# The Lattice Theory of Quark Confinement

*The force between quarks in a particle such as the proton has been simulated by imposing a discrete lattice on the structure of space and time. The results suggest why a free quark cannot be isolated*

by Claudio Rebbi

The development of quantum mechanics put to rest the uncritical acceptance of the idea that elementary particles are the "building blocks" of matter. Often such particles do not act like hard, impenetrable blocks at all, and in many circumstances they must be described as waves. Until recently, however, it still seemed that elementary particles were like building blocks at least to the extent that each particle could in principle be isolated and observed as an individual entity. The electron, the proton and the neutron, for example, can be separated from one another and observed as individual packets of waves. Even this limited interpretation of the building-block metaphor fails in the case of the quark, the supposed constituent of the proton, the neutron and many related particles. Apparently a quark cannot be isolated; although there is abundant evidence for the existence of quarks and antiquarks bound together in pairs and triplets, an individual, or free, quark has never been observed.

As the experimental evidence has accumulated, it has begun to seem that if quarks are real particles at all, they must be permanently bound within nuclear particles. Any theory of quark interactions ought to account for this phenomenon, which is called quark confinement. It is easy to construct pictorial models of particles such as the proton in which the constituent quarks are confined. For example, the quarks can be thought of as being fastened to the ends of an unbreakable string; they are then free to move about within the volume defined by the length of the string but cannot wander away from one another. It is a formidable task, however, to formulate a theory that can account for the permanent binding of quarks and the structure of nuclear particles without violating the constraints imposed by the theory of relativity, quantum mechanics and the principles of ordinary causality.

After several years of both experimental and theoretical investigations most particle physicists are confident they at last have a theory capable of explaining the interactions of quarks. One reason for confidence is that the theory is a mathematical analogue of the most successful physical theory ever developed: the quantum theory of interactions in an electromagnetic field. The latter theory is called quantum electrodynamics, or QED, and the conceptual similarity of the theory of quark interactions to QED is reflected in the name of the new theory: quantum chromodynamics, or QCD.

The difficulty that has delayed full acceptance of QCD is that its mathematical complexity makes any rigorous, analytical prediction from it exceedingly difficult. Indeed, up to now the most eagerly sought prediction of QCD, namely the demonstration of quark confinement, has not been forthcoming. Recently, however, my colleagues and I at the Brookhaven National Laboratory have applied mathematical methods that rely heavily on the capabilities of the digital computer to the problem of confinement, and a numerical breakthrough has been achieved. Because the method explores the implications of QCD by making a series of increasingly accurate approximations, the results of the calculations do not carry the same force as a logical deduction from accepted first principles. Nevertheless, the numerical results have provided strong evidence for the confinement of quarks.

The framework for the calculations is a pioneering suggestion made in 1974 by Kenneth G. Wilson of Cornell University. Wilson proposed that QCD be formulated on a cubic lattice, an array that divides space and time into discrete points. The lattice is only an approximation to real space-time, but it allows calculations to be made that would otherwise be impossible. As the mesh of the lattice is made progressively finer the values of physical quantities defined on the lattice converge to the values QCD would predict for them in ordinary, continuous space and time. Our numerical approximations show that for an extremely fine lattice, confinement is a consequence of QCD. For reasons that will become clear both QCD and QED are called gauge theories; the computational method I shall describe is therefore called a lattice gauge theory.

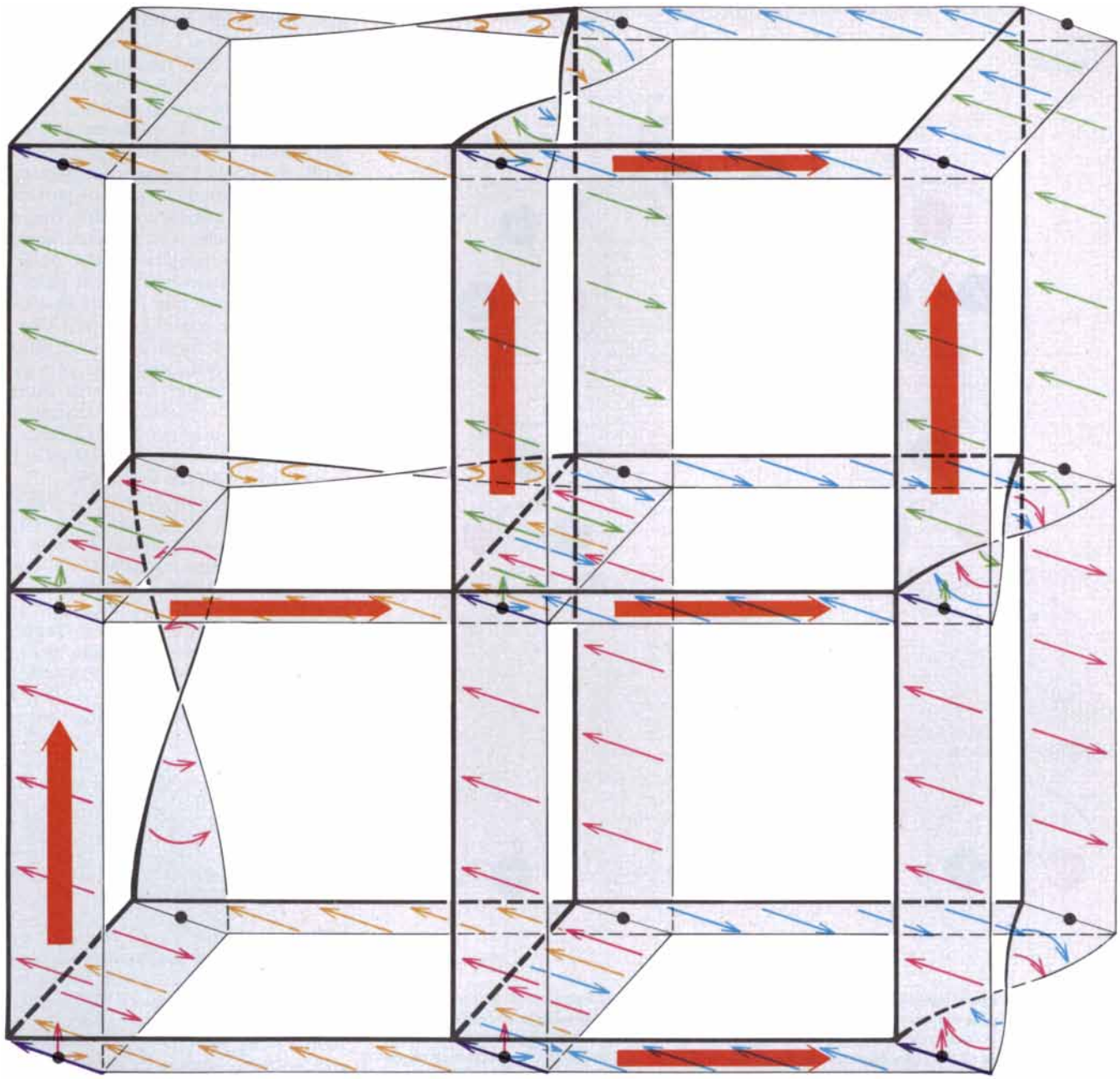
The original impetus for the quark model was the need to bring order to the large number of particles that exhibit strong interactions, or in other words those subject to the strong force. The proton and the neutron are members of this class, and indeed it is the strong force that binds them in an atomic nucleus. The existence of many other strongly interacting particles has been inferred from the decay products of collisions in accelerators. Most such particles live for an extremely short time, as short as $10^{-24}$ second, before they decay into other particles. All particles that are subject to the strong force are called hadrons, from the Greek adjective hadros, meaning robust or heavily built.

In 1962 Murray Gell-Mann of the California Institute of Technology and Yuval Ne'eman of Tel-Aviv University proposed a scheme for classifying the hadrons in symmetrical patterns. The scheme was based on the mathematical theory of groups and was called the eightfold way. A short time afterward Gell-Mann and, independently, George Zweig, also of Cal Tech, proposed a physical interpretation of the eightfold way. The mathematical classification could be explained by assuming that all hadrons are built up of more fundamental constituents, which Gell-Mann called quarks.

