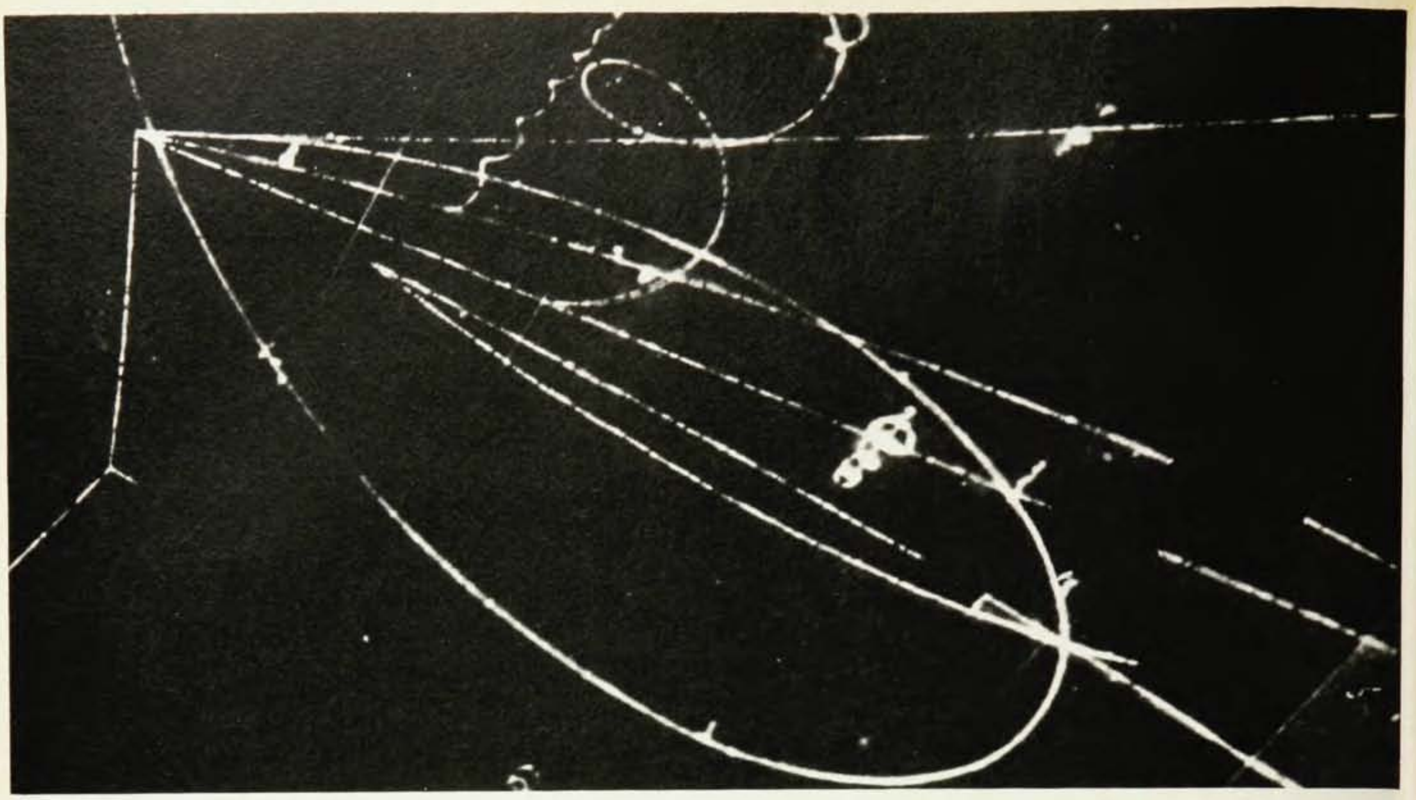


Quarks, atoms, and the 1/N expansion

Problems in quantum chromodynamics that are currently impossible to solve may have useful approximate solutions when one assumes that quarks can have a large number, N, of "colors" instead of three.

Edward Witten

Most particle physicists now believe that protons, neutrons, and other strongly interacting particles are built from more basic constituents known as "quarks" and "gluons," which interact according to the rules of a relativistic quantum field theory known as "quantum chromodynamics."

In many ways, this new theory is very similar to the much older quantum theory of electromagnetism (quantum electrodynamics, or QED). Quantum electrodynamics is the study of electric charges and the forces between them. Every electric charge, as we know from experience in macroscopic physics, creates an electromagnetic field which in turn exerts forces on other charges. Under ordinary conditions, these fields and forces are accurately described by Maxwell's equations and other macroscopic laws. However, in the microscopic world, when effects of quantum mechanics and relativity are important, an accurate description requires the use of quantum electrodynamics.

