# Einstein's impact on theoretical physics

That symmetry dictates interactions, that geometry is at the heart of physics, and that formal beauty plays a role in describing the world are insights that have had a profound effect on current thought.

Chen-Ning Yang

There occurred in the early years of this century three conceptual revolutions that profoundly changed Man's understanding of the physical universe: the special theory of relativity (in 1905), the general theory of relativity (1915) and quantum mechanics (1925). Einstein personally was responsible for the first two of these revolutions, and influenced and helped to shape the third. But it is not about his work in these conceptual revolutions that I shall write here. Much has been written about that work already. Instead, I shall discuss, in general terms, Einstein's insights on the structure of theoretical physics and their relevance to the development of physics in the second half of this century. I shall divide the discussion into four sections which are, of course, very much related.

### Symmetry dictates interaction

The first important symmetry principle discovered in fundamental physics was Lorentz invariance, which was found as a *mathematical* property of Maxwell equations, which in turn were based on the *experimental laws* of electromagnetism. In this process the invariance, or symmetry, was a secondary discovery. In his Autobiographical Notes[1] Einstein gave Hermann Minkowski credit for turning this process around. Minkowski started with Lorentz invariance, and required that field equations be covariant with respect to the invariance, as shown in the table on page 44.

Einstein himself was deeply impressed by the powerful physical consequences of symmetry principles and worked to enlarge the scope of Lorentz invariance. This idea of a more general coordinate invariance led, together with the equivalence principle, to the general theory of relativity. We might say that Einstein initiated the principle that *symmetry dictates interactions*. This principle has played an essential role in recent years in giving rise to various field theories:

- ▶ Coordinate-transformation invariance gives rise to general relativity
- ▶ Abelian gauge symmetry gives rise to electromagnetism
- ▶ Non-Abelian gauge symmetry gives rise to Non-Abelian gauge fields
- ▶ Supersymmetry gives rise to a theory with symmetry between fermions and bosons
- ▶ Supergravity symmetry gives rise to supergravity.

### Field Theory and unification

In his articles and lectures after 1920 Einstein repeatedly emphasized the concept of the field as of central importance to fundamental physics. For example, in an article published in the Journal of the Franklin Institute in 1936 he wrote:[2]

> The escape from this unsatisfactory situation by the electric field theory of Faraday and Maxwell represents probably the most profound transformation of the foundations of physics since Newton's time.

The two field theories known around that time (1936) were Maxwell's theory and Einstein's general relativity theory. Einstein devoted the last twenty years of his life striving to unify these two theories. The necessity of doing that he explained in an article published in 1934 entitled "The problems of space, ether, and the field in physics":[3]

> ... there exist two structures of space independent of each other, the metric–gravitational and the electromagnetic ... We are prompted to the belief that both sorts of fields must correspond to a unified structure of space.

In the last editions of *The Meaning of Relativity* Einstein added an appendix in which he proposed a unified theory with a non-symmetrical metric tensor $g_{\mu\nu}$. The anti-symmetrical part was to be identified with the electro-magnetic-field tensor $F_{\mu\nu}$. This effort was not particularly successful and there has been, for some time, among some people, the impression that the idea of unification was some kind of obsession affecting Einstein in his old age. Yes, it was an obsession, but an obsession with an *insight* of what the fundamental structure of theoretical physics should be. And I would add that that insight is very much the theme of the physics of today.

In any case, Einstein's emphasis on unification produced something at once. It led many distinguished mathematicians, including Tullio Levi-Ci-

Chen-Ning Yang is the Einstein Professor and the director of the Institute of Theoretical Physics at the State University of New York at Stony Brook.



0031-9228 / 80 / 060042-05/ $00.50

**The boundary of a region** has no boundary. This Möbius strip has only one surface; its boundary is a single edge, but the edge itself has no boundary. For a further explanation of this theorem, see the box on the next page. (All drawings for this article by Louis Fulgoni.)

vita, Elie Cartan and Weyl to look more deeply into possible additions to the mathematical structure of space–time.

Beginning in 1918 and 1919, Weyl made an effort to incorporate electromagnetism into gravitation. His idea led to what is called "gauge theory."[4] Since the proper treatment of coordinate invariance has produced gravity theory, Weyl thought that a new geometrical invariance could be tied to electromagnetism. His proposal was scale invariance.