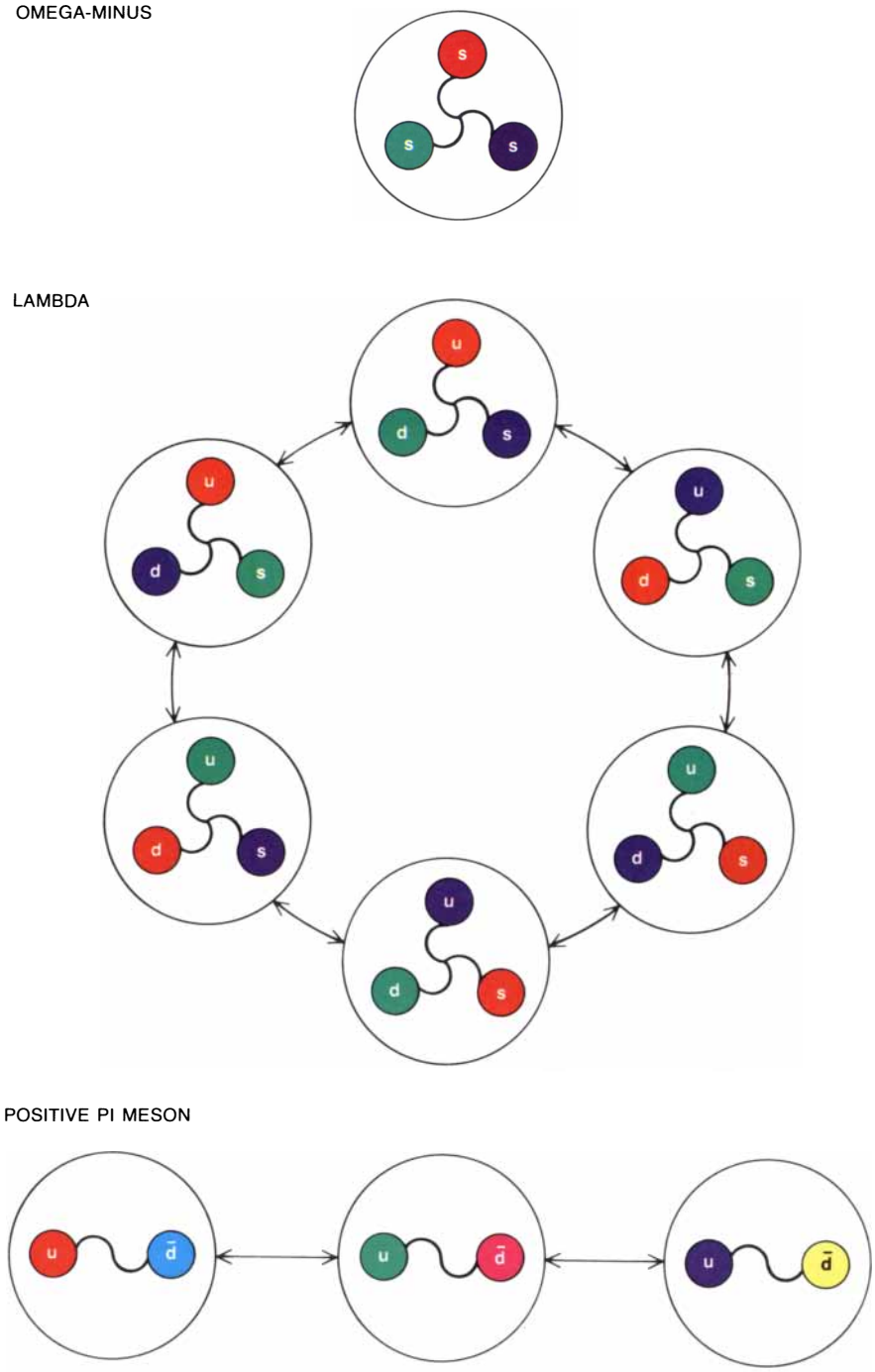
At the time every known hadron could be understood as some combination of three basic quarks (and their corresponding antiquarks): the up or

*u* quark, the down or *d* quark and the strange or *s* quark. The proton, for example, is a combination of two *u* quarks and a *d* quark, whereas the neutron is a combination of a *u* quark and two *d* quarks. The positively charged pi meson is a combination of a *u* quark and a *d* antiquark. Since the quark hypothesis was put forward more hadrons have been discovered and it has become necessary to add at least two more quarks, the charm or *c* quark and the bottom or *b* quark, to the catalogue of elementary particles. Nevertheless, the quark model remains a highly successful classification scheme: more than 100 hadrons are known, and they can all be described in terms of the quark model.

In spite of the success of the model in classifying hadrons, certain features ascribed to the quarks initially made the



**LATTICE OF POINTS** is a schematic way of representing space and time; with such a lattice physicists seek to understand the field of force that gives rise to the permanent confinement of quarks. Quarks and antiquarks, which are thought to be the constituents of the neutron and the proton and all other particles subject to the strong nuclear force, have never been observed in isolation. Instead they are bound together in triplets or pairs. The field responsible for the confinement is called the chromoelectric field, and it is classified mathematically as a gauge field. A gauge is analogous to a ruler or a pointer; it is used to compare physical quantities defined at different lattice sites. The length of the ruler and the direction of the pointer can change as they are moved about in space-time. Particles such as quarks lie only at the points, or vertexes, of the lattice. In the illustration the variables that specify the state of a particle are represented by purple arrows, which must point in one of two directions at each vertex in the front lattice plane. To compare the orientations of two purple arrows one of them must be moved next to the other. A set of rules, represented in the illustration by the bands that join neighboring vertexes, must be defined in order to specify the changes made in the orientation of the arrows during transport. The set of rules is a gauge field; it enables an arrow to be transported upward or to the right of a vertex, around a square array of four vertexes called a plaquette and back to its starting point. (The transported arrows are colored green, yellow, blue or pink.) Any set of bands that returns the transported arrow to its original orientation represents the lowest energy state of the gauge field. A set of bands that resembles a Möbius strip, however, cannot be untwisted: a transported arrow returns to its starting vertex with the opposite orientation. The energy of the gauge field defined on the lattice is stored in such plaquettes. The red arrows represent the lines of force to which the gauge field gives rise.




**COLOR is a quantum-mechanical property of quarks that was introduced to reconcile the quark model with the exclusion principle of Wolfgang Pauli. The principle states that no two quarks within a small region of space can occupy the same quantum-mechanical state. Before the introduction of the color hypothesis the quark model seemed to predict the existence of particles in which the principle is violated. For example, the constituents of the omega-minus particle are three strange or *s* quarks that were thought to be in the same quantum-mechanical state. The paradox is resolved by assuming that each of the three *s* quarks takes on one of three colors, such as red, purple and green. Since the quarks differ in color, they are in different quantum-mechanical states and the exclusion principle is saved. Color is never observed in isolation, and so the colored quarks must be combined in such a way that the omega-minus particle is colorless (or white). The color hypothesis states further that all possible combinations of the quark colors are equally likely, as long as the composite particle remains colorless. For example, the lambda particle, which is made up of three quarks, can occupy any one of six colorless states with equal probability, namely the six permutations of the three colors red, purple and green. The positive pi meson, which is made up of a quark and an antiquark, can occupy any of three colorless states with equal probability: red and cyan (antired), green and magenta (antigreen) or purple and yellow (antipurple). The confinement of quarks and of color is represented here by showing the quarks linked to one another by an unbreakable string. The quarks can move about almost freely inside a particle as long as they stay within the bounds of the string.**

physical reality of quarks difficult to accept. The most fundamental problem is the failure to detect a free quark. The proton and the neutron are strongly bound in the atomic nucleus, yet given enough energy in a nuclear collision they can be set free. Any theory that describes the interactions of quarks, however, not only must account for their binding into hadrons but also must lead to permanent confinement.

Almost equally unsettling was the fact that in certain hadrons the quark constituents seemed to violate a fundamental principle of quantum mechanics, namely the exclusion principle of Wolfgang Pauli. The exclusion principle applies to a broad category of particles including the quarks and states that no two such particles within a small region of space can simultaneously occupy the same quantum-mechanical state. In practice the principle implies that two quarks of the same kind, say two *u* quarks, cannot form a hadron unless they have opposite spins. The spin of a quark is like the spin of the earth, except that the quark's spin is quantized: it can assume only one of two values. Hence in any group of three quarks there must be at least two with the same spin.

There are several hadrons in which three identical quarks must approach one another closely enough to bind together. The omega-minus particle is one such hadron. It was predicted by the quark model, and its subsequent discovery in 1964 by Nicholas P. Samios, Ralph P. Shutt and their collaborators at Brookhaven gave strong support to the model. On the other hand, the omega-minus was also quite puzzling because the quark model predicts it must be made up of three *s* quarks, whose close association seemed to violate the exclusion principle. For these reasons and others physicists initially preferred to regard the quark as a mathematical convenience; the question of its physical existence was temporarily set aside.

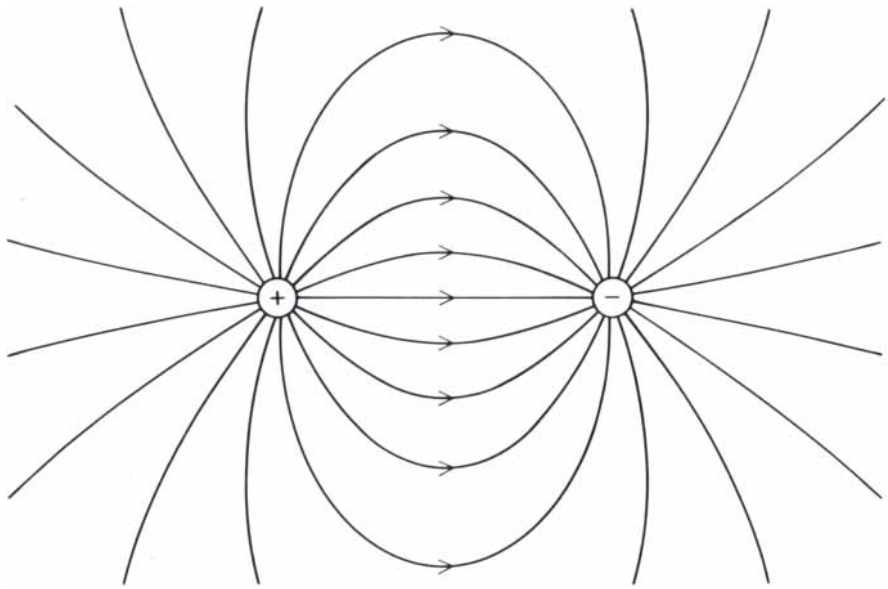
The quark model and the exclusion principle were reconciled as a result of ideas developed by Oscar W. Greenberg of the University of Maryland at College Park and, independently, by Moo-Young Han of Duke University and Yoichiro Nambu of the University of Chicago. What is needed is to assume that each kind of quark can exist in any of three states. For example, if an *s* quark in state *A* is combined with an *s* quark in state *B* and an *s* quark in state *C* to form the omega-minus particle, the exclusion principle is saved. In order to label the states of a quark physicists have whimsically taken to calling them by the names of colors: a quark can come in the three colors red, purple and green, and an antiquark can come in the three complementary colors cyan

(antired), yellow (antipurple) and magenta (antigreen). The prefix "chromo" in quantum chromodynamics refers to the color terminology.

