CHINESE JOURNAL OF PHYSICS VOL. 30, NO. 7 DECEMBER 1992

How to Quantize Yang-Mills Theory?

Hung Cheng
Depurtment of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

(Received August 15, 1992)

One spring morning thirty-five years ago, I woke up to learn that Chen Ning Yang, together with T. D. Lee, had just won the 1957 Nobel Physics Award. All of the Chinese friends I met that morning were surprised, excited, ecstatic, and proud, as Lee and Yang were the first Chinese to win this highest academic laurel. I remember that, just like everybody else, I was excited, ecstatic, and immensely proud, but unlike most people, I was not surprised. This is for a reason. About a year before this historic event took place, I ran into a schoolmate in the library of Taiwan University. As I stopped and chatted with him, he casually mentioned to me a quotation which he attributed to Oppenheimer. I was so astonished by the words that I still remember every one of them today: 五十年代的天下是楊振寧的 (the world of physics in the fifties belongs to Yang Chen Ning.) I have never been able to confirm the verity of his statements, perhaps they were no more than unfounded campus gossip. But to an undergraduate half a globe away from the U.S., these words made a deep impression. Therefore, on that memorable morning thirty-five years ago, when I heard the Nobel announcement, it seemed to me that a Nobel prize for someone who already owned the world was not excessive, and I made the wish that someday I would meet this great scientist and learn from him.

I realized the first part of my wish in the winter of 1963. There was a Christmas party given by the Chinese students of Princeton University, which Professor Yang and his wife graciously attended. It was at this party where I first saw Professor Yang in person. I said "I saw Professor Yang", because I don't think that he saw me. I was just one of the many nondescript smiling faces surrounding him, hanging onto his every word. He was telling us about his trip to Cambridge, inspecting the newly completed CEA.

I must admit that that time he did surprise me. He recited to us every parameter as well as many minute details about the accelerator, in the manner of an experimentalist, very differently from the other theoretical physicists I had come to know. But Professor Yang soon did something which surprised me even more: he began dancing with Mrs. Yang. I assure you that they were doing a most graceful waltz. I had not seen any theoretical physicist waltzing like that.

† Invited paper presented at the International Symposium in Honor of Professor C. N. Yang's 70th Birthday, July 10-11, 1992, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

I realized the second part of my wish in 1968, when I was collaborating with T.T. Wu on high-energy scattering. For about two years, T.T. Wu and I drove to Stony Brook frequently (actually, T.T. Wu did most of the driving, which I reflected upon with some horror in later years, the reason being that I was never sure if T.T. was handling the wheel or doing an integration in his head instead). We reported our latest findings on high-energy scattering to Professor Yang, and had some of the most stimulating discussions with him. I greatly appreciated his openness, his intellectual honesty, his willingness to listen to an approach which was not popular, as well as the unselfish way he bestowed his wisdom and time on us. The support and encouragement from Professor Yang were among the main reasons we were able to carry on and to reach the conclusion that the total cross section rises indefinitely as energy increases, in the face of almost universal objection.'

Today I would like to discuss another topic-a major triumph of theoretical physics in the twentieth century due mainly to Professor Yang. I am talking about the Yang-Mills theories (non-Abelian gauge theories). In particular, I would like to discuss how to quantize the field theories of Yang and Mills.

I came upon this problem in the early eighties, when I was teaching a course on quantum field theories. An important part of this course was of course Yang-Mills theories. After giving the classical description of Yang-Mills, I was eager to present to the students the quantum version of it. Like everyone else, I had, in years past, read about the standard method of quantizing Yang-Mills theories-path integration.

I considered this method powerful and elegant. However, the night when I was preparing the lectures on path integrals, I found myself puzzled over something. I was still unable to obtain an answer before the lecture was due, and was in some panic as I sought help from an expert, who assured me that the question I raised was trivial. Indeed, it was already answered in a paper written way back in 1967. I rushed over to the library to read that paper, only to find that the authors of this paper were completely oblivious of this difficulty. I went back to the expert, who listened to me patiently, and then suggested that I read another paper. This procedure repeated itself several times, and after many hours of agonizing reading, I was just as confused as I had been before. I talked to my thesis student, E.C. Tsai, about this and we both decided that, instead of reading more and more papers, we should start finding the answer on our own. It took much longer than we had anticipated. At the beginning, we tried to find the answer within the context of the path integration approach, as it had been accepted for almost two decades, and we did not doubt its validity. To our dismay we found more and more questions as time went on, and after several months of hard work, we became even more confused than before. Eventually, we had to abandon the path integral approach altogether and started anew. Needless to say, by that time the semester had already ended and I never did deliver the lectures planned for my students.

