

# Instantons in QCD

T. Schäfer

*Institute for Nuclear Theory, Department of Physics, University of Washington, Seattle, Washington 98195*

E. V. Shuryak

*Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794*

The authors review the theory and phenomenology of instantons in quantum chromodynamics (QCD). After a general overview, they provide a pedagogical introduction to semiclassical methods in quantum mechanics and field theory. The main part of the review summarizes our understanding of the instanton liquid in QCD and the role of instantons in generating the spectrum of light hadrons. The authors also discuss properties of instantons at finite temperature and how instantons can provide a mechanism for the chiral phase transition. They give an overview of the role of instantons in some other models, in particular low-dimensional sigma models, electroweak theory, and supersymmetric QCD. [S0034-6861(98)00902-7]

## CONTENTS

I. Introduction	324	IV. Towards a Theory of the Instanton Ensemble	355
A. Motivation	324	A. The instanton interaction	355
B. Physics outlook	325	1. The gauge interaction	355
1. Hadronic structure	325	2. The fermionic interaction	358
2. Scales of nonperturbative QCD	325	B. Instanton ensembles	358
3. Structure of the QCD vacuum	326	C. The mean-field approximation	359
4. QCD at finite temperature	328	D. The quark condensate in the mean-field approximation	361
C. The history of instantons	328	E. Dirac eigenvalues and chiral symmetry breaking	362
1. Discovery and early applications	328	F. Effective interaction between quarks and the mean-field approximation	363
2. Phenomenology leads to a qualitative picture	329	1. The gap equation for $N_f=1$	364
3. Technical development during the eighties	329	2. The effective interaction for two or more flavors	364
4. Recent progress	330	G. Bosonization and the spectrum of pseudoscalar mesons	366
D. Topics that are not discussed in detail	330	H. Spin-dependent interactions induced by instantons	366
II. Semiclassical Theory of Tunneling	331	V. The Interacting Instanton Liquid	367
A. Tunneling in quantum mechanics	331	A. Numerical simulations	367
1. Quantum mechanics in Euclidean space	331	B. The free energy of the instanton ensemble	368
2. Tunneling in the double-well potential	333	C. The instanton ensemble	369
3. Tunneling amplitude at one-loop order	333	D. Dirac spectra	370
4. The tunneling amplitude at two-loop order	335	E. Screening of the topological charge	371
5. Instanton/anti-instanton interaction and the ground-state energy	336	VI. Hadronic Correlation Functions	372
B. Fermions coupled to the double-well potential	338	A. Definitions and generalities	372
C. Tunneling in Yang-Mills theory	340	B. The quark propagator in the instanton liquid	374
1. Topology and classical vacua	340	1. The propagator in the field of a single instanton	374
2. Tunneling and the Belavin-Polyakov-Schwartz-Tyupkin instanton	341	2. The propagator in the instanton ensemble	375
3. The theta vacua	342	3. The propagator in the mean-field approximation	376
4. The tunneling amplitude	343	C. Mesonic correlators	377
D. Instantons and light quarks	344	1. General results and the operator product expansion	377
1. Tunneling and the $U(1)_A$ anomaly	344	2. The single-instanton approximation	378
2. Tunneling amplitude in the presence of light fermions	346	3. The random-phase approximation	379
III. Phenomenology of Instantons	347	4. The interacting instanton liquid	380
A. How often does tunneling occur in the QCD vacuum?	347	5. Other mesonic correlation functions	382
B. The typical instanton size and the instanton liquid model	348	D. Baryonic correlation functions	382
C. Instantons on the lattice	349	1. Nucleon correlation functions	383
1. The topological charge and susceptibility	349	2. Delta correlation functions	384
2. The instanton liquid on the lattice	351	E. Correlation functions on the lattice	384
3. The fate of large-size instantons and the beta function	352	F. Gluonic correlation functions	385
D. Instantons and confinement	354	G. Hadronic structure and $n$ -point correlators	387

VII. Instantons at Finite Temperature	388
A. Introduction	388
1. Finite-temperature field theory and the caloron solution	388
2. Instantons at high temperature	389
3. Instantons at low temperature	391
4. Instanton interaction at finite temperature	392
B. Chiral symmetry restoration	393
1. Introduction to QCD phase transitions	393
2. The instanton liquid at finite temperature and IA molecules	396
3. Mean-field description at finite temperature	397
4. Phase transitions in the interacting instanton model	399
C. Hadronic correlation functions at finite temperature	401
1. Introduction	401
2. Temporal correlation functions	401
3. $U(1)_A$ breaking	402
4. Screening masses	404
D. Instantons at finite temperature: Lattice studies	405
VIII. Instantons in Related Theories	406
A. Two-dimensional theories	406
1. The $O(2)$ sigma model	406
2. The $O(3)$ sigma model	407
B. Instantons in electroweak theory	408
C. Supersymmetric QCD	410
1. The instanton measure and the perturbative beta function	410
2. $N=1$ supersymmetric QCD	411
3. $N=2$ supersymmetric gauge theory	412
IX. Summary and Discussion	414
A. General remarks	414
B. Vacuum and hadronic structure	414
C. Finite temperature and chiral restoration	415
D. The big picture	416
Acknowledgments	419
Appendix: Basic Instanton Formulas	419
1. Instanton gauge potential	419
2. Fermion zero modes and overlap integrals	420
3. Properties of $\eta$ symbols	420
4. Group integration	421
References	421

## I. INTRODUCTION

### A. Motivation

Quantum chromodynamics (QCD), the field theory describing the strong interaction, is now more than 20 years old, and naturally it has reached a certain level of maturity. Perturbative QCD has been developed in great detail, with most hard processes calculated beyond leading order, compared to data, and compiled in reviews and textbooks. However, the world around us cannot be understood on the basis of perturbative QCD, and the development of nonperturbative QCD has proven to be a much more difficult task.

This is hardly surprising. While perturbative QCD could build on the methods developed in the context of quantum electrodynamics, strategies for dealing with the nonperturbative aspects of field theories first had to be developed. The gap between hadronic phenomenology on the one side and exactly solvable model field theories

on the other is still huge. While some fascinating discoveries (instantons among them) have been made, and important insights have emerged from lattice simulations and hadronic phenomenology, a lot of work remains to be done in order to unite these approaches and truly understand the phenomena involved.

Among the problems the field is faced with is a difficulty in communication between researchers working on different aspects of nonperturbative field theory, and a shortage of good introductory material for people interested in entering the field. In this review we try to provide an overview of the role of instantons in field theory, with a particular emphasis on QCD. Such a review is certainly long overdue. Many readers certainly remember learning about instantons from some of the excellent papers (Callan, Dashen, and Gross, 1978a) or introductory reviews (Coleman, 1977; Vainshtein, Zakharov, and Shifman, 1982) that appeared in the late seventies or early eighties, but since then there have been very few publications addressed to a more general audience.<sup>1</sup> The only exceptions are the book by Rajaraman (1982), which provides a general discussion of topological objects but without particular emphasis on applications, and a few chapters in the book by Shuryak (1988c), which deal with the role of instantons in hadronic physics. All of these publications are at least a decade old.

Writing this review, we had several specific goals in mind, which are represented by different layers of presentation. In the remainder of Sec. I, we provide a very general introduction into instanton physics, aimed at a very broad readership. We shall try to give qualitative answers to questions like: What are instantons? What are their effects? Why are instantons important? What is the main progress achieved during the last decade? This discussion is continued in Sec. III, in which we review the current information concerning the phenomenology of instantons in QCD.

Section II is also pedagogical, but the style is completely different. The main focus is a systematic development of the semiclassical approximation. As an illustration, we provide a detailed discussion of a classic example, the quantum mechanics of the double-well potential. However, in addition to the well-known derivation of the leading-order WKB result, we also deal with several modern developments, like two-loop corrections, instantons and perturbation theory at large orders, and supersymmetric quantum mechanics. In addition to that, we give an introduction to the semiclassical theory of instantons in gauge theory.

Sections IV–VII make up the core of the review. They provide an in-depth discussion of the role of instantons in QCD. Specifically, we try to emphasize the connection between the structure of the vacuum, hadronic correlation functions, and hadronic structure, at both zero and finite temperatures. The style is mostly

<sup>1</sup>A reprint volume that contains most of the early work and a number of important technical papers was recently published by Shifman (1994).

that of a review of the modern literature, but we have tried to make this part as self-contained as possible.

The last two sections, VIII and IX, deal with many fascinating applications of instantons in other theories and with possible lessons for QCD. The presentation is rather cursory, and our main motivation is to acquaint the reader with the significant problems and to provide a guide to the available literature.

## B. Physics outlook

### 1. Hadronic structure

In this section, we should like to provide a brief outline of the theory and phenomenology of hadronic structure and the QCD vacuum. We shall emphasize the role the vacuum plays in determining the structure of the excitations and explain how instantons come into play in making this connection.

There are two approaches to hadronic structure that predate QCD, the quark model and current algebra. The quark model provides a simple (and very successful) scheme based on the idea that hadrons can be understood as bound states of nonrelativistic constituent quarks. Current algebra is based on the (approximate)  $SU(2)_L \times SU(2)_R$  chiral symmetry of the strong interaction. The fact that this symmetry is not apparent in the hadronic spectrum led to the important concept that chiral symmetry is spontaneously broken in the ground state. Also, since the “current” quark masses appearing as symmetry-breaking terms in the effective chiral Lagrangian are small, it became clear that the constituent quarks of the nonrelativistic quark model have to be effective, composite objects.

With the advent of QCD, it was realized that current algebra is a rigorous consequence of the (approximate) chiral invariance of the QCD Lagrangian. It was also clear that quark confinement and chiral symmetry breaking are consistent with QCD, but since these phenomena occur in the nonperturbative domain, there was no obvious way to incorporate these features into hadronic models. The most popular model in the early days of QCD was the MIT bag model (DeGrand *et al.*, 1975), which emphasized the confinement property of QCD. Confinement was realized in terms of a special boundary condition on quark spinors and a phenomenological bag pressure. However, the model explicitly violated the chiral symmetry of QCD. This defect was later cured by dressing the bag with a pionic cloud in the cloudy (Thomas Theberge, and Miller, 1981) or chiral bag (Brown and Rho, 1978) models. The role of the pion cloud was found to be quite large, and it was suggested that one can do away with the quark core altogether. This idea led to (topological) soliton models of the nucleon. The soliton model was pioneered by Skyrme in a paper that appeared long before QCD (Skyrme, 1961). However, the Skyrmion was largely ignored at the time and only gained in popularity after Witten argued that in the large- $N_c$  limit the nucleon becomes a soliton, built from nonlinear meson fields (Witten, 1983).

While all these models provide a reasonable description of static properties of the nucleon, the pictures of hadronic structure they suggest are drastically different. Although it is sometimes argued that different models provide equivalent, dual descriptions of the same physics, it is clear that not all of these models can be right. In the MIT bag model, for example, everything is determined by the bag pressure, while there is no such thing in the Skyrmion, and the scale is set by chiral symmetry breaking. Quark models attribute the nucleon-delta mass splitting to perturbative one-gluon exchange, while it is due to the collective rotation of the pion field in soliton models.

In order to make progress, two shifts of emphasis have to be made. First, it is not enough just to reproduce the mass and other static properties of the nucleon. A successful model should reproduce the correlation functions (equivalent to the full spectrum, including excited states) in all relevant channels—not just baryons, but also scalar and vector mesons, etc. Second, the structure of hadrons should be understood starting from the structure the QCD vacuum. Hadrons are collective excitations, like phonons in solids, so one cannot ignore the properties of the ground state when studying its excitations.

### 2. Scales of nonperturbative QCD

In order to develop a meaningful strategy, it is important to establish whether there is a hierarchy of scales that allows the problem to be split into several independent parts. In QCD there is some evidence that such a hierarchy is provided by the scales for chiral symmetry breaking and confinement,  $\Lambda_{\chi SB} \gg \Lambda_{\text{conf}}$ .

The first hint comes from perturbative QCD. Although the perturbative coupling constant blows up at momenta given roughly by the scale parameter  $\Lambda_{\text{QCD}} \sim 0.2 \text{ GeV} \approx (1 \text{ fm})^{-1}$  (the exact value depends on the renormalization scheme), perturbative calculations are generally limited to reactions involving a scale of at least  $1 \text{ GeV} \approx (0.2 \text{ fm})^{-1}$ .

A similar estimate of the scale of nonperturbative effects can be obtained from low-energy effective theories. The first result of this kind was based on the Nambu and Jona-Lasinio (NJL) model (Nambu and Jona-Lasinio, 1961).

This model was inspired by the analogy between chiral symmetry breaking and superconductivity. It postulates a four-fermion interaction which, if it exceeds a certain strength, leads to quark condensation, the appearance of pions as Goldstone bosons, etc. The scale above which this interaction disappears and QCD becomes perturbative enters the model as an explicit UV cutoff,  $\Lambda_{\chi SB} \sim 1 \text{ GeV}$ .

It was further argued that the scales for chiral symmetry breaking and confinement are very different (Shuryak, 1981):  $\Lambda_{\chi SB} \gg \Lambda_{\text{conf}} \sim \Lambda_{\text{QCD}}$ . In particular, it was argued that constituent quarks (and pions) have sizes smaller than those of typical hadrons, explaining the success of the nonrelativistic quark model. This idea

was developed in a more systematic fashion by (Georgi and Manohar, 1984), who argue that  $\Lambda_{\chi\text{SB}} \approx 4\pi f_\pi \approx 1 \text{ GeV}$  provides a natural expansion parameter in chiral effective Lagrangians. An effective theory using pions and “constituent” quarks is the natural description in the intermediate regime  $\Lambda_{\text{conf}} < Q < \Lambda_{\chi\text{SB}}$ , where models of hadronic structure operate.

While our understanding of the confinement mechanism in QCD is still very poor, considerable progress has been made in understanding the physics of chiral symmetry breaking. The importance of instantons in this context is one of the main points of this review. Instantons are localized ( $\rho \approx 1/3 \text{ fm}$ ) regions of spacetime with very strong gluonic fields,  $G_{\mu\nu} \sim 1/(g\rho^2)$ . It is the large value of this quantity which sets the scale for  $\Lambda_{\chi\text{SB}}$ . In the regime  $\Lambda_{\text{conf}} < Q < \Lambda_{\chi\text{SB}}$ , the instanton-induced effective interaction between quarks is of the form

$$\mathcal{L} = \bar{\psi}(i\partial - m)\psi + \frac{c_\Gamma}{\Lambda^2}(\bar{\psi}\Gamma\psi)^2 + \frac{d_\Gamma}{\Lambda^5}(\bar{\psi}\Gamma\psi)^3 + \dots, \quad (1)$$

where  $\Gamma$  is some spin-isospin-color matrix,  $\Lambda \sim \rho^{-1}$  is determined by the chiral symmetry-breaking scale and higher-order terms involve more fermion fields or derivatives. The mass scale for glueballs is even higher, and these states decouple in the regime under consideration. The instanton-induced interaction is nonlocal, and when calculating higher-order corrections with the vertices in Eq. (1), loop integrals are effectively cut off at  $\Lambda \sim \rho^{-1}$ .

In addition to determining the scale  $\Lambda$ , instantons also provide an organizing principle for performing calculations with the effective Lagrangian (1). If the instanton liquid is dilute,  $\rho^4(N/V) \ll 1$ , where  $(N/V)$  is the density of instantons, then vertices with more than  $2N_f$  legs are suppressed. In addition to that, the diluteness parameter determines which diagrams involving the leading-order vertices have to be resummed. As a result, one can derive another effective theory valid at even smaller scales which describes the interaction of extended, massive, constituent quarks with pointlike pions (Diakonov, 1996). This theory is of the type considered by Georgi and Manohar (1984).

Alternatively, one can go beyond leading order in  $\rho^4(N/V)$  and study hadronic correlation functions to all orders in the instanton-induced interaction. The results are in remarkably good agreement with experiment for most light hadrons. Only when dealing with heavy quarks or high-lying resonances do confinement effects seem to play an important role. We shall discuss these questions in detail in Sec. VI.

