qquad F = (E_x\, dx + E_y\, dy + E_z\, dz)\, dt + B_x\, dydz - B_y\, dxdz + B_z\, dxdy.$$

At this stage the alternating nature of F, that is: $F_{uv} = -F_{vu}$ is really only motivated by the fact that we need to accommodate precisely six local components, but to topologists this skew symmetry is crucial, for amongst all tensor-fields on M it is precisely the "*contravariant alternating*" ones which are most naturally sensitive to the *global aspects* of M. Indeed we call the contravariant tensors with p-(lower) index the *p forms* on M, denote them by $\Omega^p(M)$ and link them by the natural extension of the total derivative "d" in (1.6) to form the "de Rham complex" of M:

$$(2.3)\qquad \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \rightarrow \ldots \xrightarrow{d} \Omega^q(M) \xrightarrow{d} \ldots$$

In a quite precise sense this complex constitutes the only "God given" set of differential equations between the first order tensor-fields on a manifold, and their existence can be traced back to the fact that the second order partials of a function commute:

$$(2.4)\qquad \frac{\delta}{\delta x^j}\left(\frac{\delta f}{\delta x^i}\right) = \frac{\delta}{\delta x^i}\left(\frac{\delta f}{\delta x^j}\right).$$

This symmetry allows one to extend the definition $df = \Sigma\, (\delta f/\delta x^i)\, dx^i$ to $d\,(\Sigma\, a_j\, dx^j) = \Sigma\, (\delta a_j/\delta x^i)\, dx^i\, dx^j$ provided one uses the anticommutative calculus of Grassmann for the dx^i's:

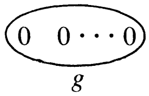
$$(2.5)\qquad dx^i\, dx^j + dx^j\, dx^i = 0.$$

It is then immediate that $d^2 f = \Sigma\, (\delta^2 f/\delta x^i \delta x^j)\, dx^i\, dx^j = 0$, and this property extends to all the d's in (2.3) so that one is led to introduce the spaces:

$$(2.6)\qquad H^q(M) = \{\operatorname{Ker}\,(d|\Omega^q)/d\Omega^{q-1}\},$$

of solutions to the equation $d\omega = 0$ in Ω^q, modulo the "trivial" solutions which are already d of something in Ω^{q-1}. These vector spaces, called the *de Rham Cohomology* of M, turn out to be finite dimensional for — say — compact manifolds, and the dimension of $H^q(M)$ — called the qth Betti number of M — is some sort of a measure on the number of "holes of dimension q" in M. Thus $\dim H^1(M) = 2g$ for the "g holed torus" below,

(2.7)



https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

while H^0 and H^2 have dimension 1 for this example.

In this context then, the electromagnetic field F is a 2-form in $\Omega^2(M)$ and we find much to our satisfaction that the "curl" part of the Maxwell-equations is expressed by:

$$dF = 0, \tag{2.8}$$

and thus makes sense on all manifolds. To obtain the remaining "divergence" part we must return to the geometry of the space, or more generally to a 4-manifold endowed with a metric tensor g_{ij} as in the previous section. We already remarked that this g gives rise to a natural volume vol(g) in $\Omega^4(M)$ and using it as well as the natural identification which the g_{ij} furnish between dual spaces, one can introduce a natural *global inner product among* all the tensor-fields of M, and in particular into the space of q-forms Ω^q on M. Thus for example for two 1-forms ω and η one has:

$$(\omega, \eta) = \int_M \omega_\alpha \eta_\beta g^{\alpha\beta} \operatorname{vol}(g). \tag{2.9}$$

Once this is understood, we see by the usual sort of integration by parts argument that the de Rham d now determines a unique adjoint operator d*, going from Ω^q to Ω^{q-1}, characterized by the adjoint property

$$(d\alpha, \beta) = (\alpha, d^*\beta). \tag{2.10}$$

In particular then the equation $d\omega = 0$ now has a natural companion $d^*\omega = 0$ and we find, again with considerable satisfaction, that in the Lorentzian flat space-time:

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2, \tag{2.11}$$

the equations

$$dF = 0 \text{ and } d^*F = 0 \tag{2.12}$$

precisely reproduce the Maxwell equations (2.1).

By the way in the world of pure mathematics these "Maxwell-equations" become the "Hodge equations" and are interesting for a quite different reason than in classical physics. In the Hodge theory one is dealing with a compact, oriented manifold, and endows it with a *positive definite* geometry g_{ij}. The forms $\omega \in \Omega^q(M)$ satisfying

$$d\omega = 0 \text{ and } d^*\omega = 0 \tag{2.13}$$

are called the *harmonic* forms, are denoted by $\mathcal{H}^q(M)$ and the famous "Hodge Theorem" asserts that the natural inclusion

$$\mathcal{H}^q(M) \hookrightarrow H^q(M) \tag{2.14}$$

is an *isomorphism onto*.

Thus in this context, the Riemann structure serves to specify canonical representatives of each cohomology class — the "harmonic" ones. This terminology arises from the fact that if $\Box$ denotes the operator

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$\square = dd^* + d^*d, \tag{2.15}$$

then $\square$ maps Ω^q into Ω^q, and the harmonic forms are precisely those which are annihilated by $\square$. Indeed

$$\square\varphi = 0 \Rightarrow (\square\varphi, \varphi) = (d^*\varphi, d^*\varphi) + (d\varphi, d\varphi) = 0. \tag{2.16}$$

Finally in flat space $\mathbf{R}^n$, $\square$ reduces to the Laplacian

$$\square = \sum_{i=1}^{n} \frac{\delta^2}{\delta x^{i^2}} \tag{2.17}$$

on the components of Ω^q, so that the term "harmonic" for its null-space is appropriate. Note that in any case $\square$ is an *elliptic* operator so that its null-space is finite-dimensional on compact manifolds. Hence the Hodge Theory brings a differential-equation rationale to the finite dimensionality of the de Rham Theory. It also serves to uncover a "hidden symmetry" in the de Rham Cohomology of compact oriented manifolds. This is the *Poincaré Duality*:

$$\dim H^q(M) \simeq \dim H^{n-q}(M), \tag{2.18}$$

which becomes apparent only in the Hodge context. Indeed, the inner product on $\Omega^q(M)$ is related to the exterior product of forms by a pointwise operator $*$: $\Omega^q \to \Omega^{n-q}$ which is characterized by

$$(\omega, \theta) = \int_M \omega \wedge *\theta. \tag{2.19}$$

It follows then that $(*)^2 = \pm 1$, and

$$d^* = \pm *d* \tag{2.20}$$

the signs depending on the dimension of the forms being considered. In any case though, it is now clear that this $*$ induces an isomorphism of the *harmonic forms* of deg q with those of $\dim(n - q)$. Q.E.D.

So much for a short excursion into cohomology and Hodge theory. Let us return now to the Maxwell equations

$$dF = 0 \qquad d{*}F = 0, \qquad F \epsilon \Omega^2(M) \tag{2.21}$$

which, in view of our above formula (2.20) — valid for all nonsingular g_{ij}'s positive definite or not — can be recast in the form:

$$dF = 0 \qquad d{*}F = 0. \tag{2.22}$$

Two natural questions now arise: (1) How do these equations fit into the Lagrangian formulation, and (2) granted that (2.22) most probably describe the electromagnetic field in a fixed gravitational "back-ground field" g_{ij}, how does the F influence the g field, or — as the physicists say, how does one *couple the F with the g*.

In most physics texts the first question is attacked by introducing an *electromagnetic*

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

potential $A = (\varphi, A_x, A_y, A_z)$, and in our notation this amounts to starting with the "ansatz":

$$dA = F, \tag{2.23}$$

where $A \in \Omega^1(M)$ is a 1-form on M. Note that this ansatz, immediately disposes of the first equation $dF = 0$, in view of the identity $d^2 = 0$. It also immediately yields a "variational principle" for the Maxwell equations.

Indeed, let us think of the *length* (F, F) of the electromagnetic field $F \in \Omega^2(M)$ as a function of the *potential* A.

Then formally

$$\delta(F, F) = \delta(dA, dA) = \delta(A, d^*dA) = 2(\delta A, d^*F), \tag{2.24}$$

so that A extremal $\Leftrightarrow$ $d^*F = 0$, with $F = dA$.

This procedure certainly brings the Maxwell Equations into the Lagrangian fold, but at a two-fold price. First of all, the equation $dA = F$ silently implies that F is *cohomologous to zero in* $H^2(M)$, and secondly, as we are now thinking of F as a function of A, builds a degeneracy into the action

$$S(A) = (dA, dA), \tag{2.25}$$

There are many A's describing the same F. In fact $S(A)$ is clearly invariant under translation by any "closed form" $\alpha \in \Omega^1(M)$ (that is one with $d\alpha = 0$).

$$S(A + \alpha) = (dA + d\alpha, dA + d\alpha) = S(A). \tag{2.26}$$

If we assume — as physics texts tend to — that $H^1(M) = 0$ then $d\alpha = 0$ implies that there is a function $\varphi \in \Omega^0(M)$ on M, with $d\varphi = \alpha$. In this instance the functions $\Omega^0(M)$ on M, are therefore seen to play the role of what we will later call the "group of gauge transformations" for the theory. In the physics literature this group is often referred to as the "*local gauge group*", and ultimately it is the nontrivial topology of this group which has been of interest to the physicists in the last ten years.

However, the first defect of the ansatz $dA = F$ troubled Dirac already in the 30's and led him to the considerations which one now lumps under the heading: "Dirac monopoles and the quantization of charge". His considerations were of course "Quantum Mechanical" so that to explain them, a tentative first step into these murky waters is now indicated.

The Feynman paradigm for "quantizing" a classical time independent system with Lagrangian density $\mathcal{L}(q, \dot{q})$ is as follows: If M is the classical configuration space of our Lagrangian system, then the "rays" in an appropriate Hilbert space $\mathcal{H}(M)$ of *complex valued* functions on M play the role of the "states" of the quantum system. Furthermore the time evolution through a time interval T of the system is represented by a unitary operator U_T in $\mathcal{H}(M)$ whose "kernel" (in the sense of an integral operator) i.e.:

$$U_T\varphi(q) = \int U_T(q, q')\varphi(q')\, dq, \tag{2.27}$$

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

is "given" by the formula:

$$(2.28) \qquad U_T(q,q') = \int_{\Omega(q,q',T)} e^{(2\pi i/h)S(\mu)}\mathcal{D}(\mu)$$

Here μ ranges over the space $\Omega(q,q';T)$ of paths from q to q' in M parametrized by the time interval $[0,T]$, $S(\mu)$ is the classical action of μ computed via the Lagrange density under consideration:

$$(2.29) \qquad S(\mu) = \int_0^T L(\mu,\dot{\mu})\,dt$$

and $\mathcal{D}(\mu)$ is an appropriate "measure" on the space $\Omega(q,q',T)$.

The underlying philosophy of this procedure is first of all that whereas in classical theory the motion proceeds only along extremals of $S(\mu)$, in quantum theory *all* paths contribute to the time evolution, and although the mathematical difficulties with making the equation (2.27) meaningful are legion, this formula does furnish one not only with a lot of intuitive insight, but also with an essentially well defined "semiclassical perturbation theory" in terms of the small paramater h.

The guiding principle for this development is in turn the finite dimensional principle of "*stationary phase*": that is, if we wish to estimate the small h behaviour of an integral

$$(2.30) \qquad \int_{[0,1]} e^{(2\pi i/h)f(x)}\,dx, \qquad x \in [0,1] \text{ say,}$$

then the leading contributions come from the places where $f'(x) = 0$. Elsewhere the function oscillates so much that it essentially cancels itself out. More precisely one has an asymptotic development for h small and > 0 of the form:

$$(2.31) \quad \int e^{(2\pi i/h)f(x)}\,dx \sim \sum e^{(2\pi i/h)f(p)}\{ih/f''(p)\}^{1/2}\{1 + a_1h^1 + a_2h^2 + \ldots\}$$

where the sum is taken over the critical points $\{p\}$ of f and the a_i are more and more complicated but explicitly computable expressions in the derivatives of f at the critical points.