If $x^\mu$ and $x^\mu + dx^\mu$ are two space-time points in the neighborhood of each other and $f$ is some physical quantity such that it is $f$ at $x^\mu$ and $f + (\partial f/\partial x^\mu)\, dx^\mu$ at $x^\mu + dx^\mu$, Weyl considered the space-time dependent rescaling of $f$, as shown in the last two rows of the table on page 48. Notice particularly the scale factor

$$1 + S_\mu dx\mu \qquad (1)$$

given in the third row.

Now Weyl observed two things about this scale factor. First, $S_\mu$ has the same number of components as the electromagnetic potential $A_\mu$. Secondly, in a further development, he proved that when one requires the theory to be invariant under the scale change, only the curl of $S_\mu$ occurs and not $S_\mu$ itself. That is also a feature of the electromagnetic potential, $A_\mu$. So he identified $S_\mu$ with $A_\mu$. This idea, however, did not work. It was discussed by several people including Einstein, who demonstrated that Weyl's theory cannot possibly describe electromagnetism, and Weyl gave it up.

Then 1925 came and quantum mechanics was invented, completely independently of this development.

We all know that in classical mechanics it is not the particle momentum $p_\mu$ that occurs, but, in presence of electromagnetism, it is always the combination:

$$\pi_\mu = p_\mu - (ie/\hbar c)A_\mu \qquad (2)$$

In quantum mechanics this is to be replaced by

$$-i\hbar[\partial_\mu - (ie/\hbar c)A_\mu] \qquad (3)$$

This was pointed out[5] by Vladimir Alexandrovitch Fock in 1927. Immediately afterwards, Fritz London compared[6] expression 3 with the increment operator $(\partial_\mu + S_\mu)$ in the last expression in the table, and concluded that $S_\mu$ is to be identified not with $A_\mu$ but with the factor $-ieA_\mu/\hbar c$. The important new point is just the insertion of an imaginary unit $i$. This has the far-reaching consequence that expression 1 becomes:

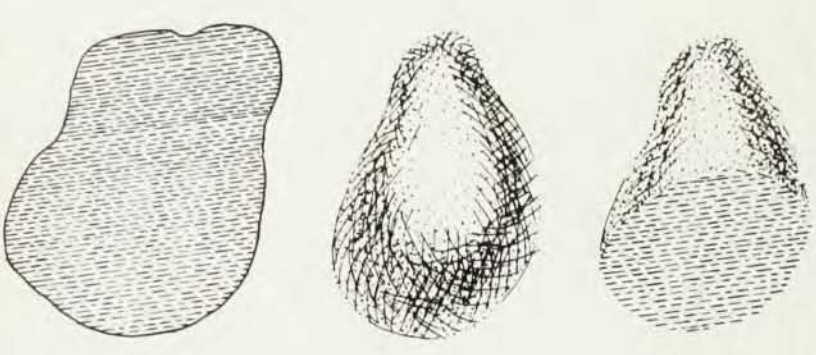
$$1 - (ie/\hbar c)A_\mu dx_\mu \rightarrow \exp[-(ie/\hbar c)A_\mu dx^\mu] \qquad (4)$$

which is a *phase* change, not a *scale* change. Therefore, local *phase invariance* is the correct quantum mechanical characterization of electromagnetism.

Weyl himself had called his idea "*Masstab Invarianz*" at first, but later changed to "*Eich-Invarianz.*" In the early 1920's it was translated into English as "gauge invariance." If we were to rename it today, it is obvious that we should call it *phase invariance*, and gauge fields should be called *phase fields*.

Once one has understood that gauge invariance is phase invariance, one finds that the key idea is a non-integrable phase factor. The substitution for the simple phase of complex numbers with a more complicated phase, namely

## A topological theorem



The boundary of a region has no boundary itself. In the example shown on the left, the shaded two-dimensional region has as its boundary a one-dimensional loop; the loop has no end, that is, it has no boundary itself.

The three-dimensional figure in the center is bounded by a closed two-dimensional surface; the surface again has no edges, that is no boundary.