The new start is based on canonical quantization, an approach which had been tried and mostly abandoned in favor of the path-integration approach. This is because most people had considered it difficult to quantize Yang-Mills theories with the former approach. We have found that such is not the case. It is simple to quantize Yang-Mills theories canonically by methods that were generally known to Fermi[2] as far back as the 1930s. Furthermore, this approach provides information which is lacking in the path-integral approach but is essential for an in-depth understanding of Yang-Mills theories. It shows that the path integral formulation of Yang-Mills theories, particularly the Faddeev and Popov formalism, is not only heuristic and imprecise, it also gives erroneous answers in many cases. It shows that certain manipulations of path integrals, now used liberally, have no basis. Indeed, I consider path integral formulation of Yang-Mills theories in its present form incorrect.

In this paper, I would like to give an elementary exposition of the study, made by E.C. Tsai and myself, sketching only the central ideas and leaving a more technical version in another paper.[3]

To understand the meaning of quantization, consider first a classical particle of mass m in potential $V(q)$, where q is the coordinate of the particle. The Langrangian for the particle is

$$L = \frac{1}{2} m \dot{q}^2 - V \tag{1}$$

where $\dot{q} \equiv dq/dt$. The Euler-Lagrange equation obtained by extremizing the action $\int L dt$ is Newton's equation. This is the essence of the Lagrangian approach to classical mechanics.

Alternatively, we may start with the Hamiltonian

$$H = \frac{p^2}{2m} + V, \tag{2}$$

where p is the momentum, the variable conjugate to q defined by

$$p \equiv \frac{\delta L}{\delta \dot{q}} \tag{3}$$

The Hamilton equations derived from H are equivalent to Newton's equation. This is the Hamiltonian approach to classical mechanics.

To quantize, we make the correspondence of

$$H \to i\hbar \frac{\partial}{\partial t} \tag{4}$$

and

$$p \to -i\hbar \frac{\partial}{\partial q} \tag{5}$$

From (2), (4) and (5), we obtain the Schrödinger equation

$$\left(i\hbar\frac{\partial}{\partial t}\right)\Psi(q,t) = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial q^2} + V\right)\Psi(q,t), \tag{6}$$

where $\Psi(q,t)$ is the wavefunction with $|\Psi(q,t)|^2$ the probability density of finding this particle at q and at time t. This is known as canonical quantization and can be considered the Hamiltonian approach to quantum mechanics.

Alternatively, we may expresss the transition amplitude as the path integral

$$\int \mathcal{D}q e^{i\int L dt}. \tag{7a}$$

The integration is over all paths beginning at q_i and ending at q_f. The amplitude in (7a) refers to that of the particle traveling from q_i to q_f. If the particle is initially described by the wavefunction $\Psi_i(q)$ (instead of being located at the point q_i) and finally by the wavefunction $\Psi_f(q)$, the transition amplitude is, instead of (7a),

$$\int \mathcal{D}q\Psi_f^*(q_f)e^{i\int L dt}\Psi_i(q_i) \tag{7b}$$

This is the path integral approach of quantization, and can be considered as the Lagrangian approach. It has been proven that, for the Lagranian in the form of (1), the path integral approach gives the same results as the canonical quantization approach. On the other hand, we must keep in mind that these two approaches are *not* necessarily equivalent if the Lagrangian is not in the form of (1). A well-known example" is $L = m(t)\dot{q}^2/2$, that of a non-relativistic particle with a mass which is a function of time.

In a field theory, the counter part of q is the field ϕ and the counterpart ofy is

$$\frac{\delta L}{\delta\dot{\phi}}. \tag{8}$$

We may quantize this field theory in either of the two ways discussed above.