Quarks and gluons

Just as quantum electrodynamics describes electric charges, so quantum chromodynamics (QCD) describes the interactions of particles, quarks and gluons, that carry a new kind of charge, the color charge. A rough translation could be made between the two theories as follows:

$$\text{electrons} \leftrightarrow \text{quarks}$$
$$\text{photons} \leftrightarrow \text{gluons}$$

The quark, like the electron, is a point particle with spin 1/2 that satisfies a Dirac equation. The electron has electric charge and, similarly, the quark carries color charge. (The quark also carries electric charge, but this is so much weaker than the color charge that we can ignore it here.) The photon is a massless spin-one particle that transmits electromagnetic forces. The gluon is a similar spin-one particle that transmits color forces. Although in many ways this analogy is quite close, there are important differences between the two theories. I will have more to say about the differences later.

When quantum electrodynamics was developed, whole new vistas came within the reach of theoretical physics. Completely new processes, such as electron–positron pair creation and annihilation, were predicted and observed. Calculations of unprecedented accura-

---

Edward Witten is a Junior Fellow in the department of physics at Harvard University.

0031-9228 / 80 / 070038-06 / $00.50

Observation of baryon charm. This bubble-chamber photo shows the result of a neutrino–proton collision that produced a μ^- and a charmed Σ_c^{++}. The Σ decayed into a pion and a charmed Λ, which in turn decayed into a proton and four pions. Events such as this are evidence that quarks come in at least four "flavors," as well as three colors. (Photo from Brookhaven National Laboratory.)

---

cy became possible, for instance, the Lamb shift of atomic energy levels and the radiative corrections to the magnetic moment of the electron. Quantum electrodynamics is accepted as the correct description of interacting electric charges because its many predictions have been accurately confirmed by experiment.

For quantum chromodynamics the situation is rather different. Because of mathematical difficulties, it has proved very difficult to extract quantitative predictions from this theory. The quantitative predictions that can be made (which mainly concern the high-energy behavior, and are based on the renormalization group and "asymptotic freedom") are often hard to check experimentally. Most of the tests of the theory that have been carried out to date are semi-quantitative.

In contrast with quantum electrodynamics, which won acceptance because of calculations which agreed with experiment to five or six decimal places, quantum chromodynamics has won acceptance because of a few semi-quantitative tests, a large weight of circumstantial evidence, and many qualitative facts about the strong interactions which it apparently explains.

In principle, quantum chromodynamics is a comprehensive theory of the strong interactions, and describes all of the properties of the strongly interacting particles. Ideally, we would like to use quantum chromodynamics to calculate the masses, lifetimes, magnetic moments, scattering rates, and other properties of the strongly interacting particles—protons, neutrons, pions, kaons, and other baryons and mesons.

In principle, all of nuclear physics is contained within quantum chromodynamics (because nuclei are bound states of protons and neutrons) just as quantum electrodynamics underlies atomic, molecular, and solid-state physics (because atoms, molecules, and solids are made from charged particles).

In practice, because of mathematical difficulties, we are not able to extract from this theory its predictions about some of the most interesting questions, for instance, the predictions concerning the masses and quantum numbers of the particles.

In fact, we aren't even able to answer some basic, preliminary questions. The most basic preliminary question that we aren't able to answer concerns what is known as "quark confinement." As we noted above, the fundamental ingredients in quantum chromodynamics are quarks and gluons. Yet quarks and gluons apparently do not exist as individual particles! Many attempts to isolate a quark have failed. Most theorists now believe that quarks and gluons are permanently bound together or confined in bound states. These bound states are presumed to be the actually observed particles—protons, neutrons, and so forth. The permanent confinement of quarks and gluons into the observed particles is believed to be a consequence of extremely strong forces that develop, via quantum effects, as a result of the nonlinear nature of the theory.

Many circumstantial arguments indicate that quantum chromodynamics really behaves in this fashion, and some very interesting new numerical calculations support this view. However, because of the great complexity of this theory, there is no analytic derivation or physical argument showing that confinement of quarks really happens in quantum chromodynamics.

The present status of the strong interactions could be compared very roughly to the theory of turbulence. There are very good reasons to believe that turbulence is described by the Navier–Stokes equations of fluid mechanics. This has been believed for many generations. But such is the mathematical complexity of the Navier–Stokes equations that—despite a great deal of very ingenious work—not much is understood about turbulence from a quantitative point of view. In the same way, quantum chromodynamics has become fairly well established as the correct theory of strong interactions even though its predictions about some of the most basic questions are clouded by the mathematical complexity.

One interesting approach to trying to circumvent the mathematical difficulties of QCD is the "$1/N$ expansion," originally suggested by Gerard 't Hooft in 1974. The $1/N$ expansion is only one of a variety of approaches, and it too is limited by mathematical difficulties; nevertheless, the $1/N$ expansion has provided some interesting insights.

Because the reasoning behind the $1/N$ expansion is a little bit abstract, let me first describe the $1/N$ expansion in some simple situations in atomic physics.