In extending this finite dimensional principle to infinite dimensions — e.g. to $\Omega(q,q',T)$ the physicists encounter and overcome, often with great ingenuity, many beautiful questions of a purely mathematical nature — such as how to define the determinant of an ordinary differential equation, etc., and the resulting perturbation theory is a perfectly well defined mathematical discipline.

In any case, all I wanted to make a little more plausible here, is that in trying to force the ansatz: $dA = F$, even when F *does not represent* 0 *in* $H^2(M)$ Dirac was led to an "integrality condition" on the cohomology class of F, which in retrospect, fits beautifully into — on the one hand the Kaluza-Klein theory — which, as we will see, geometrically couples F and g in a most satisfactory manner — and on the other hand into the purely topological problem of "killing homotopy groups".

Let us first take up this second concept. We start accordingly with the observation

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

that the second cohomology group of a 4-manifold M, can always be "destroyed" by removing certain 2-dimensional surfaces X in M. Thus on the complement of the surface X we can choose an A with $\mathrm{d}A = F$, but this A will have singularities along the surface X. Now in quantizing, say the classical charged particle moving in the electromagnetic field F, the appropriate classical action is given by:

$$S(\mu) = \int_\mu \frac{1}{2} m\dot{\mu}^2 \, \mathrm{d}t + \int_\mu \epsilon A \tag{2.32}$$

and the difficulties arise in making sense of the second term in this expression unless our path *avoids* the surface X.

But the expression:

$$\mathrm{e}^{2\pi i/h \cdot S(\mu)} \tag{2.33}$$

will be well defined as long as the ambiguity in $\int_\mu \epsilon A$ — which essentially depends on the line integral $\int_\lambda A$ over small loops λ circling X, — is an *integral multiple* of h.

Topologically, what all this amounts to is that the class of $\epsilon F \in H^2(M)$ has to be h times what we call an "integral class" in $H^2(M)$. More precisely the topologists have various methods of defining *abelian groups* $H^q(M; Z)$ called the qth *integral* cohomology groups of M, which come equipped with a natural homomorphism

$$H^q(M; Z) \rightarrow H^q(M) \tag{2.34}$$

into the de Rham groups, and whose image gives rise to a natural *geometric quantization* of the vector space $H^q(M)$.

The forms in the image of this arrow are called "integral forms" and they can be characterized by the property that their integrals over any closed hypersurface are always integers, and it is the dream of the "Geometric Quantization Program" to ultimately relate all quantization effects in physics to this topological one.[1] In the present context this aim therefore suggests that in the quantization of the electromagnetic F — which is according to Dirac an integral class — we should consider the element $\hat{F} \in H^2(M; Z)$ from which it comes. Now the striking fact is that the elements $H^2(M; Z)$ have a natural "geometric realization" in view of the following theorem — well known to all topologists and K-theorists:

[1] The term "geometric quantization" is by and large associated with a definite program of making the "canonical quantization" scheme of physics more functorial and generetic. This theory then starts with the Hamiltonian formulation of classical dynamics and the corresponding symplectic form ω on a symplectic manifold M. In this frame work the condition of ω to be an integral class is a very natural first step in the quantization. The rest of the procedure then still involves two steps: first of all the line bundle associated to ω is constructed and secondly a Lagrangian submanifold in TM is selected. This description of the quantization procedure then brings the whole quantization process into close relation with the construction of the irreducible representations of Lie groups – both in the nilpotent context of Kirilov and in the Borel-Weil context of the compact groups. This beautiful program is associated with the names of Kostant, as well as Blattner, Sternberg, Guilleimin, Surian and many others. See for instance [16]. In spite of the many virtues, their program has so far had no appreciable impact on physicists, who are loath to give up the intuition provided by the Lagrangian functional integral point of view, and that is in part why I have chosen to speak about the latter approach on this occasion.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

THEOREM: *The elements of* $H^2(M; Z)$ *are in natural* 1-1 *correspondance with the isomorphism classes of circle-bundles over M.*

Precisely, every $\hat{F}\epsilon H^2(M; Z)$ has a "flesh and blood realization" which consists of a *space* $P = P_F$ together with an action of the *circle* on S^1 on P and a natural projection

$$\begin{tikzcd} P \arrow[d, "\pi"] \\ M \end{tikzcd} \tag{2.35}$$

so that the S^1 action on P is *"free"*, preserves π, and π induces an *isomorphism* of P/S^1 *with M*.

Examples of these objects abound in geometry. For instance if M is a Riemann surface, take for P the set of its *unit tangent vectors to M*, and let π be the map which assigns to each such vector its foot. The circle S^1 then acts freely by rotating any vector counter-clockwise in its tangent plane. The simplest example of a P is of course $M \times S^1$ with S^1 acting on itself by right translation. This P is called the *trivial bundle* and it corresponds to 0 in $H^2(M; Z)$. Locally all P's are isomorphic to such a trivial one, but globally P will in general *not* equal $M \times S^1$. Indeed the "unit tangent" $-$ P of the 2-sphere S^2, is seen to be isomorphic to the group of rotations of 3-space $SO(3)$ and this P *is not trivial*. In fact it is seen to be the geometric realization of *twice* the generator of $H^2(M; Z) \simeq Z; M = S^2$. The double covering of $SO(3)$ is the 3-sphere and the resulting natural map $S^3 \to S^2$ is the famous "Hopf map" of homotopy theory which in our context geometrically represents the generator of $H^2(M; Z)$.

Explicitly this generator can be thought of as the action of S^1 on S^3 $-$ the 3-sphere:

$$|z_1|^2 + |z_2|^2 = 1, \qquad z_1, z_2 \epsilon \mathbb{C}, \tag{2.36}$$

given by $(z_1, z_2)\cdot z = (z_1 z, z_2 z)$, $z = e^{i\theta}$ or, quite equivalently, as the *right action* of the diagonal matrices S^1 on the group $SU(2)$, which of course is homeomorphic to S^3 under the map

$$(z_1, z_2) \leftrightarrow \begin{bmatrix} z_1 & z_2 \\ -\bar{z}_2 & \bar{z}_1 \end{bmatrix} \tag{2.37}$$

The bundle associated to n times the generator is furthermore paradoxically obtained by dividing this bundle by n $-$ that is to say by the action of $S^1/\mathbb{Z}_n$ on $SU(2)/\mathbb{Z}_n$, where $\mathbb{Z}_n \subset S^1$ is the set of n'th roots of 1.

Now how is all this related to the Dirac forcing of the equation $dA = F$? The answer is very beautiful and simple. Namely the geometric realization P of F, has the property that it "kills" the class F. That is, under the natural homomorphism

$$\pi^* : H^2(M; \mathbb{Z}) \to H^2(P; \mathbb{Z}) \tag{2.38}$$

$\pi^* F$ becomes "cohomologous to zero". In short *there is always a quite legitimate* 1-form A on P, such that

$$dA = \pi^* F. \tag{2.39}$$

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

On the other hand to force A back down to M one needs a "section" of π, that is, a smooth map $s: M \to P$ with $\pi \circ s = 1$. Indeed if such a section can be found then

$$\mathrm{d}s^*A = s^*\pi^*F = F \tag{2.40}$$

so that s^*A is the desired form on M. But a section of P exists only if P is trivial! (Indeed a section s clearly defines an isomorphism of $M \times S^1$ with P, by sending (m, p) to $s(m)\cdot p$. Hence for any nontrivial $P \in H^2(M; \mathbb{Z})$ the section will develop singularities — precisely along the 2-dimensional surfaces X in our M^4 we encountered earlier. By the way the "Dirac strings" are the traces of such a surface in a 3-dimensional time slice of M^4, and are the natural data for X when one is seeking time independent fields.

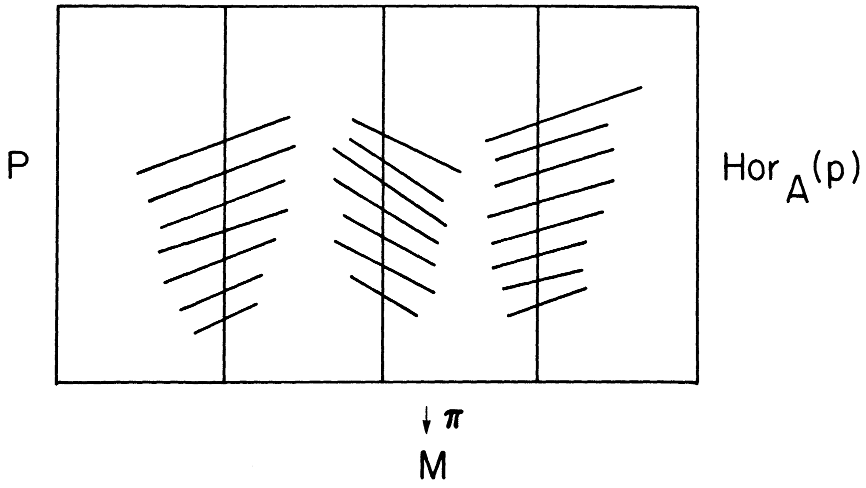
I hope that these retrospective ruminations of a topologist have convinced you that a subtle but plausible extension of the classical electromagnetic potential is the 1-form A on $P = P_F$ which has the crucial property $\pi^*F = \mathrm{d}A$. Furthermore, to restrict the ambiguity of A as much as is *naturally possible*, we may in addition stipulate that A be invariant under right translations by S^1 on P. But it then follows quite easily that:

(2.41) *the restriction of A to any fiber circle is a left invariant form on S^1 whose integral $\int_{S^1} A$ is 1, so that in particular the restriction of A to any fiber is $\neq 0$ anywhere.*

Finally, the information contained in such an A can now be coded in two very geometric and quite equivalent ways:

(2.42) *First formulation*: The electromagnetic potential A is completely determined by the subspace $\mathrm{Hor}_A(p) \in T_pP$, $p \in P$ of the tangent space to P at p on which A vanishes. Furthermore this family is always transversal to the "vertical" directions tangent to the fiber circles and moves into itself under the action of S^1.

The following schematic diagram indicates A conceived of as such a horizontal assignment.



(2.43) *Second formulation*: Let us choose a fixed invariant metric on our circle S^1, giving it length Λ-say. Also let g be a fixed metric on M. Now consider the set of S^1-invariant metrics $\hat{g}$ on P such that:

(a) $\hat{g}$ agrees with the chosen metric on every S^1-orbit of a point in P – i.e., on the fibers of π.

(b) The metric $\hat{g}$ restricted to the g-orthogonal complement of the fiber at a point $p \epsilon P$ is isomorphic to g at $\pi(p) \epsilon M$.

The equivalence of these two formulations is clear: indeed the $\hat{g}$-orthogonal complement to the fibers S^1 in P is a horizontal assignment in the sense of the first fomulation and vice versa.

Amazingly enough we now find ourselves with the geometric picture which – except for the possible twist in P over M – was already devised by Kaluza and Klein in the late 1920's to solve the second question we posed earlier concerning a natural coupling of F to g in the Einstein space-time framework.