### 3. Structure of the QCD vacuum

The ground state of QCD can be viewed as a very dense state of matter, composed of gauge fields and quarks that interact in a complicated way. Its properties are not easily accessible in experiments because we do not directly observe quark and gluon fields, only the color-neutral hadrons. Furthermore, we cannot directly determine the interaction between quarks because we

cannot measure  $qq$  and  $\bar{q}q$  scattering amplitudes as we can for the nuclear force. Instead, the main tools at our disposal are correlation functions of hadronic currents. The phenomenology of these functions was recently reviewed by Shuryak (1993). Hadronic point-to-point correlation functions were first systematically studied in the context of “QCD sum rules.” The essential point, originally emphasized by Shifman, Vainshtein, and Zakharov (1979), is that the operator product expansion relates the short-distance behavior of current correlation function to the vacuum expectation values of a small set of lowest-dimension quark and gluon operators. Using the available phenomenological information on hadronic correlation functions, Shifman, Vainshtein, and Zakharov deduced the quark and gluon condensates<sup>2</sup>

$$\langle \bar{q}q \rangle = -(230 \text{ MeV})^3, \quad \langle g^2 G^2 \rangle = (850 \text{ MeV})^4, \quad (2)$$

as well as other, more complicated, expectation values.

The significance of the quark condensate is the fact that it is an order parameter for the spontaneous breakdown of chiral symmetry in the QCD vacuum. The gluon condensate is important because the QCD trace anomaly

$$T_{\mu\mu} = \sum_f m_f \langle \bar{q}_f q_f \rangle - \frac{b}{32\pi^2} \langle g^2 G^2 \rangle \quad (3)$$

relates this quantity to the energy density  $\epsilon_0 \approx -500 \text{ MeV}/\text{fm}^3$  of the QCD vacuum. Here,  $T_{\mu\nu}$  is the energy-momentum tensor, and  $b = 11N_c/3 - 2N_f/3$  is the first coefficient of the beta function.

Any model of the QCD vacuum should explain the origin and value of these condensates, the mechanism for confinement and chiral symmetry breaking, and its relation to the underlying parameters of the theory (the scale parameter and the matter content of the theory). Most of the early attempts to understand the QCD ground state were based on the idea that the vacuum is dominated by classical gauge-field configurations, for example, constant fields (Savvidy, 1977) or regions of constant fields patched together, as in the “spaghetti vacuum” introduced by the Copenhagen group (Ambjorn and Olesen, 1977). All of these attempts were unsuccessful, however, because constant fields were found to be unstable against quantum perturbations.

Instantons are classical solutions to the Euclidean equations of motion. They are characterized by a topological quantum number and correspond to tunneling events between degenerate classical vacua in Minkowski space. As in quantum mechanics, tunneling lowers the ground-state energy. Therefore instantons provide a simple understanding of the negative nonperturbative vacuum energy density. In the presence of light fermions, instantons are associated with fermionic zero modes. Zero modes not only are crucial to our under-

<sup>2</sup>In the operator product expansion one has to introduce a scale  $\mu$  that separates soft and hard contributions. The condensates take into account soft fluctuations, and the values given here correspond to a scale  $\mu \approx 1 \text{ GeV}$ .

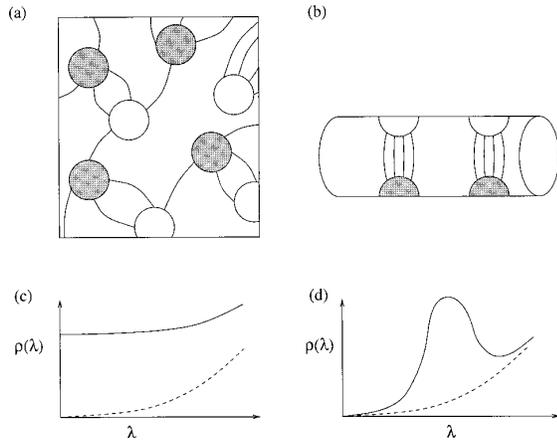


FIG. 1. Schematic picture of the instanton liquid: (a) at zero temperature; (b) above the chiral phase transition. Instantons and anti-instantons are shown as open and shaded circles. The lines correspond to fermion exchanges. Figures (c) and (d) show the schematic form of the Dirac spectrum in the configurations (a) and (b).

standing of the axial anomaly, they are also intimately connected with spontaneous chiral symmetry breaking. When instantons interact through fermion exchanges, zero modes can become delocalized, forming a collective quark condensate.

A crude picture of quark motion in the vacuum can then be formulated as follows [see Fig. 1(a)]. Instantons act as a potential well, in which light quarks can form bound states (the zero modes). If instantons form an interacting liquid, quarks can travel over large distances by hopping from one instanton to another, similar to electrons in a conductor. Just as the conductivity is determined by the density of states near the Fermi surface, the quark condensate is given by the density of eigenstates of the Dirac operator near zero virtuality. A schematic plot of the distribution of eigenvalues of the Dirac operator is shown<sup>3</sup> in Fig. 1(c). For comparison, the spectrum of the Dirac operator for noninteracting quarks is depicted by the dashed line. If the distribution of instantons in the QCD vacuum is sufficiently random, there is a nonzero density of eigenvalues near zero, and chiral symmetry is broken.

The quantum numbers of the zero modes produce very specific correlations between quarks. First, since there is exactly one zero mode per flavor, quarks with different flavors (say  $u$  and  $d$ ) can travel together, but quarks with the same flavor cannot. Furthermore, since zero modes have a definite chirality (left-handed for instantons, right-handed for anti-instantons), quarks flip their chirality as they pass through an instanton. This is very important phenomenologically because it distinguishes instanton effects from perturbative interactions, in which the chirality of a massless quark does not

<sup>3</sup>We shall discuss the eigenvalue distribution of the Dirac operator in some detail in the main part of the review; see Figs. 16 and 34 below.

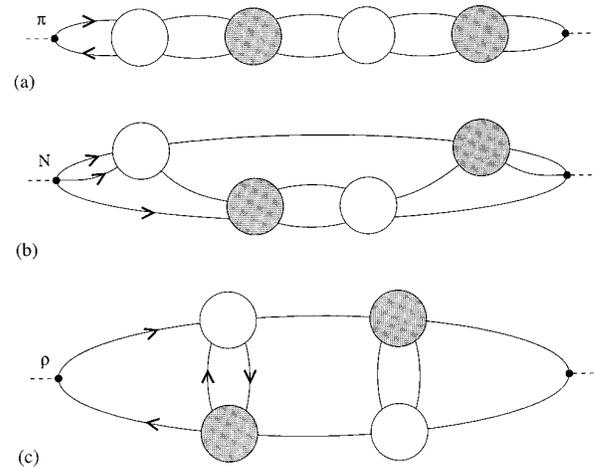


FIG. 2. Instanton contribution to hadronic correlation functions: (a) the pion; (b) the nucleon; (c) the rho-meson correlator. The solid lines correspond to zero-mode contributions to the quark propagator.

change. It also implies that quarks can only be exchanged between instantons of the opposite charge.

Based on this picture, we can also understand the formation of hadronic bound states. Bound states correspond to poles in hadronic correlation functions. As an example, let us consider the pion, which has the quantum numbers of the current  $j_\pi = \bar{u}\gamma_5 d$ . The correlation function  $\Pi(x) = \langle j_\pi(x)j_\pi(0) \rangle$  is the amplitude for an up quark and a down anti-quark with opposite chiralities created by a source at point 0 to meet again at the point  $x$ . In a disordered instanton liquid, this amplitude is large because the two quarks can propagate by the process shown in Fig. 2(a). As a result, there is a light pion state. For the  $\rho$  meson, on the other hand, we need the amplitude for the propagation of two quarks with the same chirality. This means that the quarks have to be absorbed by different instantons (or propagate in non-zero-mode states); see Fig. 2(c). The amplitude is smaller, and the meson state is much less tightly bound.

Using this picture, we can also understand the formation of a bound nucleon. Part of the proton wave function is a scalar  $ud$  diquark coupled to another  $u$  quark. This means that the nucleon can propagate as shown in Fig. 2(b). The vertex in the scalar diquark channel is identical to the one in the pion channel with one of the quark lines reversed.<sup>4</sup> The  $\Delta$  resonance has the quantum numbers of a vector diquark coupled to a third quark. Just as in the case of the  $\rho$  meson, there is no first-order instanton-induced interaction, and we expect the  $\Delta$  to be less bound than the nucleon.

The paradigm discussed here bears a striking similarity to one of the oldest approaches to hadronic structure, the Nambu–Jona-Lasinio model (Nambu and Jona-Lasinio, 1961). In analogy with the Bardeen-Cooper-

<sup>4</sup>For more than three flavors, the color structure of the two vertices is different, so there is no massless diquark in  $SU(3)$  color.

Schrieffer theory of superconductivity, it postulates a short-range attractive force between fermions (nucleons in the original model and light quarks in modern versions). If this interaction is sufficiently strong, it can rearrange the vacuum, and the ground state becomes superconducting, with a nonzero quark condensate. In the process, nearly massless current quarks become effectively massive constituent quarks. The short-range interaction can then bind these constituent quarks into hadrons (without confinement).

This brief outline indicates that instantons provide at least a qualitative understanding of many features of the QCD ground state and its hadronic excitations. How can this picture be checked and made more quantitative? Clearly, two things need to be done. First, a consistent instanton ensemble has to be constructed in order to make quantitative predictions for hadronic observables. Second, we should like to test the underlying assumption that the topological susceptibility, the gluon condensate, the chiral symmetry breaking, etc. are dominated by instantons. This can be done most directly on the lattice. We shall discuss both of these issues in the main part of this review, Secs. III–VI.

#### 4. QCD at finite temperature

Properties of the QCD vacuum, like the vacuum energy density, and the quark and gluon condensates, are not directly accessible to experiment. Measuring non-perturbative properties of the vacuum requires the ability to compare the system with the ordinary, perturbative state.<sup>5</sup> This state of matter has not existed in nature since the Big Bang, so experimental attempts at studying the perturbative phase of QCD have focused on recreating miniature Big Bangs in relativistic heavy-ion collisions.

The basic idea is that at sufficiently high temperature or density QCD will undergo a phase transition to a new state, referred to as the quark-gluon plasma, in which chiral symmetry is restored and quarks and gluon are deconfined. The temperature scale for this transition is set by the vacuum energy density and pressure of the vacuum. For the perturbative vacuum to have a pressure comparable to the vacuum pressure  $500 \text{ MeV}/\text{fm}^3$ , a temperature on the order of  $150\text{--}200 \text{ MeV}$  is required. According to our current understanding, such temperatures are reached in the ongoing or planned experiments at the Alternating-Gradient Synchrotron (AGS; about  $2+2 \text{ GeV}$  per nucleon in the center-of-mass system), CERN SPS (about  $10+10 \text{ GeV}$ ) or RHIC ( $100+100 \text{ GeV}$ ).

In order to interpret these experiments, we need to understand the properties of hadrons and hadronic matter near and above the phase transition. As for cold mat-

ter, this requires an understanding of the ground state and how the rearrangement of the vacuum takes place that causes chiral symmetry to be restored. A possible mechanism for chiral symmetry restoration in the instanton liquid is indicated in Figs. 1(b) and 1(d). At high temperature, instantons and anti-instantons have a tendency to bind in pairs that are aligned along the (Euclidean) time direction. The corresponding quark eigenstates are strongly localized, and chiral symmetry is unbroken. There is some theoretical evidence for this picture, which will be discussed in detail in Sec. VII. In particular, there is evidence from lattice simulations that instantons do not disappear at the phase transition, but only at even higher temperatures. This implies that instantons affect properties of the quark-gluon plasma at temperatures not too far above the phase transition.

### C. The history of instantons

In books and reviews, physical theories are usually presented as a systematic development, omitting the often confusing history of the subject. The history of instantons did not follow a straight path either. Early enthusiasm concerning the possibility of understanding nonperturbative phenomena in QCD, in particular confinement, caused false hopes, which led to years of frustration. Only many years later did work on phenomenological aspects of instantons lead to breakthroughs. In the following we shall try to give a brief tour of the two decades that have passed since the discovery of instantons.

#### 1. Discovery and early applications

The instanton solution of the Yang-Mills equations was discovered by Polyakov and co-workers (Belavin, Polyakov, Schwartz, and Tyupkin, 1975), motivated by the search for classical solutions with nontrivial topology in analogy with the 't Hooft-Polyakov monopole (Polyakov, 1975). Shortly thereafter, a number of authors clarified the physical meaning of the instanton as a tunneling event between degenerate classical vacua (Callan, Dashen, and Gross, 1976; Jackiw and Rebbi, 1976a; Polyakov, 1977). These works also introduced the concept of  $\theta$  vacua in connection with QCD.

Some of the early enthusiasm was fueled by Polyakov's discovery that instantons cause confinement in certain three-dimensional models (Polyakov, 1977). However, it was soon realized that this is not the case in four-dimensional gauge theories. An important development originated with the classic paper<sup>6</sup> by 't Hooft (1976a), in which he calculated the semiclassical tunneling rate. In this context he discovered the presence of zero modes in the spectrum of the Dirac operator. This result implied that tunneling is intimately connected with light fermions, in particular, that every instanton absorbs one left-handed fermion of every species, and

<sup>5</sup>A nice analogy is given by the atmospheric pressure. In order to measure this quantity directly, one has to evacuate a container filled with air. Similarly, one can measure the non-perturbative vacuum energy density by filling some volume with another phase, the quark-gluon plasma.

<sup>6</sup>In this paper, 't Hooft also coined the term instanton; Polyakov had referred to the classical solution as a pseudoparticle.

emits a right-handed one (and vice versa for anti-instantons). This result also explained how anomalies, for example the violation of axial charge in QCD and baryon number in electroweak theory, are related to instantons.

In his work, 't Hooft estimated the tunneling rate in electroweak theory, where the large Higgs expectation value guarantees the validity of the semiclassical approximation, and found it to be negligible. Early attempts to study instanton effects in QCD, where the rate is much larger but also harder to estimate, were summarized by Callan *et al.* (1978a). These authors realized that the instanton ensemble can be described as a four-dimensional “gas” of pseudoparticles that interact via forces that are dominantly of the dipole type. While they were not fully successful in constructing a consistent instanton ensemble, they nevertheless studied a number of important instanton effects: the instanton-induced potential between heavy quarks, the possibility that instantons cause the spontaneous breakdown of chiral symmetry, and instanton corrections to the running coupling constant.

One particular instanton-induced effect, the anomalous breaking of  $U(1)_A$  symmetry and the  $\eta'$  mass, deserves special attention.<sup>7</sup> Witten and Veneziano wrote down an approximate relation that connects the  $\eta'$  mass with the topological susceptibility (Veneziano, 1979; Witten, 1979a). This was a very important step, because it was the first quantitative result concerning the effect of the anomaly on the  $\eta'$  mass. However, it also caused some confusion because the result had been derived using the large- $N_c$  approximation, which is not easily applied to instantons. In fact, both Witten and Veneziano expressed strong doubts concerning the relation between the topological susceptibility and instantons,<sup>8</sup> suggesting that instantons are not important dynamically (Witten, 1979b).

## 2. Phenomenology leads to a qualitative picture

By the end of the seventies the general outlook was very pessimistic. There was no experimental evidence for instanton effects and no theoretical control over semiclassical methods in QCD. If a problem cannot be solved by direct theoretical analysis, it is often useful to turn to a more phenomenological approach. By the early eighties, such an approach to the structure of the QCD vacuum became available with the QCD sum-rule method (Shifman, Vainshtein, and Zakharov, 1979). QCD sum rules relate vacuum parameters, in particular, the quark and gluon condensates, to the behavior of hadronic correlation functions at short distances. Based

on this analysis, it was realized that “all hadrons are not alike” (Novikov *et al.*, 1981). The operator product expansion (OPE) does not give reliable predictions for scalar and pseudoscalar channels ( $\pi, \sigma, \eta, \eta'$  as well as scalar and pseudoscalar glueballs). These are precisely the channels that receive direct instanton contributions (Geshkenbein and Ioffe, 1980; Novikov *et al.*, 1981; Shuryak, 1983).