In a *gauge* field theory, to which a Yang-Mills theory belongs, there is a technical complication: the variable conjugate to A_0 does not exist, where A_μ = 0, 1, 2, 3, is the gauge field. This is because

$$\pi_0 = \frac{\delta L}{\delta\dot{A}_0} = 0 \tag{9}$$

That (9) holds is related to a property of gauge field theories: gauge invariance. This invariance implies that not all of the four components of A_μ are independent. Indeed, only two of the four components are. Which two should we choose as independent dynamical variables? In other words, what gauge should we adopt? The ready answer is that it doesn't matter, as all measurable quantities are independent of the choice. In other words, gauge invariance also implies that one gauge is just as good as the other-a statement known to every graduate student in physics.

This statement is somewhat misleading, and we need to be more precise, as we shall explain. Let us choose the temporal gauge of

‘4” =0.

This choice sidesteps the difficulty presented by (9). This is because that, with such a choice, A_0 disappears from the Lagrangian, and π_0 does not have to be defined. However, the Euler-Lagrangian equation obtained by varying A_0 is then missing from the dynamics, and must be put in by hand. In the scheme of canonical quantization it is customary to require that the quantum state $|\Psi>$ satisfies the supplementary condition of

$$G|\Psi >= 0 \tag{10a}$$

where

$$G = 0 \tag{10b}$$

is the Gauss law obtained by varying A_0.

In the scheme of path integration, the transition amplitude is taken to be

$$\int \mathcal{D}A_\mu e^{i\int L d^4x}. \tag{11}$$

As a result of gauge invariance, L remains to be of the same value if A_μ is replaced by A_μ', where A_μ and A_μ' are related by a gauge transformation. Therefore, the value of the integrand in (11) does not change as $A_\mu \to A_\mu'$. Since there are an infinite number of such A_μ' corresponding to a given A_μ, the integration in (11) is not convergent. Faddeev and Popov[5] tried to cure this infinity by inserting a delta function in the integrand of (11). The argument of the delta function is the gauge condition, which fixes the gauge. In addition, they put in a factor which is the Jacobian of transformation of the integration variables.

This approach has several difficulties:

(1) As we compare (11) with (7b), we find that the initial and the final wavefunctions are missing in (11). Therefore, (11) is in principle incomplete. In practice, the propagator for the gauge meson in the Feynman rules is not determined uniquely, if (11) is used. More precisely, the Lagrangian provides the partial differential equation satisfied by the propagator, while the initial and the final wavefunctions provide the boundary conditions. Indeed, the denominator of the propagator is usually taken to be $(k^2 + i\epsilon)$, (e.g. in the Feynman gauge) where k is the

momentum of the gauge meson, with the $i\varepsilon$ coming from the boundary conditions. How can the path integral formulation ascertain this term of $i\varepsilon$? This is the question I had the night I was preparing my lecture to my students.

To overcome this difficulty, it may seem that we merely need to insert the wavefunctions into (11). Unfortunately, the wavefunctions satisfying the supplementary condition (10) are not normalizable. Therefore, such an amended integral is infinite.

I have often been asked, "What is wrong with making (11) an ansatz, especially if we go to the Euclidean space and set the boundary values for the fields to be zero at infinity?" The answer is already obtained in the question: it is an ansatz.* Sometimes this ansatz provides the correct answer, but sometimes it does not. I have no objection to using the path integral expression for the cases in which its correctness has been established. But whenever we start with a theory, it would be unscientific to assume that (11) is correct.

(2) The path integral (11) does not incorporate operator ordering. Basically, the path integral is one involving the classical action summed over all paths. The problem is, the classical theory does not always define the quantum theory. An elementary example is the theory for a non-relativistic particle in two spatial dimensions. In polar coordinates, the classical Hamiltonian is

$$\frac{1}{2}\left(p_r^2 + \frac{1}{r^2}p_\theta^2\right).$$

As we all know, the quantum Hamiltonian is not equal to this expression with p_r replaced by $-i\hbar\,\partial/\partial r$. Rather, we must replace p_r^2 by $-\hbar^2\, 1/r\,\partial/\partial r\; r\,\partial/\partial r$; the order of the operators cannot be ignored.