The introduction of color had to be supplemented by another hypothesis if the successful classification of hadrons was to be maintained. Although the new color degree of freedom made possible a quark model of particles such as the omega-minus, it also led to a multiplication of other hadrons. The lambda particle, for example, is made up of a *u*, a *d* and an *s* quark; if each quark can exist in any of three colors, it would seem there should be nine lambda particles, one for each color combination, rather than the single particle that is observed. To avoid such redundancy one adds the hypothesis that the quarks in a hadron can assume only those combinations of colors that leave the hadron colorless, or white, if the rules of color addition (with ordinary light) are assumed. The three quarks in a proton or a lambda particle must include one red, one purple and one green, whereas the quark and the antiquark in a pi meson can be red and cyan, purple and yellow or green and magenta. Because the "total" color is always the same, in the quantum-mechanical sense that each colorless state can occur with equal probability, there is effectively only one lambda particle and only one positive pi meson.

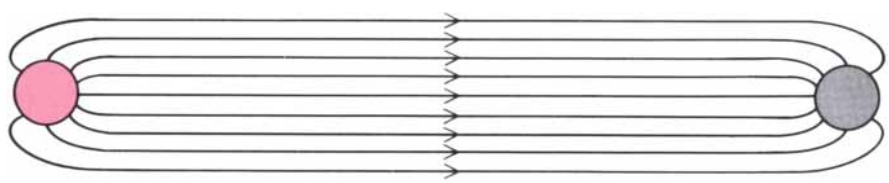
In the late 1960's strong evidence that the quarks in hadrons are real particles instead of mere mathematical entities came from a variety of experimental results. Of notable importance was a series of experiments done at the Stanford Linear Accelerator Center (SLAC) by Jerome I. Friedman and Henry W. Kendall of the Massachusetts Institute of Technology and Richard E. Taylor of SLAC. High-energy electrons were directed against a fixed target of protons in order to probe the protons for internal structure. By examining the decay products of the collisions it was possible to show that inside the proton there are constituents with all the properties attributed to quarks.

Moreover, although no free quarks were detected, the experiments showed that within the proton the quarks are in a nearly free state of motion. This result was quite puzzling: how could forces strong enough to keep quarks permanently bound together also allow them to move about almost freely when they are at close quarters inside a proton? The three quarklike objects in the proton appeared bound to one another like the three stones tied together in a bola, the South American hunting device. The stones in the bola move freely as long as they remain within the limits of the connecting string; the string, however, keeps them from flying apart.




**LINES OF FORCE of the electromagnetic field spread out in space. The field shown is the one generated by two particles of opposite electric charge; although the lines of force are densest in the region between the particles, they extend in other directions as well. The intensity of the field at any point (that is, the strength of the force "felt" at a given point by a unit electric charge) is proportional to the number of lines crossing a surface of unit area, orthogonal to the lines of force, that passes through the point. The electromagnetic force that is generated by a single source of electric charge diminishes as the square of the distance from the source.**

The new experimental evidence for quarks combined with the introduction of color gave strong impetus to the formulation of a theory of quark dynamics. Color could serve as a source for a new field called the chromoelectric field, which would give rise to a new kind of interaction among colored particles. In 1973 H. David Politzer of Cal Tech and, independently, David Gross of Princeton University and Frank Wilczek of the University of California at Santa Barbara realized that a dynamic interaction based on the chromoelectric field would lead to a progressively weaker force between quarks as they approached one another. The prediction explained the almost free motion of quarks inside protons that had been observed in the SLAC experiments. It was then conjectured that the same interaction could be responsible for the confinement of quarks, although at the time there were no theoretical results to support such an appealing idea. Nevertheless, once the chromoelectric field is introduced the rather ad hoc assumption that all hadrons are colorless and the observed confinement of quarks can be understood as two aspects of the same phenomenon. If one quark, say a red one, is pulled away from a hadron, both the quark and the fragment left behind are colored. If color, like electric charge, is the source of a field, there could be an attractive force between the two colored fragments. Confinement of the quarks might then result if the attraction between the two



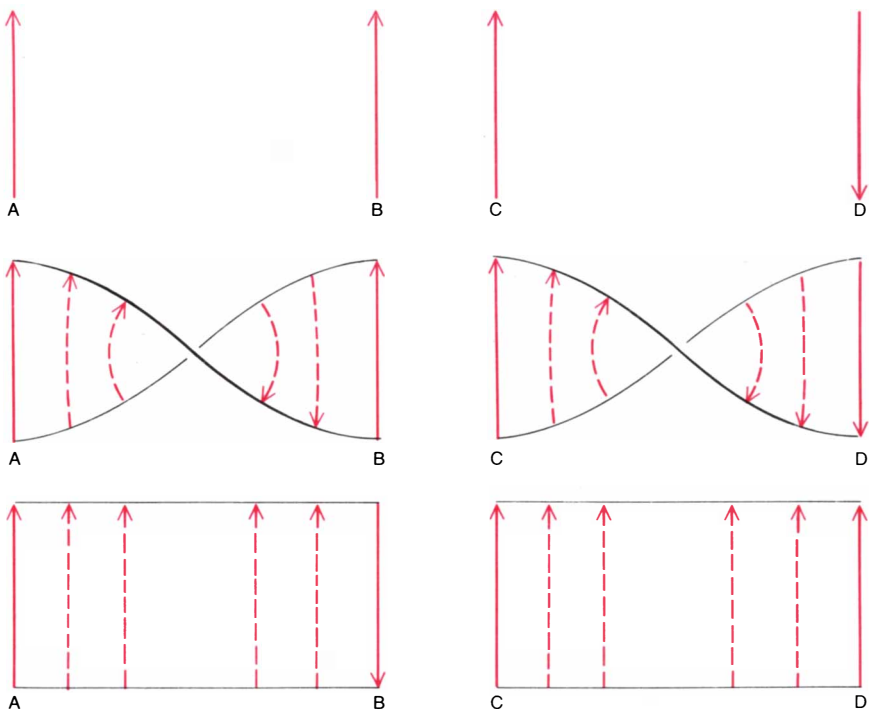
**COMPRESSION of the lines of force between two particles into a thin tube of uniform cross section would make the force between the particles constant, regardless of the distance between them. A surface of unit area orthogonal to the tube would always meet the same number of lines of force, no matter where the surface was placed along the tube. Because the force that binds the particles remains constant, increasing the separation of the particles by a given increment would always require the same amount of energy, no matter how far apart they were at the outset. An infinite amount of energy would be needed to free one of the particles from the other. Such compression of the lines of force would therefore lead to the permanent confinement of the two particles. If the radius of the tube is negligible, the bundle of lines of force resembles a string. A model explaining the confinement of quarks by this principle was formulated mathematically in 1968 by Gabriele Veneziano of the European Organization for Nuclear Research (CERN) and interpreted as a string by Yoichiro Nambu of the University of Chicago. The idea that the string could be physically realized as a bundle of lines of force was proposed in 1973 by Holger B. Nielsen and Paul Olesen of the Niels Bohr Institute in Copenhagen.**

fragments is so strong that it is impossible to separate them beyond some limit.

In the early 1970's a dynamic model independent of the quark model was developed to account for certain properties of hadrons that had not been explained by quarks. According to the dynamic model, the hadron is not a point or a spherical particle but can better be understood as a string. The string can rotate or vibrate in ways prescribed by the laws of relativistic dynamics, and its end points are required to move at the speed of light. Calculations showed that the force acting along the string must be enormous: about 14 tons. The quantized vibrations of the taut string give rise to various states that could be identified with certain hadrons.

It is evident that the string model of the hadron and the bola analogy can be combined. If three quarks or a quark and an antiquark are placed at the end points of the relativistic string, the tension of the string could explain the permanent binding of the particles. The string model, however, like the original concept of the quark, is itself a mere mathematical abstraction: the string is a one-dimensional object. Could it nonetheless be an approximation of some other structure that is more acceptable from a physical point of view?

In 1973 Holger B. Nielsen and Paul Olesen of the Niels Bohr Institute in Copenhagen pointed out that the string could be interpreted as a bundle of lines of force of a suitable field. In an electromagnetic field, lines of force have the familiar patterns made by iron filings on a sheet of paper held over a magnet. The intensity of the field is proportional to the density of the lines of force. Thus when the lines of force spread out, as they do at points increasingly distant from the poles of an ordinary magnet, the intensity of the field diminishes. If the lines of force are squeezed into a tube of uniform cross section, however, the intensity of the field remains constant all along the tube. The force needed to separate a quark and an antiquark at opposite ends of such a tube would also remain constant no matter how far apart the two particles were placed. In order to liberate one of the quarks an infinite amount of energy would have to be supplied.