In order to understand the available phenomenology, a qualitative picture, later termed the instanton liquid model, was proposed by Shuryak (1982a). In this work, the two basic parameters of the instanton ensemble were suggested: the mean density of instantons is  $n \approx 1 \text{ fm}^{-4}$ , while their average size is  $\rho \approx 1/3 \text{ fm}$ . This means that the spacetime volume occupied by instantons  $f \sim n\rho^4$  is small; the instanton ensemble is dilute. This observation provides a small expansion parameter that we can use in order to perform systematic calculations.

Using the instanton liquid parameters  $n \approx 1 \text{ fm}^{-4}$ ,  $\rho \approx 1/3 \text{ fm}$ , one can reproduce the phenomenological values of the quark and gluon condensates. In addition to that, one can calculate direct instanton corrections to hadronic correlation functions at short distances. The results were found to be in good agreement with experiment in both attractive ( $\pi, K$ ) and repulsive ( $\eta'$ ) pseudoscalar meson channels (Shuryak, 1983).

## 3. Technical development during the eighties

Despite its phenomenological success, there was no theoretical justification for the instanton liquid model. The first steps towards providing some theoretical basis for the instanton model were taken by Ilgenfritz and Müller-Preussker (1981) and Diakonov and Petrov (1984). These authors used variational techniques and the mean-field approximation to deal with the statistical mechanics of the instanton liquid. The ensemble was stabilized using a phenomenological core (Ilgenfritz and Müller-Preussker, 1981) or a repulsive interaction derived from a specific ansatz for the gauge-field interaction (Diakonov and Petrov, 1984). The resulting ensembles were found to be consistent with the phenomenological estimates.

The instanton ensemble in the presence of light quarks was studied by Diakonov and Petrov (1986). This work introduced the picture of the quark condensate as a collective state built from delocalized zero modes. The quark condensate was calculated using the mean-field approximation and found to be in agreement with experiment. Hadronic states were studied in the random-phase approximation. At least in the case of pseudoscalar mesons, the results were also in good agreement with experiment.

In parallel, numerical methods for studying the instanton liquid were developed (Shuryak, 1988a). Numerical simulations allow one to go beyond the mean-field and random-phase approximations, and include the 't Hooft interaction to all orders. This means that one can also study hadronic channels that, like vector mesons, do not have first-order instanton-induced interactions, or chan-

<sup>7</sup>There is one historical episode that we should mention briefly. Crewther and Christos (Christos, 1984) questioned the sign of the axial charge violation caused by instantons. A rebuttal of these arguments can be found in 't Hooft (1986).

<sup>8</sup>Today, there is substantial evidence from lattice simulations that the topological susceptibility is dominated by instantons; see Sec. III.C.2.

nels, like the nucleon, that are difficult to treat in the random-phase approximation.

Nevertheless, many important aspects of the model remain to be understood. This applies in particular to the theoretical foundation of the instanton liquid model. When the instanton anti-instanton interaction was studied in more detail, it became clear that there is no classical repulsion in the gauge-field interaction. A well-separated instanton anti-instanton pair is connected to the perturbative vacuum by a smooth path (Balitsky and Yung, 1986; Verbaarschot, 1991). This means that the instanton ensemble cannot be stabilized by purely classical interactions because, in general, it is not possible to separate nonperturbative (instanton-induced) and perturbative effects. Only in special cases, like quantum mechanics (Sec. II.A.5) and supersymmetric field theory (Sec. VIII.C), has this separation been accomplished.

#### 4. Recent progress

In the past few years a great deal has been learned about instantons in QCD. The instanton liquid model with the parameters mentioned above, now referred to as the random instanton liquid model, has been used for large-scale, quantitative calculations of hadronic correlation functions in essentially all meson and baryon channels (Shuryak and Verbaarschot, 1993a; Schäfer, Shuryak, and Verbaarschot, 1994). Hadronic masses and coupling constants for most of the low-lying mesons and baryon states have been shown to be in quantitative agreement with experiment.

The next surprise came from a comparison of the correlators calculated in the random model and the first results from lattice calculations (Chu *et al.*, 1993a). The results agree quantitatively, not only in channels that were already known from phenomenology, but also in others (such as the nucleon and delta) where no previous information (except for the particle masses, of course) existed.

These calculations were followed up by direct studies of the instanton liquid on the lattice. Using a procedure called cooling, one can extract the classical content of strongly fluctuating lattice configurations. Using cooled configurations, the MIT group determined the main parameters of the instanton liquid (Chu *et al.*, 1994). Within 10% error, the density and average size coincided with the values suggested a decade earlier. In the meantime, these numbers had been confirmed by other calculations (see Sec. III.C). In addition to that, it was shown that the agreement between lattice correlation functions and the instanton model was not a coincidence: the correlators were essentially unaffected by cooling. This implied that neither perturbative (removed by cooling) nor confinement (strongly reduced) forces are crucial for hadronic properties.

Technical advances in numerical simulation of the instanton liquid led to the construction of a self-consistent, interacting instanton ensemble, which satisfied all the general constraints imposed by the trace anomaly and chiral low-energy theorems (Shuryak and Verbaarschot,

1995; Schäfer and Shuryak, 1996a, 1996b). The corresponding unquenched (with fermion vacuum bubbles included) correlation functions significantly improved the description of the  $\eta'$  and  $\delta$  mesons, which are the two channels in which the random model fails completely.

Finally, significant progress was made in understanding the instanton liquid at finite temperature and the mechanism for the chiral phase transition. It was realized earlier that, at high temperature, instantons should be suppressed by Debye screening (Shuryak, 1978b; Pisarski and Yaffe, 1980). Therefore it was generally assumed that chiral symmetry restoration is a consequence of the disappearance of instantons at high temperature.

More recently it was argued that, up to the critical temperature, the density of instantons should not be suppressed (Shuryak and Velkovsky, 1994). This prediction was confirmed by direct lattice measurements of the topological susceptibility (Chu and Schramm, 1995), which indeed found little change in the topological susceptibility for  $T < T_c$  and the expected suppression for  $T > T_c$ . If instantons are not suppressed around  $T_c$ , a different mechanism for the chiral phase transition is needed. It was suggested that chiral symmetry is restored because the instanton liquid is rearranged, going from a random phase below  $T_c$  to a correlated phase of instanton anti-instanton molecules above  $T_c$  (Ilgenfritz and Shuryak, 1994; Schäfer, Shuryak, and Verbaarschot, 1995). This idea was confirmed in direct simulations of the instanton ensemble, and a number of consequences of the scenario were explored.

#### D. Topics that are not discussed in detail

There is a vast literature on instantons (the SLAC database lists over 3000 references, which probably does not include the more mathematically oriented works), and limitations of space and time, as well as of our expertise, have forced us to exclude many interesting subjects from this review. Our emphasis in writing this review has been on the theory and phenomenology of instantons in QCD. We discuss instantons in other models only to the extent that they are pedagogically valuable or provide important lessons for QCD. Let us mention a few important omissions and give a brief guide to the relevant literature:

- (1) Direct manifestations of small-size instantons in high-energy baryon-number-violating reactions. The hope is that in these processes one may observe rather spectacular instanton effects in a regime where reliable semiclassical calculations are possible. In the electroweak theory, instantons lead to baryon-number violation, but the amplitude for this reaction is strongly suppressed at low energies. It was hoped that this suppression could be overcome at energies on the order of the sphaleron barrier  $E \approx 10$  TeV, but the emerging consensus is that this dramatic phenomenon will not be observable. Some of the literature is mentioned in Sec. VIII.B; see also the recent review of Aoyama *et al.* (1997).

- (2) The transition from tunneling to thermal activation and the calculation of the sphaleron rate at high temperature. This question is of interest in connection with baryogenesis in the early universe and axial charge fluctuations in the quark-gluon plasma. A recent discussion can be found in the review of Smilga (1996).
- (3) The decay of unstable vacua in quantum mechanics or field theory (Coleman, 1977). A more recent review can be found in Aoyama *et al.* (1997).
- (4) Direct instanton contributions to deep-inelastic scattering and other hard processes in QCD. See Balitskii and Braun (1993; 1995) and the review of Ringwald and Schrempp (1994).
- (5) Instanton-inspired models of hadrons or phenomenological Lagrangians supplemented by the 't Hooft interaction. These models include the Nambu–Jona-Lasinia models (Hatsuda and Kunihiro, 1994), soliton models (Diakonov, Petrov, and Poblitsa, 1988; Christov *et al.*, 1996), potential models (Blask *et al.*, 1990), bag models (Dorokhov, Zubov, and Kochelev, 1992), etc.
- (6) Mathematical aspects of instantons (Eguchi, Gilkey, and Hanson, 1980), the construction of the most general  $n$ -instanton solution (Atiyah, Hitchin, Drinfeld, and Manin, 1977), constrained instantons (Auffleck, 1981), instantons and four-manifolds (Fried and Uhlenbeck, 1984), and the connection between instantons and solitons (Atiyah and Manton, 1989). For a review of known solutions of the classical Yang-Mills field equations in both Euclidean and Minkowski space, we refer the reader to Actor (1979).
- (7) Formal aspects of the supersymmetric instanton calculus, spinor techniques, etc. This material is covered by Novikov, Shifman, Voloshin, and Zakharov (1983), Novikov, Shifman, Vainshtein, and Zakharov (1985a), Novikov (1987), and Amati *et al.* (1988).
- (8) The strong  $CP$  problem, bounds on the theta parameter, the axion mechanism (Peccei and Quinn, 1977), etc. Some remarks on these questions can be found in Sec. II.C.3.

## II. SEMICLASSICAL THEORY OF TUNNELING

### A. Tunneling in quantum mechanics

#### 1. Quantum mechanics in Euclidean space

This section serves as a brief introduction to path-integral methods and can easily be skipped by readers familiar with the subject. We shall demonstrate the use of Feynman diagrams in a simple quantum-mechanical problem, which does not suffer from any of the divergencies that occur in field theory. Indeed, we hope that this simple example will find its way into introductory field theory courses.

Another point we should like to emphasize in this section is the similarity between quantum and statistical mechanics. Qualitatively, both quantum and statistical

mechanics deal with variables that are subject to random fluctuations (quantum or thermal), so that only ensemble-averaged quantities make sense. Formally, the connection is related to the similarity between the statistical partition function  $\text{tr}[\exp(-\beta H)]$  and the generating functional (6) (see below), describing the dynamic evolution of a quantum system.

Consider the simplest possible quantum-mechanical system, the motion of a particle of mass  $m$  in a time-independent potential  $V(x)$ . The standard approach is based on an expansion in terms of stationary states  $\psi_n(x)$ , given as solutions of the Schrödinger equation  $H\psi_n = E_n\psi_n$ . Instead, we shall concentrate on another object, the Green's function<sup>9</sup>

$$G(x, y, t) = \langle y | \exp(-iHt) | x \rangle, \quad (4)$$

which is the amplitude for a particle to go from point  $x$  at time  $t=0$  to point  $y$  at time  $t$ . The Green's function can be expanded in terms of stationary states

$$G(x, y, t) = \sum_{n=1}^{\infty} \psi_n^*(x) \psi_n(y) \exp(-iE_n t). \quad (5)$$

This representation has many nice features that are described in standard text books on quantum mechanics. There is, however, another representation that is more useful in introducing semiclassical methods and in dealing with systems containing many degrees of freedom, the Feynman path integral (Feynman and Hibbs, 1965),

$$G(x, y, t) = \int Dx(t) \exp(iS[x(t)]). \quad (6)$$

Here, the Green's function is given as a sum over all possible paths  $x(t)$  leading from  $x$  at  $t=0$  to  $y$  at time  $t$ . The weight for the paths is provided by the action  $S = \int dt [m\dot{x}^2/2 - V(x)]$ . One way to provide a more precise definition of the path integral is to discretize the path. Dividing the time axis into  $N$  intervals,  $a = t/N$ , the path integral becomes an  $N$ -dimensional integral over  $x_n = x(t_n = an)$   $n=1, \dots, N$ . The discretized action is given by

$$S = \sum_n \left[ \frac{m}{2a} (x_n - x_{n-1})^2 - aV(x_n) \right]. \quad (7)$$

This form is not unique; other discretizations with the same continuum limit can also be used. The path integral is now reduced to a multiple integral over  $x_n$ , where we have to take the limit  $n \rightarrow \infty$ . In general, only Gaussian integrals can be done exactly. In the case of the harmonic oscillator  $V(x) = m\omega^2 x^2/2$  one finds (Feynman and Hibbs, 1965)

$$G_{osc}(x, y, t) = \left( \frac{m\omega}{2\pi i \sin \omega t} \right)^{1/2} \exp \left[ \left( \frac{im\omega}{2 \sin \omega t} \right) \times ((x^2 + y^2) \cos(\omega t) - 2xy) \right]. \quad (8)$$

<sup>9</sup>We use natural units  $\hbar = h/2\pi = 1$  and  $c = 1$ . Mass, energy, and momentum all have dimensions of inverse length.

In principal, the discretized action (7) should be amenable to numerical simulations. In practice, the strongly fluctuating phase in (6) renders this approach completely useless. There is, however, a simple way to get around the problem. If one performs an analytic continuation of  $G(x, y, t)$  to imaginary (Euclidean) time  $\tau = it$ , the weight function becomes  $\exp(-S_E[x(\tau)])$  with the Euclidean action  $S_E = \int dt [m(dx/d\tau)^2/2 + V(x)]$ . In this case, we have a positive-definite weight, and numerical simulations, even for multidimensional problems, are feasible. Note that the relative sign of the kinetic- and potential-energy terms in the Euclidean action has changed, making it look like a Hamiltonian. In Euclidean space, the discretized action (7) looks like the energy functional of a one-dimensional spin chain with nearest-neighbor interactions. This observation provides the formal link between an  $n$ -dimensional statistical system and Euclidean quantum (field) theory in  $(n-1)$  dimensions.

Euclidean Green's functions can be interpreted in terms of "thermal" distributions. If we use periodic boundary conditions ( $x=y$ ) and integrate over  $x$ , we obtain the statistical sum

$$\int dx G(x, x, \tau) = \sum_{n=1}^{\infty} \exp(-E_n \tau), \quad (9)$$

where the time interval  $\tau$  plays the role of an inverse temperature  $\tau = T^{-1}$ . In particular,  $G(x, x, 1/T)$  has the physical meaning of a probability distribution for  $x$  at temperature  $T$ . For the harmonic oscillator mentioned above, the Euclidean Green's function is

$$G_{osc}(x, y, \tau) = \left( \frac{m\omega}{2\pi \sinh \omega\tau} \right)^{1/2} \exp \left[ - \left( \frac{m\omega}{2 \sinh \omega\tau} \right) \times ((x^2 + y^2) \cosh(\omega\tau) - 2xy) \right]. \quad (10)$$

For  $x=y$ , the spatial distribution is Gaussian at any  $T$ , with a width  $\langle x^2 \rangle = 1/2m\omega \coth(\omega/2T)$ . If  $\tau$  is very large, the effective temperature is small and the ground state dominates. From the exponential decay, we can read off the ground-state energy  $E_0 = \omega/2$ , and from the spatial distribution, the width of the ground-state wave function  $\langle x^2 \rangle = (2m\omega)^{-1}$ . At high  $T$  we get the classical result  $\langle x^2 \rangle = T/(m\omega^2)$ .