For this purpose consider the family of metrics $\hat{g}$ of our second formulation and the corresponding Einstein-Hilbert-Lagrangian on P:

$$S(\hat{g}) = \int_P R(\hat{g}) \operatorname{vol}(\hat{g}) \tag{2.44}$$

One now computes $R(\hat{g})$ in terms of g and the 1-form A describing $\hat{g}$, to be:

$$R(\hat{g}) = \pi^*\{R(g) - \frac{1}{2}\Lambda^2|F|^2\} \tag{2.45}$$

where $\pi^*F = \mathrm{d}A$, so that by Fubini's theorem, (2.44) translates into:

$$S(\hat{g}) = \Lambda \int_M \{R(g) - \frac{1}{\Lambda^2}|F|^2\} \operatorname{vol}(g). \tag{2.46}$$

It then follows that the corresponding variational equations naturally fall into two sets. First of all one has, corresponding to variations in A, the Maxwell equations:

$$\mathrm{d}F = 0 \text{ and } \mathrm{d}^*F = 0, \tag{2.47}$$

and corresponding to the variation of the "base" g_{ij}'s the Eisenstein equation

$$R^{uv} - \frac{1}{2}g^{uv}R = \frac{1}{\Lambda^2}T^{uv} \tag{2.48}$$

where T^{uv} is the "Maxwell stress tensor of A":

$$T^{uv} = F^{ua}F^{vb}g_{ab} - \frac{1}{4}g^{uv}|F|^2. \tag{2.49}$$

In short then — Hamilton's principle applied to $S(g)$ in the Kaluza-Klein P, produces the correct Einstein-Maxwell equations (2.48) which couple the electromagnetic field

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

F and the metric with the length Λ of the circle playing the role of the coupling constant between these two fields.[2]

To sum up, we see that the evolution of the electromagnetic potential from a 1-form on space time to a special sort of 1-form on a possibly nontrivial circle bundle P over M — is in every way a satisfactory one, both from the classical and the quantum point of view. And in fact once grasped, the purely mathematical evolution from this example to the general Yang-Mills theory is also "in retrospect" — very natural. The watchword is simply this: in the above replace S^1 by an arbitrary compact Lie group G.

But before I come to speak about this step and the remarkable consequences the consideration of the resulting equations has had in pure mathematics, let me say a few words concerning the "Dirac monopoles" or rather the quantization of the motion of a classical charged particle under its influence, from the above "string-free" point of view.

First of all, what is the classical Dirac monopole? In terms of $\boldsymbol{E}$ and $\boldsymbol{B}$ in $\mathbf{R}^3$, this field is given by

$$\text{(2.50)} \qquad \boldsymbol{E} = 0, \quad \boldsymbol{B} = \frac{1}{4\pi}\frac{\boldsymbol{x}}{|x|^3},$$

so that for F we have the expression:

$$\text{(2.51)} \qquad F = \frac{1}{4\pi}\frac{x^1\,\mathrm{d}x^2\,\mathrm{d}x^3 - x^2\,\mathrm{d}x^1\,\mathrm{d}x^3 + x^3\,\mathrm{d}x^1\,\mathrm{d}x^2}{|x|^3},$$

and consequently $\int_{S^2} F = 1$. In short F generates the integral cohomology of $H^2(\mathbf{R}^3 - \mathrm{pt})$.

Now consider the polar decomposition $\mathbf{R}^3 - 0 = S^2 \times \mathbf{R}^+$, let $S^3 \xrightarrow{\pi} S^2$ be the Hopf-fibring with fiber circle S^1 and let

$$\text{(2.52)} \qquad P = S^3 \times \mathbf{R}^+ \xrightarrow{\pi \times 1} S^2 \times \mathbf{R}^+ = \mathbf{R}^3 - 0$$

be the corresponding nontrivial circle bundle over $\mathbf{R}^3 - 0$. Next choose a left and right invariant metric on S^3 which reduces to the standard one on the fiber S^1, and let A be the corresponding orthogonal projection on the fibers in P. The resulting F on $S^2 \times \mathbf{R}^+ = \mathbf{R}^3 - 0$ is then precisely the monopole field described earlier in its $\mathbf{R}^3 - 0$ manifestation.

An interesting step occurs now in the quantization of the effect of this field on a charged particle. Classically the action $S(\mu)$ of a particle of mass m and charge ϵ moving under the influence of a field $F = \mathrm{d}A$ along a path μ is given by (2.32), that is, by:

[2] In terms of a standard set of units for gravitation and electromagnetism the length Λ is seen to be very small indeed and this is often interpreted by the physicists as an explanation of why we do not observe the fifth dimension of P. The fiber-direction is so "curled up" that we cannot observe it.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$S(\mu) = \int_\mu \left\{\frac{1}{2} m\dot{\mu}^2 + \epsilon A(\dot{\mu})\right\} dt. \tag{2.53}$$

Manifestly in our generalized setting where A is only defined on P and not on M, the second term — which describes the effect of F — needs rethinking before the Feynman Paradigm can be applied. The upshot of this rethinking is again mathematically very satisfactory.

Namely, one concludes that the appropriate generalization of the Hilbert space of states, when dealing with a field F coming from a nontrivial integral class F, is *not* the space of L_2-functions on the base manifold $M = \mathbf{R}^3 - 0$, but rather a possibly twisted version of this space with the twist determined by the charge.

To explain these matters consider, quite generally, the space $\Gamma(L^n)$ of complex valued functions f on P which are equivariant under the action of S^1 on P relative to the n'th power representation of S^1 on $\mathbb{C}$, that is the space:

$$\Gamma(L^n) = \{f: P \to \mathbb{C} | f(p\cdot z) = z^{-n}f(p)\} \qquad z\epsilon S^1 \subset \mathbb{C}^*. \tag{2.54}$$

Note that for $n = 0$, $\Gamma(L^n)$ reduces to the ordinary complex valued functions on M, so that these $\Gamma(L^n)$ are all of roughly the "same size" as the complex valued functions on M. In fact when P is trivial, which means that P has a global section s, the assignment $f \to s^*f$ is in fact an isomorphism of $\Gamma(L^n)$ with $\Omega^0(M)$, so that the question which of the $\Gamma(L^n)$ to use is interesting only when P is nontrivial. At the same time this construction puts the physicists' insistence that the phase of a state function φ is physically meaningless into its proper mathematical context. Indeed one may — as the notation already indicated — consider $\Gamma(L^n)$ as the *spaces of sections of a line-bundle L^n over M* — associated to P via the representation $\rho : z \to z^n$ of S^1. Technically, L^n is the space of orbits of the S^1-action on $P \times \mathbb{C}$ given by $(p, \omega)\cdot z = (pz, z^n\omega)$, and the mathematicians denote it by

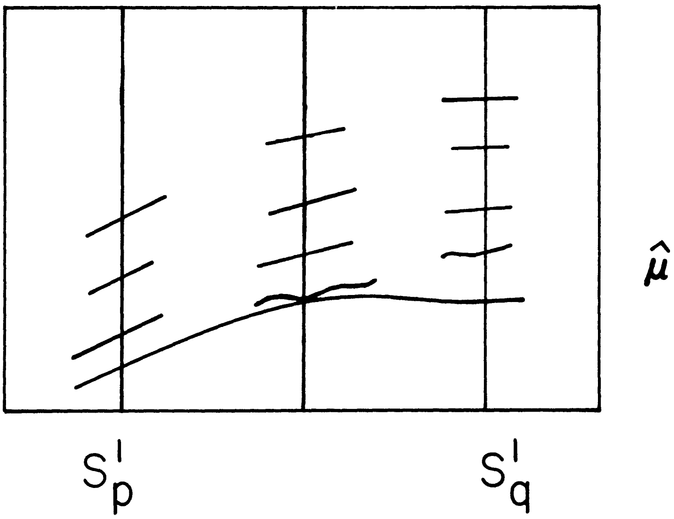
$$L^n = P \underset{\rho}{\times} \mathbb{C}. \tag{2.55}$$

As all P's are locally trivial, the sections $s\epsilon\Gamma(L^n)$ certainly locally look like complex valued functions, in the sense that if s is any nonvanishing section of P over U, then all other sections over U, are multiples fs with f a complex valued function on U. But certainly the sections have no well defined phases!

Granting all this, how do these considerations help in making sense of the term $\epsilon\int_\mu A$ in $S(\mu)$? Clearly there is no help for it; this term can be made meaningful only by "lifting" the curve μ to a curve $\hat{\mu}$ on P, and then considering the integral $\epsilon \int_{\hat{\mu}} A$. But this integral can be made to vary widely — depending on which lifting μ of M is chosen. In fact $\int_{\hat{\mu}} A$ can be made to vanish *identically*, by *choosing a horizontal lift* $\hat{\mu}$ of μ.

That is, we choose $\hat{\mu}$ in such a way that $d\hat{\mu}/dt$ always projects on du/dt and at the same time is annihilated by A for all t, i.e. is A-horizontal. In terms of our schematic representation, such a horizontal lift is indicated below

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press



and it should be clear that these horizontal lifts are now unique once an initial point for $\hat{\mu}$ has been chosen, and that the assignment:

$$\text{initial point of } \hat{\mu} \to \text{final point of } \hat{\mu} \tag{2.56}$$

defines an equivariant map

$$h_\mu : S^1_p \to S^1_q \tag{2.57}$$

of the fiber at $p = \mu(0)$ to the fiber at $q = \mu(1)$.

Thus, the only geometric sense one can make of the term $\int_\mu A$ is really this isomorphism h_μ, which by the construction of L, also induces linear isomorphisms, denoted by $\rho(h_\mu)$:

$$\rho(h_\mu) : L^n_p \to L^n_q \tag{2.58}$$

of the fibers of the line bundle L^n at p and q respectively. Now the values of a section $s\epsilon\Gamma(L^n)$ at different points cannot be compared as there is no *a priori* isomorphism between L_p and L_q; on the other hand we see that with the aid of the electromagnetic potential A and a path μ from p to q we have a natural comparison of $s(p)$ with $s(q)$, namely a comparison relative to $\rho(h_\mu)$. It is for this reason that the mathematicians have called the data incorporated in A a "connection" on P, *it connects the fibers at different points* — once a path *between the points is chosen.*

But returning to the problem at hand; how does this geometry fit with the Feynman ansatz for quantizing, according to which the electromagnetic contribution should be

$$e^{(2\pi i/h)\varepsilon\int_\mu A}. \tag{2.59}$$

It fits beautifully provided ϵ, the electric charge, is an integral multiple of $h/2\pi$:

$$\epsilon = n(h/2\pi). \tag{2.60}$$

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

for then and only then can this expression be unambiguously identified with the isomorphism $\rho(h_u)$ and become lift independent.

Indeed, consider two liftings $\hat{\mu}$ and $\hat{\mu}'$ of μ joining $\hat{p}$ in S^1_p to $\hat{q}$ in S^q_1. Setting out along $\hat{\mu}$ and returning via $\hat{\mu}'$ then defines a map α of the circle S^1 into P,

$$\alpha : S^1 \to P, \tag{2.61}$$

such that:

$$\int_{\hat{\mu}} A - \int_{\hat{\mu}'} A = \int_{\alpha} A \tag{2.62}$$

Now a little geometry (the covering homotopy theorem) teaches us that α can be deformed to a map α' which wraps S^1 around the fiber circle S^1_p of P, a certain number of times, and furthermore that this deformation takes place along a 2-surface N in P which projects onto μ and whose boundary is given by α.

Now by assumption, the integral $\int_\alpha A$ is an integer; furthermore by Stokes on the one hand and the fact that πN is one-dimensional on the other:

$$\int_{\alpha} A - \int_{\alpha'} A = \int_N \mathrm{d}A = \int_N \pi^* F = 0. \tag{2.63}$$

Thus the ambiguity in (2.59) will cancel out if and only if

$$\epsilon = n(h/2\pi). \tag{2.64}$$

Having fixed n, one now argues that the remaining ambiguity in μ — the choice of initial points — cancels out only if we consider the space $\Gamma(L^n)$ as our state space and once this is done — the formula

$$U_T(p, q) = \int e^{(2\pi i/h)S(\mu)} \mathcal{D}(\mu) \tag{2.65}$$

also makes good formal sense provided the kernel $U_T(p, q)$ is properly interpreted as an integral operator in $\Gamma(L^n)$, rather than on $\mathcal{F}(M)$.