**DIRECTIONS OF TWO ARROWS defined only at the vertexes of a lattice cannot be compared unless there is some way to transport the arrows from one vertex to the other. In the diagram at the top the arrows at vertexes *A* and *B* seem to point in the same direction, whereas the arrows at *C* and *D* seem to point in opposite directions. The comparison, however, presupposes the existence of the paper on which the diagram is drawn, which acts as an intermediary along which one arrow can be moved next to the other. One might just as well assume that the points in each pair are connected by a half-twisted ribbon of paper, as in the middle diagram; the comparison then gives the contrary result. Without a gauge field to supply a rule of transport from point to point the orientations of the arrows at adjacent points cannot be compared. The diagram at the bottom shows that the result of a comparison need not change when the direction of an arrow is reversed at any point; the comparison remains valid if the ribbon is correspondingly twisted or untwisted. The property of the gauge field that makes it possible to compensate for a change in the direction of any individual arrow is called local gauge invariance.**

The quantum-mechanical reality of quarks and strings requires that the lines of force associated with the color interaction of quarks act quite differently from the lines of force associated with the electromagnetic interaction of electrically charged particles. Because both forces propagate in the vacuum, one might assume that any differences between them would be caused by the intrinsic nature of the forces themselves and not by the interaction of the forces with the vacuum. In classical, or Newtonian, mechanics the assumption would be sound; indeed, there can be no interaction between a field and the classical vacuum because the classical vacuum is by definition a state with no matter and no energy in it. In quantum mechanics, however, even the vacuum has a structure, which can alter the propagation of fields and forces.

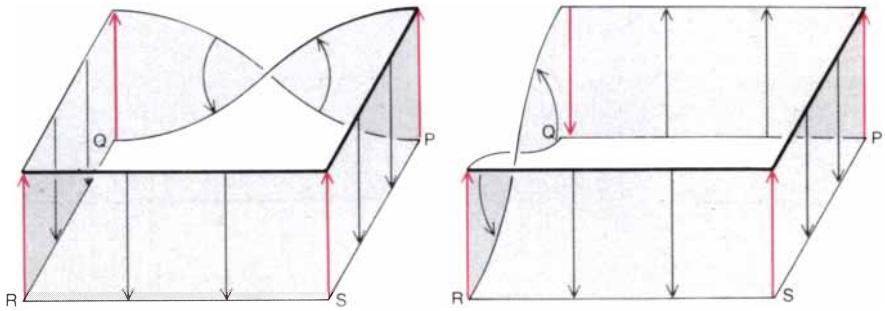
The structure of the vacuum is a consequence of the uncertainty principle of Werner Heisenberg. One version of the uncertainty principle states that for any physical event there is an uncertainty about the energy released during the event that is related to an uncertainty about the exact time of its occurrence. More precisely, the product of the uncertainty about the energy and the uncertainty about the time is not less than some numerical constant. For an event confined to an extremely short interval there is a correspondingly large uncertainty about its energy. During any short interval, therefore, there is a substantial probability that the quantum-mechanical vacuum has some nonzero energy.

The energy of the vacuum can manifest itself in the spontaneous creation or annihilation of a particle and its antiparticle or in the appearance and disappearance of an electric or a chromoelectric field throughout various regions of space. Such variations of a quantum field are called fluctuations. In the electromagnetic field between two electrically charged particles, for example, the presence of quantum fluctuations implies that the interactions between the two charges are not strictly determined by the classical field predicted by Maxwell's equations. Instead the measured electromagnetic field is the average of all the fields that can be generated by the quantum fluctuations, weighted according to the probability that a given fluctuation will occur.

For most practical applications of electrodynamics the effects of the quantum fluctuations are very small. A measurement of the field between two macroscopic electrically charged objects would be consistent with the value predicted by the classical theory. In high-energy collisions of charged particles, however, the quantum-mechanical fluctuations become much more important

and must be taken into account to calculate the electromagnetic effects. In the standard method (which for many problems is quite successful) one first calculates the properties of the field in the classical vacuum. One then builds on the result of the classical calculation by correcting it for quantum-mechanical fluctuations of a progressively higher complexity, according to what is called a perturbative expansion. In the quantum theory of electromagnetism the larger, or the more complex, a fluctuation, the less the probability that it will take place. Hence almost all the corrections to the classical electromagnetic field that must be made in the quantum-mechanical calculations are the outcome of small fluctuations.

One might suppose the properties of the chromoelectric field between two quarks could be deduced in a closely analogous way. It would seem, at least in principle, that a perturbative expansion could give the strength of the field at any point to any degree of accuracy needed. It turns out, however, that the method of perturbative expansion works only if the field calculated for the classical vacuum is the dominant effect. In other words, the method works only if the corrections that must be made to take account of the fluctuations are small and become smaller still as fluctuations of increasing size are considered. For quantum-mechanical phenomena that depend primarily on the effects of large fluctuations the perturbative expansion does not converge, that is, the series of calculations does not approach a constant, finite value. Such phenomena are said to be nonperturbative. The confinement of the lines of force of the chromoelectric field between two quarks, and hence the permanent binding of the two quarks, is a nonperturbative phenomenon.




**FRUSTRATED PLAQUETTE is given a unit twist by the four gauge fields that make up its sides. The gauge fields, represented by ribbons, can be twisted or untwisted locally to compensate for the reversal of an arrow at any point, but the overall twist in the four ribbons of the plaquette cannot be removed. The twist between *P* and *Q* (*left*) can be removed by reversing the direction of the arrow at *Q*. The effect, however, is merely to transfer the twist to the part of the ribbon between *Q* and *R* (*right*). No matter which arrows are reversed, or in other words no matter how many local gauge transformations are performed, an odd number of the four ribbons must stay twisted. Hence the plaquette remains frustrated in spite of local gauge invariance.**

How, then, can confinement be demonstrated? Workers in theoretical physics recognized that a new approach had to be devised, in which large fluctuations of the quantum-mechanical field are considered at the outset of the calculations. The lattice method suggested by Wilson is such an approach.

The lattice is generally a cubic one and can be thought of as the edges and vertexes of a collection of densely stacked cubes. The lattice extends in time as well as in space, that is, each point on the lattice designates both a position in space and a moment in time. To visualize the lattice one can think of an array of cubes in which two of the axes are labeled with spatial coordinates and the third axis is labeled with temporal coordinates; the full lattice has three spatial dimensions as well as the time axis, and so it is a four-dimensional structure. Between any two neighboring vertexes of the lattice there is a link, which can be pictured as a line connecting the two vertexes. A small square bounded by four links is called a plaquette. In Wilson's formulation the vertexes, links and plaquettes of the lattice are all that is left of ordinary physical space and time.

Links and plaquettes on the lattice must be regarded as entities at a different level of abstraction from the vertexes, or lattice points. Although links and plaquettes are defined by lattice points, there are no additional lattice points along a link or within a plaquette. In other words, space and time on the lattice are quite unlike ordinary space and time, which always include an infinite number of points between any two given points.

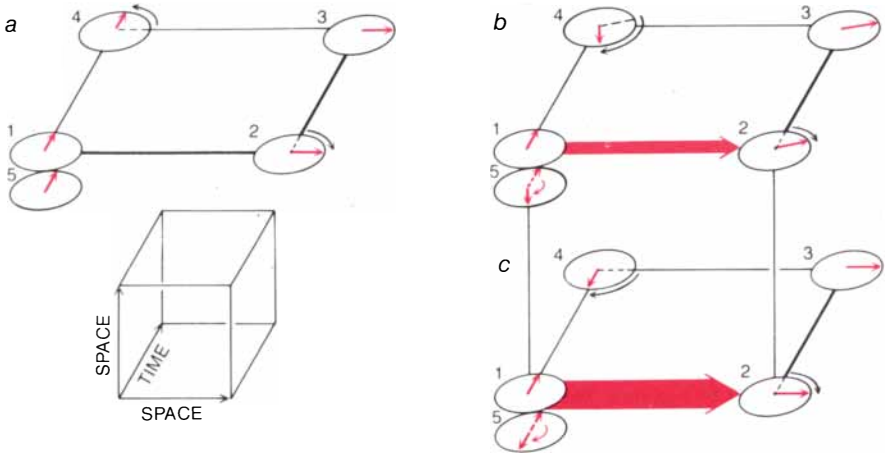
Wilson's introduction of a space-time lattice is not meant to imply that physical processes really take place on a lattice. Space-time, according to all current evidence, is continuous. Instead the lattice represents what theoretical physicists call a regularization, a temporary artifact for making calculations that would otherwise be impossible. Applied to the problem of confinement, the strategy is as follows. All particles are defined only at the vertexes of the lattice, and the strength of the field is defined only along the links of the lattice. (Actually what is defined at each vertex is the probability that a particle will be found there. The probability of finding a particle between two adjacent vertexes is not defined.) When no particles are present, the symmetry of the fluctuations on the lattice dramatically simplifies the calculation of the average electric or chromoelectric field generated by strong fluctuations. The fields, which are vector quantities and therefore have both magnitude and orientation, are just as likely to point in one direction along a link as they are to point in the opposite one. Hence in the vacuum state, with no particles, the mean value of the electric or the chromoelectric field throughout the lattice is zero.

A similar although slightly more elaborate calculation can be done for large fluctuations of the field when a single particle and its corresponding antiparticle are defined on the lattice. On the average the fluctuations of the field again cancel except along the links of the lattice that make up the shortest path between the particle and the antiparticle. The results do not depend on the kind of field defined on the lattice; the particles can be a quark and its antiquark or an electron and a positron. Thus confinement is a natural outcome of defining the field on the lattice.

The next step in the strategy is to remove the lattice and regain ordinary, continuous space-time. The lattice spacing is made progressively smaller so that the vertexes of the lattice become closer and denser in space-time. If the reduction in the lattice spacing proceeds to the mathematical limit, continuous space-time is recovered. At the limit the procedure gives the average field after all the quantum-mechanical fluctuations have been taken into account.

