INTRODUCTION

Yang and Mills' paper, originally published in *The Physical Review* [1] has been reproduced several times, such as in C. N. Yang's collection of Selected Papers [2]. Let us here briefly summarize the idea.

It had been established that the strong interactions appear to be fairly accurately invariant under isospin transformations. These transformations, which mathematically form the Lie group $SO(3)$, act as rotations in an internal space, such that the three pion states, (π^+, π^0, π^-), form a *vector*, and the nucleons, (p, n) form a *spinor*. If all states involved in a strong interaction process are rotated by the same angle around the same axis in isospin space, the laws of physics are observed to be practically the same as before.

This concept of isospin appears to make sense only if the states *after* an interaction are rotated the same way as the states *before* the interaction. We call such s symmetry a *global* symmetry: the $SO(3)$ rotation has to be performed everywhere in space-time in the same way. Yang and Mills pointed out that this seems odd. "It seems that this is not consistent with the localized field concept that underlies the usual physical theories." At this point, not everyone agrees. It is easy enough to write down equations for perfectly localized fields that show only global, continuous symmetries such as isospin. As often happens with obviously false statements, they were eagerly embraced by some enthusiastic followers; yet this cannot be the real reason why the theory became as important as it is today. I think that what Yang and Mills really wanted to say is this: global symmetries are fine, but could it not be so that there exist more delicate varieties of symmetries? Could one have an isospin-*like* symmetry that allows the $SO(3)$ rotation to be different at different points in space-time? There were at least two examples known of forces in Nature that indeed are associated with *local* symmetries: electromagnetism (where we have a local $U(1)$ invariance), and gravity (where the group of Lorentz transformations is replaced by general, *local* coordinate transformations).

And so it happened that, by asking a rather ill-posed question, Yang and Mills made a momentous discovery: electro-magnetism and gravity are

50 YEARS OF YANG-MILLS THEORY

http://www.worldscibooks.com/physics/5601.html

not the only force laws one can write down that have a local symmetry at their basis. One can start with any compact Lie group (of which $SO(3)$ may be regarded as the prototype), and build a generalized theory of electromagnetism, now called Yang–Mills theory.

In electromagnetism, the group of transformations considered consists of multiplying the wave functions $\psi(x)$ for charged particles by an arbitrary phase factor $e^{i\Lambda}$, and if Λ is allowed to be x-dependent, then the gradient of $\Lambda(x)$ is to be added to the vector potential field $A_\mu(x)$. In the Yang–Mills theory, we consider multiplets of fields $\psi^i(x)$, where i stands for some 'internal' index counting different species of particles. The group of gauge transformations consists of that of the unitary transformations

$$\psi^i(x) = S^i_j \psi'^j(x)\,, \tag{1}$$

or a subgroup of these transformations. In the simplest non-trivial case, we have the group $SU(2)$ of transformations on doublet states ($i = 1, 2$).

In analogy with the electro-magnetic case, we modify the field equations for these multiplets of fields, by replacing all gradients $\partial_\mu \psi^i$, wherever they occur in the field equations or in the Lagrangian, by the so-called covariant derivative:

$$D_\mu \psi^i(x) \stackrel{\text{def}}{=} \partial_\mu \psi^i(x) - ie B_{\mu}{}^i_j \psi^j(x)\,, \tag{2}$$

where $B_{\mu}{}^i_j$ is a new set of vector fields, transforming as the adjoint representation under the global gauge transformations. Invariance under *local* gauge transformations requires that

$$(\partial_\mu - ieB_\mu)\psi = S(\partial_\mu - ieB'_\mu)\psi'\,, \tag{3}$$

where we suppressed the indices.

Combining Eqs. (1) and (3), the isotopic gauge transformation on B_μ is obtained:

$$B_\mu = SB'_\mu S^{-1} + \frac{i}{e} S\partial_\mu S^{-1}\,. \tag{4}$$

Noticing that

$$[D_\mu, D_\nu]\psi = -ie(\partial_\mu B_\nu - \partial_\nu B_\mu - ie[B_\mu, B_\nu])\psi\,, \tag{5}$$

one finds that the field combination

$$F_{\mu\nu} = \partial_\nu B_\mu - \partial_\mu B_\nu + ie[B_\mu, B_\nu] \tag{6}$$

transforms covariantly under local gauge transformations:

$$F_{\mu\nu} = SF_{\mu\nu}{}'S^{-1}\,, \tag{7}$$

which of course can also be verified directly by applying Eq. (4). This field looks very similar to the Maxwell field $F_{\mu\nu}$, apart from the commutator term. The commutator term is one of the prime novelties of the Yang–Mills theory.