Having come this far a final remark concerning these $\Gamma(L^n)$ for the monopole may be in order. Recall that in our "string-free" treatment

$$\Gamma(L^n) = \{\rho\text{-equivariant maps from } SU(2) \text{ to } \mathbb{C}\} \times (\text{functions of } r \epsilon \mathbf{R}^+), \tag{2.66}$$

so that $SU(2)$ acts naturally on the first factor by left multiplication, reflecting the spatial symmetry of the problem.

Furthermore as the representation theory of $SU(2)$ is so well known, one may in a certain sense compare the spaces $\Gamma(L^n)$ by counting how often a given irreducible representation η of $SU(2)$ occurs in $\Gamma(L^n)$. This is a famous question in the general theory of induced representations with a famous answer the "Frobenius reciprocity theorem", according to which η occurs in $\Gamma(L^n) = \mathrm{Map}_\rho(SU(2), \mathbb{C})$ as often as the representation ρ occurs in η restricted to $S^1 \subset SU(2)$.

Now for $SU(2)$ the irreducible representations are in 1-1 correspondence with the

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

nonnegative integers $n \leftarrow \pi_n$, and the above criterion easily leads us to the conclusion that if ρ is given by $z \rightarrow z^n$, then $\mathrm{Map}_\rho(SU(2); \mathbb{C})$ contains π_k only if $k \geq n$, and then precisely with multiplicitly 1.

3. **The Yang-Mills equations**. In retrospect the mathematical evolution from the example just considered to the general Yang-Mills is nearly inevitable. All that has to be done is to replace the group S^1 with a more general compact Lie group G, promote P to a G-bundle and — to recapture the state space $\Gamma(L^n)$ — specify a representation

$$\rho : G \rightarrow \mathrm{Aut}(V), \tag{3.1}$$

of G on some finite dimensional vector space V.

This vector space V should be thought as the "internal state space" of the corresponding "ρ-particle", and the analogue of the electromagnetic potential A is now by definition the *horizontal assignment* $\mathrm{Hor}_A(p)$ in the tangent bundle to P, satisfying the same axioms relative to G as the A in our first formulation of the last section did relative to S^1. Thus the assignment,

$$p \mapsto \mathrm{Hor}_A(p) \in T_p(M)$$

is to be:

(3.2) transversal to the fibers G_p through p, and

(3.3) *the right action of G should preserve this assignment.*

The second formulation of the last section makes equally good sense, so that once a right and left invariant metric for G has been selected the Yang-Mills field A may also be thought of as being given by a G-invariant metric g on P, so that finally the old Lagrangian

$$S(\hat{g}) = \int_P R(\hat{g}) \operatorname{vol}(\hat{g}) \tag{3.4}$$

makes sense in the new context, and can be, just as before, "descended" to the base M, of P, where it now takes the form

$$S(\hat{g}) = \mathrm{Vol}(G) \int_M \left\{ R(g) - \frac{1}{2} \|F\|^2 + R^G \right\} \operatorname{vol}(g), \tag{3.5}$$

with F, the so-called "curvature of A", given by a formula we will write as

$$dA + \frac{1}{2} [A, A] = \pi^* F, \tag{3.6}$$

while R^G is the scalar curvature of G.

The variation of these equations finally lead one to the Einstein-Yang-Mills equations; which we can again separate into two parts and write down more or less schematically as:

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$d_A F = 0 \text{ and } d_A^* F = 0, \tag{3.7}$$

and

$$R^{uv}(g) - \frac{1}{2} R(g) g^{uv} = T^{uv}(F). \tag{3.8}$$

The first of these can be taken to be the definition of the *Yang-Mills* equation.

I will not be able to go into detail concerning these equations here and I refer the interested reader to the very fine set of lectures by R. Palais [14], for a derivation of the Einstein-Yang-Mills equations in this context, as well as for an extensive bibliography. But a few general remarks are certainly in order.

REMARKS. (1) The infinitesimal implication of the step from S^1 to G is to pass from the Lie algebra of S^1 – that is $\mathbf{R}$ – to the Lie algebra $\boldsymbol{g}$ of G. Thus in (3.6) the Yang-Mills field A has to be interpreted as a 1-form on P with values in $\boldsymbol{g}$ and the $\pi^* F$ must be similarly construed as a 2-form on P with values in $\boldsymbol{g}$. The proper "place" of F itself is then the space of 2-forms on M with values in the vector bundle "ad P" associated to P via the adjoint representation

$$G \xrightarrow{Ad} \operatorname{Aut}(\boldsymbol{g}). \tag{3.9}$$

Thus technically $F \in \Omega^2(M; \operatorname{ad} P)$. In terms of a local gauge — i.e. sections of P, and a basis $\{e_\alpha\}$ for $\boldsymbol{g}$, the pull-back s^*A is then given by a set of ordinary 1-forms A^α, and (3.6) takes the form

$$dA^\alpha + \frac{1}{2} C^\alpha_{\beta j} A^\beta \wedge A^j = F^\alpha, \tag{3.10}$$

with $C^\alpha_{\beta j}$ the structure constants of $\boldsymbol{g}$. The mathematicians like to think in terms of promoting the whole de Rham complex $\Omega^*(M)$ to the complex $\Omega^*(M; \operatorname{ad} P)$ of forms with values in the vector bundle $\operatorname{ad} P$, and then argue that once an A is specified, then this complex is equipped with a well-defined operator d_A

$$\Omega^q(M; \operatorname{ad} P) \to \Omega^{q+1}(M; \operatorname{ad} P) \tag{3.11}$$

but unlike the de Rham d, one does not have $d_A^2 = 0$ but rather the identity

$$d_A^2 = \text{Multiplication by } F_A. \tag{3.12}$$

However, in analogy to the implication $F = dA \Rightarrow dF = 0$, one does find that (3.6) implies the first part $d_A F = 0$, of (3.7). The * operation extends without trouble to $\Omega^*(M; \operatorname{ad} P)$ — but now involves both the metric g_{ij} on M and a left and right invariant metric on G. The upshot is then that though (3.7) is in many ways the analogue of the Hodge theory, the final equation $d_A^* F = 0$ is third order in A.

Let me finally say a few words concerning the motivation of the generalization from S^1 to G, both from the mathematical and the physical point of view. The joy is of course that they are so very different.

From the Differential Topology point of view the "Frame bundle" $F(M)$ of a man-

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

ifold is the simplest tangible topological consequence of the definition of a differential structure on M. It is a smooth G-bundle over M, with G the full linear group $GL(n)$, $n = \dim M$, and its point can be thought of as pairs $\{p; f\}$ with p a point in M, and f a basis $\{f_1, \ldots f_n\}$ for the tangent space T_pM to M at p. This bundle sits naturally as a differentiable principal $GL(n)$-bundle over M:

$$F(M) \xrightarrow{\pi} M, \tag{3.13}$$

with $\pi\{p; f\} = p$, and with $GL(n)$ acting on the basis part of $\{p; f\}$ in the obvious manner. At first sight $F(M)$ seems an unwieldy and redundant object, but conceptually it has many virtues. For instance the tensor-fields on M of various types are all seen to be ρ-equivariant functions on $F(M)$ with ρ taking values in some representation of $GL(n)$.

From the more topological point of view the existence of $F(M)$ immediately raises the question of how "nontrivial" this bundle is. Note that $F(M)$ trivial $\Leftrightarrow$ π has a section, which means that one can find n vector fields $X_1 \ldots, X_n$, on M which span the tangent space to M at every point $p \in M$. The manifolds for which $F(M)$ admits a global section are therefore called parallelisable and they are the only manifolds on which such a global notion of parallelism is well defined. In general M will of course not have this property — for example among the spheres only S^1, S^3 and S^7 do, but for some manifolds M, the $F(M)$ might still be in some sense less twisted then it is in general. These concepts are all clarified by the concept of a "reduction of the structure group" which also plays a role in physics — so that it is appropriate to explain it here briefly. Given a general G-bundle P over M and a closed subgroup $H \subset G$, we can "divide" P by H, that is consider the space of H orbits on P, to obtain a new bundle projection π_H

$$P/H \xrightarrow[\pi_H]{} M, \tag{3.14}$$

whose fiber is the coset space G/H. Now any section of π_H is called "a reduction of the structure group from G to H". From this point of view then a section of P itself corresponds to "a reduction of the structure group to the identity". All questions of the sort — does M admit a nonvanishing vector-field, or a k-plane field, or an almost complex structure, correspond to the topological question of whether $F(M)$ admits reductions to appropriate $H \subset GL(n, \mathbf{R})$. For instance for the almost complex case the reduction has to be to $GL(n, \mathbb{C}) \subset GL(2n, \mathbf{R})$.

Now although all these questions arising naturally in geometry relate primarily to the frame bundle $F(M)$ over M, to deal with them at all efficiently it became necessary to consider the totality of possible G-bundles over M, and to understand something concerning their classification.

To start with we define a bundle isomorphism f: $P \rightarrow P'$ between two bundles P and P' over M to be a diffeomorphism which

(3.15) *commutes with the action of G on P, and*

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$(3.16)\qquad \textit{induces the identity map on } M.$$

Note here that an f subject to (3.15) does preserve fibers and hence does induce a map f on M so that (3.16) makes sense.

The first great discovery of the topologists was that the isomorphism classes into which the notion of isomorphism divides the G-bundles, and which we denote by $\mathscr{E}(M; G)$ are relatively small in number, i.e. they are countable and to distinguish them is as hard as to distinguish maps of M into a certain space *up to homotopy*.

Precisely, one has the following theorem.

THEOREM. *Every connected compact Lie group G determines a space BG and the isomorphism classes of G-bundles over M are in* 1-1 *correspondence with the homotopy classes of maps of M to BG:*

$$(3.17)\qquad \mathscr{E}(M; G) \simeq [M; BG].$$

In short the isomorphism classes of bundles are truly objects of "homotopy theory" proper.

REMARKS. (1) To explain the mechanism of the 1-1 correspondence note that like cohomology, G-bundles move backwards: given a map $f: N \rightarrow M$ and a G-bundle P over M, the subset of $N \times P$ consisting of pairs (n, p) with $f(A) = \pi(p)$ naturally defines a G-bundle $f^{-1}P$ over N. Thus (3.17) is induced by a fixed G-bundle EG over BG — called the universal bundle.

(2) The space BG is only defined up to homotopy type, but convenient models for it, as well as for EG, can easily be constructed for the classical compact groups, e.g. $U(n)$, $SO(n)$, $SP(n)$ with the aid of a countably infinite complex Hilbert space H.

Indeed for $U(n)$, one sets:

$$(3.18)\qquad E(U_n) = \{e_i, \ldots, e_n; (e_i, e_j) = \delta_{ij}, e_i \in H\}$$

equal to the space of orthonormal n-frames in H. The action of U_n on EU_n is then the obvious one, and $E(U_n)/U_n = BU_n$ becomes identical with the "Grassmanian of n-planes" in $\mathscr{H}$:

$$(3.19)\qquad BU(n) = \{A \subset \mathscr{H}, \dim A = n).$$

In particular then we have for $n = 1$, that $EU(1) = S_1(\mathscr{H})$ the unit sphere of H and $BU(1) = \mathbb{C}P$ the infinite dimensional complex projective space. Thus in this case we have, in view of our earlier theorem in section 2, quite distinct – *but true* – descriptions of the isomorphism classes of S^1 bundles:

$$(3.20)\qquad H^2(M; \mathbb{Z}) \simeq \mathscr{E}(M; S^1) \simeq [M; \mathbb{C}P].$$

For a nonabelian group G, the description $[M; BG]$ persists as we saw, but there is in general no purely *cohomological* description of $\mathscr{E}(M; G)$!