Atomic physics

Let us consider the familiar Hamiltonian of the hydrogen atom:

$$H = \frac{p^2}{2m} - \frac{e^2}{r} \qquad (1)$$

This Hamiltonian consists of the free kinetic energy $p^2/2m$ plus a potential energy $-e^2/r$. One might think that for small e^2 one could understand the hydrogen atom by treating the potential energy as a perturbation.

This doesn't work because e^2 is not dimensionless and it doesn't make sense to say that e^2 is "large" or "small"—the value of e^2 just depends on the choice of units. After a rescaling $r \to tr$, $p \to p/t$, with $t = 1/me^2$, the Hamiltonian becomes

$$H = (me^4)\left(\frac{p^2}{2} - \frac{1}{r}\right) \qquad (2)$$

and one sees that the "coupling constant" e^2 appears only in an overall factor me^4 multiplying the whole Hamiltonian, which merely helps set the overall scale of energies. Therefore, except for the overall scale of lengths and energies (or times) the physics of the hydrogen atom—and atomic and molecular physics in general—is independent of e^2, and perturbation theory in e^2 is unenlightening.

Otherwise, atomic and molecular physics would be completely different! Weak coupling calculations of, for instance, the structure of the iron atom would be a basic technique in theoretical physics.

The hydrogen atom is a simple example of a problem without a free parameter, because it can be described by the reduced Hamiltonian

$$\hat{H} = \frac{p^2}{2} - \frac{1}{r} \qquad (3)$$

in which there is no free parameter. Likewise, other atoms and molecules can be described by reduced Hamiltonians with e^2 scaled out.

Without a free parameter there is no perturbation expansion. In such a case what can one do? Apparently we have no option except to look for an exact solution, or to seek a numerical solution on a computer. For atomic physics, this is reasonable, because one can solve the hydrogen problem exactly, and computer solutions are feasible for more complicated atoms.

But suppose—and this is the case in QCD and in many other analogous problems—that we were unable to diagonalize the Hamiltonian exactly, and that even a computer solution were formidably difficult or impossible. How then might we proceed?

To make progress, we must make an expansion of some kind. Since there is no obvious expansion parameter we must find a hidden one. That is, we must find a quantity one usually regards as given and fixed that we may treat as a free, variable parameter.

For instance, we may take a cue from the spectacular developments of this decade in critical phenomena. After decades in which the study of critical phenomena was thwarted by the absence of an expansion parameter, Kenneth Wilson and Michael Fisher suggested that to introduce a parameter, one should regard the number of spatial dimensions not as a fixed number, three, but as a variable parameter. They showed that critical phenomena are simple in four dimensions and that in $4-\epsilon$ dimensions critical phenomena can be understood by perturbation theory in ϵ. Even at $\epsilon = 1$, the original three-dimensional problem, this perturbation theory is quite successful.

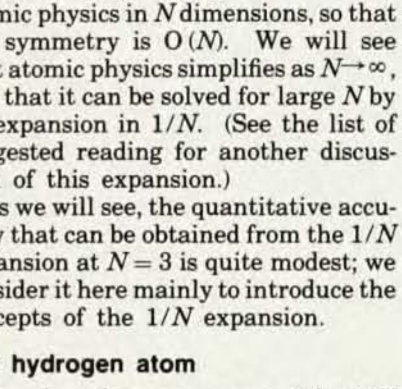
How, by analogy, can we create an expansion parameter "from thin air" for atomic physics?

Instead of studying atomic physics in three dimensions, where it possesses an O (3) rotation symmetry, let us consider atomic physics in N dimensions, so that the symmetry is O (N). We will see that atomic physics simplifies as $N \to \infty$, and that it can be solved for large N by an expansion in $1/N$. (See the list of suggested reading for another discussion of this expansion.)

As we will see, the quantitative accuracy that can be obtained from the $1/N$ expansion at $N = 3$ is quite modest; we consider it here mainly to introduce the concepts of the $1/N$ expansion.

The hydrogen atom

Now, how does one carry out the $1/N$ expansion for atoms? We will first consider hydrogen. Of course, the hydrogen atom is exactly soluble, so it will not be startling to see it solved by an expansion in powers of $1/N$. But, as we shall see, the expansion can also be carried out for more complex atoms, such as helium.



1

For hydrogen we would like to solve the Schrödinger equation

$$\left(-\frac{1}{2m}\nabla^2 - \frac{e^2}{r}\right)\psi = E\psi \qquad (4)$$

(in units in which $\hbar = 1$). For simplicity, let us consider the s-wave states only, although it is not difficult to include orbital angular momentum in this procedure. For s-wave states, ψ is a function of r only, and the Schrödinger equation can be written

$$\left[-\frac{1}{2m}\left(\frac{d^2}{dr^2} + \frac{N}{r}\frac{d}{dr}\right) - \frac{e^2}{r}\right]\psi$$
$$= E\psi \qquad (5)$$

(Actually, the "N" in equation 5 should be $N-1$. For large N, the difference between N and $N-1$ is negligible. Similar approximations will be made in some of the equations that follow.)