What is gained by introducing the lattice? The strong fluctuations that must be responsible for squeezing together the lines of the color force are considered from the outset along each lattice link. On the other hand, there is a price that must be paid: as long as the lattice is relatively coarse the quantum fluctuations give rise to the confinement of the electromagnetic field as well as the chromoelectric field.

It is well known that electric charges, unlike color charges, do exist in isolation. Since the lattice method predicts the confinement of electric charge, one must be skeptical about the lattice approach unless it can be shown that as the mesh of the lattice is made finer the elec-




**GAUGE-FIELD CONCEPT can be generalized by allowing the quantities defined at each lattice point to vary over a continuous range. For example, the direction of an arrow at each point could be allowed to form an arbitrary angle with the vertical. A gauge field makes it possible to compare the angles at different lattice points. If an arrow does not return to its original orientation after a complete circuit of a plaquette, the angular difference between the initial and the final orientations (a quantity called the phase angle) measures the degree of frustration of the plaquette. The electric field is such a gauge field, and the strength of the field in a given direction at any point is measured by the degree of frustration of an associated plaquette. The illustration shows how the frustration of plaquettes is associated with a single component of the electric field, namely the spatial component in the left-right direction. Strictly speaking, the lattice extends in four dimensions, the three spatial ones and time, but one can visualize the lattice as having only two spatial coordinates and a temporal one. As a dial is transported counterclockwise around a plaquette that includes the left-right spatial component and the time dimension, the gauge field causes an arrow on the dial to rotate. If the arrow returns to its starting position after the dial makes a full circuit of the plaquette, the strength of the field is zero in the spatial direction and there is no field line (*a*). As the phase angle increases, the strength of the field along the spatial direction increases as well (*b–c*), which is represented by the increasing thickness of the field lines (*colored arrows*) along the spatial coordinates. Mathematically the rotation of the arrow on the dial can be identified with the rotation of an arrow in the plane of complex numbers, that is, numbers that have both a real part and an imaginary part.**

tromagnetic field begins to act in accordance with its well-established properties. In other words, what one would like to show is that at some stage in the shrinking of the lattice the electromagnetic lines of force break out of their confinement to a line of lattice links, whereas the chromoelectric lines of force remain squeezed together all the way to the limit of continuous space-time. It is not a straightforward matter to demonstrate that things happen in precisely this way, but in the past few years the demonstration has been achieved by making elaborate numerical calculations with the aid of high-speed computers.

Theoretical physicists express the fundamental difference between the electromagnetic field and the chromoelectric field by saying that QED is an Abelian gauge theory whereas QCD is a non-Abelian gauge theory. The terms refer to the Norwegian mathematician Niels Henrik Abel. The distinction between Abelian and non-Abelian is drawn from the mathematical theory of groups, which describes the symmetries inherent in a sequence of operations, such as a sequence of rotations. If the operations that are members of a group can be carried out in any sequence with the same final result, the group is Abelian. For example, the group of rotations about a single axis is Abelian, because such operations have the same effect regardless of their sequence. On the other hand, if the sequence in which two or more operations are carried out does affect the final outcome, the group of operations is a non-Abelian one. The rotations of a cube about its three axes form a non-Abelian group: when the cube is turned about a vertical axis and a horizontal one, the result depends on which operation is done first.

In order to understand how ideas from group theory apply to QCD and QED, one must understand the concept of a gauge field. The concept can best be illustrated for isolated points of space and time arranged in a lattice. Particles can rest on any of the lattice vertexes or hop from one vertex to another; as the particles move through the space-time lattice they can change their state, where a state is defined by quantities that can vary over a certain range of values. I shall make the simplifying assumption that the state of a particle is described by just one variable, which can take on exactly two values. For example, at each vertex of the lattice there might be a variable whose value is either $+1$ or $-1$, indicating the sign of the electric charge. The two possible values can be represented by an arrow that points either up or down.

In describing how interactions propagate in space and time it is essential to be able to compare the values of variables at neighboring points. To compare the length of two objects in separate regions of space one needs a measuring stick, or gauge, that can be moved next to one object, marked and then moved to the second object, where the comparison is made. Similarly, on the lattice one must be able to compare the orientation of the arrows at neighboring vertexes. At first such a comparison seems trivial. Suppose two vertexes of the lattice are represented on a sheet of paper and at each vertex there is an arrow directed upward. Is it not obvious that the two arrows point in the same direction? The question, however, presupposes the existence of the sheet of paper on which the lattice is drawn. The paper acts as an intermediary that allows the two orientations to be compared. In a sense, one transports the arrow from one vertex to the other with the eye, and one concludes that the two arrows had the same orientation before the transport because after the transport they match.

Suppose the sheet of paper between the two arrows had been given a half twist [*see illustration on page 58*]. An upward-pointing arrow transported along the twisted paper would point downward when it reached the second vertex. Because the direction of an arrow is defined only at the individual, isolated vertexes of the lattice, there is no way to decide which of the two methods of transport is correct. Indeed, without the sheet of paper or some similar assumption about the effects of transport no comparison of directions at different vertexes is possible.

In field theory a gauge is any standard of measurement, analogous to the distance between two marks on a metal bar or the direction of an arrow with respect to a dial, that can change under the influence of a field as the gauge is moved about in space and time. A field that can effect such changes is called a gauge field, and it specifies explicitly the assumptions that must be made about the transport of the gauge. In my simple example the gauge field is a set of rules for transporting the arrows along the links of the lattice from one vertex to the next. One can think of the gauge field as a ribbon that connects neighboring lattice sites and thereby allows the arrows at different points to be transported and compared.

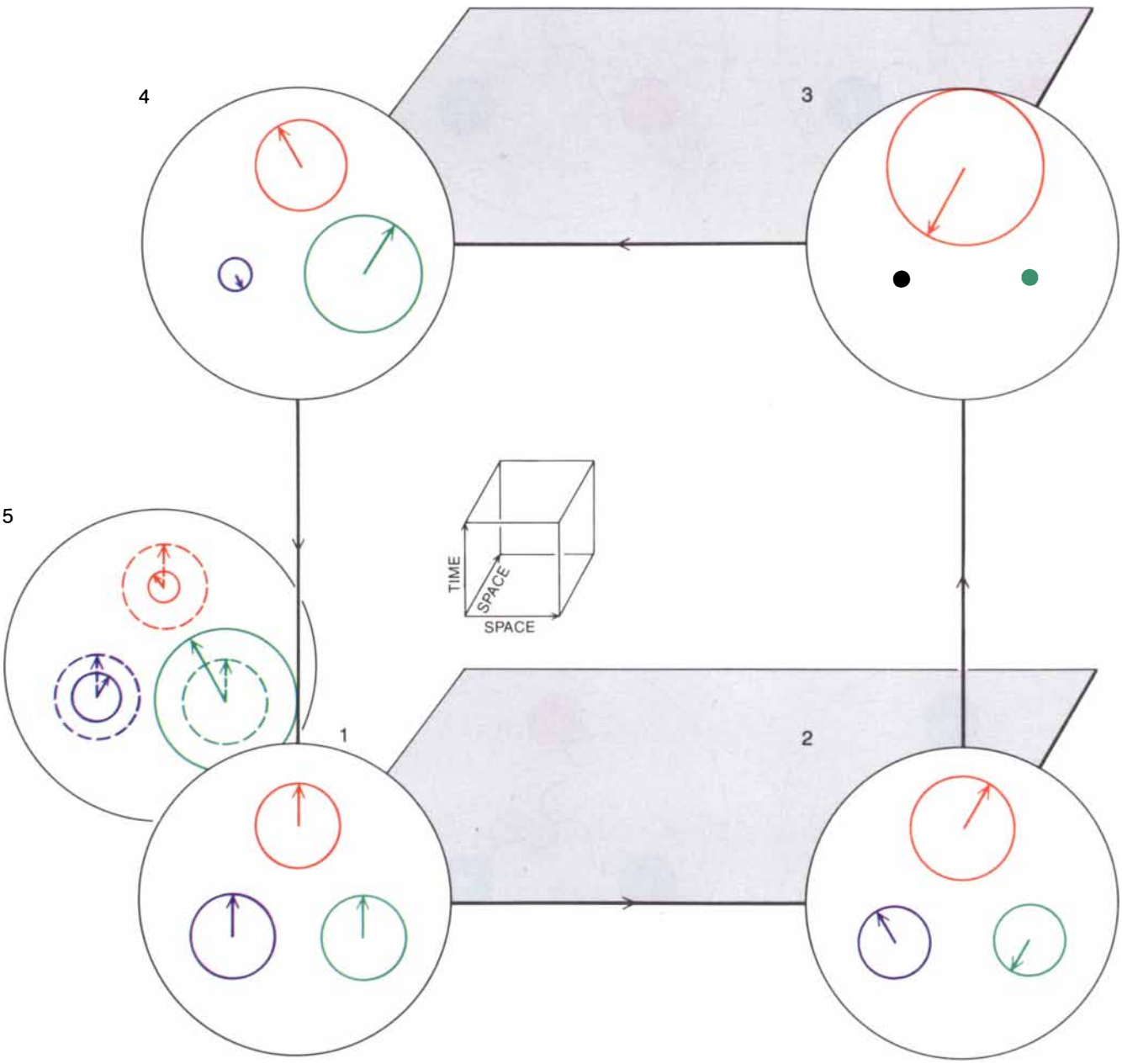
It is important to note that the arrow representing the state at any lattice vertex can be reversed, provided the ribbon representing the gauge field is correspondingly twisted or untwisted. If the arrows and the ribbons are adjusted in

coordination, the result of any comparison between state variables does not change, nor does the physical information represented by the system. The possibility of reversing the arrows without modifying the physical information is called local gauge invariance.