Non-Gaussian path integrals cannot be done exactly. As long as the nonlinearities are small, we can use perturbation theory. Consider an anharmonic oscillator with (Euclidean) action

$$S_E = \int d\tau \left[ \frac{\dot{x}^2}{2} + \frac{\omega^2 x^2}{2} + \alpha x^3 + \beta x^4 \right]. \quad (11)$$

Expanding the path integral in powers of  $\alpha$  and  $\beta$ , one can derive the Feynman rules for the anharmonic oscillator. The free propagator is given by

$$G_0(\tau_1, \tau_2) = \langle x(\tau_1)x(\tau_2) \rangle = \frac{1}{2\omega} \exp(-\omega|\tau_1 - \tau_2|). \quad (12)$$

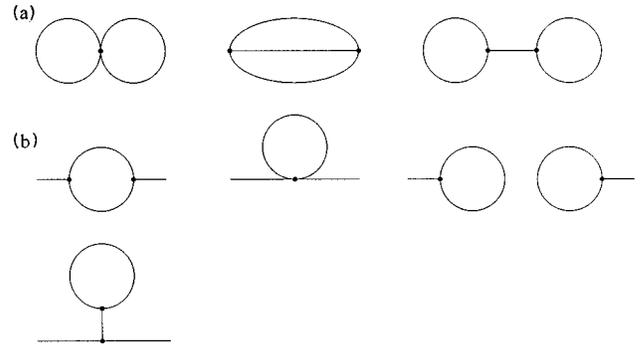


FIG. 3. Feynman diagrams for the anharmonic oscillator: (a) energy; (b) Green's function.

In addition to that, there are three- and four-point vertices with coupling constants  $\alpha$  and  $\beta$ . To calculate an  $n$ -point Green's function, we have to sum over all diagrams with  $n$  external legs and integrate over the time variables corresponding to internal vertices.

The vacuum energy is given by the sum of all closed diagrams. At one-loop order, there is only one diagram, the free-particle loop diagram. At two-loop order, there are two  $O(\alpha^2)$  and one  $O(\beta)$  diagram [see Fig. 3(a)]. The calculation of the diagrams is remarkably simple. Since the propagator is exponentially suppressed for large times, everything is finite. Summing all the diagrams, we get

$$\langle 0 | \exp(-H\tau) | 0 \rangle = \sqrt{\frac{\omega}{\pi}} \exp\left(-\frac{\omega\tau}{2}\right) \times \left[ 1 - \left( \frac{3\beta}{4\omega^2} - \frac{11\alpha^2}{8\omega^4} \right) \tau + \dots \right]. \quad (13)$$

For small  $\alpha^2$ ,  $\beta$ , and  $\tau$  not too large, we can exponentiate the result and read off the correction to the ground-state energy

$$E_0 = \frac{\omega}{2} + \frac{3\beta}{4\omega^2} - \frac{11\alpha^2}{8\omega^4} + \dots \quad (14)$$

Of course, we could have obtained the result using ordinary Rayleigh-Schrödinger perturbation theory, but the method discussed here proves to be much more powerful when we come to nonperturbative effects and field theory.

One more simple exercise is worth mentioning: the evaluation of first perturbative corrections to the Green's function. The diagrams shown in Fig. 3(b) give

$$\Delta G_0(0, \tau) = \frac{9\alpha^2}{4\omega^6} + \frac{\alpha^2}{2\omega^6} e^{-2\omega\tau} + \frac{15\alpha^2}{4\omega^5} \tau e^{-\omega\tau} - \frac{3\beta}{2\omega^3} \tau e^{-\omega\tau}. \quad (15)$$

Comparing the result with the decomposition in terms of stationary states

$$G(0, \tau) = \sum_{n=0}^{\infty} e^{-(E_n - E_0)\tau} |\langle 0 | x | n \rangle|^2, \quad (16)$$

we can identify the first (time-independent) term with the square of the ground-state expectation value  $\langle 0|x|0\rangle$  (which is nonzero due to the tadpole diagram). The second term comes from the excitation of two quanta, and the last two (with extra factors of  $\tau$ ) are the lowest-order “mass renormalization,” or corrections to the zero-order gap between the ground and first excited states,  $E_1 - E_0 = \omega$ .

## 2. Tunneling in the double-well potential

Tunneling phenomena in quantum mechanics were discovered by George Gamow in the late 1920’s in the context of alpha decay. He introduced the exponential suppression factor that explained why a decay governed by the Coulomb interaction (with a typical nuclear time scale of  $10^{-22}$  sec) could lead to lifetimes of millions of years. Tunneling is a quantum-mechanical phenomenon, a particle penetrating a classically forbidden region. Nevertheless, we shall describe the tunneling process using classical equations of motion. Again, the essential idea is to continue the transition amplitude to imaginary time.

Let us give a qualitative argument for why tunneling can be treated as a classical process in imaginary time. The energy of a particle moving in the potential  $V(x)$  is given by  $E = p^2/(2m) + V(x)$ , and in classical mechanics only regions of phase space where the kinetic energy is positive are accessible. In order to reach the classically forbidden region  $E < V$ , the kinetic energy would have to be negative, corresponding to imaginary momentum  $p$ . In the semiclassical WKB approximation to quantum mechanics, the wave function is given by  $\psi(x) \sim \exp[i\Phi(x)]$  with  $\Phi(x) = \pm \int^x dx' p(x') + O(\hbar)$  where  $p(x) = (2m)^{1/2}[E - V(x)]^{1/2}$  is the local classical momentum. In the classically allowed region, the wave function is oscillatory, while in the classically forbidden region (corresponding to imaginary momenta) it is exponentially suppressed.

Another way to introduce imaginary momenta, which is more easily generalized to multidimensional problems and field theory, is by considering motion in imaginary time. Continuing  $\tau = it$ , the classical equation of motion is given by

$$m \frac{d^2x}{d\tau^2} = + \frac{dV}{dx}, \tag{17}$$

where the sign of the potential-energy term has changed. This means that classically forbidden regions are now classically allowed. The distinguished role of the classical tunneling path becomes clear if one considers the Feynman path integral. Although any path is allowed in quantum mechanics, the path integral is dominated by paths that maximize the weight factor  $\exp(-S[x_{cl}(\tau)])$  or minimize the Euclidean action. The classical path is the path with the smallest possible action.

Let us consider a widely used toy model, the double-well potential,

$$V = \lambda(x^2 - \eta^2)^2, \tag{18}$$

with minima at  $\pm \eta$ , the two “classical vacua” of the system. Quantizing around the two minima, we would find two degenerate states localized at  $x = \pm \eta$ . Of course, we know that this is not the correct result. Tunneling will mix the two states; the true ground state is (approximately) the symmetric combination, while the first excited state is the antisymmetric combination of the two states.

It is easy to solve the equations of motion in imaginary time and obtain the classical tunneling solution<sup>10</sup>

$$x_{cl}(\tau) = \eta \tanh\left[\frac{\omega}{2}(\tau - \tau_0)\right], \tag{19}$$

which goes from  $x(-\infty) = -\eta$  to  $x(\infty) = \eta$ . Here,  $\tau_0$  is a free parameter (the instanton center) and  $\omega^2 = 8\lambda\eta^2$ . The action of the solution is  $S_0 = \omega^3/(12\lambda)$ . We shall refer to the path (19) as the instanton, since (unlike a soliton) the solution is localized in time.<sup>11</sup> An anti-instanton solution is given by  $x_{cl}^A(\tau) = -x_{cl}^I(\tau)$ . It is convenient to rescale the time variable such that  $\omega = 1$  and to shift  $x$  such that one of the minima is at  $x = 0$ . In this case, there is only one dimensionless parameter,  $\lambda$ , and since  $S_0 = 1/(12\lambda)$ , the validity of the semiclassical expansion is controlled by  $\lambda \ll 1$ .

The semiclassical approximation to the path integral is obtained by systematically expanding around the classical solution

$$\begin{aligned} \langle -\eta | e^{-H\tau} | \eta \rangle &= e^{-S_0} \int Dx(\tau) \\ &\times \exp\left(-\frac{1}{2} \delta x \left. \frac{\delta^2 S}{\delta x^2} \right|_{x_{cl}} \delta x + \dots\right). \end{aligned} \tag{20}$$

Note that the linear term is absent because  $x_{cl}$  is a solution of the equations of motion. Also note that we implicitly assume  $\tau$  to be large, but smaller than the typical lifetime for tunneling. If  $\tau$  is larger than the lifetime, we have to take into account multi-instanton configurations (see below). Clearly, the tunneling amplitude is proportional to  $\exp(-S_0)$ . The pre-exponent requires the calculation of fluctuations around the classical instanton solution. We shall study this problem in the following section.

## 3. Tunneling amplitude at one-loop order

In order to take into account fluctuations around the classical path, we have to calculate the path integral

$$\int [D\delta x] \exp\left(-\frac{1}{2} \int d\tau \delta x(\tau) O \delta x(\tau)\right) \tag{21}$$

<sup>10</sup>This solution is most easily found using energy conservation  $m\dot{x}^2/2 - V(x) = \text{const}$  rather than the (second-order) equation of motion  $m\ddot{x} = V'$ . This is analogous to the situation in field theory, where it is more convenient to use self-duality than the equations of motion.

<sup>11</sup>In (1+1)-dimensional  $\phi^4$  theory, there is a soliton solution with the same functional form, usually referred to as the kink solution.

where  $O$  is the differential operator,

$$O = -\frac{1}{2} \frac{d^2}{d\tau^2} + \frac{d^2 V}{dx^2} \Big|_{x=x_{cl}}. \tag{22}$$

This calculation is somewhat technical, but it provides a very good illustration of the steps that are required to solve the more difficult field-theory problem. We follow here the original work (Polyakov, 1977) and the review (Vainshtein *et al.*, 1982). A simpler method for calculating the determinant is described in the appendix of Coleman's lecture notes (Coleman, 1977).

The integral (21) is Gaussian, so it can be done exactly. Expanding the differential operator  $O$  in some basis  $\{x_i(\tau)\}$ , we have

$$\int \left( \prod_n dx_n \right) \exp \left( -\frac{1}{2} \sum_{ij} x_i O_{ij} x_j \right) = (2\pi)^{n/2} (\det O)^{-1/2}. \tag{23}$$

The determinant can be calculated by diagonalizing  $O$ ,  $Ox_n(\tau) = \epsilon_n x_n(\tau)$ . This eigenvalue equation is just a one-dimensional Schrödinger equation,<sup>12</sup>

$$\left( -\frac{d^2}{d\tau^2} + \omega^2 \left[ 1 - \frac{3}{2 \cosh^2(\omega\tau/2)} \right] \right) x_n(\tau) = \epsilon_n x_n(\tau). \tag{24}$$

There are two bound states plus a continuum of scattering states. The lowest eigenvalue is  $\epsilon_0 = 0$ , and the other bound state is at  $\epsilon_1 = \frac{3}{4}\omega^2$ . The eigenfunction of the zero-energy state is

$$x_0(\tau) = \sqrt{\frac{3\omega}{8}} \frac{1}{\cosh^2(\omega\tau/2)}, \tag{25}$$

where we have normalized the wave function,  $\int d\tau x_n^2 = 1$ . There should be a simple explanation for the presence of a zero mode. Indeed, the appearance of a zero mode is related to translational invariance, the fact that the action does not depend on the location  $\tau_0$  of the instanton. The zero-mode wave function is just the derivative of the instanton solution over  $\tau_0$ ,

$$x_0(\tau) = -S_0^{-1/2} \frac{d}{d\tau_0} x_{cl}(\tau - \tau_0), \tag{26}$$

where the normalization follows from the fact that the classical solution has zero energy. If one of the eigenvalues is zero, this means that the determinant vanishes and the tunneling amplitude is infinite. However, the presence of a zero mode also implies that there is one direction in functional space in which fluctuations are large, so the integral is not Gaussian. This means that the integral in that direction should not be performed in a Gaussian approximation, but has to be done exactly.

This can be achieved by replacing the integral over the expansion parameter  $c_0$  associated with the zero-mode direction [we have parametrized the path by

$x(\tau) = \sum_n c_n x_n(\tau)$ ] with an integral over the collective coordinate  $\tau_0$ . Using

$$dx = \frac{dx_{cl}}{d\tau_0} d\tau_0 = -\sqrt{S_0} x_0(\tau) d\tau_0 \tag{27}$$

and  $dx = x_0 dc_0$ , we have  $dc_0 = \sqrt{S_0} d\tau_0$ . The functional integral over the quantum fluctuation is now given by

$$\int [D\delta x(\tau)] \exp(-S) = \left[ \prod_{n>0} \left( \frac{2\pi}{\epsilon_n} \right) \right]^{1/2} \sqrt{S_0} \int d\tau_0, \tag{28}$$

where the first factor, the determinant with the zero mode excluded, is often referred to as  $\det' O$ . The result shows that the tunneling amplitude grows linearly with time. This is as it should be; there is a finite transition probability per unit time.

The next step is the calculation of the non-zero-mode determinant. For this purpose we make the spectrum discrete by considering a finite time interval  $(-\tau_m/2, \tau_m/2)$  and imposing boundary conditions at  $\pm \tau_m/2$ :  $x_n(\pm \tau_m/2) = 0$ . The product of all eigenvalues is divergent, but the divergence is related to large eigenvalues, independent of the detailed shape of the potential. The determinant can be renormalized by taking the ratio over the determinant of the free harmonic oscillator. The result is

$$\begin{aligned} & \left( \frac{\det \left[ -\frac{d^2}{d\tau^2} + V''(x_{cl}) \right]}{\det \left[ -\frac{d^2}{d\tau^2} + \omega^2 \right]} \right)^{-1/2} \\ &= \sqrt{\frac{S_0}{2\pi}} \omega \int d\tau_0 \left( \frac{\det' \left[ -\frac{d^2}{d\tau^2} + V''(x_{cl}) \right]}{\omega^{-2} \det \left[ -\frac{d^2}{d\tau^2} + \omega^2 \right]} \right)^{-1/2} \end{aligned} \tag{29}$$

where we have eliminated the zero mode from the determinant and replaced it by the integration over  $\tau_0$ . We also have to extract the lowest mode from the harmonic-oscillator determinant, which is given by  $\omega^2$ . The next eigenvalue is  $3\omega^2/4$ , while the corresponding oscillator mode is  $\omega^2$  (up to corrections of order  $1/\tau_m^2$ , that are not important as  $\tau_m \rightarrow \infty$ ). The rest of the spectrum is continuous as  $\tau_m \rightarrow \infty$ . The contribution from these states can be calculated as follows.

The potential  $V''(x_{cl})$  is localized, so for  $\tau \rightarrow \pm \infty$  the eigenfunctions are just plane waves. This means we can take one of the two linearly independent solutions to be  $x_p(\tau) \sim \exp(ip\tau)$  as  $\tau \rightarrow \infty$ . The effect of the potential is to give a phase shift

$$x_p(\tau) = \exp(ip\tau + i\delta_p) \quad \tau \rightarrow -\infty, \tag{30}$$

where, for this particular potential, there is no reflected wave. The phase shift is given by (Landau and Lifshitz, 1959)

$$\exp(i\delta_p) = \frac{1 + ip/\omega}{1 - ip/\omega} \frac{1 + 2ip/\omega}{1 - 2ip/\omega}. \tag{31}$$

<sup>12</sup>This particular Schrödinger equation is discussed in many textbooks on quantum mechanics, e.g., Landau and Lifshitz (1959).