If we cut the region to give an edge to the surface we create a second surface to complete the boundary of the smaller volume, as shown in the sketch on the right. If the edge of this cut is oriented, say, in the clockwise sense, then the edge of the part of the original surface must be oriented in the counterclockwise sense. Thus, although each of the two parts of the surface bounding the three-dimensional region has a boundary, the whole surface of the volume has no boundary; the edge is included twice, oriented in opposite directions, and therefore cancelled out.

an element of a Lie group, leads one to non-Abelian gauge theories, which were first fomulated in 1954.

We should emphasize here that the concept of phase is of great practical importance in contemporary physics. For example, the theories of superconductivity and superfluidity, the Josephson effect, holography, masers and lasers are all fundamentally based on various aspects of the concept of phase.

In 1967 Steven Weinberg and Abdus Salam independently proposed a model for a unified theory of electromagnetism with the weak interactions.The model is based on two key concepts: *Non-Abelian gauge field* and *broken symmetry*. An important further idea, due to Sheldon Glashow, was needed to eliminate contradictions with experiments. In the last six years this model has gathered amazing experimental support. The success has in turn given rise to exuberant efforts at a larger unification of strong, electromagnetic and weak interactions. I am afraid we are still some ways from a successful larger unification, and even further from a holistic unification of these interactions with general relativity. But there is little doubt that Einstein's insistence on the importance of unification was a deep insight, which he had courageously defended, against all spoken and unspoken criticism.

### Geometrization of physics

Another recurrent theme in Einstein's perception about the fundamentals of theoretical physics derived from his partiality to geometrical concepts. This is not surprising since he himself created the profound concept that gravity and mechanics should be described in terms of Riemannian geometry. That he regarded electromagnetism as also geometrical was evident from the earlier quote taken from his 1934 article. He stated there that electromagnetism is a "structure" of space. If one accepts the thesis that Einstein was partial to geometrical concepts, then one might perhaps even advance the view that he liked wave mechanics because it is more geometrical and disliked matrix mechanics because it is more algebraic.

Einstein strived to find the geometrical structure that gives rise to electromagnetism. He was aware of the fact that Lorentz invariance was not enough to give Maxwell equations:[7]

> Maxwell equations imply the "Lorentz group," but the "Lorentz group" does not imply Maxwell's equations.

For example, scalar fields are seemingly simpler than Maxwell's electromagnetic field, and are consistent with Lorentz invariance, but are not the basis of electromagnetism.

Einstein was also deeply aware of the necessity to have geometrical structures that give rise to nonlinear equations:[1]

> The true laws cannot be linear nor can they derived from such.

It turns out that the structure that Einstein was seeking was the gauge field: It is a geometrical structure, as we shall presently discuss; the simplest Abelian gauge field is Maxwell's electromagnetic field and a non-Abelian gauge field is necessarily nonlinear.

We had earlier referred to the early history of gauge fields. That gauge fields are deeply related to the geometrical concept of connections on fiber bundles has been appreciated by physicists only in recent years.

To illustrate the geometrical nature of gauge fields, let us write the Gauss and Faraday laws of electromagnetism in the following well-known form:

$$\partial_\lambda f_{\mu\nu} + \partial_\mu f_{\nu\lambda} + \partial_\nu f_{\lambda\mu} = 0$$

where $f_{\mu\nu}$ is the electromagnetic field. This equation turns out to be deeply related to the theorem that the boundary of a region has no boundary itself, which is, of course, a geometrical statement (see the box on this page). Another illustration of the geometrical nature of gauge fields can be found in the fact that global considerations have become important for gauge fields through the following theoretical and experimental developments:

- Dirac's magnetic monopole (1931)
- Bohm–Aharonov experiment (1960)
- 't Hooft–Polyakov monopole (1974)
- instantons (1975)

These ideas are described in the box on page 48.