To eliminate terms with first derivatives from the Hamiltonian, we make the transformation

$$H \to r^{N/2} H r^{-N/2}$$

(this is equivalent to redefining the wavefunction by $\psi = r^{-N/2}\tilde{\psi}$) whereupon the Hamiltonian becomes

$$H = -\frac{1}{2m}\frac{d^2}{dr^2} + \frac{N^2}{8mr^2} - \frac{e^2}{r} \qquad (6)$$

If now we rescale the radial coordinate, defining $r = N^2 R$, then in terms of R

$$H = \frac{1}{N^2}\left(-\frac{1}{2mN^2}\frac{d^2}{dR^2} + \frac{1}{8R^2 m} - \frac{e^2}{R}\right) \qquad (7)$$

Apart from the overall factor of $1/N^2$, which only determines the overall scale of energy or time, the only N in this Hamiltonian is the N^2 that appears with the mass in the kinetic energy term. This is a Hamiltonian for a particle with an effective mass $M_{\text{eff}} = mN^2$, moving in an effective potential

$$V_{\text{eff}} = \frac{1}{8R^2 m} - \frac{e^2}{R} \qquad (8)$$

For large N the effective mass is very large, so that the particle simply sits in the bottom of the effective potential well—the quantum fluctuations are negligible. The ground state energy is simply the absolute minimum of V_{eff}.

To calculate the excitation spectrum, one may, for large N, simply make a quadratic approximation to the effective potential near its minimum, because the large effective mass ensures that the particle stays very close to the minimum of V_{eff}. The anharmonic terms in the expansion of V_{eff} around its minimum can be included as perturbations; this leads to an expansion in powers of $1/N$.

Consider next helium. If **x** and **y** are the positions of the two electrons (coordinates x_i and y_i), the Hamiltonian is

$$H = -\frac{1}{2m}\sum_i\left(\frac{d^2}{dx_i{}^2} + \frac{d^2}{dy_i{}^2}\right)$$
$$-2e^2\left(\frac{1}{x} + \frac{1}{y}\right) + \frac{e^2}{|\mathbf{x} - \mathbf{y}|} \qquad (9)$$

Let us once again consider s-waves, or states of zero angular momentum. The wavefunction of such a state is a function only of the rotationally invariant quantities, which (figure 1) are the distances x and y of the two electrons from the nucleus, and the angle θ between the two electrons. Acting on such a wavefunction, the Hamiltonian takes a rather awkward-looking form, but it can be simplified by manipulations similar to those we carried out for hydrogen. After a transformation

$$H \to (xy \sin\theta)^{n/2} H (xy \sin\theta)^{-N/2}$$

and a rescaling $x = N^2 X$, $y = N^2 Y$, the Hamiltonian becomes

$$H = \frac{1}{N^2}\left[\frac{1}{2mN^2}\left(-\frac{d^2}{dX^2} - \frac{d^2}{dY^2} - (1/X^2 + 1/Y^2)\frac{d^2}{d\theta^2}\right) + V_{\text{eff}}(X,Y,\theta)\right] \qquad (10)$$

where

$$V_{\text{eff}}(X,Y,\theta)$$
$$= \frac{1}{8m}\left(\frac{1}{X^2} + \frac{1}{Y^2}\right)\frac{1}{\sin^2\theta}$$
$$+ e^2(X^2 + Y^2 - 2XY\cos\theta)^{-1/2}$$
$$- 2e^2(1/X + 1/Y) \qquad (11)$$

Once again, for large N, the kinetic energy is suppressed, and the particle simply sits in the minimum of the effective potential. The ground state energy is the absolute minimum of V_{eff}, and one could compute the excitation energies from a quadratic approximation to V_{eff} near its minimum.

Notice that for large N the helium atom is very much like a molecule. There are well-defined "bond lengths," the equilibrium values of x and y, and a well-defined "bond angle," the equilibrium value of θ. These bond lengths and bond angles are well defined, with negligible quantum fluctuations, because the effective mass, $M_{\text{eff}} = N^2 m$, is large, just as in a molecule the bond parameters are well defined and the quantum fluctuations small because the nuclear masses are large.

As far as quantitative results are concerned, the accuracy that can be obtained at $N = 3$ is quite modest, as I said before. For hydrogen it is quite easy to show, by finding the minimum of the effective potential, that the ground state binding energy, to lowest order in $1/N$, is $(2/N^2)me^4$. At $N = 3$, this is $\frac{2}{9}\,me^4$, while the actual value in three dimensions is $\frac{1}{2}\,me^4$.