Because of local gauge invariance one might think a gauge field is nothing but an unnecessary complication. What is the point of introducing the twisted ribbon if it can then be untwisted by a local gauge transformation? The objection is valid as long as one considers only pairs of lattice points. The importance of the local gauge invariance becomes apparent, however, when one considers a plaquette on the lattice: a set of four points in a square, connected along the lattice links by four ribbons that constitute the gauge field. Suppose one of the ribbons is twisted and the other three are not [*see illustration on page 59*]. The single twisted ribbon can be untwisted (and one of the arrows can be reversed), but not without introducing a half twist in one of the other three ribbons. No matter which arrows are reversed and no matter how many local gauge transformations are carried out, an odd number of the four ribbons must be twisted. A plaquette whose ribbons cannot all be untwisted is called a frustrated plaquette.

The frustration of the plaquette manifests itself in another way. An upward-pointing arrow that is transported along the ribbons around the frustrated plaquette will return to its starting vertex pointing down. Thus on the lattice the frustrated plaquettes are those that generate a mismatch in orientation when an arrow is transported all the way around.

It is not difficult to apply the idea of a

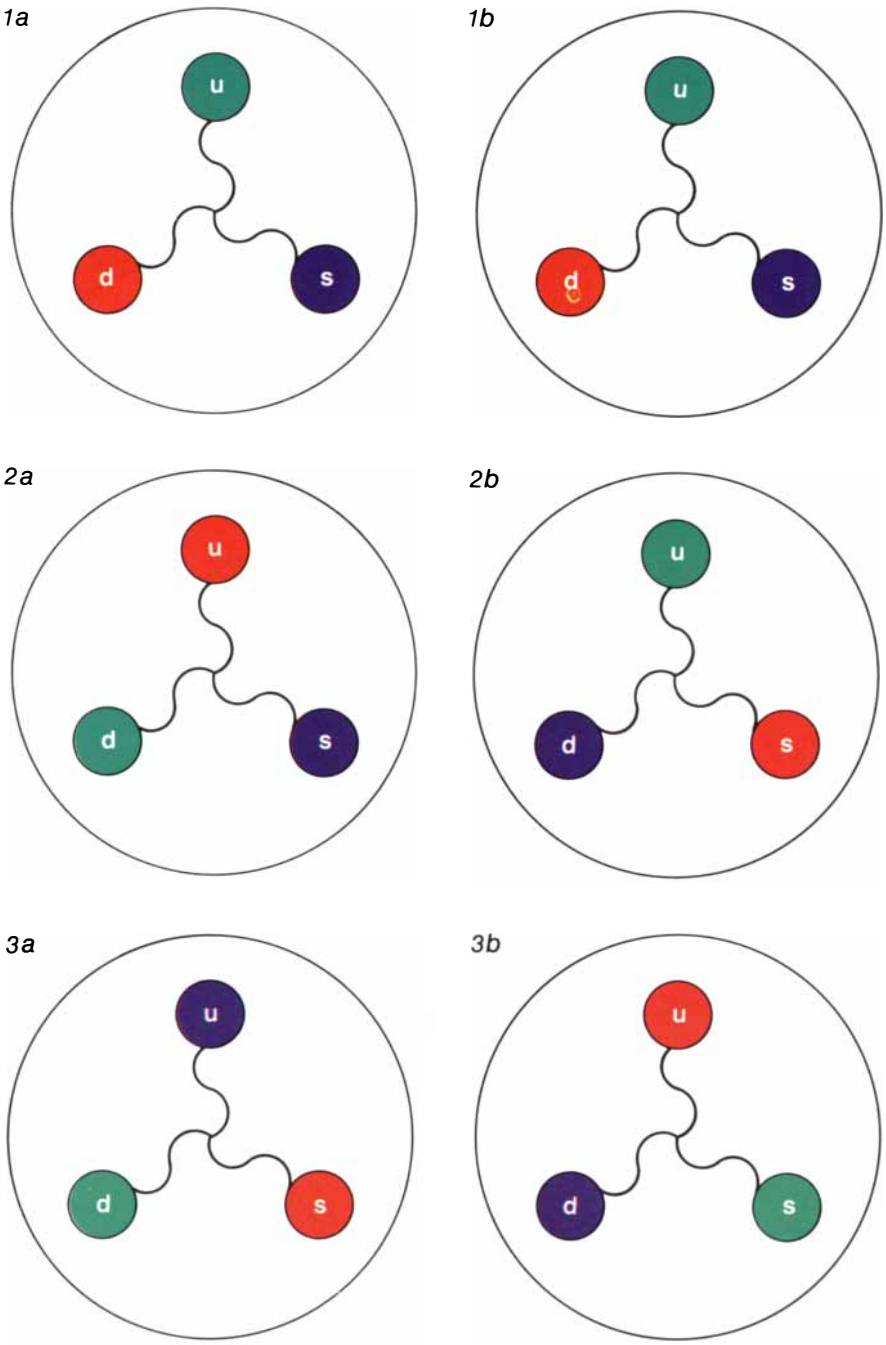



**CHROMOELECTRIC FIELD is a gauge field similar in principle to the electromagnetic field but more complicated mathematically. At every point on a lattice there are three arrows instead of one; they correspond to the color charges of a quark. Moreover, the color gauge field affects not only the direction of each arrow but also its length. The lengths of the arrows are not independent of one another: the square root of the sum of the squares of the lengths must equal 1. The strength of the chromoelectric field along each link of the lattice is dependent on the phase angles and the change in the configuration of the three arrows after a complete circuit of a plaquette.**

frustrated plaquette to ordinary space-time. In physical space-time the ribbons along which the arrows move cannot be visualized directly; just as the sheets of paper that define the gauge field on the lattice are not themselves part of the lattice, so a gauge field defined in ordinary space-time is not itself part of space-time. Mathematically the higher-dimensional abstract space that specifies the rotations of the arrow is called a connection in a fiber bundle [see "Fiber Bundles and Quantum Theory," by Herbert J. Bernstein and Anthony V. Phillips; SCIENTIFIC AMERICAN, July, 1981]. Nevertheless, it is possible without complete mathematical understanding to imagine that an arrow moved about along some closed path in space-time could return to its starting point with its direction changed.

The idea of a frustrated plaquette has an important physical interpretation. In any gauge field the energy of the field resides precisely in the plaquettes that are frustrated. The plaquettes that are not frustrated, in which all the ribbons can be untwisted and all the arrows can be oriented in the same direction, are associated with the vacuum state of a physical system, the configuration with no energy. Frustrated plaquettes are the sites of the fluctuations in a quantum-mechanical field.

The idea of a gauge field can readily be generalized to more complex situations. The arrows, for example, can be allowed to form an arbitrary angle with respect to a fixed direction rather than being constrained to point up or down. The gauge field now specifies the angle by which the arrow rotates when it is transported along a link. Consequently the frustration of a plaquette can take on a continuous range of values, measured by the angular difference in the direction of the arrow after it is transported around the entire plaquette. The degree of frustration of a plaquette, expressed in suitable mathematical units, is called its action. The action of the entire system is the sum of the actions of all the individual plaquettes.




**GROUP OF COLOR TRANSFORMATIONS in the chromoelectric field is classified mathematically as a non-Abelian one. A non-Abelian group is an abstract set of operations (together with the objects on which the operations are carried out) in which the final outcome of several operations is dependent on the sequence in which they are done. For the chromoelectric gauge theory the operations are interchanges of colors among quarks. In the illustration two interchanges of color are applied to a given quark configuration (*1a, 1b*). At the left an exchange of red and green (*2a*) is followed by an exchange of purple and red (*3a*). At the right the order of the two exchanges is reversed (*2b, 3b*). The final configurations of the quarks are not the same.**

It turns out the electric field is a gauge field that specifies the continuous rotations of an arrow about a single axis, although this fact is not apparent in any but the most unified and sophisticated formulations of the concept of the electric field. It is more common to think of the electric field as a set of vectors, with one vector at each point in space giving the magnitude and direction of the field at that point. Understood according to the more comprehensive framework of the gauge theory, however, the magnitude of each vector is directly proportional to the frustration of an associated plaquette.