Gauge fields are also intrinsically

### Symmetry and physical laws

**Before Einstein and Minkowski**

experiment → field equations → symmetry (invariance)

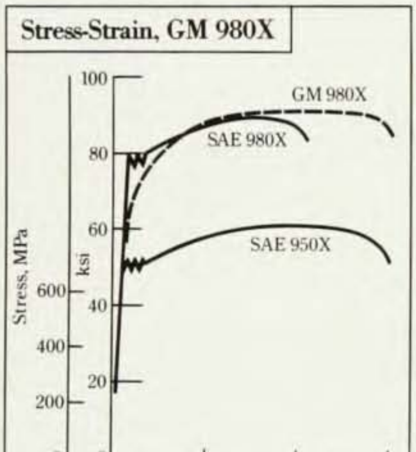
**After Einstein and Minkowski**

symmetry → field equations

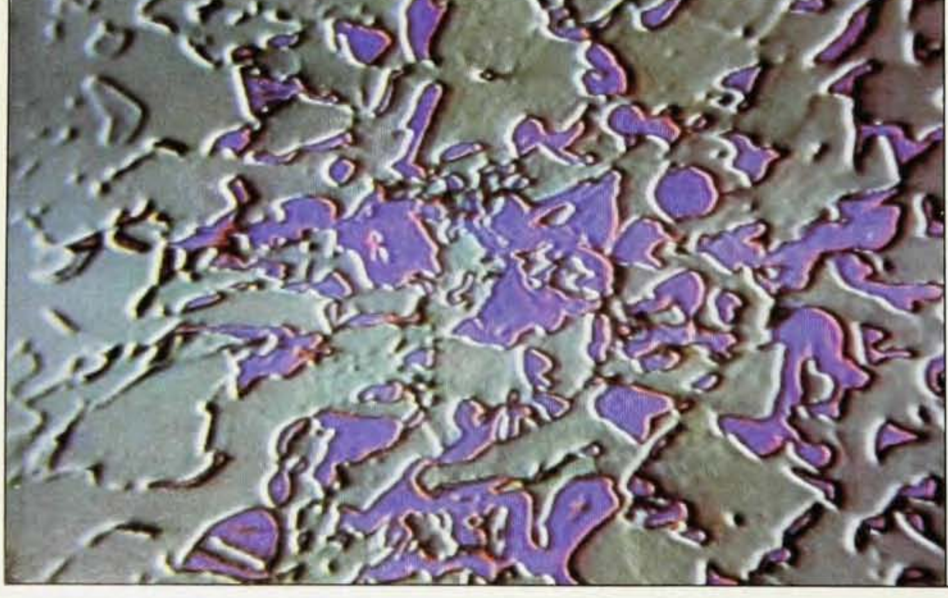

The Ductility Factor

# The Ductility Factor

*The use of high strength, low alloy steel has been severely limited, due to its low ductility. Now, a simple heat treating and controlled cooling process, developed at the General Motors Research Laboratories, has successfully enhanced formability properties without sacrificing strength.*




*A comparison of the stress-strain behavior of GM 980X, SAE 980X, and SAE 950X steels. GM 980X offers greater ductility at the same strength as SAE 980X, and greater strength at the same ductility as SAE 950X.*



*Scanning electron microscope micrograph of dual phase steel at a magnification of 2,000. The matrix (background) is ferrite; the second phase is martensite.*

FOR SOME TIME, automotive engineers and designers have been faced with the challenge of building cars light enough to get good gas mileage, but still roomy enough to comfortably transport four or five passengers. One technique which has proved fruitful is materials substitution.

Lighter materials, such as aluminum alloys and plastics and high strength, low alloy steels (HSLA), are being phased into new vehicle designs to replace certain plain carbon steel components. Each, though, has displayed inherent problems which limit its utilization.

Unlike plastics and aluminum, however, HSLA steels have the same density as plain carbon steel. Weight reduction is achieved because thinner sections (less volume) can be used to carry the same load. Since the formability (ductility) of most high strength steels is poor, though, it has only been possible to form simple shapes from it. This has severely limited the widespread use of HSLA steels (such as SAE 980X) for auto components. New hope for the increased utilization of HSLA steel has arisen, however, with the development of a new dual-phase steel, GM 980X, at the General Motors Research Laboratories.

General Motors is not in the steel business, and GM 980X is not a brand of steel. GM 980X is the designation for a type of steel displaying mechanical properties similar to those of the samples first formulated at the General Motors Research Laboratories. "GM" in the designation indicates that the steel is a variation of the conventional SAE 980X grade. In the standard SAE system for material identification, "9" designates that the steel is HSLA. "80" is the nominal yield strength of the metal in thousands of pounds per square inch. The "X" denotes a micro-alloyed steel—one containing on the order of 0.1% of other metals such as vanadium, columbium, titanium, or zirconium as a strengthening agent.