The second independent solution is obtained by  $\tau \rightarrow -\tau$ . The spectrum is determined by the quantization condition  $x(\pm\tau_m/2) = 0$ , which gives

$$p_n \tau_m - \delta_{p_n} = \pi n, \quad (32)$$

while the harmonic-oscillator modes are determined by  $p_n \tau_m = \pi n$ . If we denote the solutions of Eq. (32) by  $\tilde{p}_n$ , the ratio of the determinants is given by

$$\begin{aligned} \prod_n \left[ \frac{\omega^2 + \tilde{p}_n^2}{\omega^2 + p_n^2} \right] &= \exp \left( \sum_n \log \left[ \frac{\omega^2 + \tilde{p}_n^2}{\omega^2 + p_n^2} \right] \right) \\ &= \exp \left( \frac{1}{\pi} \int_0^\infty \frac{2p dp \delta_p}{p^2 + \omega^2} \right) = \frac{1}{9}, \end{aligned} \quad (33)$$

where we have expanded the integrand in the small difference  $\tilde{p}_n - p_n = \delta_{p_n} / \tau_m$  and changed from summation over  $n$  to an integral over  $p$ . In order to perform the integral, it is convenient to integrate by part and use the result for  $(d\delta_p)/(dp)$ . Collecting everything, we finally get

$$\begin{aligned} \langle -\eta | e^{-H\tau_m} | \eta \rangle &= \left[ \sqrt{\frac{\omega}{\pi}} \exp \left( -\frac{\omega\tau_m}{2} \right) \right] \\ &\quad \times \left[ \sqrt{\frac{6S_0}{\pi}} \exp(-S_0) \right] (\omega\tau_m), \end{aligned} \quad (34)$$

where the first factor comes from the harmonic-oscillator amplitude and the second is the ratio of the two determinants.

The result shows that the transition amplitude is proportional to the time interval  $\tau_m$ . In terms of stationary states this is due to the fact that the contributions from the two lowest states almost cancel each other. The ground-state wave function is the symmetric combination  $\Psi_0(x) = [\phi_{-\eta}(x) + \phi_\eta(x)]/\sqrt{2}$ , while the first excited state  $E_1 = E_0 + \Delta E$  is antisymmetric,  $\Psi_1(x) = [\phi_{-\eta}(x) - \phi_\eta(x)]/\sqrt{2}$ . Here,  $\phi_{\pm\eta}$  are the harmonic-oscillator wave functions around the two classical minima. For times  $\tau \ll 1/\Delta E$ , the tunneling amplitude is given by

$$\begin{aligned} \langle -\eta | e^{-H\tau_m} | \eta \rangle &= \Psi_0^*(-\eta) \Psi_0(\eta) e^{-E_0\tau_m} \\ &\quad + \Psi_1^*(-\eta) \Psi_1(\eta) e^{-E_1\tau_m} + \dots \\ &= \frac{1}{2} \phi_{-\eta}^*(-\eta) \phi_\eta(\eta) (\Delta E \tau_m) e^{-\omega\tau_m/2} \\ &\quad + \dots \end{aligned} \quad (35)$$

Note that the validity of the semiclassical approximation requires  $\tau \gg 1/\omega$ .

We can read off the level splitting from Eqs. (34) and (35). The result can be obtained in an even more elegant way by going to large times  $\tau > 1/\Delta E$ . In this case, multi-instanton paths are important. If we ignore the interaction between instantons, multi-instanton contributions can easily be summed up

$$\begin{aligned} \langle -\eta | e^{-H\tau_m} | \eta \rangle &= \sqrt{\frac{\omega}{\pi}} e^{-\omega\tau_m/2} \sum_{n \text{ odd}} \int_{-\tau_m/2 < \tau_1 < \dots < \tau_m/2} \\ &\quad \times \left[ \prod_{i=1}^n \omega d\tau_i \right] \left( \sqrt{\frac{6S_0}{\pi}} \exp(-S_0) \right)^n \\ &= \sqrt{\frac{\omega}{\pi}} e^{-\omega\tau_m/2} \sum_{n \text{ odd}} \frac{(\omega\tau_m d)^n}{n!} \\ &= \sqrt{\frac{\omega}{\pi}} e^{-\omega\tau_m/2} \sinh(\omega\tau_m d), \end{aligned} \quad (36)$$

where  $d = (6S_0/\pi)^{1/2} \exp(-S_0)$ . Summing over all instantons simply leads to the exponentiation of the tunneling rate. Now we can directly read off the level splitting

$$\Delta E = \sqrt{\frac{6S_0}{\pi}} \omega \exp(-S_0). \quad (37)$$

If the tunneling rate increases,  $1/\Delta E \approx 1/\omega$ , interactions between instantons become important. We shall study this problem in Sec. II.A.5.

#### 4. The tunneling amplitude at two-loop order

The WKB method can be used to calculate higher orders systematically in  $1/S_0$ . Beyond leading order, however, the WKB method becomes quite tedious, even when applied to quantum mechanics. In this subsection we show how the  $1/S_0$  correction to the level splitting in the double well can be determined using a two-loop instanton calculation. We follow here Wohler and Shuryak (1994), who corrected a few mistakes in the earlier literature (Aleinikov and Shuryak, 1987; Olejnik, 1989). Numerical simulations were performed by Shuryak (1988b), while the correct result was first obtained using different methods by Zinn-Justin (1981).

To next order in  $1/S_0$ , the tunneling amplitude can be decomposed as

$$\begin{aligned} \langle -\eta | e^{-H\tau} | \eta \rangle &\approx |\psi_0(\eta)|^2 \left( 1 + \frac{2A}{S_0} + \dots \right) \\ &\quad \times \exp \left( -\frac{\omega\tau}{2} \left[ 1 + \frac{B}{S_0} + \dots \right] \right) \\ &\quad \times \Delta E_0 \left( 1 + \frac{C}{S_0} + \dots \right) \tau, \end{aligned} \quad (38)$$

where we are interested in the coefficient  $C$ , the next order correction to the level splitting. The other two corrections,  $A$  and  $B$ , are unrelated to tunneling, and we can get rid of them by dividing the amplitude by  $\langle \eta | \exp(-H\tau) | \eta \rangle$ ; see Eq. (15).

In order to calculate the next-order correction to the instanton result, we have to expand the action beyond order  $(\delta x)^2$ . The result can be interpreted in terms of a new set of Feynman rules in the presence of an instanton (see Fig. 4). The triple and quartic coupling constants are  $\alpha = 4\lambda x_{cl}(t)$  and  $\beta = \lambda$  (compared to  $\alpha_0 = 4\lambda \eta = \sqrt{2\lambda}$  and  $\beta_0 = \lambda$  for the anharmonic oscillator).

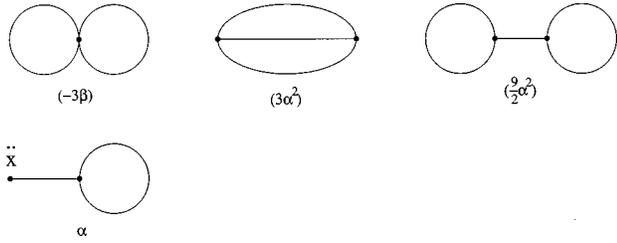


FIG. 4. Feynman diagrams for the two-loop correction to the tunneling amplitude in the quantum-mechanical double-well potential. The first three correspond to the diagrams in Fig. 3(a) but with different propagators and vertices, while the fourth diagram contains a new vertex, generated by the collective coordinate Jacobian.

The propagator is the Green's function of the differential operator (22). There is one complication due to the fact that the operator  $O$  has a zero mode. The Green's function is uniquely defined by requiring it to be orthogonal to the translational zero mode. The result is (Olejnik, 1989)

$$(2\omega)G(x,y) = g_0(x,y) \left[ 2 - xy + \frac{1}{4}|x-y|(11-3xy) + (x-y)^2 \right] + \frac{3}{8}(1-x^2)(1-y^2) \times \left[ \log[g_0(x,y)] - \frac{11}{3} \right], \tag{39}$$

$$g_0(x,y) = \frac{1 - |x-y| - xy}{1 + |x-y| - xy}, \tag{40}$$

where  $x = \tanh(\omega t/2)$ ,  $y = \tanh(\omega t'/2)$ , and  $g_0(x,y)$  is the Green's function of the harmonic oscillator (12). There are four diagrams at two-loop order (see Fig. 4). The first three diagrams are of the same form as the anharmonic-oscillator diagrams. Subtracting these contributions, we get

$$a_1 = -3\beta_0 \int_{-\infty}^{\infty} dt (G^2(t,t) - G_0^2(t,t)) = -\frac{97}{1680} S_0^{-1}, \tag{41}$$

$$b_{11} = 3\alpha_0^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dt dt' (\tanh(t/2) \times \tanh(t'/2) G^3(t,t') - G_0^3(t,t')) = -\frac{53}{1260} S_0^{-1}, \tag{42}$$

$$b_{12} = \frac{9}{2} \alpha_0^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dt dt' (\alpha^2 \tanh(t/2) \times \tanh(t'/2) G(t,t) G(t,t') G(t',t') - G_0(t,t) G_0(t,t') G_0(t',t')) = -\frac{39}{560} S_0^{-1}. \tag{43}$$

The last diagram comes from expanding the Jacobian in

$\delta x$ . This leads to a tadpole graph proportional to  $\ddot{x}_{cl}$ , which has no counterpart in the anharmonic-oscillator case. We get

$$c_1 = -9\beta_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dt dt' \frac{\tanh(t/2)}{\cosh^2(t/2)} \times \tanh(t'/2) G(t,t') G(t',t') = -\frac{49}{60} S_0^{-1}. \tag{44}$$

The sum of the four diagrams is  $C = (a_1 + b_{11} + b_{12} + c_1) S_0 = -71/72$ . The two-loop result for the level splitting is

$$\Delta E = \sqrt{\frac{6S_0}{\pi}} \omega \exp\left(-S_0 - \frac{71}{72} \frac{1}{S_0} + \dots\right), \tag{45}$$

in agreement with the WKB result obtained by Zinn-Justin (1981). The fact that the next-order correction is of order one and negative is significant. It implies that the one-loop result becomes inaccurate for moderately large barriers ( $S \sim 1$ ) and that it overestimates the tunneling probability. We have presented this calculation in detail in order to show that the instanton method can be systematically extended to higher orders in  $1/S$ . In field theory, however, this calculation is sufficiently difficult that it has not yet been performed.

### 5. Instanton/anti-instanton interaction and the ground-state energy

Up to now we have focused on the tunneling amplitude for transitions between the two degenerate vacua of the double-well potential. This amplitude is directly related to the gap  $\Delta E$  between the ground state and the first excited state. In this subsection we wish to discuss how the semiclassical theory can be used to calculate the mean  $E_{ctr} = (E_0 + E_1)/2$  of the two levels. In this context, it is customary to define the double-well potential by  $V = (x^2/2)(1 - gx)^2$ . The coupling constant  $g$  is related to the coupling  $\lambda$  used above by  $g^2 = 2\lambda$ . Unlike the splitting, the mean energy is related to topologically trivial paths connecting the same vacua. The simplest nonperturbative path of this type is an instanton/anti-instanton pair.

In Sec. II.A.3 we calculated the tunneling amplitude using the assumption that instantons do not interact with each other. We found that tunneling makes the coordinates uncorrelated and leads to a level splitting. If we take the interaction among instantons into account, the contribution from instanton/anti-instanton pairs is given by

$$\langle \eta | e^{-H\tau_m} | \eta \rangle = \tau_m \int \frac{d\tau}{\pi g^2} \exp[S_{IA}(\tau)], \tag{46}$$

where  $S_{IA}(\tau)$  is the action of an instanton/anti-instanton pair with separation  $\tau$ , and the prefactor  $(\pi g^2)^{-1}$  comes from the square of the single-instanton density. The action of an instanton/anti-instanton (IA) pair can be calculated given an ansatz for the path that goes from one

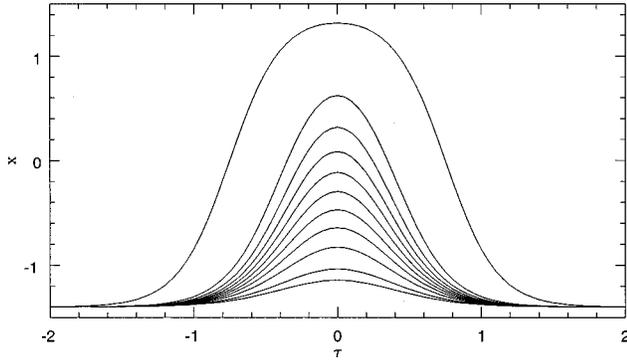


FIG. 5. Streamline configurations in the double-well potential for  $\eta=1.4$  and  $\lambda=1$  ( $\omega=4$ ,  $S_0 \approx 5$ ), adapted from Shuryak, 1988b. The horizontal axis shows the time coordinate and the vertical axis the amplitude  $x(\tau)$ . The different paths correspond to different values of the streamline parameter  $\lambda$  as the configuration evolves from a well-separated pair to an almost perturbative path. The initial path has an action  $S=1.99S_0$ , the other paths correspond to a fixed reduction of the action by  $0.2S_0$ .

minimum of the potential to the other and back. An example of such a path is the “sum ansatz” (Zinn-Justin, 1983),

$$x_{sum}(\tau) = \frac{1}{2g} \left[ 2 - \tanh\left(\frac{\tau - \tau_I}{2}\right) + \tanh\left(\frac{\tau - \tau_A}{2}\right) \right]. \quad (47)$$

This path has the action  $S_{IA}(T) = 1/g^2 [1/3 - 2e^{-T} + O(e^{-2T})]$ , where  $T = |\tau_I - \tau_A|$ . It is qualitatively clear that if the two instantons are separated by a large time interval  $T \gg 1$ , the action  $S_{IA}(T)$  is close to  $2S_0$ . In the opposite limit  $T \rightarrow 0$ , the instanton and the anti-instanton annihilate, and the action  $S_{IA}(T)$  should tend to zero. In that limit, however, the IA pair is at best an approximate solution of the classical equations of motion, and it is not clear how the path should be chosen.

The best way to deal with this problem is the “streamline” or “valley” method (Balitsky and Yung, 1986). In this approach one starts with a well-separated IA pair and lets the system evolve using the method of steepest descent. This means that we have to solve

$$f(\lambda) \frac{dx_\lambda(\tau)}{d\lambda} = \frac{\delta S}{\delta x_\lambda(\tau)}, \quad (48)$$

where  $\lambda$  labels the path as we proceed from the initial configuration  $x_{\lambda=0}(\tau) = x_{sum}(\tau)$  down the valley to the vacuum configuration, and  $f(\lambda)$  is an arbitrary function that reflects the reparametrization invariance of the streamline solution. A sequence of paths obtained by solving the streamline equation (48) numerically is shown in Fig. 5 (Shuryak, 1988b). An analytical solution to first order in  $1/S_0$  can be found in Balitsky and Yung (1986). The action density  $s = \dot{x}^2/2 + V(x)$  corresponding to the paths in Fig. 5 is shown in Fig. 6. We can see clearly how the two localized solutions merge and eventually disappear as the configuration progresses down the valley. Using the streamline solution, we find the instanton/anti-instanton action for large  $T$  (Faleev and Silvestrov, 1995)

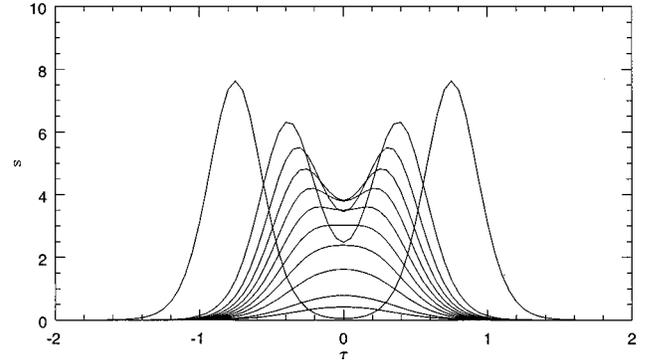


FIG. 6. Distribution of the action density  $s = \dot{x}^2/2 + V(x)$  for the streamline configurations shown in Fig. 5.

$$S(T) = \frac{1}{g^2} \left[ \frac{1}{3} - 2e^{-T} - 12Te^{-2T} + O(e^{-2T}) \right] + [24Te^{-T} + O(e^{-T})] + \left[ \frac{71g^2}{6} + O(g^4) \right], \quad (49)$$

where the first term is the classical streamline interaction up to next-to-leading order, the second term is the quantum correction (one-loop) to the leading-order interaction, and the last term is the two-loop correction to the single-instanton interaction.