The B field transforms as the adjoint representation of the global part of the gauge group, so, in the simplest non-Abelian case of $SO(3)$, it forms a triplet. Such an isospin-vector, Lorentz-vector field was new to Yang and Mills, so they turned their attention to the physical significance of this field. One *could* regard it as a mere mathematical artifact; today we would call such a field a 'background field'. They emphasize that this would be physically unacceptable. If these fields exist at all, they must be endowed with dynamical properties and obey field equations. Fortunately, many clues could be found in Maxwell's equations, which we are only too familiar with. Take as our Lagrangian:

$$\mathcal{L} = -\tfrac{1}{4}\,\mathrm{Tr}\,(F_{\mu\nu}F_{\mu\nu})\,. \tag{8}$$

This is by far the simplest expression one can write down to generalize the Maxwell equations in a gauge-invariant way, and one has to accept the presence of the commutator term in Eq. (6). Indeed, when the issue of renormalizability is raised, it is the *only* acceptable kinetic term for the lagrangian. Adding the Dirac Lagrangian for a fermion doublet, with the gradients duly modified into covariant derivatives, one has

$$\mathcal{L} = -\tfrac{1}{4}\,\mathrm{Tr}\,(F_{\mu\nu}F_{\mu\nu}) - \bar{\psi}\gamma_\mu(\partial_\mu - ieB_\mu)\psi - m\bar{\psi}\psi\,. \tag{9}$$

The field equations are

$$\partial_\nu F_{\mu\nu} - ie[B_\nu, F_{\mu\nu}] = -J_\mu\,; \quad J_{\mu}{}_i^j = ie\bar{\psi}_i\gamma_\mu\psi^j\,, \tag{10}$$

$$\gamma_\mu(\partial_\mu - ieB_\mu)\psi + m\psi = 0\,. \tag{11}$$

It was realized, from the start, that this system of equations should be subject to quantization, and the quanta of the B field should be added to the existing spectrum of elementary particles. The B quanta would be expected to be exchanged between any pair of particles carrying isospin, generating not only a force much like the electro-magnetic force, but also a force that rotates these particles in isospin space, which means that elementary reactions involving the transmutation of particles into their isospin partners will result. A novelty in the Yang–Mills theory was that the B quanta are predicted to interact directly with one another. These interactions originate from the commutator term in the $F_{\mu\nu}$ field in Eq. (10), but one can also understand physically why such interactions have to occur: in contrast with ordinary photons, the Yang–Mills quanta also carry isospin, so they will undergo isospin transitions themselves, and furthermore, some of them are charged, so the neutral components of the Yang–Mills fields cause Coulomb-like interactions between these charged objects.

Two fundamental problems were duly recognized by Yang and Mills in their paper. First, we have the divergences. Primitive Feynman diagrams tend to lead to divergent integrals, so some kind of renormalization procedure is required. This problem was a familiar one, at that time, and it was known that it could be addressed, at least in the electro-magnetic case. At that time, however, it was generally believed that more advanced theories would be developed in the future, where, somehow, the difficulty of the infinities would be avoided altogether. Many theoreticians expected that these smarter theories would completely replace our 'primitive' quantum field theories.

A second problem was a novel one, and it was very disconcerting. The Yang–Mills equations resemble the Maxwell equations a bit too much: just like photons, the Yang–Mills B quanta would be massless particles. It seemed that there simply exists no mass term compatible with local gauge symmetry. The lightest particles with isospin are the pions, and they are copiously produced in high-energy collisions between nucleons. The only limiting factor appears to be the energy required for their production, and an essential part of this energy is in the mass of these pions. If Yang–Mills particles would be massless while carrying isospin, they would have to be produced even more abundantly than the pions. In reality, no such particles are ever seen to be produced. We only have the pions, and it was established, beyond any doubt, that they have zero spin, unlike the Yang–Mills bosons, which should have spin one. Thus, abundant experimental evidence appeared to indicate that Yang–Mills particles do not exist.

The only remedy to this problem appeared to be that, somehow, some dynamical mechanism generates mass terms for the Yang–Mills quanta. If one tries to write down such terms, they invariably violate gauge-invariance. This had already been noted by Pauli, as Yang recalls in his comments added in his collected papers. Could one combine this problem with the first, that of the divergent integrals? Questions of this sort were justly postponed for future generations to investigate. Indeed, we now have most of the answers to these questions, and, with the proper adjustments, Yang–Mills theory is now recognized as an essential ingredient in all our theories for sub-atomic particles. We know that these fields are there, that our theories not only look elegant with these fields incorporated, but that these fields *have* to be included in any system of particles as soon as the interactions tend to become strong.

Indeed, one of the reasons why, up till the early '70s, the notion of quantized fields was rejected by many experts in particle theory, was the so-called Landau ghost. It was the 'certainty' that, when extrapolated to higher energies, the interactions among the fields, due to non-linearities, would explode to infinity. Not only would this render any decent calculation hopelessly complicated, but it would even jeopardize the very foundation of such a theory, since one would have expected that, at least at the very tiniest distance scales, the interactions should be under control, to some extent. Well, today we still think that this objection holds, but only if one excludes Yang–Mills fields. The Yang–Mills field interactions tend to extenuate this divergence, through a mechanism called 'asymptotic freedom'. But, this would not be known for nearly twenty years to come. In 1954, most of those investigators who still did adhere to quantum field theory were either stubborn, or ignorant, or both. Serendipity? Perhaps.

C. N. Yang's Earliest Calculations

Yang now was so kind as to offer copies of his 1947 notes. They were clearly unfinished, and reproduced in Chapter 1. These were the notes of a graduate student still struggling with the concept of gauge invariance, a long way off from the masterpiece of 1954.

Robert L. Mills passed away on October 27, 1999. Mills had developed an excellent reputation as a mathematical physicist while studying at Columbia and at Cambridge University. He was still a PhD student while writing his paper with C. N. Yang. Chapter 1 also contains a letter written by Frank Yang, for *Physics Today.*

[1] C. N. Yang and R. L. Mills, *Phys. Rev.* **95**, 631 (1954).
[2] C. N. Yang, *Collected Papers 1945–1980*, with Commentary, W. H. Freeman and Co., San Francisco, 1983.