As L. D. Mlodinow and N. Papanicolaou have pointed out, the accuracy can be improved a great deal by including several terms in the series. The first few terms in the $1/N$ expansion give a binding energy for hydrogen of

$$E = (\tfrac{1}{2}me^4)\,4N^{-2} \times\left(1 + \frac{2}{N} + \frac{3}{N^2} + \cdots\right) \qquad (12)$$

The factor in brackets should equal 1 at $N = 3$, but the sum of the first three terms in the series is $\frac{8}{9}$. The exact binding energy of hydrogen, known from the exact diagonalization of the Hamiltonian is, incidentally,

$$E = (\tfrac{1}{2}me^4)\,4(N-1)^{-2} \qquad (13)$$
$$= (\tfrac{1}{2}me^4)\,4N^{-2} \times\left(1 + \frac{2}{N} + \frac{3}{N^2} + \cdots\right)$$

For helium it is necessary to minimize the effective potential numerically. A simple calculation shows that the equilibrium "bond length" is approximately

$$x = y = 0.1517\,N^2\,(1/me^2) \qquad (14)$$

while the "bond angle" of the atom, to lowest order in $1/N$, is approximately

$$\theta = 95.30^\circ \qquad (15)$$

For the binding energy of the atom one finds $(10.95/N^2)(me^4)$, which at $N=3$ is $1.217\ (me^4)$. The observed binding energy in Nature is $2.90\ (me^4)$. The agreement is clearly not very good. It isn't known how much one could improve it by including higher-order terms.

One can also use the $1/N$ expansion to calculate quantities other than binding energies. For instance, it is interesting to try to calculate the magnetic susceptibility of the helium atom. A standard quantum mechanical formula relates the magnetic susceptibility χ (defined in terms of the energy shift in a magnetic field by $\Delta E = -\tfrac{1}{2}\chi H^2$) to the expectation value of the square of the distance from the electrons to the nucleus. To lowest order in $1/N$, this is just the square of the "bond length" of equation 14, so very little additional calculation is needed. One finds that to lowest order in $1/N$ the susceptibility is

$$\chi = -(0.1517)^2 N^3 \left(\frac{e^2}{\hbar c}\right)^2 \left(\frac{\hbar^2}{me^2}\right)^3 \qquad (16)$$

At $N=3$ this is

$$\chi = -3.31 \times 10^{-5} \left(\frac{\hbar^2}{me^2}\right)^3 \qquad (17)$$

while the experimental value is

$$\chi = -2.11 \times 10^{-5} \left(\frac{\hbar^2}{me^2}\right)^3 \qquad (18)$$

The accuracy obtained from the first term in the $1/N$ expansion is, again, rather modest.

Quantum chromodynamics

Now let us return to our original subject of interest, quantum chromodynamics. The basic fact that makes QCD a difficult theory to understand is that in QCD, as in atomic physics, the coupling constant can be scaled out of the problem. Because the coupling constant can be scaled out, perturbation theory (although it is very important and is the source of most of what we know about this theory) does not illuminate the key unsolved questions, such as understanding confinement and predicting the mass spectrum.

In atomic physics, it is obvious, on dimensional grounds, that the coupling constant can be scaled out of the problem. In quantum chromodynamics this is far from obvious, but it is true. Scaling out the coupling constant in QCD involves the use of a mathematical machinery known as the "renormalization group" (which has also been used in statistical mechanics to describe anomalous scaling laws). The fact that the coupling constant can be scaled out in QCD is probably one of the most subtle discoveries to have been made in particle physics. It was properly appreciated only after the discovery of "asymptotic freedom" (the weakness of the QCD interaction at very high energies) in 1973, and it played a great role in pinpointing quantum chromodynamics as the correct theory.

In quantum chromodynamics the probability amplitude for a quark to emit a gluon (which is one of the basic processes in this theory) is proportional to the "color charge" g of the quark. This quantity, which is rather analogous to the electric charge of the electron in quantum electrodynamics, is known as the QCD "coupling constant." The renormalization group can be used to show that this coupling constant g does not have a characteristic value; rather, its value depends on the energy scale of the processes one considers—or on the units in which one measures energy.

This can be a great advantage. The variable nature of the QCD coupling constant has made possible reliable predictions concerning the high-energy behavior of this theory. These predictions are the basis for most attempts to test QCD experimentally.

But the variable nature of the QCD coupling constant also means that perturbation theory cannot answer the central unsolved questions. Because the coupling constant is variable and depends on energy, its numerical value, just as in atomic physics, can be absorbed by properly defining the overall scale of energies. As in atomic physics, nothing depends on the coupling constant except this overall scale of energies, and therefore perturbation theory cannot answer such unsolved problems as explaining confinement and predicting the mass spectrum.

To solve these problems, we must somehow circumvent the apparent absence in QCD of a relevant expansion parameter. The $1/N$ expansion of QCD, originally suggested by 't Hooft in 1974, is an attempt to do this.

To describe the $1/N$ expansion, it is necessary to describe QCD in somewhat more detail. In QCD there is actually not just one type of quark, but three types or "colors" of quarks. These are sometimes denoted as "red," "blue" and "green" quarks (the primary colors); here we will label the quark colors by number: q^i is the quark of type (color) i, where i may equal 1, 2, or 3.