(3) Note that an immediate consequence of the classification theorem is that over a contractible space M all G-bundles are trivial. Hence the bundles over the upper and

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

lower hemispheres of an n-sphere are trivial, and the homotopy class of a P over S^n is defined by the gluing together over the equator $S^{n-1} \subset S^n$. In this way one sees that when G is connected

$$\mathscr{E}(S^n, G) = [S^{n-1}; G] = \pi_{n-1}(G). \tag{3.21}$$

Thus over the spheres the isomorphisms of classes of G-bundle inherit a group structure from the "homotopy groups" $\pi_{n-1}(G)$ – as the homotopy classes of maps of a sphere into a space are called. This description is again compatible with (3.17) according to which

$$\mathscr{E}(S^n, G) = [M; BG] = \pi_n(BG). \tag{3.22}$$

Indeed the homotopy groups of G and BG are always related by a shift of one:

$$\pi_k(BG) = \pi_{k-1}(G). \tag{3.23}$$

This is a consequence of the characterization of the universal bundle EG as that bundle for which all π_k's, $k > 0$, vanish, see for instance [1].

I turn next to the much thornier path which leads modern field theory to have anything to do with the nonabelian groups, let alone a G-bundle's topological properties.

The beginning of this story is certainly Noether's Theorem in classical mechanics, which asserts that any infinitesimal symmetry of a Lagrangian $L(q, \dot{q})$ gives rise to a conserved quantity during the motion. Thus, for instance, the symmetries under translation: $\partial/\partial q(m\dot{q}^2) = 0$ leads to the conservation of momentum, and similarily, the time independence of a Lagrangian leads to the conservation of energy. In the inverse direction this has led the physicists to look "behind" any empirical conservation law for a symmetry in the Lagrangian density which describes it. Thus from conserved properties and their behaviour they were led to a hierarchy of Lagrangians for these phenomena which fit precisely into the context of G-bundles as described above. I am certainly not competent to say much about this development, but in any case let me show you how subtle the physicists craft is, by explaining the prize-winning Lagrangian of the Weinberg-Salam model describing the unification of the electromagnetic and the "weak" force. But let me give a precise and global definition which is hopefully intelligible to mathematicians, rather than the usual infinitesimal formulae found in the physicists texts.

The base manifold M is Minkowski space and let SOF be the principal $SO(3, 1)$ bundle given by the reduction of the frame bundle FM induced by the Minkowski metric. Finally let P_1 be the double covering of SOF. Thus the structure group of P_1 is $SL(2, \mathbb{C})$ the double cover of the Lorentz-group. Note that the metric induces a canonical connection (Yang-Mills field) on SOF – which being infinitesimal in character lifts naturally to the double cover P_1. This field only plays the role of a background field in the model but it is still very much part of the set up. Next let P_2 be a principal bundle over M with structure group $SU(2) \times U(1)$, which I suspect should be thought of as the double cover of $U(2)$.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

We now fully consider the bundle

$$(3.24)\qquad P_1 \times {}_M P_2 = P$$

over M given by that part of the product $P_1 \times P_2$ which projects to the diagonal $M \subset M \times M$. The total structure group is therefore

$$(3.25)\qquad G = SU(2) \times U(1) \times SL(2, \mathbb{C}).$$

The Lagrangian in question will involve the following fields: first of all there is a Yang-Mills field A for P_2 with corresponding $F = dA + 1/2\,[A, A]$. Next there are three fields φ, ψ_L and e_R modeled by certain ρ-equivariant maps of P into appropriate vector spaces. More precisely let α, ξ, and Δ denote the standard representations of $SU(2)$ on $\mathbb{C}^2$, $U(1)$ on $\mathbb{C}$, and $SL(2, \mathbb{C})$ on $\mathbb{C}^2$ respectively.

$$(3.26)\qquad \begin{aligned} &\alpha\colon SU(2) \dashrightarrow \mathrm{Aut}(\mathbb{C}^2)\\ &\xi\colon U(1) \dashrightarrow \mathrm{Aut}(\mathbb{C})\\ &\Delta\colon SL(2, \mathbb{C}) \dashrightarrow \mathrm{Aut}(\mathbb{C}^2). \end{aligned}$$

With this understood and abbreviating the space of ρ-equivariant functions on P by $\Gamma(\rho)$, we have:

$$(3.27)\qquad \begin{aligned} \varphi &\in \Gamma(\alpha \otimes 1 \otimes 1)\\ \psi_L &\in \Gamma(\alpha \otimes \xi \otimes \Delta)\\ e_R &\in \Gamma(1 \otimes \xi^* \otimes \Delta), \end{aligned}$$

while the Weinberg-Salam-Lagrangian density takes the form:

$$(3.28)\qquad \begin{aligned} L(\varphi, \psi_L, e_R) = -\frac{1}{4} F^a_{uv} F^{uv}_a - D_\mu{}^* D_\mu - h(|\varphi|) - \psi_L^* \gamma D \psi_L\\ - e_R^* - \kappa e_R \varphi^* \psi_L - \kappa \psi_L \varphi e_R. \end{aligned}$$

Here κ is a coupling constant, $D\varphi^*$ denotes the covariant derivative of φ relative to the Yang-Mills field A, whereas γD involves the covariant derivative relative to the connection induced on P via A on the first two factors and the fixed metric connection on the other factor and is the appropriate form of the Dirac operator in this context. To check that this density is well defined one finally has to know something about how the representations involved behave under the tensor-product.

For example $e_R \otimes \varphi^* \otimes \psi_L$ is in the representation:

$$(3.29)\qquad \begin{aligned} (1 \otimes \xi \otimes \Delta) \otimes (\alpha^* \otimes 1 \otimes 1) \otimes (\alpha \otimes \xi \otimes \Delta)\\ \simeq (\alpha \otimes \alpha^*) \otimes (\xi \otimes \xi^*) \otimes (\Delta \otimes \Delta) \end{aligned}$$

which has a natural projection to the trivial representation — given by the hermitian inner product in the first two factors, and the determinant in the last one, and $e_R \varphi^* \psi_L$ denotes $e_R \otimes \varphi^* \otimes \psi_L$ followed by this projection. And so on. The fundamental feature of this and all models in these gauge theories is then that the coupling constants correspond to projections onto the trivial representation of appropriate tensor product representations.

I have neither the time nor the expertise to bring you to an understanding of how this Lagrangian leads to predictions which are being verified today — at astronomical expense by the way — in the largest accelerators of the world. All I can comment on are some of the trivial features shared by all the gauge-theoretic Lagrangians describing the fundamental processes of nature. First of all, in all these theories there are two kinds of "particles", there are the "gauge particles" — represented by Yang-Mills fields A, on an appropriate principal G-bundle P over space time, and there are the particles proper — modeled by the ρ-equivariant maps of P to V relative to some representations of

$$(3.30) \qquad \rho: G \times SL(2, \mathbb{C}) \to \operatorname{Aut} V$$

The irreducible finite dimensional (holomorphic) representations of $SL(2, \mathbb{C})$ are given by the various symmetric powers $S^k(\Delta)$ of the standard representation Δ of $SL(2, \mathbb{C})$ this set is indexed by the $1/2$ integers by the physicists — and if ρ is of the form $\alpha \otimes S^k\Delta$ then $k/2$ denotes the "spin" of the "particle" in question. The "Higgs term" $h(|\varphi|)$ of (3.28) is a real valued function on V which is G-invariant and has its minimum $= 0$ on a nontrivial orbit G/H of G acting on V. Its function is to describe the "spontaneous breakdown of symmetry" which occurs near the vacuum, i.e., near the lowest energy state of the system. The analogy usually given for this phenomenon is that an assembly of regularly spaced unaligned magnets in the plane, pointing freely about their centers will align themselves in "some" direction; however, which direction they choose depends on matters of chance.

All these strands are now woven into a Lagrangian $L(A, \varphi)$ which above all has to satisfy (1) local gauge-invariance and (2) which, by and large, involves the fields to "order ≤ 4". By gauge invariance we mean the following. Consider any automorphism of P,

$$(3.31) \qquad f: P \to P,$$

that is then an isomorphism of P with itself in the sense of (3.15, 3.16). Clearly such an f induces an automorphism f^*, on the ρ-equivariant functions from P to V as well as on the Yang-Mills fields on P. A transverse, G-invariant field, A, goes over into another such field, f_*A, under f. Hence it makes sense to demand that

$$(3.32) \qquad L(A, \varphi) = L(f_*A, f^*\varphi)$$

and this is the requirement of "local gauge invariance". We met this invariance already in the electromagnetic Lagrangian — and it was there referred to as the price we had to pay to bring the Maxwell equations into the fold of Lagrangian theories.

Conceptually "local gauge invariance" is just another manifestation of the principle that the fundamental equations of nature must be intrinsic. They cannot depend on coordinates — nor on an explicit model for P. In short the equations must depend only

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

on the "isomorphism class of P" — or equivalently they must be invariant under the group $G(P)$ of "automorphisms of P".

An attentive reader might object at this point, that in view of (3.17) and the contractability of Minkowski space all the P's in the present context must be isomorphic to the trivial bundle, so that all this attention to bundle concepts is useless mathematical pedantry. But this is not so; in many contexts, natural boundary conditions on A or φ are best understood by interpreting them as defining bundles on S^4, or, when a time slice is involved, as bundle on S^3, or yet again in the case of periodic boundary conditions in $\mathbf{R}^4$, as bundle on the torus T_4. In these cases the bundle classification theorems become important and — what is in a sense even more interesting — the topology of $G(P)$ becomes nontrivial and in fact naturally involves rather higher homotopy groups of G and M than the physicists by and large have got used to. (We will discuss this point in greater detail in the next section).

But to return to the second condition concerning the "degrees" in $L(A, \varphi)$. This condition emerges from the requirement that the resulting quantum theory be "renormalizable". This concept is rather beyond the scope of these lectures and certainly this lecturer. But for those of you willing to put on your safety belts, let me say two words about the beginnings of this flight of ideas.

Corresponding to a Lagrangian $L(A, \varphi)$ such as we have been discussing, we will often find the expression

$$(3.33) \qquad Z = \int e^{-(2\pi/h)S(A,\varphi)}\, \mathcal{D}(A) \times \mathcal{D}(\varphi)$$

in the physics texts, as describing the partition function of the energy states of the theory. Here — God bless them — the integration is to be carried out over all the fields — but in the "Euclidean" version of the theory, i.e. with t replaced by — it!

Let us follow the path to such a formula in the simplest classical quantum case. Here we have a particle of mass m moving under the influence of a potential $V(q)$ so that

$$(3.34) \qquad L(q, \dot{q}) = +\frac{1}{2} m\dot{q}^2 - V(q)$$

with $V(q)$ something like $q^2/2$ say. The corresponding quantum state space $L^2(\mathbf{R})$ then breaks into a discrete set of eigenstates with eigenvalues E_n, and corresponding orthonormal basis φ_n.

Now under thermal equilibrium at a temperature T and according to very general principles, the probability of finding this system in an energy state E is taken to be proportional to $e^{-E/kT}$ where k is Boltzmann's constant. Hence the so called "partition function" of the system

$$(3.35) \qquad Z(\beta) = \sum e^{-E_n\beta}, \qquad \beta = 1/kT$$

becomes a crucial object in the statistical study of the system.