While it is easy to take care of operator ordering in this elementary example, it is not easy to do so in non-abelian gauge field theories. As an example, the formulation of non-abelian field theories in the Coulomb gauge was first correctly done by Schwinger,[6,7,8] using the formulation of canonical quantization. The subsequent path-integral formulation of the same problem' differs in content from Schwinger's formulation: some interaction terms are missing. This difference is due to the failure of the path integral approach to take care of operator ordering, and, as it turns out, the resulting theory does not have relativistic covarance.

(3) The transition amplitude is not always given by (11). As we have mentioned, an elementary example can already be found in ordinary quantum mechanics: that of a particle of time-dependent mass. In this case, we all know that a determinant has to be added to the expression in (1). A more modern example can be found in the Skyrme model, the Lagrangian of which is in the form of

*It can be shown that the argument used by Abers and Lee to justify this ansatz fails, as the quantum state wavefunctions are not normalizable.

$$L = \frac{1}{2}\dot{\phi}_a M_{ab}\dot{\phi}_b - V(\phi_a, \vec{\nabla}\phi_a),$$

where the time derivatives of the fields are exhibited explicitly, with M_{ab} dependent on ϕ and $\nabla\phi$. Again, the determinant of M_{ab} must be inserted to (11). Since it depends on $\vec{\nabla}\phi$ but not on $\dot{\phi}$, the resulting path integral is not relativistically invariant, even though L is.[10]

This is one example in which a theory which is covariant at the classical level cannot be made so at the quantum level. While the necessity of modifying (11) is only too well-known, I bring up this example to prove that we tend to forget to do so. We do not always recognize that (11) is not the starting point of a quantum theory, and that it is derived from the canonical approach, valid only for theories with a Lagrangian of certain specified form. For a Lagrangian not in this form, it is not always clear what modification one must make without going through the canonical derivation. If one chooses to ignore this fact and used (11) regardless, then the theory obtained may not be unitary, to name one possible difficulty. Indeed, the path integral in (11) may not even be finite.

As we have mentioned, a test of the path integral formulation presented itself in the quantization in the Coulomb gauge. The canonical quantization in the Coulomb gauge was first done by Schwinger,[6] while the path-integral quantization in the Coulomb gauge was first done by Abers and Lee.' These two approaches give incompatible results. Some people told me that they ignored the result of Abers and Lee, if for no better reason than their trust in Schwinger's power of deduction. However, Abers and Lee cannot be so easily dismissed. For one thing, their result is on the basis of path integration. The path integration formulation also produces the result in the Feynman gauge, which is unitary and explicitly covariant, and was first suggested by no less respected a figure than Feynman himself. Furthermore, manipulation of path integrals popularly used by workers in the field also "proves" that the result in the Coulomb gauge via path integrals is equivalent to the result in the Feynrnan gauge. This equivalence can also be "proved" by diagrammatic techniques. On the strength of these two "proofs", some people may then be persuaded to discard Schwinger's result, if for no better reason than their trust in Feynman's power in making guesses.

Science should not be a popularity contest. Tsai and I therefore took it upon ourselves to find a definitive proof. We have found that Schwinger is right, but so is Feynman. Their results are in fact equivalent, despite the two "proofs" mentioned in the preceeding paragraph, which are both wrong. This means that path integral formulation of Yang-Mills theories cannot be relied upon to produce correct answers. It is a formalism without a valid basis.

Tsai and I arrived at these conclusions with the use of the theorem of equivalence, which we have proved diagrammatically: Let the gauge meson propagator be

$$-i\frac{g_{\mu\nu} - a_\mu(k)k_\nu + a_\nu(-k)k_\mu}{k^2 + i\epsilon}$$

let the ghost-ghost gauge meson vertex be

$$\left[a(q)\cdot q-1\right]q^{\mu}-q^{2}a^{\mu}(q),$$

where q is the outgoing ghost momentum, and let the ghost propagator be

$$\frac{}{q^{2}+i\epsilon}$$

then the on-shell scattering amplitude (as well as the matrix elements of gauge invariant operators) is independent of a_{μ}.