If one tries to use the instanton result (46) in order to calculate corrections to  $E_{ctr}$ , one encounters two problems. First, the integral diverges at large  $T$ . This is simply related to the fact that IA pairs with large separation should not be counted as pairs, but as independent instantons. This problem is easily solved; all one has to do is subtract the square of the single-instanton contribution. Second, once this subtraction has been performed, the integral is dominated by the region of small  $T$ , where the action is not reliably calculable. This problem is indeed a serious one, stemming from the fact that  $E_{ctr}$  is not directly related to tunneling, but is dominated by perturbative contributions. In general, we expect  $E_{ctr}$  to have an expansion

$$E_{ctr} = \sum_k g^{2k} E_{ctr,k}^{(0)} + e^{-2/(6g^2)} \sum_k g^{2k} E_{ctr,k}^{(2)} + \dots, \quad (50)$$

where the first term is the perturbative contribution, the second corresponds to one IA pair, and so on. However, perturbation theory in  $g$  is divergent (not even Borel summable), so the calculation of the IA contribution requires a suitable definition of perturbation theory.

One way to deal with this problem (going back to Dyson’s classical work on QED) is analytic continuation in  $g$ . For  $g$  imaginary ( $g^2$  negative), the IA contribution is well defined [the integral over  $T$  is dominated by  $T \sim -\log(-g^2)$ ]. The IA contribution to  $E_{ctr}$  is (Bogomolny, 1980; Zinn-Justin, 1983)

$$E_{ctr}^{(2)} = \frac{e^{-1/(3g^2)}}{\pi g^2} \left[ \log\left(-\frac{2}{g^2}\right) + \gamma + O(g^2 \log(g^2)) \right], \quad (51)$$

where  $\gamma=0.577\dots$  is Euler's constant. When we now continue back to positive  $g^2$ , we get both real and imaginary contributions to  $E_{\text{ctr}}$ . Since the sum of all contributions to  $E_{\text{ctr}}$  is certainly real, the imaginary part has to cancel against a small  $O(e^{-1/(3g^2)})$  imaginary part in the perturbative expansion. This allows us to determine the imaginary part  $\text{Im} E_{\text{ctr}}^{(0)}$  of the analytically continued perturbative sum.<sup>13</sup>

From the knowledge of the imaginary part of perturbation theory, one can determine the large-order behavior of the perturbation series  $E_{\text{ctr}}^{(0)} = \sum_k g^{2k} E_{\text{ctr},k}^{(0)}$  (Brezin, Parisi, and Zinn-Justin, 1977; Lipatov, 1977). The coefficients are given by the dispersion integrals

$$E_{\text{ctr},k}^{(0)} = \frac{1}{\pi} \int_0^\infty \text{Im}[E_{\text{ctr},k}^{(0)}(g^2)] \frac{dg^2}{g^{2k+2}}. \quad (52)$$

Since the semiclassical result (51) is reliable for small  $g$ , we can calculate the large-order coefficients. Including the corrections calculated by Faleev and Silvestrov (1995), we have

$$E_{\text{ctr},k}^{(0)} = \frac{3^{k+1}k}{\pi} \left( 1 - \frac{53}{18k} + \dots \right). \quad (53)$$

This result can be compared with the exact coefficients (Brezin *et al.*, 1977). For small  $k$  the result is completely wrong, but already for  $k=5,6,7,8$  the ratio of the asymptotic result to the exact coefficients is 1.04, 1.11, 1.12, 1.11. We conclude that instantons determine the large-order behavior of the perturbative expansion. This is in fact a generic result: the asymptotic behavior of perturbation theory is governed by semiclassical configurations (although not necessarily involving instantons).

In order to check the instanton/anti-instanton result (51) against the numerical value of  $E_{\text{ctr}}$  for different values of  $g$ , we have to subtract the perturbative contribution to  $E_{\text{ctr}}$ . This can be done using analytic continuation and the Borel transform (Zinn-Justin, 1982), and the result is in very good agreement with the instanton calculation. A simpler way to check the instanton result was proposed by Faleev and Silvestrov (1995). These authors simply truncated the perturbative series at the  $N$ th term. In this case, the best accuracy occurs when  $|N-1/3g^2| \sim 1$  and the estimate for  $E_{\text{ctr}}$  is given by

$$E_{\text{ctr}} = \sum_{n=0}^N g^{2n} E_{\text{ctr},n}^{(0)} + \frac{3Ne^{-N}}{\pi} \left[ \log(6N) + \gamma + \frac{1}{3} \sqrt{\frac{2\pi}{N}} \left( 1 - \frac{53}{18N} \right) \right], \quad (54)$$

which is compared to the exact values in Table I. We observe that the result (54) is indeed very accurate, and that the error is on the order of  $e^{-N} = e^{-1/3g^2}$ .

<sup>13</sup>How can the perturbative result develop an imaginary part? After analytic continuation, the perturbative sum is Borel summable because the coefficients alternate in sign. If we define  $E_{\text{ctr}}^{(0)}$  by analytic continuation of the Borel sum, it will have an imaginary part for positive  $g^2$ .

TABLE I. Exact center-of-the-band energies  $E_{\text{ctr}} = (E_0 + E_1)/2$  for different values of  $g^2$  [expressed in terms of  $N = 1/(3g^2)$ ] compared to the semiclassical estimate discussed in the text.

$N \equiv 1/(3g^2)$	4	6	8	10	12
$E_{\text{ctr}}^{\text{ex}}$	0.4439	0.43797	0.44832	0.459178	0.467156
$E_{\text{ctr}}^{\text{th}}$	0.4367	0.44367	0.44933	0.459307	0.467173

In summary,  $E_{\text{ctr}}$  is related to configurations with no net topology, and in this case the calculation of instanton effects requires a suitable definition of the perturbation series. This can be accomplished using analytic continuation in the coupling constant. After analytic continuation, we can perform a reliable interacting instanton calculation, but the result has an imaginary part. This shows that the instanton contribution by itself is not well determined—it depends on the definition of the perturbation sum. However, the sum of perturbative and nonperturbative contributions is well defined (and real) and agrees very accurately with the numerical value of  $E_{\text{ctr}}$ .

In gauge theories the situation is indeed very similar: there are both perturbative and nonperturbative contributions to the vacuum energy, and the two contributions are not clearly separated. However, in the case of gauge theories, we do not know how to define perturbation theory, so we are not yet able to perform a reliable calculation of the vacuum energy similar to Eq. (54).

## B. Fermions coupled to the double-well potential

In this section we consider one fermionic degree of freedom  $\psi_\alpha$  ( $\alpha=1,2$ ) coupled to the double-well potential. This model provides additional insight into the vacuum structure, not only of quantum mechanics, but also of gauge theories. We shall see that fermions are intimately related to tunneling and that the fermion-induced interaction between instantons leads to strong instanton/anti-instanton correlations. Another motivation for studying fermions coupled to the double-well potential is that for a particular choice of the coupling constant, the theory is supersymmetric. This means that perturbative corrections to the vacuum energy cancel, and the instanton contribution is more easily defined.

The model is defined by the action

$$S = \frac{1}{2} \int dt (\dot{x}^2 + W'^2 + \psi \dot{\psi} + c W'' \psi \sigma_2 \psi), \quad (55)$$

where  $\psi_\alpha$  ( $\alpha=1,2$ ) is a two component spinor, dots denote time and primes spatial derivatives, and  $W' = x(1 - gx)$ . We shall see that the vacuum structure depends crucially on the Yukawa coupling  $c$ . For  $c=0$  fermions decouple, and we recover the double-well potential studied in the previous sections, while for  $c=1$  the classical action is supersymmetric. The supersymmetry transformation is given by

$$\delta x = \zeta \sigma_2 \psi, \quad \delta \psi = \sigma_2 \zeta \dot{x} - W' \zeta, \quad (56)$$

where  $\zeta$  is a Grassmann variable. For this reason,  $W$  is usually referred to as the superpotential. The action (55) can be rewritten in terms of two bosonic partner potentials (Salomonson and Holton, 1981; Cooper, Khare, and Sukhatme, 1995). Nevertheless, it is instructive to keep the fermionic degree of freedom, because the model has many interesting properties that also apply to QCD, where the action cannot be bosonized.

As before, the potential  $V = \frac{1}{2}W'^2$  has degenerate minima connected by the instanton solution. The tunneling amplitude is given by

$$\text{Tr}(e^{-\beta H}) = \int d\tau J \frac{1}{\sqrt{\det \mathcal{O}_B}} \sqrt{\det \mathcal{O}_F} e^{-S_{\text{cl}}}, \quad (57)$$

where  $S_{\text{cl}}$  is the classical action,  $\mathcal{O}_B$  is the bosonic operator (22), and  $\mathcal{O}_F$  is the Dirac operator,

$$\mathcal{O}_F = \frac{d}{dt} + c\sigma_2 W''(x_{\text{cl}}). \quad (58)$$

As explained in Sec. II.A.3,  $\mathcal{O}_B$  has a zero mode, related to translational invariance. This mode has to be treated separately, which leads to a Jacobian  $J$  and an integral over the corresponding collective coordinate  $\tau$ . The fermion determinant also has a zero mode,<sup>14</sup> given by

$$\chi^{I,A} = N \exp\left(\mp \int_{-\infty}^t dt' c W''(x_{\text{cl}})\right) \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \mp i \end{pmatrix}. \quad (59)$$

Since the fermion determinant appears in the numerator of the tunneling probability, the presence of a zero mode implies that the tunneling rate is zero.

The reason for this is simple: the two vacua have different fermion number, so they cannot be connected by a bosonic operator. The tunneling amplitude is nonzero only if a fermion is created during the process  $\langle 0,+ | \psi_+ | 0,- \rangle$ , where  $\psi_{\pm} = (1/\sqrt{2})(\psi_1 \pm i\psi_2)$  and  $|0,\pm\rangle$  denote the corresponding eigenstates. Formally, we get a finite result because the fermion creation operator absorbs the zero mode in the fermion determinant. As we shall see later, this mechanism is completely analogous to the axial  $U(1)_A$  anomaly in QCD and baryon-number violation in electroweak theory. For  $c=1$ , the tunneling rate is given by (Salomonson and Holton, 1981)

$$\langle 0,+ | \psi_+ | 0,- \rangle = \frac{1}{\sqrt{\pi g^2}} e^{-1/6g^2}. \quad (60)$$

This result can be checked by performing a direct calculation using the Schrödinger equation.

Let us now return to the calculation of the ground-state energy. For  $c=0$ , the vacuum energy is the sum of perturbative contributions and a negative nonperturbative shift  $O(e^{-1/(6g^2)})$  due to individual instantons. For  $c \neq 0$ , the tunneling amplitude [Eq. (60)] will only enter squared, so one needs to consider instanton/anti-

instanton pairs. Between the two tunneling events, the system has an excited fermionic state, which causes a new interaction between the instantons. For  $c=1$ , supersymmetry implies that all perturbative contributions (including the zero-point oscillation) to the vacuum energy cancel. Using supersymmetry, one can calculate the vacuum energy from the tunneling rate<sup>15</sup> [Eq. (60)] (Salomonson and Holton, 1981). The result is  $O(e^{-1/(3g^2)})$  and positive, which implies that supersymmetry is broken.<sup>16</sup> While the dependence on  $g$  is what we would expect for a gas of instanton/anti-instanton molecules, understanding the sign in the context of an instanton calculation is more subtle (see below).

It is an instructive exercise to calculate the vacuum energy numerically (which is quite straightforward; we are still dealing with a simple quantum-mechanical toy model). In general, the vacuum energy is nonzero, but for  $c \rightarrow 1$ , the vacuum energy is zero up to exponentially small corrections. Varying the coupling constant, one can verify that the vacuum energy is smaller than any power in  $g$ , showing that supersymmetry breaking is a nonperturbative effect.

For  $c \neq 1$ , the instanton/anti-instanton contribution to the vacuum energy has to be calculated directly. Even for  $c=1$ , where the result can be determined indirectly, this is a very instructive calculation. For an instanton/anti-instanton path, there is no fermionic zero mode. Writing the fermion determinant in the basis spanned by the original zero modes of the individual pseudoparticles, we have

$$\det(\mathcal{O}_F)_{ZMZ} = \begin{pmatrix} 0 & T_{\text{IA}} \\ T_{\text{AI}} & 0 \end{pmatrix}, \quad (61)$$

where  $T_{\text{IA}}$  is the overlap matrix element,

$$T_{\text{IA}} = \int_{-\infty}^{\infty} dt \chi_A(\partial_t + c\sigma_2 W''(x_{\text{IA}}(t))) \chi_I. \quad (62)$$

Clearly, mixing between the two zero modes shifts the eigenvalues away from zero, and the determinant is nonzero. As before, we have to choose the correct instanton/anti-instanton path  $x_{\text{IA}}(t)$  in order to evaluate  $T_{\text{IA}}$ . Using the valley method introduced in the last section, we find the ground-state energy (Balitsky and Yung, 1986)

$$E = \frac{1}{2} [1 - c + O(g^2)] - \frac{1}{2\pi} e^{-1/3g^2} \left(\frac{g^2}{2}\right)^{c-1} 2c^c \int_0^{\infty} d\tau \times \exp\left(-2c(\tau - \tau_0) + \frac{2}{g^2} e^{-2\tau}\right), \quad (63)$$

<sup>15</sup>The reason is that, for supersymmetry theories, the Hamiltonian is the square of the supersymmetry generators  $Q_{\alpha}$ ,  $H = \frac{1}{2}\{Q_+, Q_-\}$ . Since the tunneling amplitude  $\langle 0,+ | \psi_+ | 0,- \rangle$  is proportional to the matrix element of  $Q_+$  between the two different vacua, the ground-state energy is determined by the square of the tunneling amplitude.

<sup>16</sup>This was indeed the first known example of nonperturbative supersymmetry breaking (Witten, 1981).

<sup>14</sup>In the supersymmetric case, the fermion zero mode is the superpartner of the translational zero mode.

where  $\tau$  is the instanton/anti-instanton separation and  $\exp(-2\tau_0) = cg^2/2$ . The two terms in the exponent inside the integral correspond to the fermionic and bosonic interactions between instantons. One can see that fermions cut off the integral at large  $\tau$ . There is an attractive interaction, which grows with distance and forces instantons and anti-instantons to be correlated. Therefore, for  $c \neq 0$ , the vacuum is no longer an ensemble of random tunneling events, but consists of correlated instanton/anti-instanton molecules.

The fact that both the bosonic and the fermionic interaction are attractive means that the integral (63), just like (46), is dominated by small  $\tau$  where the integrand is not reliable. This problem can be solved as outlined in the last section, by analytic continuation in the coupling constant. As an alternative, Balitsky and Yung suggested shifting the integration contour in the complex  $\tau$  plane,  $\tau \rightarrow \tau + i\pi/2$ . On this path, the sign of the bosonic interaction is reversed, and the fermionic interaction picks up a phase factor  $\exp(ic\pi)$ . This means that there is a stable saddle point, but the instanton contribution to the ground-state energy is in general complex. The imaginary part cancels against the imaginary part of the perturbation series, and only the sum of the two contributions is well defined.

A special case is the supersymmetric point  $c=1$ . In this case, perturbation theory vanishes and the contribution from instanton/anti-instanton molecules is real,

$$E = \frac{1}{2\pi} e^{-1/3g^2} [1 + O(g^2)]. \quad (64)$$

This implies that at the supersymmetry point  $c=1$  there is a well-defined instanton/anti-instanton contribution. The result agrees with what one finds from the  $H = \frac{1}{2}\{Q_+, Q_-\}$  relation or directly from the Schrödinger equation.