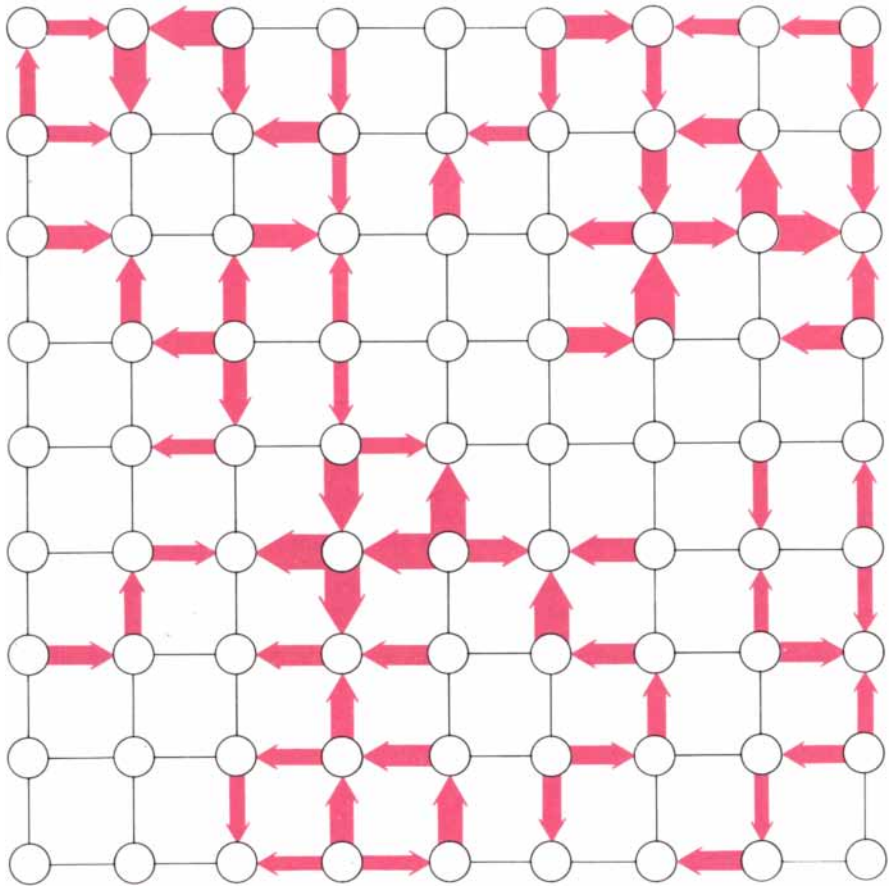
The rotating arrow of the gauge field associated with electromagnetism actually represents a complex number (one with both a real part and an imaginary part), which changes in value as the arrow is carried around the plaquette. (The transported arrow and the complex number must not be confused with the vector that represents the electric field itself at each link on the lattice.) The plaquette is a loop in space-time, and so it is necessary to imagine the transported arrow as moving forward and backward in time as well as in space. The arrow is carried along a certain axis in space, say the positive $x$ axis, then forward in time, then back to its starting point on the $x$ axis and finally backward to its starting point in time. The amount

by which the direction of the transported arrow changes as a result of the transport is called the phase angle. According to gauge theory, the magnitude of the $x$ component of the vector that ordinarily gives the strength of the electric field at a point is a measure of the phase angle generated around a space-time loop that begins in the $x$ direction. The quantum-mechanical fluctuations at a point in the electric field can therefore be thought of as fluctuations in the amount of rotation an arrow would undergo if the arrow were transported around a plaquette extended along one spatial dimension and along the temporal dimension [*see illustration on page 60*].

It is difficult to suppress a feeling of unreality when one is asked to regard the electric field—an entity that can become quite tangible when one gets a shock—as the abstract space of phase rotations. Nevertheless, as one goes more deeply into the study of the physical world, tangible facts and mathematical concepts become intertwined. The abstract idea of phase space at last makes contact with the distinction between QED and QCD, that is, with the distinction between an Abelian gauge theory and a non-Abelian one. The fundamental group operations in QED are the rotations of phase angles. The result of two successive phase rotations, say one of 30 degrees and one of 50 degrees, does not depend on the sequence in which they are done. The net result in either case is a rotation by 80 degrees. Such an outcome is characteristic of an Abelian group and QED is therefore an Abelian gauge theory.

If the state variables defined at the vertexes of a lattice are the colors of quarks, the transported gauge is made up of three arrows instead of just one, and the arrows can vary in length as well as in angle [*see illustration on page 61*]. Each arrow represents one of three color charges; the color charges together determine the color of the quark. As the quark is transported along a link of the lattice it can change its color from red to green; hence a red quark can exchange colors with a green quark, and then the newly green quark can exchange colors with a purple quark. The final result of the exchanges, however, depends on the sequence of events. Since an outcome that depends on sequence is characteristic of a non-Abelian group, QCD is a non-Abelian gauge theory.

The non-Abelian nature of QCD introduces an extra degree of freedom into the fluctuations of the chromoelectric field. Moreover, the existence of three kinds of color also implies that the plaquettes of the chromoelectric field can be frustrated in many more ways than the plaquettes of the electromagnetic field can. It is likely that the extra degree of freedom and the additional disorder of the non-Abelian theory are responsible for the confinement of quarks within hadrons.

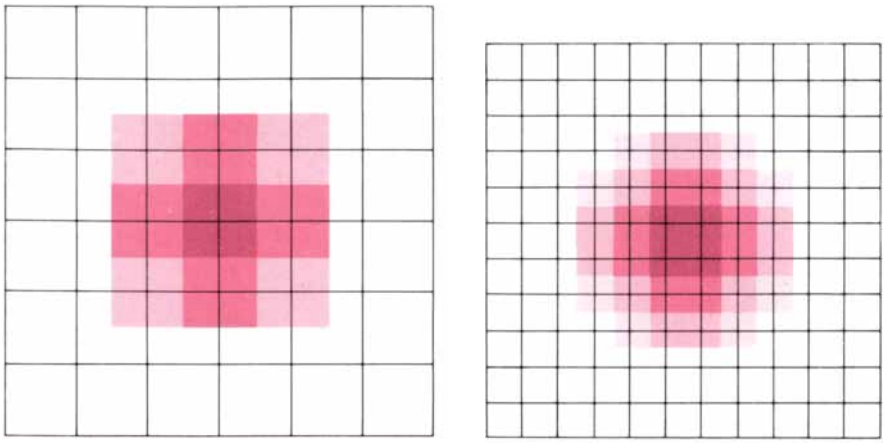


**FLUCTUATIONS of the lines of force on a lattice are characteristic of the quantum-mechanical vacuum. The fluctuations are a consequence of the uncertainty principle of Werner Heisenberg. The principle states that the product of the uncertainty about the energy of a system and the uncertainty about the time of an event is not less than a numerical constant. Hence during any short interval there is a correspondingly large uncertainty about the energy of the system, which manifests itself in spontaneous fluctuations of the quantum field. The fluctuations cancel one another in the absence of charged particles, and so a measurement of, say, the electric field in the vacuum would give the value zero because it would determine only the average of all the fluctuations. Nevertheless, the fluctuations of the lines of force in the vacuum must be carefully considered in calculating the field of force between two charged particles.**

I have already stated that confinement on a coarse lattice can be demonstrated by a relatively simple calculation for large fluctuations of the quantum-mechanical field. For smaller fluctuations on a finer lattice, however, the calculation of the field becomes much more difficult. The value of any physical quantity measured in an experiment is the quantum-expectation value, which is a weighted average of all possible values the quantity can have. The measured electric or chromoelectric field is the average of all the possible configurations of the field to which fluctuations can give rise. Not all configurations contribute equally to the average, and so each configuration must be weighted, or multiplied by some factor based on the probability of the configuration. In principle, therefore, the quantum-expectation value of a physical quantity defined on the lattice can be calculated in two steps. First the value is calculated for each configuration of the field fluctuations and multiplied by the weight factor for that configuration. Then the products are added and the result is divided by the sum of the weight factors.

Even for a lattice defined over a small volume of space-time, however, the number of possible configurations is so large that a complete summation is out of the question. For the simplest-possible gauge theory, in which the gauge field between two lattice vertexes can be either twisted or untwisted, the number of configurations on a lattice extended for 10 sites in each direction is $2^{40,000}$, or more than $10^{12,041}$.

To calculate the quantum-expectation value of the field on the lattice one must therefore resort to techniques of statistical sampling. The techniques are analogous to the ones employed in opin-



**SHRINKING A LATTICE on which particles and fields have been defined must be carried out in order to approximate the effects of the field in ordinary, continuous space and time. As the lattice spacing is reduced, however, the size of a composite particle such as a proton must be preserved, that is, the probability that a quark will be found at a given lattice vertex within the proton must be spread over an increasing number of vertexes. The illustration at the right shows the result of shrinking the lattice at the left to half its previous size. The color density represents the relative probability that a quark will be found at a given vertex. The shrinking can be accomplished by systematically changing the value of the quantity called the coupling constant, which determines the strength of the fluctuations in the quantum-mechanical field.**

ion polling. One cannot ask the opinion of every person in the U.S. in order to determine how an issue is perceived by various groups in the population, and so a sample is selected. The result should reflect the actual opinion of the groups if the probability of selecting a respondent from a given group matches the portion, or weight, of the group in the population as a whole.

In a similar way the configurations of the field fluctuations are sampled by a computer program. The computer generates a large number of configurations (but many fewer than $10^{12,041}$), and the probability that a particular configuration is generated is set equal to the quantum-mechanical weight factor for that configuration. The sample of configurations tends to give the same average value for the quantum field as the total population of configurations does.

The importance of a configuration in calculating the quantum-expectation value is determined by its action, which is generally designated $S$; the weight factor is therefore given by a mathematical function of the action $S$. An elegant formula for the weight factor in quantum-mechanical systems was introduced by Richard P. Feynman of Cal Tech. First the total action of the configuration is divided by a constant, $g^2$, then the exponential of this quantity is determined; in other words, the number $e$ (equal to approximately 2.7) is raised to the power $S/g^2$. The weight factor is inversely proportional to the result. Thus the weight factor is proportional to $1/\exp(S/g^2)$.

The formula for the weight factor implies that the higher the action of a configuration is, the less the configuration is weighted in calculating the average quantum field. Because the action $S$ is not itself exponentiated but instead is divided by $g^2$, changing the value of $g$ can have a significant effect on the weight factor of a given configuration. If $g$ is large, $S/g^2$ is small and $1/\exp(S/g^2)$ is larger than it would be if $g$ were small. The quantity $g$ is called the coupling constant; hence when the coupling constant is large, the weight factor of a quantum-mechanical fluctuation with large action is higher than it is when the coupling constant is small.

The idea of a coupling constant may be familiar as a measure of the intrinsic strength of a force. In electromagnetic theory the coupling constant is an important physical quantity with a value of about 1/137. From the rather abstract perspective of the lattice gauge theory, however, $g$ is to be understood as a quantity that takes on a fixed value for a given lattice spacing but can vary with the spacing. Once confinement is demonstrated on a coarse lattice the mesh of the lattice is made finer by carefully decreasing the value of the coupling constant. The process by which the lattice is made progressively finer until continuous space-time is recovered is called renormalization.