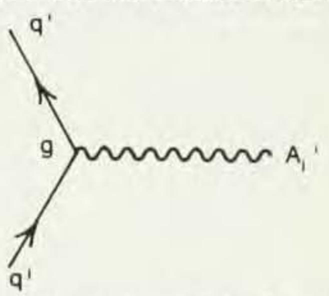
Historically, the fact that there are three colors was discovered from the fact that the proton (or neutron) seems to contain three quarks (one of each color). Apart from color, quarks can also be distinguished by another property known whimsically as "flavor" (up, down, strange, charm, etc.). The quark flavor is very important in weak and electromagnetic interactions but unimportant for strong interactions, so we will ignore the existence of flavor and simply think of quarks as coming in three colors.

Each color of quark participates equally in strong interactions. This is expressed mathematically by saying that there is a symmetry group, denoted by SU(3), relating the three kinds (colors) of quark; mathematically, the group has properties similar to the rotation group. The fact that the symmetry is SU(3) and not SU(N) for some other N expresses the fact that the number N of quark colors is equal to 3.

One of the basic processes in QCD is the process in which a quark emits a gluon

$$\text{quark} \rightarrow \text{quark} + \text{gluon}$$

The initial and final quarks have three color states each; the gluon field is a 3×3 matrix $A_j^{\,i}$, in color space. Thus, the most general allowed process is that a quark of type q^i emits a gluon of type $A_j^{\,i}$ and becomes a quark of type q^j



2

Because the 3×3 matrix for the gluon field is required to be a traceless matrix, it has not 9 but only 8 independent components. This fact plays no role in the large-N expansion, and we may simply think of the gluon field as a 3×3 matrix.

At this point enters the innovation that was introduced by 't Hooft in 1974. As in the atomic physics example considered in the previous section, 't Hooft wanted to introduce a free expansion parameter in a theory that appeared to have no such parameter. He suggested that one generalize from three quark colors to N colors. We still label the quarks as q^i, but now i runs, not from 1 to 3, but from 1 to N. The symmetry group becomes SU(N) rather than SU(3). The gluon field is now an $N\times N$ rather than 3×3 matrix.

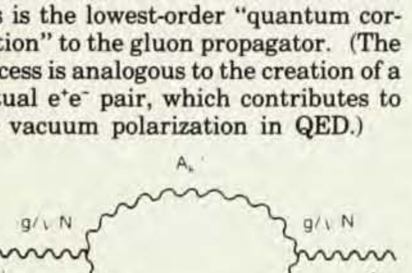
This step is similar to the method, in atomic physics, of generalizing from 3 to N dimensions and from an O(3) to O(N) rotation symmetry.

As in the case of atomic physics, the generalization has its benefits: 't Hooft showed that quantum chromodynamics also simplifies for large N. The simplification has led to some rather interesting insights, even though so far we are still very far from being able to use the $1/N$ expansion to solve the problems we would most like to solve.

The basic reason that quantum chromodynamics simplifies for large numbers of colors is very simple. For large N, the gluon field $A_j{}^i$ has N^2 components (actually N^2-1, but the difference between N^2 and N^2-1 is unimportant if N is large). Therefore there are N^2 (or N^2-1) gluon particles which might appear as intermediate states in Feynman diagrams.

For large N, the Feynman diagrams contain large combinatoric factors, arising from the large number of possible intermediate states. Only the diagrams with the largest possible combinatoric factors need to be included when N is large. So only a subclass of diagrams is relevant, and the theory simplifies.

To see how this works in more detail, let us consider the lowest-order contribution to the gluon "vacuum polarization." One gluon can split into two gluons which then recombine into one; this is the lowest-order "quantum correction" to the gluon propagator. (The process is analogous to the creation of a virtual e^+e^- pair, which contributes to the vacuum polarization in QED.)



3

It is not hard to see that for any choice of initial and final states, there are N possibilities for the intermediate state in the diagram. If the initial state gluon is of type $A_j{}^i$, it can split into a pair of gluons, one of type $A_k{}^i$ and one of type $A_j{}^k$, where k is arbitrary. Since there are N possible values of k, there are N possibilities for the intermediate state.

In quantum mechanics one is always required to sum over all possible intermediate states. Therefore the contribution of this diagram is proportional to a combinatoric factor of N from a sum over N different intermediate states.

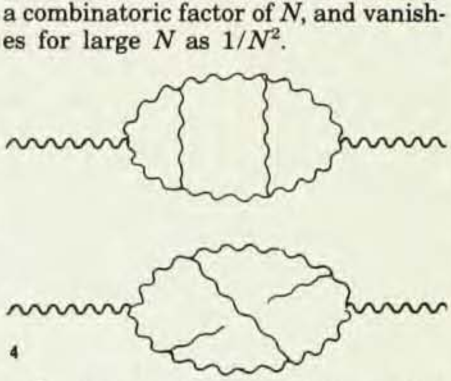
If quantum chromodynamics is to have a smooth limit for large N, this factor of N must somehow be canceled. If the correction to the propagation of the gluon were to diverge for large N in proportion to N, all the other calculations would also give divergent results, and we could not construct a useful quantum chromodynamics for large numbers of colors.