I am of the opinion that this theorem is as close to a statement on (local) gauge invariance as anything we know. As the reader may recall, I warned in the paragraphs following eq. (9) that we need to be more precise about the meaning of gauge invariance in the quantum theory. Let me now explain why. First of all, the concept of gauge in the quantum theory is different from that in the classical theory. I will give two arguments to illustrate this point: (a) The Feynman gauge is the most useful gauge in quantum theory. Yet it is not a gauge in the classical sense. This is because, instead of having A_{μ} satisfy a gauge condition, the four components of A_{μ} in the Feynman gauge are treated as independent. (b) In canonical quantization, we choose the temporal gauge $A_0 = 0$; however, as we further utilize the Gauss law explicitly and eliminate $\nabla\cdot A$, as Schwinger has done, we call the results those in the Coulomb gauge, not those in the temporal gauge!

With this arguments, I believe that the definition of gauge in quantum theory merely refers to the form we choose for the gauge meson propagator. As we change this form, the scattering amplitude is actually not invariant. It then becomes necessary to introduce in addition a fictitious field called a ghost field with the proper vertices and propogator as prescribed in the theorem of equivalence. The contribution of the ghost field cancels the unwanted contribution of the gauge field, and the scattering amplitude is invariant in this sense.

There is one thing to keep in mind when we apply the theorem of equivalence. The function $a_{\mu}(k)$ in this theorem may depend on one or more parameters. As one of these parameters takes a special value, the Feynman integrals may become singular. An example is the propagator of the gauge meson in the axial gauge, in which $a_{\mu}(k) = n_{\mu}/k\cdot n - k_{\mu}/2(k\cdot n)^2$. We may define $(k\cdot n)^{-1}$ as $\lim_{\varepsilon\to 0}(k\cdot n+i\varepsilon)^{-1}$, or some other way. The point is that the propagator is singular as $\varepsilon\to 0$. The correct scattering amplitude is obtained by first choosing a non-zero ε, and taking the limit $\varepsilon\to 0$ after the integration has been carried out, not before. In other words, the limit $\varepsilon\to 0$ should be taken for the Feynman integral, not the integrand. With this point in mind, we were able to prove that the Feynman rules in the Coulomb gauge obtained via path integration is incorrect, and that the Feynman rules in the axial gauge with the principal-value prescription is also incorrect.[12] In particular, there are contributions of ghost-loops in the latter set of Feynman rules. This means that, among others, the light-cone gauge formulation many people use to analyze inelastic electron scattering is incorrect. Such analyses should be re-done.

The quantum theory of Yang-Mills is so full of hazardous traps that even experts make common mistakes. I would therefore like to remind the readers of the following points:

(i) It is not always legal to use the classical Euler-Lagrange equation to eliminate the two superfluous components of A_μ. What we should watch out for is to make sure that operator ordering is properly kept in such an elimination. An illegal use of the Gauss law was the trap Abers and Lee fell into.

(ii) In the quantum theory, the equivalence of a gauge with another gauge must be proven. Therefore, we cannot, for example, just add to the Lagrangian some gauge-fixing terms together with the associated ghost terms, and consider the mission accomplished if the contribution to the unitarity condition by the ghosts cancels that by the longitudinal gauge mesons, as Feynman had done.[11] A demonstration of equivalence of two gauges must include a proof that the scattering amplitudes in these two gauges are equal. This remark applies to any new scheme of quantization people may devise in the future. It also applies to the quantization of other theories such as the theory of gravitation.

(iii) The quantum state wavefunctions depend on the gauge, and in a proof of equivalence of two gauges, the relationship betweeen the wavefunctions in these gauges must be specified. This relationship between the Feynman gauge and the temporal gauge is especially non-trivial. This is because there are ghost variables introduced in the Feynman gauge, which are absent in the original theory. We have found that the relationship between the wavefunctions in the Feynman gauge and those in the temporal gauge is[13]

$$|\psi_F> = e^{\theta}|\psi_W>, \tag{12}$$

where

$$\begin{aligned}
\theta &= \theta_1 + \theta_2, \\
\theta_1 &= -i\int d^3x\eta^a \frac{1}{\sqrt{-\nabla^2}}(\vec{\nabla}\cdot\vec{D}\xi)^a, \\
\theta_2 &= \frac{1}{2}\int d^3x(A_0^a - iA_L^a)\sqrt{-\nabla^2}(A_0^a - iA_L^a).
\end{aligned}$$

In the above $|\psi_F>$ and $|\psi_W>$ are the quantum states in the Feynman gauge and in the temporal gauge, respectively, η^a and ξ^a are hermitian ghost field variables (with a the group index), $A_L{}^a \equiv 1/\sqrt{-\nabla^2}\,(\nabla\cdot A^a)$, and D is the gauge-covariant gradient.