In summary, in the presence of light fermions, tunneling is possible only if the fermion number changes during the transition. Fermions create a long-range attractive interaction between instantons and anti-instantons, and the vacuum is dominated by instanton/anti-instanton “molecules.” It is nontrivial to calculate the contribution of these configurations to the ground-state energy because topologically trivial paths can mix with perturbative corrections. The contribution of molecules is most easily defined if one allows the collective coordinate (time separation) to be complex. In this case, there exists a saddle point where the repulsive bosonic interaction balances the attractive fermionic interaction and molecules are stable. These objects give a nonperturbative contribution to the ground-state energy, which is in general complex, except in the supersymmetric case where it is real and positive.

### C. Tunneling in Yang-Mills theory

#### 1. Topology and classical vacua

Before we study tunneling phenomena in Yang-Mills theory, we have to become more familiar with the clas-

sical vacuum of the theory. In the Hamiltonian formulation, it is convenient to use the temporal gauge  $A_0=0$  (here we use matrix notation  $A_i = A_i^a \lambda^a/2$ , where the  $SU(N)$  generators satisfy  $[\lambda^a, \lambda^b] = 2if^{abc}\lambda^c$  and are normalized according to  $\text{Tr}(\lambda^a \lambda^b) = 2\delta^{ab}$ ). In this case, the conjugate momentum to the field variables  $A_i(x)$  is just the electric field  $E_i = \partial_0 A_i$ . The Hamiltonian is given by

$$H = \frac{1}{2g^2} \int d^3x (E_i^2 + B_i^2), \quad (65)$$

where  $E_i^2$  is the kinetic and  $B_i^2$  the potential-energy term. The classical vacuum corresponds to configurations with zero field strength. For non-Abelian gauge fields this does not imply that the potential has to be constant, but limits the gauge fields to be “pure gauge,”

$$A_i = iU(\vec{x})\partial_i U(\vec{x})^\dagger. \quad (66)$$

In order to enumerate the classical vacua, we have to classify all possible gauge transformations  $U(\vec{x})$ . This means that we have to study equivalence classes of maps from 3-space  $R^3$  into the gauge group  $SU(N)$ . In practice, we can restrict ourselves to matrices satisfying  $U(\vec{x}) \rightarrow 1$  as  $x \rightarrow \infty$  (Callan *et al.*, 1978a). Such mappings can be classified using an integer called the winding (or Pontryagin) number, which counts how many times the group manifold is covered,

$$n_W = \frac{1}{24\pi^2} \int d^3x \epsilon^{ijk} \text{Tr}[(U^\dagger \partial_i U)(U^\dagger \partial_j U) \times (U^\dagger \partial_k U)]. \quad (67)$$

In terms of the corresponding gauge fields, this number is the Chern-Simons characteristic

$$n_{CS} = \frac{1}{16\pi^2} \int d^3x \epsilon^{ijk} \left( A_i^a \partial_j A_k^a + \frac{1}{3} f^{abc} A_i^a A_j^b A_k^c \right). \quad (68)$$

Because of its topological meaning, continuous deformations of the gauge fields do not change  $n_{CS}$ . In the case of  $SU(2)$ , an example of a mapping with winding number  $n$  can be found from the “hedgehog” ansatz,

$$U(\vec{x}) = \exp[if(r)\tau^a \hat{x}^a], \quad (69)$$

where  $r = |\vec{x}|$  and  $\hat{x}^a = x^a/r$ . For this mapping, we find

$$n_W = \frac{2}{\pi} \int dr \sin^2(f) \frac{df}{dr} = \frac{1}{\pi} \left[ f(r) - \frac{\sin[2f(r)]}{2} \right]_0^\infty. \quad (70)$$

In order for  $U(\vec{x})$  to be uniquely defined,  $f(r)$  has to be a multiple of  $\pi$  at both zero and infinity, so that  $n_W$  is indeed an integer. Any smooth function with  $f(r \rightarrow \infty) = 0$  and  $f(0) = n\pi$  provides an example of a function with winding number  $n$ .

We conclude that there is an infinite set of classical vacua enumerated by an integer  $n$ . Since they are topologically different, one cannot go from one vacuum to another by means of a continuous gauge transformation. Therefore there is no path from one vacuum to another such that the energy remains zero all the way.

2. Tunneling and the Belavin-Polyakov-Schwartz-Tyupkin instanton

Two important questions concerning the classical vacua immediately come to mind. First, is there some physical observable that distinguishes between them? Second, is there any way to go from one vacuum to another? The answer to the first question is positive but most easily demonstrated in the presence of light fermions, so we shall come to it later. Let us now concentrate on the second question.

We are going to look for a tunneling path in gauge theory which connects topologically different classical vacua. From the quantum-mechanical example, we know that we have to look for classical solutions of the Euclidean equations of motion. The best tunneling path is the solution with minimal Euclidean action connecting vacua with different Chern-Simons numbers. To find these solutions, it is convenient to exploit the identity

$$S = \frac{1}{4g^2} \int d^4x G_{\mu\nu}^a G_{\mu\nu}^a = \frac{1}{4g^2} \int d^4x \left[ \pm G_{\mu\nu}^a \tilde{G}_{\mu\nu}^a + \frac{1}{2} (G_{\mu\nu}^a \mp \tilde{G}_{\mu\nu}^a)^2 \right], \quad (71)$$

where  $\tilde{G}_{\mu\nu} = 1/2 \epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}$  is the dual field strength tensor (the field tensor in which the roles of electric and magnetic fields are interchanged). Since the first term is a topological invariant (see below) and the last term is always positive, it is clear that the action is minimal if the field is (anti) self-dual,

$$G_{\mu\nu}^a = \pm \tilde{G}_{\mu\nu}^a. \quad (72)$$

One can also show directly that the self-duality condition implies the equations of motion,<sup>17</sup>  $D_\mu G_{\mu\nu} = 0$ . This is a useful observation because, in contrast to the equation of motion, the self-duality equation (72) is a first-order differential equation. In addition to that, one can show that the energy-momentum tensor vanishes for self-dual fields. In particular, self-dual fields have zero (Minkowski) energy density.

The action of a self-dual field configuration is determined by the topological charge (or four-dimensional Pontryagin index)

$$Q = \frac{1}{32\pi^2} \int d^4x G_{\mu\nu}^a \tilde{G}_{\mu\nu}^a. \quad (73)$$

From Eq. (71), we have  $S = (8\pi^2|Q|)/g^2$  for self-dual fields. For finite-action field configurations,  $Q$  has to be an integer. This can be seen from the fact that the integrand is a total derivative,

$$Q = \int d^4x \partial_\mu K_\mu = \int d\sigma_\mu K_\mu, \quad (74)$$

$$K_\mu = \frac{1}{16\pi^2} \epsilon_{\mu\alpha\beta\gamma} \left( A_\alpha^a \partial_\beta A_\gamma^a + \frac{1}{3} f^{abc} A_\alpha^a A_\beta^b A_\gamma^c \right). \quad (75)$$

<sup>17</sup>The reverse is not true, but one can show that non-self-dual solutions of the equations of motion are saddle points, not local minima of the action.

For finite-action configurations, the gauge potential has to be pure gauge at infinity,  $A_\mu \rightarrow iU\partial_\mu U^\dagger$ . In analogy with the arguments given in the last section, all maps from the 3-sphere  $S_3$  (corresponding to  $|x| \rightarrow \infty$ ) into the gauge group can be classified by a winding number  $n$ . Inserting  $A_\mu = iU\partial_\mu U^\dagger$  into Eq. (74), one finds that  $Q = n$ .

Furthermore, if the gauge potential falls off rapidly at spatial infinity, then

$$Q = \int dt \frac{d}{dt} \int d^3x K_0 = n_{CS}(t=\infty) - n_{CS}(t=-\infty), \quad (76)$$

which shows that field configurations with  $Q \neq 0$  connect different topological vacua. In order to find an explicit solution with  $Q = 1$ , it is useful to start from the simplest winding number  $n = 1$  configuration. As in Eq. (69), we can take  $A_\mu = iU\partial_\mu U^\dagger$  with  $U = i\hat{x}_\mu \tau_\mu^+$ , where  $\tau_\mu^\pm = (\vec{\tau}, \mp i)$ . Then  $A_\mu^a = 2\eta_{a\mu\nu} x_\nu / x^2$ , where we have introduced the 't Hooft symbol  $\eta_{a\mu\nu}$ ,

$$\eta_{a\mu\nu} = \begin{cases} \epsilon_{a\mu\nu}, & \mu, \nu = 1, 2, 3, \\ \delta_{a\mu}, & \nu = 4, \\ -\delta_{a\nu}, & \mu = 4. \end{cases} \quad (77)$$

We also define  $\bar{\eta}_{a\mu\nu}$  by changing the sign of the last two equations. Further properties of  $\eta_{a\mu\nu}$  are summarized in the Appendix, Sec. A.3. We can now look for a solution of the self-duality equation (72) using the ansatz  $A_\mu^a = 2\eta_{a\mu\nu} x_\nu f(x^2)/x^2$ , where  $f$  has to satisfy the boundary condition  $f \rightarrow 1$  as  $x^2 \rightarrow \infty$ . Inserting the ansatz in Eq. (72), we get

$$f(1-f) - x^2 f' = 0. \quad (78)$$

This equation is solved by  $f = x^2/(x^2 + \rho^2)$ , which gives the Belavin-Polyakov-Schwartz-Tyupkin instanton solution (Belavin *et al.*, 1975),

$$A_\mu^a(x) = \frac{2\eta_{a\mu\nu} x_\nu}{x^2 + \rho^2}. \quad (79)$$

Here  $\rho$  is an arbitrary parameter characterizing the size of the instanton. A solution with topological charge  $Q = -1$  can be obtained by replacing  $\eta_{a\mu\nu}$  with  $\bar{\eta}_{a\mu\nu}$ . The corresponding field strength is

$$(G_{\mu\nu}^a)^2 = \frac{192\rho^4}{(x^2 + \rho^2)^4}. \quad (80)$$

In our conventions, the coupling constant appears only as a factor in front of the action. This convention is very convenient in dealing with classical solutions. For perturbative calculations, it is more common to rescale the fields as  $A_\mu \rightarrow gA_\mu$ . In this case, there is a factor  $1/g$  in the instanton gauge potential, which shows that the field of the instanton is much stronger than an ordinary, perturbative field.

Note that  $G_{\mu\nu}^a$  is well localized (it falls off as  $1/x^4$ ) despite the fact that the gauge potential is long range,  $A_\mu \sim 1/x$ . The invariance of the Yang-Mills equations under coordinate inversion (Jackiw and Rebbi, 1976b) implies that the singularity of the potential can be shifted from infinity to the origin by means of a (singular) gauge transformation  $U = i\hat{x}_\mu \tau^+$ . The gauge potential in singular gauge is given by

$$A_\mu^a(x) = 2 \frac{x_\nu}{x^2} \frac{\bar{\eta}_{a\mu\nu} \rho^2}{x^2 + \rho^2}. \tag{81}$$

This singularity at the origin is not physical; the field strength and topological charge density are smooth. However, in order to calculate the topological charge from a surface integral over  $K_\mu$ , we need to surround the origin by a small sphere. The topology of this configuration is therefore located at the origin, not at infinity. In order to study instanton/anti-instanton configurations, we shall work mainly with such singular configurations.

The classical instanton solution has a number of degrees of freedom known as collective coordinates. In the case of SU(2), the solution is characterized by the instanton size  $\rho$ , the instanton position  $z_\mu$ , and three parameters which determine the color orientation of the instanton. The group orientation can be specified in terms of the SU(2) matrix  $U$ ,  $A_\mu \rightarrow UA_\mu U^\dagger$ , or the corresponding rotation matrix  $R^{ab} = \frac{1}{2} \text{tr}(U \tau^a U^\dagger \tau^b)$ , such that  $A_\mu^a \rightarrow R^{ab} A_\mu^b$ . Due to the symmetries of the instanton configuration, ordinary rotations do not generate new solutions.

SU(3) instantons can be constructed by embedding the SU(2) solution. For  $|Q|=1$ , there are no genuine SU(3) solutions. The number of parameters characterizing the color orientation is seven, not eight, because one of the SU(3) generators leaves the instanton invariant. For SU( $N$ ), the number of collective coordinates (including position and size) is  $4N$ . There exist exact  $n$ -instanton solutions with  $4nN$  parameters, but they are difficult to construct in general (Atiyah *et al.*, 1977). A simple solution in which the relative color orientations are fixed was given by 't Hooft (unpublished), see Witten (1977), Jackiw, Nohl, and Rebbi (1977), and the Appendix Sec. A.1.

We have explicitly constructed the tunneling path that connects different topological vacua. The instanton action is given by  $S = (8\pi^2|Q|)/g^2$ , implying that the tunneling probability is

$$P_{\text{tunneling}} \sim \exp(-8\pi^2/g^2). \tag{82}$$

As in the quantum-mechanical example, the coefficient in front of the exponent is determined by a one-loop calculation.

### 3. The theta vacua

We have seen that non-Abelian gauge theory has a periodic potential and that instantons connect the different vacua. This means that the ground state of QCD cannot be described by any of the topological vacuum

states but has to be a superposition of all vacua. This problem is similar to the motion of an electron in the periodic potential of a crystal. It is well known that the solutions form a band  $\psi_\theta$  characterized by a phase  $\theta \in [0, 2\pi]$  (sometimes referred to as quasimomentum). The wave functions are Bloch waves satisfying the periodicity condition  $\psi_\theta(x+n) = e^{i\theta n} \psi_\theta(x)$ .

Let us see how this band arises from tunneling events. If instantons are sufficiently dilute, then the amplitude to go from one topological vacuum  $|i\rangle$  to another  $|j\rangle$  is given by

$$\langle j | \exp(-H\tau) | i \rangle = \sum_{N_+} \sum_{N_-} \frac{\delta_{N_+ - N_- - j + i}}{N_+! N_-!} (K\tau e^{-S})^{N_+ + N_-}, \tag{83}$$

where  $K$  is the pre-exponential factor in the tunneling amplitude, and  $N_\pm$  are the numbers of instantons and anti-instantons. Using the identity

$$\delta_{ab} = \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{i\theta(a-b)}, \tag{84}$$

we can write the sum over instantons and anti-instantons as

$$\begin{aligned} \langle j | \exp(-H\tau) | i \rangle &= \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{i\theta(i-j)} \\ &\quad \times \exp[2K\tau \cos(\theta) \exp(-S)]. \end{aligned} \tag{85}$$

This result shows that the true eigenstates are the theta vacua  $|\theta\rangle = \sum_n e^{in\theta} |n\rangle$ . Their energy is

$$E(\theta) = -2K \cos(\theta) \exp(-S). \tag{86}$$

The width of the zone is on the order of the tunneling rate. The lowest state corresponds to  $\theta=0$  and has negative energy. This is as it should be; tunneling lowers the ground-state energy.

Does this result imply that in QCD there is a continuum of states without a mass gap? Not at all. Although one can construct stationary states for any value of  $\theta$ , they are not excitations of the  $\theta=0$  vacuum because in QCD the value of  $\theta$  cannot be changed. As far as the strong interaction is concerned, different values of  $\theta$  correspond to different worlds. Indeed, we can fix the value of  $\theta$  by adding an additional term,

$$\mathcal{L} = \frac{i\theta}{32\pi^2} G_{\mu\nu}^a \tilde{G}_{\mu\nu}^a \tag{87}$$

to the QCD Lagrangian.

Does physics depend on the value of  $\theta$ ? Naively, the interaction (87) violates both  $T$  and  $CP$  invariance. On the other hand, Eq. (87) is a surface term, and one might suspect that confinement somehow screens the effects of the  $\theta$  term. A similar phenomenon is known to occur in three-dimensional compact electrodynamics (Polyakov, 1977). In QCD, however, one can show that if the  $U(1)_A$  problem is solved (there is no massless  $\eta'$  state in the chiral limit) and none of the quarks is mass-

less, a nonzero value of  $\theta$  implies that  $CP$  is broken (Shifman, Vainshtein, and Zakharov, 1980a).