GM 980X displays the same strength, after strain hardening, as SAE 980X steel, but has far more ductility. This characteristic allows it to be formed into various complex shapes which were previously thought to be impossible with HSLA steels. The superior formability of GM 980X has substantially increased the utilization of HSLA steel in the manufacturing of automotive components such as wheel discs and rims, bumper face bars and reinforcements, control arms, and steering coupling reinforcements.

Dr. M.S. Rashid, discoverer of

the technique to make GM 980X steel, comments, "I was working on another project using HSLA steel, when I noticed that if SAE 980X steel is heated above its eutectoid temperature (the temperature at which the crystalline structure of metal is transformed) for a few minutes, and cooled under controlled conditions, the steel developed significantly higher ductility and strain-hardening characteristics, with no reduction in tensile strength."

FURTHER experiments proved that the key variables to make GM 980X are steel chemistry, heating time and temperature, and the rate at which the steel is cooled. Specimens of SAE 980X were heated in a neutral salt bath, then cooled to room temperature with cooling rates ranging from 5° to 14°C/sec. (9° to 26°F/sec.). Dr. Rashid notes, "We found that the maximum total elongation resulted when the cooling rate was 9°C/sec. (16°F), and the lowest total elongation resulted from the highest cooling rate (14°C or 26°F/sec.)."

GM 980X steel has a high strain-hardening coefficient or n value, accompanied by a large total elongation. The n value gives a measure of the ability of the metal to distribute strain. The higher the n value, the more uniform the strain distribution and the greater the resistance of the metal to necking (localized hour-glass-shaped thinning that stretched metals display just prior to breaking). Tests have proved that GM 980X distributes strain more uniformly than SAE 980X, has a greater resistance to necking, and thus has far superior formability.

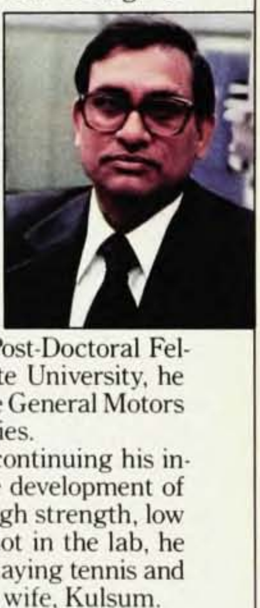
"The superior formability of GM 980X compared to SAE 980X steel appears to depend on the nature of two microstructural constituents, a ferrite matrix (the principal microstructural component) with a very high strain-hardening coefficient, and a deformable martensite (the other crystalline structure) phase. In the SAE 980X, failure occurs after the ferrite becomes highly strained, but when the GM 980X ferrite is highly strained, strain is apparently transferred to the martensite phase, and it also deforms.

"Therefore, voids leading to failure do not form until after more extensive deformation has occurred and the martensite phase is also highly strained. Obviously, the exact nature of these constituents must be important, and any variations in the nature of these constituents could influence formability. This is the subject of ongoing research."

Dr. Rashid's discovery represents a significant breakthrough in the area of steel development. His findings have opened the door to a new class of materials and have completely disproved the commonly held belief that high strength steel is not a practical material for extensive automotive application. "At GM, we've done what was previously thought to be impossible," says Dr. Rashid, "and now we're hard at work to find an even stronger and more ductile steel to meet the needs of the future."

## THE MAN BEHIND THE WORK

M.S. Rashid is a Senior Research Engineer in the Metallurgy Department at the General Motors Research Laboratories. He was born in the city of Vellore in Tamil Nadu (Madras), India, and attended the College of Engineering at the University of Madras–Guindy. He came to the United States in 1963 and was awarded a Ph.D. in Metallurgical Engineering from the University of Illinois at Urbana-Champaign in 1969.



After a three year Post-Doctoral Fellowship at Iowa State University, he joined the staff of the General Motors Research Laboratories.