We note that $|\psi_F>$ given by (12) is normalizable, although $|\psi_W>$ which satisfies (10a) is not. (That $|\psi_W>$ is not normalizable, i.e., $<\psi_W|\psi_W> = \infty$, gives rise to some technical difficulty in extracting scattering amplitudes of the temporal gauge.)

I have recently extended the study to gravitation, in collaboration with S. P. Li. The quanturn theory of gravitation in the temporal gauge has been set up, but the work in the harmonic gauge is still in progress. Our study indicates that almost everything in quantum gravity has a counterpart in Yang-Mills theories. This suggests that general relativity may be classified as a special case of Yang-Mills theories.

Yang-Mills theories were first proposed in 1954, but it took almost thirty years before people were able to acquire a complete understanding of its quantum content. This is a stunning testimony to Professor Yang's creativity and intellectual power. Yet Professor Yang's wisdom is not restricted to science. Anyone who has ever had the privilege of talking to Professor Yang is invariably struck by the breadth of his interests and the depth of his thinking, and always walks away feeling enriched as a person. In the twenty-nine years since I first saw him, I have learned from him not only physics, but also various issues of life ranging from what are the great questions facing China, why my TV had good resolution, as well as how poorly I painted my kitchen wall. I admire him not just for his contributions to physics, which are truly revolutionary, but also for his compassion, his generosity, his honesty and integrity – in short, his whole being as a person. I would like to dedicate to him the following classic poem written hundreds of years ago, which expresses most aptly my feelings toward him:

雲山蒼蒼，江水泱泱，先生之風，山高水長。

I venture to translate this into English:

The green mountain fades high into the clouds,
The Yantze flows endlessly,
The humanity in you, Professor Yang,
Is as high as the mountain and as endless as the river.

Happy seventieth birthday, Professor Yang!

ACKNOWLEDGMENT

I would like to thank Dr. S. P. Li for a critical reading of this manuscript.

REFERENCES

1. See, for example, the third paragraph of Chapter 11 in *Basics of Functional Methods and Eikonal Models,* by H. M. Fried, Editorial Frontieres (1990).
2. E. Fermi, Rev. of Mod. Phys. 4,105 (1932).
3. H. Cheng, *How to Quantize Yang-Mills Theories Correctly,* (Proceeding of the Conference on the Interface between Mathematics and Physics, Academia Sinica, Taipei, Taiwan, 1992) 10.5 (1932); see also Proceedings of the Vienna Conference on Physical and Nonstandard gauges, Gaigg et.

al (Eds.) Springer-Verlang (1989).
4. T. D. Lee and C. N. Yang, Phys. Rev. 128,885 (1962).
5. L. S. Fadeev and V. N. Popov, Phys. Lett. **25B**, 29 (1967).
6. J. Schwinger, Phys. Rev. 127 (1962) and 130,402 (1963).
7. N. H. Christ and T. D. Lee, Phys. Rev. **D22**, 939 (1980).
8. H. Cheng and E. C. Tsai, Phys. Rev. Lett. 57,511 (1986).
9. E. S. Abers and B. W. Lee, Phys. Report. **9C**, 5 (1973).
10. H. Cheng, K. L. Chang, E. C. Tsai, and X. Zhu, *Difficulty with the Quantum Skyrme Method* (unpublished).
11. R. P. Feynman, Acta. Phys. Pol. 24,697 (1963).
12. H. Cheng and E. C. Tsai, Phys. Rev. Lett. **57**, 511 (1986).
13. H. Cheng and E. C. Tsai, Phys. Rev. **D40**, 1246 (1989).