There is only one way to cancel the combinatoric factor of N. We must remember that in our calculations, for each of the two vertices where a gluon splits into two gluons or where two gluons recombine into one, there is a factor of the coupling constant. If we choose the coupling constant to be $g/\sqrt{N}$, where g is to be held fixed as $N \to \infty$, then the factors of N cancel out of figure 3 because $N(g/\sqrt{N})^2 = g^2$, independent of N. So the choice of the coupling as $g/\sqrt{N}$ gives a smooth limit to the one-loop diagram of figure 3.

Moreover, this is the only choice of coupling constant that gives a smooth limit to this one-loop diagram. With any other choice, the coupling constant factor will not cancel the combinatoric factor, and the large-N limit of QCD will not exist.

But choosing the coupling constant as $g/\sqrt{N}$ is a fateful choice. Complicated diagrams will have factors of $g/\sqrt{N}$ at each vertex, and so will vanish for large N unless, like the simple one-loop diagram, they have combinatoric factors large enough to cancel the factors at the vertex. It turns out that a certain class of diagrams with the largest combinatoric factors survives for large N, while all other diagrams vanish as $N \to \infty$.

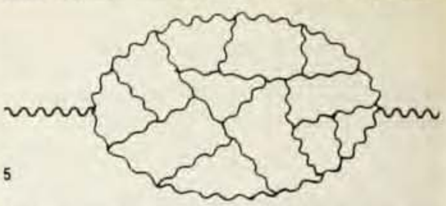
For example, the three-loop diagrams in figure 4 each have factors of $(g/\sqrt{N})^6$ from the six vertices. The first turns out to have a combinatoric factor of N^3 from summing over the various intermediate states. Since $N^3(g/\sqrt{N})^6 = g^6$, which is independent of N, the first diagram survives and has a smooth limit for large N. However, the second diagram in figure 4 has only a combinatoric factor of N, and vanishes for large N as $1/N^2$.



4

The general class of diagrams that survives for large N was originally determined by 't Hooft. The diagrams that survive are the "planar" diagrams. A planar diagram is a diagram that can be drawn in the plane with no two lines crossing. The first diagram of figure 4 is a planar diagram, and survives as $N \to \infty$. But the second diagram in figure 4 is not a planar diagram, since two gluon lines cross at the center of the diagram, and it vanishes for large N.

The planar diagrams are a vast class of diagrams. A typical planar diagram with many loops is indicated in figure 5. Summing the planar diagrams is clearly a very ambitious task. And this—or something equivalent—would have to be accomplished if the $1/N$ expansion is to become a really important tool for theoretical particle physics.



5

Since 1974, when t'Hooft first proposed the $1/N$ expansion, this problem has been the subject of some fairly intensive study. We have been able to gain a certain amount of insight, but so far there has not been much tangible progress.

If one could sum the planar diagrams, one could predict the particle lifetimes, masses, magnetic moments, and everything else, to lowest order in $1/N$. One might think that without being able to sum the planar diagrams, one could learn very little from the $1/N$ expansion. However, at this point we encounter a surprise.

Even without being able to sum the planar diagrams, we can get a considerable insight into the phenomenology of quantum chromodynamics. The reason for this is that there are certain "selection rules" or qualitative properties that are preserved, diagram by diagram, by each of the planar diagrams, but which are violated by the nonplanar diagrams.

These selection rules can be regarded as predictions or tests of the $1/N$ expansion, and are observed to be fairly well satisfied in Nature. If, therefore, we assume that the $1/N$ expansion is a good approximation to the real world, which has $1/N = 1/3$, then we obtain by means of the $1/N$ expansion an understanding of some qualitative properties of the strong interactions that are not otherwise well understood. Conversely, the success of the selection rules encourages us to believe that the $1/N$ expansion is relevant to the real world.

The selection rules of the $1/N$ expansion are roughly analogous to selection rules in atomic or nuclear physics. Even without a full understanding of the wavefunctions or interactions of nuclei, one can predict, for example, that electric dipole transitions are faster than quadrupole transitions on the basis of general properties of the interactions which we do understand. Likewise, even though we cannot sum the planar diagrams, we can make certain predictions about the behavior of QCD as $N \to \infty$ simply from the fact that it is the planar diagrams that dominate.

For a full treatment, the reader is referred to the literature. But let me give a few examples here.

In nature there are observed three spin-one mesons of different charge, the ρ^+, ρ^0, and ρ^-, which have approximately the same mass (about 770 MeV). We understand why their masses are equal: This follows from the fact that the strong interactions remain invariant under rotations of isospin, and the three mesons differ only by the orientation, not the magnitude, of their isospin.