Dr. Rashid is continuing his investigations into the development of even more ductile high strength, low alloy steels. When not in the lab, he enjoys relaxing by playing tennis and racquetball with his wife, Kulsum.

related to general relativity which is founded on geometrical concepts. The precise relationship is, however, quite subtle and is still being explored.

### On the method of theoretical physics

In his Herbert Spencer lecture of 1933, bearing the title that I have taken as the title of this section, Einstein analyzed the meaning of theoretical physics and its development. The following are some striking quotes from that lecture:[8]

> . . . the axiomatic basis of theoretical physics cannot be extracted from experience but must be freely created . . .
> Experience may suggest the appropriate mathematical concepts, but they most certainly cannot be deduced from it . . .
> But the creative principle resides in mathematics. In a certain sense, therefore, I hold it true that pure thought can grasp reality, as the ancients dreamed.

Was Einstein saying that fundamental theoretical physics is a part of mathematics? Was he saying that fundamental theoretical physics should have the tradition and style of mathematics? The answers to these questions are no. Einstein was a physicist and not a mathematician. Furthermore, he considered himself a physicist, and not a mathematician. He gave the reasons in a very penetrating way in his Autobiographical Notes:[1]

> . . . This was obviously due to the fact that my intuition was not strong enough in the field of mathematics in order to differentiate clearly the fundamentally important, that which is really basic, from the rest of the more or less dispensable erudition. Beyond this, however, my interest in the knowledge of nature was also unqualifiedly stronger; and it was not clear to me as a student that the approach to a more profound knowledge of the basic principles of physics is tied up with the most intricate mathematical methods. This dawned upon me only gradually after years of independent scientific work. True enough, physics also was divided into separate fields, each of which was capable of devouring a short lifetime of work without having satisfied the hunger for deeper knowledge. The mass of insufficiently connected experimental data was overwhelming here also. In this field, however, I soon learned to scent out that which was able to lead to fundamentals and to turn aside from everything else, from the multitude of things which clutter up the mind and divert it from the essential . . .

But he realized, from his own experience, and from the great revolutions in physics in the early years of this century, that although physics is and remains rooted in experimental laws, yet more and more, mathematical simplicity and beauty are playing a role in the formation of concepts in fundamental physics. He compared[7] theories that are "close to experience" with more mathematical ones:

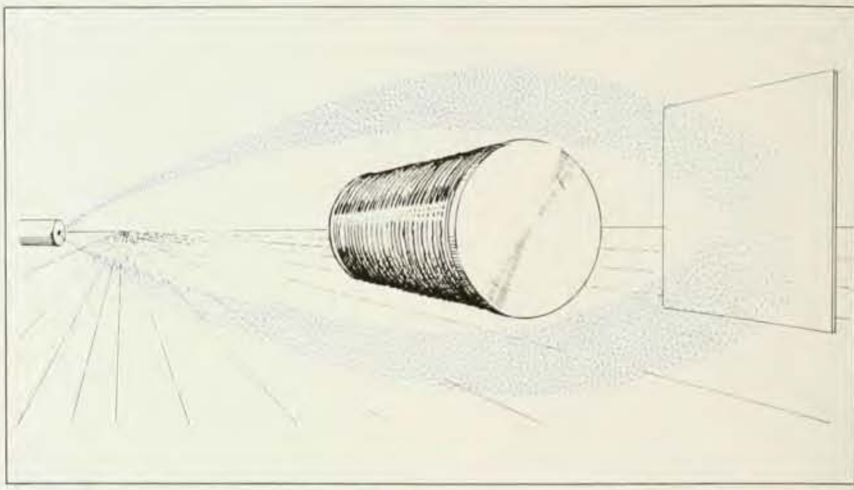
> On the other hand, it must be conceded that a theory has an important advantage if its basic concepts and fundamental hypotheses are "close to experience," and greater confidence in such a theory is certainly justified. There is less danger of going completely astray, particularly since it takes so much less time and effort to disprove such theories by experience. Yet more and more, as the depth of our knowledge increases, we must give up this advantage in our quest for logical simplicity and uniformity in the foundations of physical theory . . .