Strictly speaking, it is not yet possible to renormalize the lattice for fluctuations of the chromoelectric field, or in other words to shrink the lattice spacing to zero. Nevertheless, it is possible to make the lattice spacing smaller while still keeping it greater than zero and to search for indications that the lines of force do or do not remain confined on finer lattices. As the lattice spacing is reduced all physical objects should remain the same size. On a coarse lattice the probability distribution for a proton might be defined as nonzero across only three lattice spacings. When the lattice spacing is reduced to half its original value, the probability distribution must stretch over six lattice spacings. Since reducing the value of $g$ lowers the probability of lattice configurations that have a large action, the decrease in $g$ has the effect of "zeroing in" for a closer look at the fluctuations of the field, at the lines of force and at the particles defined on the lattice. Thus reducing the value of $g$ makes the proton look larger on the lattice.

The numerical investigation of the properties of a gauge field on the lattice proceeds by limiting the lattice to a large but finite volume. The number of lattice vertexes, links, plaquettes and state variables is therefore finite, although it may be larger than 100,000. Initial values of all the variables are stored in the memory of a large computer. By randomly varying the elements in the starting configuration according to a suitable algorithm the computer generates a sample of as many as 100,000 configurations. Finally it calculates the average quantum-mechanical effects to which the configurations in its sample give rise.

Because of the element of randomness in the calculation, the method is called a Monte Carlo simulation. Before my own work and that of my colleagues on quark confinement the Monte Carlo method had been applied with considerable success to the analysis of the properties of thermodynamic systems, and Wilson had emphasized its suitability for the analysis of lattice gauge theories in quantum mechanics. In 1979, working at Brookhaven, Michael J. Creutz, Laurence A. Jacobs and I first applied Monte Carlo simulation to the study of Abelian gauge theories. We wanted to test whether or not the confinement of particles on the lattice observed at large values of the coupling constant in QED disappears as it should when the continuum limit is approached. The results were spectacular; they showed clearly that at some stage in the reduction of the coupling constant the lines of force on the lattice suddenly undergo a transition. The electric field, which for large values of $g$ is confined to the lattice links between two electric charges, suddenly spreads out all around the charges.

It is useful to compare the sudden deconfinement of the electric field with the sudden change in properties observed when a solid changes to the liquid phase. What Creutz, Jacobs and I observed in our numerical calculations was a phase transition. Everything happened in the computer model as if we had a four-dimensional crystal that we could heat or cool by changing the value of $g$. At

a certain point the crystal underwent a phase transition; on one side of the transition electric charges are confined, whereas on the other side they are not. Our results were later confirmed independently by an extensive numerical simulation done at the European Organization for Nuclear Research (CERN) by Benny Lautrup of the Niels Bohr Institute and Michael Nauenberg of the University of California at Santa Barbara and by other investigators.

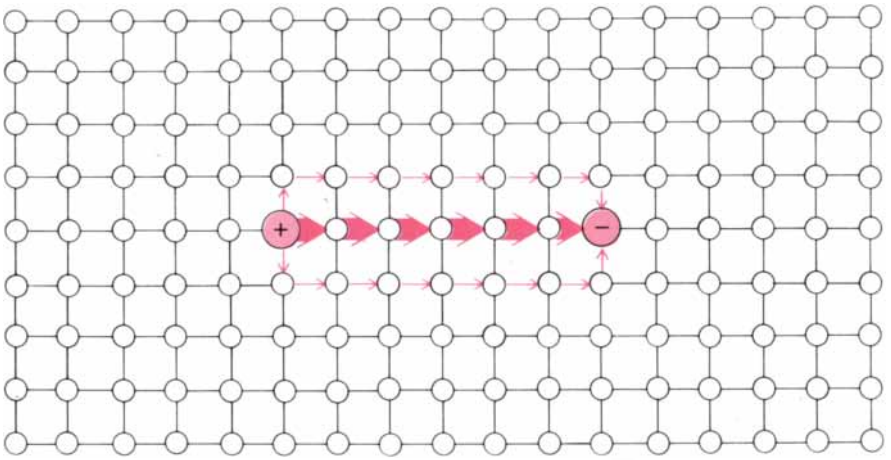
Soon after we got our results for the Abelian lattice gauge theory Creutz extended the Monte Carlo simulation to the non-Abelian model. His results were just as spectacular and physically more interesting; with QED the correct answer was known beforehand and the simulation was a test of the Monte Carlo method rather than of QED, but the outcome for a non-Abelian system was entirely unknown. Creutz's simulation showed that, contrary to the Abelian case, the non-Abelian lattice gauge theory undergoes no phase transition as the value of the coupling constant is gradually lowered. Thus what had long been sought is now finally obtained: a demonstration, albeit by numerical methods, that quantum chromodynamics leads to the confinement of quarks.

Monte Carlo simulation makes it possible to explore the predictions of QCD for many physical processes. For example, after Creutz showed that there is no deconfining phase transition in QCD he was also able to estimate the force holding the quarks together. The result is in excellent agreement with the prediction of the string model, which makes no pretense of describing the dynamics of quarks and hadrons in their full generality. Creutz's value for the string tension was confirmed independently by Wilson, by Gyan Bhanot of CERN, by me and by several others.

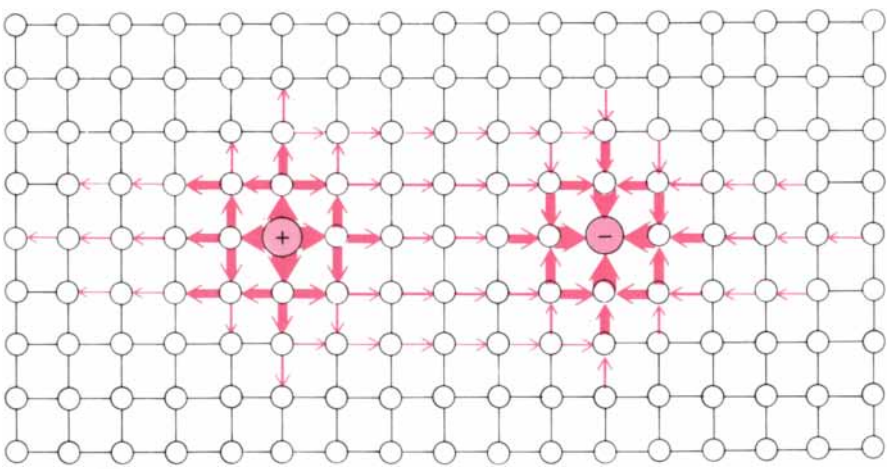
Another prediction of Monte Carlo simulation is the surprising result that at extremely high temperature quarks become deconfined and can move about freely. The temperature must be about two trillion degrees Celsius, much too high to be created in the laboratory but perhaps not too high for free quarks to be found in the interior of hot stars. The prediction also suggests that free quarks existed shortly after the big bang.

The current frontier in the understanding of quark interactions is the numerical simulation of changes in the chromoelectric field as the quarks move about. All the investigations described so far evaluate the chromoelectric field by assuming that the quarks are stationary. It is possible to employ the Monte Carlo method to simulate the quantum mechanics of moving quarks in a gauge field, but the amount of computation required puts the simulation almost out of reach, even for the most powerful computers. Herbert W. Hamber of the Institute for Advanced Study, Enzo Marinari and Georgio Parisi of the University of Rome and I, and independently Donald H. Weingarten of Indiana University, have recently proposed an approximation scheme whereby the computations become feasible. The approximation has been applied to a calculation of the masses of several hadrons, and the theoretical results are in good agreement with experimental values.

Lattice gauge theory has at last brought QCD to the stage where one can calculate its predictions and compare them with experiment. It is an elegant theory, based on relatively simple mathematical concepts yet rich in consequences. For the moment all its verifiable predictions are passing the test of experiment. What remains to be done is to learn how its consequences can be proved by logical deduction. Although a numerical approximation cannot satisfy the need for logical demonstration, it does augur well for such a demonstration in the future. In other branches of theoretical physics there have been many instances in which a definite result, obtained numerically at first, was soon proved by analytic methods. Several of my colleagues and I are confident that QCD will soon pass this test as well.




**CONFINEMENT OF A FIELD, whether the field is the electric or the chromoelectric one, can be demonstrated by fairly simple calculations when the field is defined on a coarse lattice and the fluctuations of the lines of force are strong. The strong fluctuations cancel one another everywhere on the lattice except in the region between two charged particles. The lines of force are compressed into a thin tube, and so the two particles cannot be moved indefinitely far apart.**




**CALCULATION OF THE ELECTRIC FIELD generated by two oppositely charged particles defined on a lattice was done by the author with a Control Data Corporation CDC-7600 computer at the Brookhaven National Laboratory. The illustration shows the outcome of thousands of calculations, done for progressively finer lattice spacings. As the spacing is reduced the lines of force that are confined on the coarse lattice (*see upper illustration on this page*) suddenly undergo a phase transition and break out in all directions. The deconfinement of the electric field on the lattice confirms the plausibility of the lattice approach, because it is well known that the electric field spreads out in ordinary, continuous space. Similar calculations done by Michael J. Creutz of Brookhaven have demonstrated that as the lattice spacing is reduced for the chromoelectric field the lines of force remain confined, just as they are on the coarse lattice. Thus the confinement of quarks can be derived from the theory of the chromoelectric field.**