Now the time evolution of this system through a time t is given by the unitary operator U_t sending φ_n *to* $e^{-(2\pi i/h)E_n t}\varphi_n$, so that up to a rescaling, $Z(\beta)$ corresponds to the trace of U_t analytically continued to $t = -i\beta$: we therefore define $Z(\beta)$ by

$$Z(\beta) = \text{Trace } U_t;\ t = -i\beta. \tag{3.36}$$

But according to Feynman

$$\text{Trace } U_t = \int_{\Lambda M} e^{(2\pi i/h)S(\mu)}\mathcal{D}(\mu) \tag{3.37}$$

where now $S(\mu)$ has been analytically continued to the *imaginary time interval* $[0, -i\beta]$, and the integral is taken over the space ΛM, of *all maps of the circle to M*.

Thus in view of the fact that

$$S(\mu) = \int_0^t \left(\frac{1}{2}m\dot{\mu}^2 - V(\mu)\right) dt \tag{3.38}$$

one obtains

$$Z(\beta) = \int_{\Lambda M} e^{-(2\pi/h)\int_0^{\beta h}\{\frac{1}{2}m\dot{\mu}^2 + V(\mu)\}dt}\,\mathcal{D}(\mu) \tag{3.39}$$

By the way these purely heuristic and formal considerations have now brought us to a formula in which the i of the exponent has disappeared, and is for this reason much more accessible to rigorous mathematical treatment than the original Feynman integral.

In fact if (3.39) is rewritten in the form:

$$Z(\beta) = \int_{\Lambda M} e^{-(2\pi/h)\int_0^{\beta h} V(\mu)dt}\, e^{-(2\pi/h)\int_0^{\beta h} m\dot{\mu}dt}\,\mathcal{D}(\mu) \tag{3.40}$$

then the last two terms combine to define a true measure on ΛM — the famous "Wiener measure" on this space of paths. Thus (3.40) is mathematically well-defined and is therefore also the usual starting point of the "constructive field theory" pioneered by Jaffe and Glimm (see, for instance [7]).

But to return to our main concern, I hope that we can now discern the heuristics which lead the physicists to pass from (3.39) to (3.32) once they applied to (3.39) the basic principle of field-theory acccording to which the individual values of the fields in question $A(q)$, are to be considered as the generalized coordinates — that is the q's of the theory. From that perspective one can then attempt to bring a precise definition of (3.32) by starting from a lattice version of this integral, and then letting the lattice spacing tend to 0. However, in making sense of (3.32) there is still a difficulty of another sort to be overcome, namely that the theory is invariant under the action of the local gauge group $G(P)$.

Note that this is a "large group". For instance when P is trivial, this group can be thought of as the space of all smooth maps of M to G:

$$G(P) = \text{Map }(M; G), \tag{3.41}$$

made into a group under pointwise multiplication. Hence if we try to compute an integral of the sort

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$(3.42)\qquad \int \Phi \mathfrak{D}A \times \mathfrak{D}\varphi, \qquad \Phi = \Phi(A, \varphi)$$

over the space fields in question, that is the space $\Gamma(P,.V)$ of the Yang-Mills fields $A(P)$ cross the section $\Gamma(\rho)$:

$$(3.43)\qquad \Gamma(P, V) = A(P) \times \Gamma(\rho)$$

of a $\boldsymbol{G}(P)$-invariant function Φ, then we must first renormalize relative to "Haar measure" of $\boldsymbol{G}(P)$. Indeed the "heuristic measure" $A \times \mathfrak{D}\varphi$ is of course invariant under $\boldsymbol{G}(P)$ because a gauge transformation is seen to induce an isometry in the values of the fields at any particular point, so that the integral of a $\boldsymbol{G}(P)$-invariant function Φ has to be renormalized by the Haar-measure of $\boldsymbol{G}(P)$ before one can expect a reasonable answer.

The physicists deal with this problem by means of the Fadeev-Popov ansatz. This involves choosing an auxiliarly function $f(A, \varphi)$ which is as far from gauge invariant as possible, i.e. so that the equation $f(A, \varphi) = 0$ specifies a single A in each orbit of $G(P)$, and then reducing the integral to this subset. In this procedure the "Fadeev-Popov determinant" has to be introduced to make the computation independent of which f was used, and the physicists have very ingenious ways of reinterpreting the effect of this determinant as "ghosts".

The finite dimensional analogue of this procedure is the following. Suppose that the Lie group G acts smoothly and freely on a manifold W and that this action preserves a smooth volume ω on W.

Then the choice of a left invariant volume v on G induces a well defined volume ω/v on W/G — which, given a section of the action, can be computed in terms of ω, the δ-function of the section, and the Fadeev-Papov correction.

Thus this step in the literature can be interpreted as saying that the "heuristic volume" $\mathfrak{D}A \times \mathfrak{D}\varphi$ on $\Gamma(P, \rho)$ descends to a "heuristic volume" on $\Gamma(P, V)/\boldsymbol{G}(P)$, and that by this step the $\boldsymbol{G}(P)$-degeneracy of our Lagrangian $L(A, \varphi)$ is cured. Here the "Lie group aspects" of $\boldsymbol{G}(P)$ are therefore used to the full. *There would be no intrinsic way of curing a degeneracy which is not essentially group theoretical.* And of course there are difficulties with this procedure even in the finite dimensional case. For instance the moment that the action fails to be free, M/G fails to be manifold etc. In the infinite dimensional case a rigorous mathematical treatment is more difficult, involves the whole machinery of L^p spaces, etc., but can be carried on rigorously to a certain extent. See for instance [10].

In any case I hope that all this prepares us for the conceptual leap according to which classical field theory corresponds to ordinary Lagrangian theory on $\Gamma(P, V)/\boldsymbol{G}(P)$ and that quantum field theory is then the corresponding quantum theory on this space — e.g., the formula (3.32) for the energy partition function! This formula stands as a succinct and beautiful reminder of this analogy and points the way in which all the tricks of the trade of classical quantum mechanics might carry over to the field theoretic case, and just as in the classical theory the temperature behaviour of Z leads to different

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

phases, so here the magnitudes of the various coupling constants in the Lagrangian lead the physicists to speculate about radically different phases of the field theory.

Finally just one word about the road that lead the physicists to the models we have been discussing. They were certainly not thinking of bundles, geometry, etc.! Rather they kept their eyes on experiments, and the masses of particles generated in accelerators. Now in the field theory transcription of particles, masses are associated with that part of the Lagrangian which is "quadratic in the fields". Noether's theorem now led them to write Lagrangians with large symmetry groups, but they found as a consequence that this also forces large degeneracies in the quadratic terms and predicted corresponding massless particles — the "Goldstone bosons" which were not observed in experiment. Essentially at this stage they were writing Lagrangians involving only our $\Gamma(\rho)$ space, i.e. they were from the mathematical point of view *fixing the principal bundle P*, i.e. writing equations *not invariant* under *automorphisms* of P. To get out of this dilemma and looking back on electromagnetism then led to the brilliant idea of "gauging the equations" — i.e. introducing the Yang-Mills fields as part of the Lagrangian — but at the same time demanding gauge invariance for all physically meaningful concepts. This freedom of choosing the appropriate gauge, finally led them via the "Higgs mechanism," to realize that in an intelligent gauge — corresponding to a reduction of the group G to H where $f(\varphi) = 0$ — some of the gauge fields "acquire masses" and thus did away with the unwanted Goldstone bosons, provided the center of H was at most 1-dimensional. These considerations together with the renormalization condition then seriously restricts the range of the possible models and so on.

Inadequate though they are, I hope that these remarks have opened up some new associations for the mathematicians and hopefully even for the physicists. To me it has always been a source of great satisfaction, that in these models the particles of nature are modeled by fields which "tell you how to differentiate" — that is the Yang-Mills fields, as well as by "fields which are differentiated" — that is the elements of $\Gamma(\rho)$.

Philosophically it is also very attractive to hold with the grand unified theory of Georgi, Glashow and others — according to which the symbiotic relationship of the biological world is already foreshadowed in the fundamental forces of nature in the sense that they all combine into one Yang-Mills field governed by a large Lie group.

4. **Applications to geometry and topology**. In mathematics the Yang-Mills fields A of a principal bundle are called "connections", and are there thought of as the infinitesimal expression of the fact that a principal G- bundle P is locally the product of G and M. Indeed at a point $p = (g, m)$ of a product $G \times M$, the tangent space splits canonically into the direct sum of a vertical space $\simeq T_gG$, and a horizontal space $H_p = T_mM$. For a possibly twisted P over M one still has a natural vertical component at a point $p \epsilon P$, but a G-invariant complement to it is not specified canonically. There are many such possible and the assignments of G invariant fields of such complements comprise the space $\mathscr{A}(P)$ of connections on P. This space naturally inherits the structure of an affine space: given two assignments A and A' of "horizontality" in P,

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

their difference $A - A'$ can be identified canonically with a 1-form η with values in the bundle ad P, and conversely, if A is a connection and $\eta \in \Omega^1(M, \operatorname{ad} P)$ then $A + t\eta$ defines a "line through A" in the direction η. In short $\mathcal{A}(P)$ is an affine space whose "tangent space" at any point $A \in r\mathcal{A}(P)$ is canonically isomorphic to $\Omega^1(M; \operatorname{ad} P)$.

Now just as in the global theory the nontriviality of P finds its expression in the fact that $P \xrightarrow{\pi} M$ has no global section, so in the infinitesimal theory, the 2-form characterized by $\pi^*F = dA + 1/2\,[A, A]$ is the appropriate measure of whether A *locally admits sections which are everywhere A-horizontal.*

In the language of differential geometry — F measures the extent to which the subbundle of A-horizontal spaces in P fails to be integrable. Thus $F = F(A)$ is a natural measure of the "curvature" of the assignment A of horizontality in P and the question arises whether one can conclude something about the "curvature" of P over M, i.e. concerning the nontriviality of P as a G-bundle over M — form the knowledge of a single $F(A)$. The theory of characteristic classes answers this question in the affirmative. Indeed it asserts that any polynomial function

$$(4.1) \qquad \varphi\colon \boldsymbol{g} \to R, \qquad \varphi \in S^q(\boldsymbol{g}^*)$$

from the Lie algebra $\boldsymbol{g}$ of G to $\mathbf{R}$, which is *invariant under the adjoint representation of G on* $\boldsymbol{g}$, can be meaningfully "applied" to the curvature $F \in \Omega^2(M; \operatorname{ad} P)$ to yield an (ordinary) differential form

$$(4.2) \qquad \varphi(F) \in \Omega^{2q}(M)$$

on M, of dimension twice the degree of φ. Furthermore $\varphi(F)$ is closed, and its cohomology class $[\varphi(F)]$ are a genuine obstruction to the triviality of P and as such they have played a fundamental role in every aspect of modern topology and geometry.

For instance, we already remarked that the first manifestation of a differential structure on a manifold M is the existence of frame bundle $F(M)$ over M. Its structure group is $GL(n, \mathbf{R})$, $n = \dim M$, so that its Lie algebra is isomorphic to the space $g\ell(n)$ of all $n \times n$ matrices and the adjoint representation is given by conjugation $a \to gag^{-1}$ of the matrix a by a nonsingular matrix g. Plausibly enough the coefficients of the characteristic polynomial of the matrix a now turn out to be a generating set for the invariants polynomials on $g\ell(n)$. Thus if we define $p_k(a)$ by

$$(4.3) \qquad \det\left(1 + \frac{it}{2\pi} a\right) = \sum t^k p_k(a)$$

then the p_i generate the ring of invariant polynomials on $g\ell(n)$ and the corresponding classes $[p_i(F)]$ where F is the curvature of some connection A on $F(M)$ are some sort of a cohomological measure of the nontriviality of $F(M)$. These, up to sign, are the "Pontryagin classes" of M.