As a defense against misunderstanding by his fellow physicists, he pleaded:[3]

> The theoretical scientist is compelled in an increasing degree to be guided by purely mathematical, formal considerations . . . The theorist who undertakes such a labor should not be carped at as "fanciful"; on the contrary, he should be granted the right to give free rein to his fancy, for there is no other way to the goal.

The relationship between fundamental theoretical physics and mathematics is a fascinating subject. Perhaps I

## Global effects of gauge fields



A magnetic monopole of strength $g$ is a simple and natural idea. Dirac pointed out in 1931 that in quantum mechanics the magnitude of $g$ must be related to the electric charge $e$ through the condition $2eg/\hbar c$ = integer. It turns out that this condition is the simplest example of a very general and profound topological theorem: the Chern–Weil theorem. The next simplest example of the theorem is the so-called "instanton" solution of $SU_2$ gauge fields discovered in 1975. The 't Hooft–Polyakov monopole is a singularity-free solution for certain gauge fields, the existence of which is dependent on topological properties.

The Bohm–Aharonov experiment was proposed and performed in 1959–1960. As sketched in the figure above, electrons from the source go past a long solenoid, but are excluded from its inside. They produce an interference pattern on the screen. There is no electric or magnetic field outside of the solenoid, so the electrons suffer no local forces. Yet the interference pattern depends on the magnetic flux *inside* the solenoid, which shows that the effect of electromagnetism is not entirely local.

## Scale transformations

| Quantity | Value at first point | Value at neighboring point |
|---|---|---|
| coordinate | $x^\mu$ | $x^\mu + dx^\mu$ |
| field | $f$ | $f + (\partial_\mu f)dx^\mu$ |
| scale | 1 | $1 + S_\mu dx^\mu$ |
| scaled field | $f$ | $f + (\partial_\mu + S_\mu)f dx^\mu$ |

We use the notation $\partial_\mu = (\partial/\partial x^\mu)$ and the summation convention.

can be allowed to tell you a story at this point.

In 1975, impressed with the fact that gauge fields are connections on fiber bundles, I drove to the house of Shiing-Shen Chern in El Cerrito, near Berkeley. (I had taken courses with him in the early 1940's when he was a young professor and I an undergraduate student at the National Southwest Associated University in Kunming, China. That was before fiber bundles had become important in differential geometry and before Chern had made history with his contributions to the generalized Gauss–Bonnet theorem and the Chern classes.) We had much to talk about: friends, relatives, China. When our conversation turned to fiber bundles, I told him that I had finally learned from Jim Simons the beauty of fiber-bundle theory and the profound Chern–Weil theorem. I said I found it amazing that gauge fields are exactly connections on fiber bundles, which the mathematicians developed *without reference to the physical world*. I added, "this is both thrilling and puzzling, since you mathematicians dreamed up these concepts out of nowhere." He immediately protested, "No, no. These concepts were not dreamed up. They were natural and real."

Deep as the relationship is between mathematics and physics, it would be wrong, however, to think that the two disciplines overlap that much. They do not. And they have their separate aims and tastes. They have distinctly different value judgments, and they have different traditions. At the fundamental conceptual level they *amazingly* share some concepts, but even there, the life force of each discipline runs along its own veins.

* * *

*This article is adapted from a talk given at the Second Marcel Grossman meeting, held in Trieste, Italy, July 1979, in honor of the hundredth anniversary of the birth of Albert Einstein.*

## References


1. A. Einstein, "Autobiographical Notes" in *Albert Einstein, Philosopher-Scientist*, P. A. Schilpp, ed., Open Court, Evanston, Ill. (1949).
2. A. Einstein, J. Franklin Inst. **221**, 43 (1936).
3. A. Einstein in *Mein Weltbild*, Querido, Amsterdam (1934), translated in *Ideas and Opinions*, Bonanza, New York (1954).
4. For a short description of the history of gauge theory, see C.-N. Yang, Ann. N.Y. Acad. Sci. **294**, 86 (1977).
5. V. I. Fock, Z. Phys. **39**, 226 (1927).
6. F. London, Z. Phys. **42**, 375 (1927).
7. A. Einstein, Sci. Am., April 1950, page 13.
8. A. Einstein, *On the Method of Theoretical Physics*, Clarendon, Oxford (1933); reprinted in ref.3. □