But there is a fourth spin-one particle, the ω, which has almost the same mass (about 784 MeV) as the ρ^+, ρ^0, and ρ^-. The ω is not related to the ρ by isospin; it is an "isosinglet." There is no reason of symmetry for the ω to have equal mass with the ρ, so why are the masses so close?

The $1/N$ expansion provides a possible answer. We are not able to sum the planar diagrams so as to determine what masses the ρ and ω have to lowest order in $1/N$. But it is easy to show that the planar diagrams do not distinguish the ρ from the ω and therefore give them the *same* mass. So the ρ and ω have equal masses to lowest order in $1/N$. That their masses are approximately equal in the real world indicates that the $1/N$ expansion may be a good approximation even at $1/N = 1/3$.

As another example, let us consider the B meson, which weighs 1237 MeV and decays (figure 6) to four pions:

$$\mathrm{B} \rightarrow \pi\pi\pi\pi \qquad (20)$$

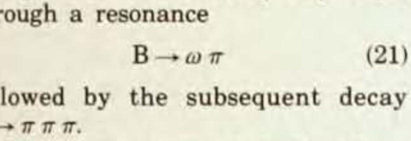
It is possible for this decay to proceed through a resonance

$$\mathrm{B} \rightarrow \omega\pi \qquad (21)$$

followed by the subsequent decay $\omega \rightarrow \pi\pi\pi$.

Now the question arises, what fraction of B decays are resonant and what fraction non-resonant? We do not know how to sum the planar diagrams so as to predict the rate of either decay 20 or 21. However, it is easy to show, by counting powers of N, that the probability amplitude for process 21 scales with N like $1/N^{1/2}$, while that for 20 scales like $1/N^{3/2}$. Therefore, we predict that, at least for large N, process 21 will dominate. There should be of order N^2 resonant decays for every one non-resonant decay. In fact, the resonant process is observed to dominate by a large margin. Again, this encourages us to hope that the $1/N$ expansion is a good approximation.

6



The last example that we will consider here is this: why does there not exist in Nature a meson–meson bound state? Why is there, for instance, no $\pi^+\pi^+$ bound state, which would be a meson of charge two, or a K^0K^0 bound state, which would have charge zero and strangeness two? (Such states are called "exotics" by particle physicists.)

We can argue that such bound states should be absent at least for large N, because it is simply not possible to draw a planar diagram describing the scattering, for example of two π^+ mesons. Any diagram describing the scattering of two π^+ mesons must involve nonplanar gluon exchange, and would be of order $1/N^2$. Although we cannot sum the diagrams with nonplanar gluon exchange, we *can* say that the potential of interaction between two π^+ mesons is of order $1/N^2$. This potential is, therefore, weak if N is large.

The $\pi^+\pi^+$ potential is a short-range potential, because it involves exchange of strongly interacting particles, all of which have masses. Because a weak, short-range potential will not produce a bound state, there are no $\pi^+\pi^+$ bound states, at least if N is large. This argument doesn't depend on any special properties of pions and is valid in all exotic channels. The absence of exotic bound states is again an encouraging sign for the $1/N$ expansion.

In short, the $1/N$ expansion has made a respectable contribution to our understanding of the phenomenology of quantum chromodynamics. The success of the selection rules gives basis for hope that the $1/N$ expansion may be a reasonably good approximation in QCD (better, let us hope, than in atomic physics!). And finally, the $1/N$ expansion also seems to be a reasonable framework for thinking about the main unsolved problems of quantum chromodynamics.

Suggested reading

- **The 1/N expansion** developed from the spherical model of T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952), and was formulated in a more modern version by H. E. Stanley, Phys. Rev. **176**, 718 (1968).
- **QCD:** The first application to QCD was by G. 't Hooft, Nucl. Phys. **B72**, 461 (1974); **B75**, 461 (1974).
  Some recent reviews of the $1/N$ expansion in QCD, from different points of view, are:
  G. Rossi and G. Veneziano, Nucl. Phys. **B123**, 507 (1977); G. F. Chew and C. Rosenzweig, Phys. Reports **41C**, No. 5 (1978); A. de Rújula, talk at the 1979 European Physical Society Conference (CERN preprint); S. Coleman, Erice Lectures, 1979 (to appear); E. Witten, Nucl. Phys. **B160**, 57 (1979).
- **Planar diagrams:** There is a certain amount of literature by now on the problem of trying to sum planar diagrams. One particularly interesting paper is E. Brezin, C. Itzykson, G. Parisi, and J. B. Zuber, Comm. Math. Phys. **59**, 35 (1978).
- **Atomic physics:** A recent study of the $1/N$ expansion in atomic physics is L. D. Mlodinow and N. Papanicolaou, "SO(2,1) Algebra and the Large N Expansion in Quantum Mechanics," Berkeley preprint, November 1979. ☐