Similarly if P is a $U(n)$ bundle, the Lie algebra of the structure group consists of the skew hermitian matrices $u(n)$, and the $\mathbf{R}$-valued invariant polynomials of $u(n)$ are defined by:

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$(4.4) \qquad \det\left(1 + \frac{it}{2\pi}a\right) = \sum t^k c_k(a)$$

and the corresponding classes $[c_k(F)]$ referred to as the Chern-classes of P. These constructions were in the air since the thirties and forties but the final links in the general theory of characteristic classes for G-bundles were being forged in Chicago largely under the impetus of A. Weil only in the early fifties and I recall vividly trying to understand it all during my stay at that time in Princeton at the Institute for Advanced Study. Ironically I also saw a good deal of Yang in those days. We used to meet on Saturday mornings to paint the fence of the nursery school. Unfortunately it never occurred to us to talk shop, and so it took well into the seventies until I, as well as most of my friends became aware of the pertinence of the Yang-Mills equations not only to the theory of characteristic classes, but to seemingly quite unrelated matters.

In my own case — and that is the first application I would like to discuss here — I became truly aware of these equations only in the summer of 1977 when I joined M. Atiyah in Oxford after a sabbatical stay at the Tata Institute in India. There I had become fascinated with the question the moduli-spaces of "stable bundles over Riemann surfaces" which had been initiated by D. Mumford, Seshadri, Narasimhan, Ramanan, and other algebraic geometers. In Oxford I found M. Atiyah very much involved in trying to generalize the t'Hooft solutions of the Yang-Mills equations over S^4 — a project in which he, Ward and Hitchin on the one hand and Drinfeld and Manin on the other succeeded a few months later — and it soon became apparent to us that this moduli space of stable bundles over a compact Riemann surface M, was precisely identical with the space of *minimal moduli* for the Yang-Mills equations:

$$(4.5) \qquad dA = 0 \qquad d^*A = 0$$

for a fixed $U(n)$-bundle P over M. Precisely then, the problem turned out to be to determine the algebraic topology of the space $\mathrm{Min}(P)$ consisting of the absolute minimum of our Lagrangian:

$$(4.6) \qquad S(A) = \int_M F \wedge *F, \qquad \text{as a set in } \mathcal{A}(P)/\boldsymbol{G}(P).$$

The local theory over a 2-dimensional base of these equations is rather trivial but the global structure of $\mathrm{Min}(P)$ is far from easy to determine and some nontrivial topology occurs there even in the simple case of the trivial $U(1)$-bundle.

Indeed in that case we may choose a section and a corresponding trivial connection A_0 on P, and then identify $\mathcal{A}(P)$ with $\Omega^1(M)$ via the map

$$(4.7) \qquad \Omega^1(M) \to A(P)$$

which sends α to $A_0 + \alpha$. The Yang-Mills equations now read $d^*F = 0$ with $F = d\alpha$. Hence $*F \in \Omega^0(M)$ is a constant — and therefore zero because M is compact and $\int_M F = 0$ by Stokes. In short the solution space of Y-M in $\Omega^1(M)$ simply consists of the linear subspace $Z^1(M)$ of closed 1-forms on M. So far so good — but we must next

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

divide $Z^1(M)$ by $\boldsymbol{G}(P)$! A little topological reflection shows that $\boldsymbol{G}(P)$ falls into components which are indexed by the elements of $H^1(M; Z)$, the integral 1st cohomology group of M. A little geometry then shows further that to an element λ in the identity component $\boldsymbol{G}_0(P)$ of $G(P)$ there is associated a function f on M so that λ acts on α by sending α to $\alpha + \mathrm{d}f$. Thus dividing $Z^1(M)$ by $\boldsymbol{G}_0(P)$ we already obtain the finite dimensional space $H^1(M)$ and if we next divide by $\boldsymbol{G}(P)/\boldsymbol{G}_0(P)$ one finds the true quotient of moduli to be the $2g$ dimensional torus:

$$T^{2g} = H^1(M; \mathbf{R})/H^1(M; Z) \tag{4.8}$$

Actually this torus, T^{2g} inherits a "complex structure" from the Riemann structure on M and so endowed is called the *Picard variety* of the Riemann surface M, and is a central object in much of algebraic geometry.

In any case the above analysis extends to all $U(1)$-bundle types over M. Topologically these P's are classified by one integer $c_1(P) \epsilon H^2(M)$ — their first Chern class — and in every case, the space $\mathrm{Min}(P)$ turns out to be a torus T^{2g}.

For the higher $U(n)$-bundles this analysis becomes considerably more difficult. Let me illustrate with the $U(2)$ case. The topological classification of the possible P's is still given by $c_1(P)$ — that is by one integer, but the behaviour of $\mathrm{Min}(P)$ turns out to depend only on the parity of $c_1(P)$. Where $c_1(P)$ is odd, $\mathrm{Min}(P)$ is seen to be a smooth variety — while in the other case it acquires singularities. In either case the Yang-Mills functional is easily seen to have other than minimal solutions, but these nonminimal ones are quite transparent. They correspond to the a direct sum of the $U(1)$ cases. That is, they are constructed in terms of two $U(1)$ bundles P' and P'' with $c_1(P') + c_1(P'') = c_1(P)$ by counting their minimal solutions to obtain a family of "decomposable solutions of Yang-Mills" on P. Thus the topological types of these extrema are all given by $4g$ dimensional tori

$$T^{4g} = T^{2g} \times T^{2g}. \tag{4.9}$$

Abstractly speaking our findings so far amount to the following data: We are given one infinite dimensional manifold

$$\mathcal{M} = \mathcal{A}/G \tag{4.10}$$

and a function f — that is the Yang-Mills functional — concerning whose extremal behaviour we know that: (1) If $c_1(P)$ is odd the minimum, $\mathrm{Min}(f)$, of f is a smooth manifold and (2) that all of f's higher critical sets are explicitly known manifolds say C_k, $k = 1, 2 \ldots$. Further a local computation shows that these C_k have well defined finite "indexes of instability λ_k". That is C_k has λ_k independent directions of steepest descent relative to f.

Now if all these data were finite dimensional and compact say, then the algebraic topology of the data would not be independent. Indeed the "Morse theory" then establishes the following inequality between them: let us write $P_t(X)$ for the Poincaré polynomial of a space X.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$P_t(X) = \sum t^k \dim H^k(X). \tag{4.11}$$

Then under generic nondegeneracy conditions the Morse inequalities read:

$$P_t(\mathrm{Min}\, f) + \sum t_k^\lambda P_t(C_k) = P_t(\mathcal{M}) + (1 + t)Q(t), \tag{4.12}$$

where $Q(t) = a_0 + a_1 t + \ldots$ is some polynomial with *nonnegative coefficients*.

Conceptually these inequalities teach us that the algebraic topology of $\mathcal{M}$ forces extrema on any nondegenerate function on $\mathcal{M}$, and if the function f fits $\mathcal{M}$ "like a glove" in the sense that the error term $Q(t)$ vanishes identically, we call f a "*perfect Morse function*" on $\mathcal{M}$.

For such functions the formula (4.12) can of course be solved for $P_t(\mathrm{Min}(f))$, so that for *perfect functions one obtains the expression*:

$$P_t\{\mathrm{Min}(f)\} = P_t(\mathcal{M}) - \sum_{k \geq 1} t_k^\lambda P_t(C_k). \tag{4.13}$$

It is precisely the analogue of this formula for the infinite dimensional case with $\mathcal{M} = \mathcal{A}(P)$, and f the Yang-Mills-Lagrangian $S(A) = \int_M F \wedge F^*$ which is proved in [1]. For instance when P is a $U(2)$-bundle with $c_1(P)$ odd, over a Riemann surface of genus g, this analogue takes the form:

$$\frac{P_t(\mathrm{Min}(S))}{(1 - t^2)} = \frac{((1 + t)^{4g}(1 + t^3)^{2g})}{(1 - t^2)^2(1 - t^4)} - \frac{t^{2g}(1 + t)^{4g}}{(1 - t^2)^2(1 - t^4)}, \tag{4.14}$$

and should be thought of as a consequence of the fact that the Yang-Mills Lagrangian does indeed fit the space $\mathcal{A}(P)$ to perfection, or to put it more clearly, the critical behaviours of the Yang-Mills Lagrangian $S(A)$ is *the minimal one consistent with the invariance of* S *under the local gauge group* $\boldsymbol{G}(P)$.

To explain this a little more precisely, recall that $\mathcal{A}(P)$ is a contractible space. Thus the Morse theory as such, predicts only a single critical point! What forces extrema on $S(A)$ must therefore be the fact that S has the large symmetry group $\boldsymbol{G}(P)$. Thus if $\boldsymbol{G}(P)$ acted freely on $\mathcal{A}(P)$ the algebraic topology of $\mathcal{A}(P)/\boldsymbol{G}(P)$ would presumably be the correct "measure" of the forced extrema on S. But it does not act freely — indeed the stability groups of the action at various points $A \in \mathcal{A}(P)$ vary from S^1 — the center of $U(2)$ — on generic orbits, to $S^1 \times S^1$ on the orbits where A decomposes the bundle P into a product of $U(1)$ bundles, e.g., on all the higher critical sets C_k.

But we have already discussed the algebraic topology prescription for using a nonfree action on a space X. Namely one passes from the action of G on X to the action of G on the space $X \times E(G)$.

In our situation we therefore pass from $\mathcal{A}(P)$ to $\mathcal{A}(P) \times E\boldsymbol{G}$, $\boldsymbol{G} = \boldsymbol{G}(P)$, and consider $S(A)$ as a function on this product invariant under $\boldsymbol{G}(P)$. It then descends to a function S_G on the quotient $\mathcal{A}(P) \times E\boldsymbol{G}$, $\boldsymbol{G}$, which in view of the contractability of $\mathcal{A}(P)$, is homotopically equivalent to the classifying space of $B\boldsymbol{G}(P)$:

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

$$BG(P) \simeq EG(P)/G(P). \tag{4.15}$$

In short then, what is established in [1] is that the function S_G induced by S on $BG(P)$ is perfect in the sense of Morse, so that (4.14) becomes a precise transcription of (4.13) with $\mathcal{M} = BG(P)$. In fact the first term on the right is precisely the Poincaré – now series – of the classifying space of the local gauge group of any $U(2)$-bundle:

$$P_t(BG) = \frac{(1+t)^{4g}(1+t^3)^{2g}}{(1-t^2)(1-t^4)}, \tag{4.16}$$

and here, as in all these formulas, the rational functions just stand for the formal power series they determine.

Thus the second term in (4.14) arises from the summation of all the contributions of the C_k's:

$$\frac{t^{2g}(1+t^4)^{4g}}{(1-t^2)(1-t^4)} \equiv \sum_k t^{2g+4k}\frac{(1+t^4)^{2g}}{(1-t^2)^2}, \tag{4.17}$$

and we also discern here the correct "equivalent way" of counting the contribution of C_k; it is simply: $t_k^{\lambda}P_t(C_k)\cdot P_t(BH_k)$, where H_k is the stability group of C_k. Recall that $P_t(BS^1) = 1/1 - t^2$ and hence that $P_t\{B(S' \times S^1)\} = 1/(1-t^2)^2$.

The development just sketched in for $U(2)$ generalizes to arbitrary $U(n)$-bundles P over Riemann surfaces and produces an explicit — though rather formidable — inductive formula for $P_t(\text{Min } P)$ in terms of the $P_t(\text{Min } P')$ with P' a $U(k)$-bundle with $k < n$, as well as setting the stage for an, in principle, complete description of the cohomology ring $H^*(\text{Min } P)$. In particular one concludes from this approach to the problem that these spaces never have any torsion.

Actually the formula (4.10) and its generalization to $P_t(\text{Min } (P))$ was not new when Atiyah and I noted it in the Y-M framework. However, the alternate derivations were so different in character that they in no way diminished our pleasure. For $U(2)$ the formula (4.13) is contained in Newstead's work [13] which was directly topological and could not be extended to $U(k)$. But the general case pioneered by Harder in [5], was also in the literature, and there was derived as a check of the "Weil conjectures" in algebraic geometry against Newsteads results. Of course, since that time these famous conjectures have been solved by Deligne and the Harder formula was extended to $U(n)$ by Harder and Narasimhan, [6], so that the formula (4.13) and its generalizations can now also be derived from this formidable result. Thus Harder starts from the algebraic geometry description of stability, counts rational points on the variety of stable bundles by means of a famous formula of C. L. Siegel, then translates the result to $P_t(\text{Min } P)$ now over the complex field – according to the Weyl prescription.

From our point of view the beauty of this derivation is that it fits so naturally with the Yang-Mills one. In particular $P_t(\text{Min } P)$ again appears by subtraction just as in (4.13), each contribution now having a number theoretic interpretation. *Roughly our t occurs as q^{ℓ} — with q the number of elements in the finite field.*

There is another point worth making about the Y-M theory in this first nontrivial

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

dimension: it also fits naturally into the theory of the "moment map" which occupies so much in the literature of the "symplectic school" of dynamics these days.

Indeed the perfection of Y-M as a function on $\mathscr{A}$ is considerably clarified if one notes — as is done in our paper — that the function $A \rightarrow F(A)$ is precisely the moment map of the action of $G(P)$ on $\mathscr{A}(P)$ relative to the "symplectic form" ω on (P) given by

$$\omega(\alpha, \beta) = \int_M \alpha \wedge \beta \tag{4.18}$$

(Recall here that the tangent space of $\mathscr{A}(P)$ is naturally isomorphic to $\Omega^1(M; \operatorname{ad} P)$, and that $\Omega^2(M; \operatorname{ad} P)$ where F takes values, can be identified with the "Lie algebra" of $G(P)$ – which should be thought of as $\Omega^0(M; \operatorname{ad} P)$.

Thus the Yang-Mills Lagrangian becomes the norm squared of the moment map of this action, and this interpretation leads one to conjecture that corresponding perfection theorems enable one to compute $P_t(\operatorname{Min} f)$ where f is the moment map for a quite general symplectic action. This also turns out to be the case, as was shown quite recently by F. Kirwan in [8].

But let me leave the two-dimensional world now, to say a few words about a spectacular recent advance in the topology of four manifolds, which occured as a consequence of the Yang-Mills theory. This is the new "smoothability obstruction" of Simon Donaldson [2]. His work is a brilliant application of some of the fundamental existence techniques largely due to K. Uhlenbeck [18], and C. Taubes [17], and combined with the equally brilliant work of Michael Freedman [4] finally leads to the even more unexpected conclusion that the differentiable structure on $\mathbf{R}^4$ is not unique. There is now even speculation that the different such structures on $\mathbf{R}^4$ might be uncountable and might "fit" into a smooth moduli-manifold. This would indeed be a new phenomenon because all previous "exotic differentiability" classes were naturally "quantized" and eventually could be detected in the context of characteristic classes.

A fine up-to-date reference, hot off the press for this whole story, is the set of notes by Dan Fried [3] on a seminar on "Gauge theories and four manifolds" organized by Mike Freedman and Karen Uhlenbeck at the Berkeley Institute. Here let me just give you the barest feeling for Donaldson's theorem.

A first step in this direction is to acquaint you with the "intersection form" Q_M which, in view of the general Poincaré duality on topological manifolds, is defined on the middle dimensional cohomology $H^n(M)$ of a compact oriented manifold of dimension $2n$. Indeed Q_M is there naturally defined as a consequence of the product structure of the cohomology by the formula:

$$Q_M(u, v) = \int_M u \wedge v \qquad u, v \epsilon H^n(M). \tag{4.19}$$

Here the integral sign should be thought of as the isomorphism of $H^{2n}(M) \simeq Z$ given by the orientation on M and which, on the de Rham level, is precisely furnished by integration.

On a 4-dimension manifold Q_M is then a symmetric quadratic form defined on the

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

Z-module $H^2(M; Z)$ and its equivalence class amongst the totality of such forms is a delicate and elusive topological invariant of M which is the corner stone of any classification theory.

Now for compact, simply connected topological four manifolds Freedman's recent complete classification [4] shows that in this context *any unimodular intersection form* can occur.

In contrast, Donaldson's theorem asserts that if such an M admits a smooth structure – then *only those positive definite forms can occur which are diagonalizable over the integers.*

The strength of this restriction is brought home to one by the fact noted in [4] that, say, for rank 40, there are more than 10^{51} inequivalent positive definite forms over the integers.

How does the Donaldson's theorem come about? Let us start with the easy lemma that a positive definite unimodular form Q is diagonalizable over Z if and only if the number of solutions of the equation

$$Q(\alpha, \alpha) = 1 \tag{4.20}$$

is equal to twice the rank of Q.

This, combined with the invariance of the index of Q (i.e., the number of positive minus the number of negative eigenvalues of Q over $\mathbf{R}$) under 'cobordism' leads to the conclusion that if we can construct a "manifold with boundary" W, such that the boundary ∂W falls into components which consists entirely of M and a disjoint union of complex projective planes $\mathbb{C}P_2$;

$$\partial W = M \cup \mathbb{C}P_2 \cup \ldots \cup \mathbb{C}P_2. \tag{4.21}$$

then the theorem will follow.

Amazingly enough a component of the space of moduli of the "self-dual Y-M equations" over M relative to a $SU(2)$ bundle P with $C_2(P) = -1$ furnishes one with precisely such a manifold W.

The reason for this is to be found first of all in the many special features which magically occur in four dimensions and which make all the various topological and Yang-Mills notions interact most dramatically. Indeed in this dimension the * operator maps 2-forms into 2-forms:

$$* : \Omega^2(M) \to \Omega^2(M) \tag{4.22}$$

with $*^2 = 1$, and thus decomposes this space into the ± eigenspaces of *:

$$\Omega^2(M) = \Omega_2^+ \oplus \Omega_2^- \tag{4.23}$$

This decomposition in turn leads to the notion of "self dual solution" of Yang-Mills. These are connections A such that their curvatures F satisfy the equation:

$$*F = F. \tag{4.24}$$

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

They correspond to special solutions to Yang-Mills because $d_A F \equiv 0$ so that (4.24) implies $d_A {*F} = 0$. One similarly has anti-self dual solutions given by

$$*F = -F. \tag{4.25}$$

Self-dual solutions were of course first constructed by the physicists Polyakov and t'Hopft (see [3]), for $SU(2)$ bundles over S^4, and the *A-H-D-M* construction gives a beautiful and very explicit extension of their work.

They also fit into the present context due to a fundamental result of C. Taubes [17]. Namely he showed how to "transplant" the solution from S^4 to any manifold M with positive definite Q_M, and in particular he showed that on such a manifold M, every P with $c_2(P) = -1$ has nontrivial self dual solutions of Yang-Mills, which naturally form a 5-dimensional manifold in $\mathrm{Min}(P)$, diffeomorphic to $M \times (0, 1)$. Thus in Donaldson's scheme of things Taubes essentially constructed a "collar" about that part of the boundary of W which contains M.

On the other hand there are also the more trivial "decomposable solutions" where the connection A naturally splits the $SU(2)$ bundle P into a direct sum of $U(1)$-bundles

$$P = P' \underset{M}{\times} P''. \tag{4.26}$$

At these degenerate solutions one then easily finds that

$$0 = c_1(P) = c_1(P') + c_1(P''), \tag{4.27}$$

In short recalling that $c_2(P) = -1$ the degenerate solutions are parametrized by half the number of solutions of the equations $Q_M(\alpha, \alpha) = 1$.

Finally Donaldson argues that for generic metrics on M, the moduli space looks like a cone over $\mathbb{C}P_2$ near these singular points. Q.E.D.

Of course here I am glossing over the many technical difficulties that beset this program and which can be surmounted only because of the basic convergence theorems of K. Uhlenbeck concerning sequences of connections with bounded curvatures. The delicacy of the situation is exemplified by the fact that one does not know whether the minimum of Yang-Mills, that is $\mathrm{Min}(P)$, is connected!

All in all this then is as beautiful and successful a confluence of different mathematical ideas as one can hope for and an appropriate place to end my lecture. Still in the perverse world of mathematics where questions are by and large more welcome than solutions I should reassure the reader that beautiful as these developments are, our general knowledge concerning the Yang-Mills equations even in four dimensions is still pitifully inadequate. Even for the 4-sphere where the *AHDM* construction provides one with an explicit set of equations for the moduli of the self dual solutions, we know hardly anything about the topology of the solution space, or more precisely how badly the Morse theory fails and in what dimension it fails. At this writing we do not even know whether there are solutions to Y-M other than the self dual ones over S^4. In short there still is here a great opportunity for cross fertilization between physics and mathematics which I hope will continue to strengthen the bonds between these two great disciplines in the future.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press

REFERENCES

1. M. F. Atiyah and R. Bott, *The Yang-Mills equations over Riemann surfaces*. Phil. Trans. Roy. Soc. Lond. A**308** (1982), pp. 523–615.

2. S. K. Donaldson, *An application of gauge theory to four dimensional topology*, J. Diffl. Geom. **18**, No. 2 (1983), pp. 279–315.

3. D. Fried, *Gauge theories and four manifolds*, Math. Sci. Research Inst., Berkeley, (1983).

4. M. H. Freedman, *The topology of four dimensional manifolds*, J. Diffl. Geom. **17** (1982), pp. 357–453.

5. G. Harder, *Eine Bemerkung zu einer Arbeit von P. E. Newstead*, J. Math. **242** (1970), pp. 16–25.

6. G. Harder and M. S. Narasimhan, *On the cohomology groups of moduli spaces of vector bundles over curves*, Math. Ann. **212** (1975), pp. 215–248.

7. A. Jaffe and G. Glimm, *Quantum physics*, New York: Springer Verlag, (1981).

8. F. C. Kirwin, *The cohomology of quotient spaces in algebraic and symplectic geometry I*, Thesis, Oxford (1982), to appear in Math. Notes, Princeton Univ. Press Yellow Series.

9. P. K. Mitter and C. M. Viallet, *On the bundle of connections and the gauge orbit manifold in Yang-Mills theory*, Commun. Math. Phys. **79** (1981), pp. 457–472.

10. D. Mumford, *Geometric invariant theory*, Berlin: Springer-Verlag, (1965).

11. M. S. Narasimhan and T. R. Ramadas, *Geometry of SU(2) gauge fields*, Commun. Math. Phys. **67** (1979), pp. 121–136.

12. M. S. Narasimhan and C. S. Seshadri, *Stable and unitary vector bundles on a compact Riemann surface*, Ann. Math. **82** (1965), pp. 540–567.

13. P. E. Newstead, *Characteristics classes of stable bundles over an algebraic curve*, Trans. Am. Math. Soc. **169** (1972), pp. 337–345.

14. R. S. Palais, *The geometrization of physics*, Lecture notes in Math. Inst. of Math. National Tsing Hua Univ. 1981.

15. C. S. Seshadri, *Space of unitary vector bundles on a compact Riemann surface*, Ann. Math. **85** (1967), pp. 303–336.

16. J. Snatycki, *Geometric quantization and quantum mechanics*, Appl. Math. Sci. **30**, Berlin: Springer-Verlag, (1980).

17. C. H. Taubes, *Self-dual connections on non-self-dual 4-manifolds*, J. Diffl. Geom. **17** (1982), pp. 139–170.

18. K. Uhlenbeck, *Connections with L^p bounds on curvature*, Commun. Math. Phys. **83** (1982), pp. 31–42.

https://doi.org/10.4153/CMB-1985-016-3 Published online by Cambridge University Press