Consider the expectation value of the  $CP$ -violating observable  $\langle G\tilde{G} \rangle$ . Expanding the partition function in powers of  $\theta$ , we have  $\langle G\tilde{G} \rangle = \theta(32\pi^2)\chi_{\text{top}}$ . Furthermore, in Sec. V.E we shall prove an important low-energy theorem that determines the topological susceptibility for small quark masses. Using these results, we have

$$\langle G\tilde{G} \rangle = -\theta(32\pi^2)f_\pi^2 m_\pi^2 \frac{m_u m_d}{(m_u + m_d)^2} \quad (88)$$

for two light flavors to leading order in  $\theta$  and the quark masses. Similar estimates can be obtained for  $CP$ -violating observables that are directly accessible to experiment. The most severe limits on  $CP$  violation in the strong interaction come from the electric dipole of the neutron. Current experiments imply that (Baluni, 1979; Crewther, *et al.*, 1979)

$$\theta < 10^{-9}. \quad (89)$$

The question of why  $\theta$  is so small is known as the strong  $CP$  problem. The status of this problem is unclear. As long as we do not understand the source of  $CP$  violation in nature, it is not clear whether the strong  $CP$  problem is expected to have a solution within the standard model, or whether there is some mechanism outside the standard model that adjusts  $\theta$  to be small.

One possibility is provided by the fact that the state with  $\theta=0$  has the lowest energy. This means that if  $\theta$  becomes a dynamic variable, the vacuum can relax to the  $\theta=0$  state (just as electrons can drop to the bottom of the conduction band by emitting phonons). This is the basis of the axion mechanism (Peccei and Quinn, 1977). The axion is a hypothetical pseudoscalar particle that couples to  $G\tilde{G}$ . The equations of motion for the axion field automatically remove the effective  $\theta$  term, which is now a combination of  $\theta_{\text{QCD}}$  and the axion expectation value. Experimental limits on the axion coupling are very severe, but an “invisible axion” might still exist (Kim, 1979; Shifman, Vainshtein, and Zakharov, 1980b; Zhitnitsky, 1980; Dine and Fischler, 1983; Preskill, Wise, and Wilczek, 1983).

The simplest way to resolve the strong  $CP$  problem is to assume that the mass of the  $u$  quark vanishes (presumably because of a symmetry not manifest in the standard model). Unfortunately, this possibility appears to be ruled out phenomenologically, but there is no way to know for sure before this scenario is explored in more detail on the lattice. More recently, it was suggested that QCD might undergo a phase transition near  $\theta=0, \pi$ . In the former case, there is some support for this idea from lattice simulations (Schierholz, 1994), but the instanton model and lattice measurements of the topological susceptibility do not suggest any singularity around  $\theta=0$ . The latter limit  $\theta=\pi$  also conserves  $CP$  and has a number of interesting properties (Snyderman and Gupta, 1981); in this world the instanton-induced interaction

between quarks would change sign. Clearly, it is important to understand the properties of QCD with a nonzero  $\theta$  angle in more detail.

#### 4. The tunneling amplitude

The next natural step is the one-loop calculation of the pre-exponent in the tunneling amplitude. In gauge theory, this is a rather tedious calculation, which was done in the classic paper by 't Hooft (1976b). Basically, the procedure is completely analogous to what we did in the context of quantum mechanics. The field is expanded around the classical solution,  $A_\mu = A_\mu^{\text{cl}} + \delta A_\mu$ . In QCD, we have to make a gauge choice. In this case, it is most convenient to work in a background field gauge  $D_\mu(A_\nu^{\text{cl}})\delta A_\mu = 0$ .

We have to calculate the one-loop determinants for gauge fields, ghosts, and possible matter fields (which we shall deal with later). The determinants are divergent both in the ultraviolet, like any other one-loop graph, and in the infrared, due to the presence of zero modes. As we shall see below, the two are actually related. In fact, the QCD beta function is partly determined by zero modes (while in certain supersymmetric theories, the beta function is completely determined by zero modes; see Sec. VIII.C.1).

We already know how to deal with the  $4N_c$  zero modes of the system. The integral over the zero mode is traded for an integral over the corresponding collective variable. For each zero mode, we get one factor of the Jacobian  $\sqrt{S_0}$ . The integration over the color orientations is compact, so it just gives a factor, but the integral over size and position, we have to keep. As a result, we get a differential tunneling rate,

$$dn_I \sim \left(\frac{8\pi^2}{g^2}\right)^{2N_c} \exp\left(-\frac{8\pi^2}{g^2}\right) \rho^{-5} d\rho dz, \quad (90)$$

where the power of  $\rho$  can be determined from dimensional considerations.

The ultraviolet divergence is regulated using the Pauli-Villars scheme, which is the most convenient method when dealing with fluctuations around non-trivial classical field configurations (the final result can be converted into any other scheme). This means that the determinant  $\det O$  of the differential operator  $O$  is divided by  $\det(O+M^2)$ , where  $M$  is the regulator mass. Since we have to extract  $4N_c$  zero modes from  $\det O$ , this gives a factor  $M^{4N_c}$  in the numerator of the tunneling probability.

In addition to that, there will be a logarithmic dependence on  $M$  coming from the ultraviolet divergence. To one-loop order, it is just the logarithmic part of the polarization operator. For any classical field  $A_\mu^{\text{cl}}$  the result can be written as a contribution to the effective action (Brown and Creamer, 1978; Vainshtein *et al.*, 1982),

$$\delta S_{NZM} = \frac{2}{3} \frac{g^2}{8\pi^2} \log(M\rho) S(A^{\text{cl}}). \quad (91)$$

In the background field of an instanton, the classical ac-

tion cancels the prefactor  $g^2/8\pi^2$ , and  $\exp(-\delta S_{NZM}) \sim (M\rho)^{-2/3}$ . Now we can collect all terms in the exponent of the tunneling rate,

$$\begin{aligned} dn_I &\sim \exp\left(-\frac{8\pi^2}{g^2} + 4N_c \log(M\rho)\right. \\ &\quad \left. - \frac{N_c}{3} \log(M\rho)\right) \rho^{-5} d\rho dz_\mu \\ &\equiv \exp\left(-\frac{8\pi^2}{g^2(\rho)}\right) \rho^{-5} d\rho dz_\mu, \end{aligned} \quad (92)$$

where we have recovered the running coupling constant  $(8\pi^2)/g^2(\rho) = (8\pi^2)/g^2 - (11N_c/3)\log(M\rho)$ . Thus the infrared and ultraviolet divergent terms combine to give the coefficient of the one-loop beta function,  $b = 11N_c/3$ , and the bare charge and the regulator mass  $M$  can be combined into a running coupling constant. At two-loop order, the renormalization group requires the miracle to happen once again, and the nonzero mode determinant can be combined with the bare charge to give the two-loop beta function in the exponent and the one-loop running coupling in the pre-exponent.

The remaining constant was calculated by 't Hooft (1976b) and Bernard *et al.* (1977). The result is

$$\begin{aligned} dn_I &= \frac{0.466 \exp(-1.679N_c) \left(\frac{8\pi^2}{g^2}\right)^{2N_c}}{(N_c-1)!(N_c-2)!} \\ &\quad \times \exp\left(-\frac{8\pi^2}{g^2(\rho)}\right) \frac{d^4z d\rho}{\rho^5}. \end{aligned} \quad (93)$$

The tunneling rate  $dn_A$  for anti-instantons is of course identical. Using the one-loop beta function, one can write the result as

$$\frac{dn_I}{d^4z} \sim \frac{d\rho}{\rho^5} (\rho\Lambda)^b, \quad (94)$$

and because of the large coefficient  $b = (11N_c/3) = 11$ , the exponent overcomes the Jacobian and small-size instantons are strongly suppressed. On the other hand, there appears to be a divergence at large  $\rho$ . This is related to the fact that the perturbative beta function is not applicable in this regime. We shall come back to this question in Sec. III.

## D. Instantons and light quarks

### 1. Tunneling and the $U(1)_A$ anomaly

When we considered the topology of gauge fields and the appearance of topological vacua, we posed the question whether the different vacua could be physically distinguished. In the presence of light fermions, there is a simple physical observable that distinguishes between the topological vacua, the axial charge. This observation helped to clarify the mechanism of chiral anomalies and showed how perturbative and nonperturbative effects are related in the breaking of classical symmetries.

Anomalies first appeared in the context of perturbation theory (Adler, 1969; Bell and Jackiw, 1969), when it

became apparent that loop diagrams involving external vector and axial-vector currents could not be regulated in such a way that all the currents remained conserved. From the triangle diagram involving two gauge fields and the flavor singlet axial current, one finds

$$\partial_\mu j_\mu^5 = \frac{N_f}{16\pi^2} G_{\mu\nu}^a \tilde{G}_{\mu\nu}^a, \quad (95)$$

where  $j_\mu^5 = \bar{q} \gamma_\mu \gamma_5 q$  with  $q = (u, d, s, \dots)$ . This result is not modified at higher orders in the perturbative expansion. At this point, the gauge field on the right-hand side is some arbitrary background field. The fact that the flavor singlet current has an anomalous divergence was quite welcome in QCD because it seemed to explain the absence of a ninth Goldstone boson, the so-called  $U(1)_A$  puzzle.

Nevertheless, there are two apparent problems with Eq. (95) if we want to understand the  $U(1)_A$  puzzle. The first is that the right-hand side is proportional to the divergence of the topological current,  $\partial_\mu K_\mu$ , so it appears that we can still define a conserved axial current. The other is that, since the right-hand side of the anomaly equation is just a surface term, it seems that the anomaly does not have any physical effects.

The answer to the first problem is that, while the topological charge is gauge invariant, the topological current is not. The appearance of massless poles in correlation functions of  $K_\mu$  does not necessarily correspond to massless particles. The answer to the second question is of course related to instantons. Because QCD has topologically distinct vacua, surface terms are relevant.

In order to see how instantons can lead to the non-conservation of axial charge, let us calculate the change in axial charge,

$$\Delta Q_5 = Q_5(t=+\infty) - Q_5(t=-\infty) = \int d^4x \partial_\mu j_\mu^5. \quad (96)$$

In terms of the fermion propagator,  $\Delta Q_5$  is given by

$$\Delta Q_5 = \int d^4x N_f \partial_\mu \text{tr}(S(x,x) \gamma_\mu \gamma_5). \quad (97)$$

The fermion propagator is the inverse of the Dirac operator,  $S(x,y) = \langle x | (i\mathcal{D})^{-1} | y \rangle$ . For any given gauge field, we can determine the propagator in terms of the eigenfunctions  $i\mathcal{D} \psi_\lambda = \lambda \psi_\lambda$  of the Dirac operator,

$$S(x,y) = \sum_\lambda \frac{\psi_\lambda(x) \psi_\lambda^\dagger(y)}{\lambda}. \quad (98)$$

Using the eigenvalue equation, we can now evaluate  $\Delta Q_5$ ,

$$\Delta Q_5 = N_f \int d^4x \text{tr} \left( \sum_\lambda \frac{\psi_\lambda(x) \psi_\lambda^\dagger(x)}{\lambda} 2\lambda \gamma_5 \right). \quad (99)$$

For every nonzero  $\lambda$ ,  $\gamma_5 \psi_\lambda$  is an eigenvector with eigenvalue  $-\lambda$ . But this means that  $\psi_\lambda$  and  $\gamma_5 \psi_\lambda$  are orthogonal, so only zero modes can contribute to Eq. (99):

$$\Delta Q_5 = 2N_f(n_L - n_R), \quad (100)$$

where  $N_{L,R}$  is the number of left and right-handed zero modes, and we have used the fact that the eigenstates are normalized.

The crucial property of instantons, originally discovered by 't Hooft, is that the Dirac operator has a zero mode  $i\mathcal{D}\psi_0(x)=0$  in the instanton field. For an instanton in the singular gauge, the zero-mode wave function is

$$\psi_0(x) = \frac{\rho}{\pi} \frac{1}{(x^2 + \rho^2)^{3/2}} \frac{\gamma \cdot x}{\sqrt{x^2}} \frac{1 + \gamma_5}{2} \phi, \quad (101)$$

where  $\phi^{am} = \epsilon^{am}/\sqrt{2}$  is a constant spinor in which the  $SU(2)$  color index  $\alpha$  is coupled to the spin index  $m = 1, 2$ . Let us briefly digress in order to show that Eq. (101) is indeed a solution of the Dirac equation. First, observe that<sup>18</sup>

$$(i\mathcal{D})^2 = \left( -D^2 + \frac{1}{2} \sigma_{\mu\nu} G_{\mu\nu} \right). \quad (102)$$

We can now use the fact that  $\sigma_{\mu\nu} G_{\mu\nu}^{(\pm)} = \mp \gamma_5 \sigma_{\mu\nu} G_{\mu\nu}^{(\pm)}$  for (anti) self-dual fields  $G_{\mu\nu}^{(\pm)}$ . In the case of a self-dual gauge potential, the Dirac equation  $i\mathcal{D}\psi=0$  then implies ( $\psi = \chi_L + \chi_R$ ),

$$\left( -D^2 + \frac{1}{2} \sigma_{\mu\nu} G_{\mu\nu}^{(+)} \right) \chi_L = 0, \quad -D^2 \chi_R = 0, \quad (103)$$

and vice versa ( $+\leftrightarrow-, L\leftrightarrow R$ ) for anti-self-dual fields. Since  $-D^2$  is a positive operator,  $\chi_R$  has to vanish, and the zero mode in the background field of an instanton has to be left handed, while it is right handed in the case of an anti-instanton.<sup>19</sup> A general analysis of the solutions of Eq. (103) was given by 't Hooft (1976b) and Jackiw and Rebbi (1977). In practice, the zero mode is most easily found by expanding the spinor  $\chi$  as  $\chi_\alpha^m = M_\mu (\tau_\mu^{(+)})^{am}$ . For (multi) instanton gauge potentials of the form  $A_\mu^a = \bar{\eta}_{\mu\nu}^a \partial_\nu \log \Pi(x)$  (see Appendix A1), it is convenient to make the ansatz (Grossman, 1977)

$$\chi_\alpha^m = \sqrt{\Pi(x)} \partial_\mu \left( \frac{\Phi(x)}{\Pi(x)} \right) (\tau_\mu^{(+)})^{am}. \quad (104)$$

Substituting this ansatz in Eq. (103) shows that  $\Phi(x)$  has to satisfy the Laplace equation  $\square\Phi(x)=0$ . A solution that leads to a normalizable zero mode is given by  $\Phi(x) = \rho^2/x^2$ , from which we finally obtain Eq. (101). Again, we can obtain an  $SU(3)$  solution by embedding the  $SU(2)$  result.

We can now see how tunneling between topologically different configurations (described semiclassically by in-

stantons) explains the axial anomaly. Integrating the anomaly equation (95), we find that  $\Delta Q_5$  is related to the topological charge  $Q$ . On the other hand, from Eq. (100) we know that  $\Delta Q_5$  counts the number of left-handed zero modes minus the number of right-handed zero modes. But this is exactly what instantons do: every instanton contributes one unit to the topological charge and has a left-handed zero mode, while anti-instantons have  $Q = -1$  and a right-handed zero mode.

There is another way to look at this process, known as the ‘‘infinite hotel story’’ (Gribov, 1981; see also Shifman, 1989). Let us consider the gauge potential of an instanton as a fixed background field and diagonalize the time-dependent Dirac Hamiltonian  $i\vec{a} \cdot \vec{D}$  (again, it is most convenient to work in the temporal gauge). The presence of a four-dimensional, normalizable zero mode implies that there is one left-handed state that crosses from positive to negative energy during the tunneling event, while one right-handed state crosses the other way. This can be seen as follows: In the adiabatic approximation, solutions of the Dirac equation are given by

$$\psi_i(\vec{x}, t) = \psi_i(\vec{x}, t = -\infty) \exp\left( - \int_{-\infty}^t dt' \epsilon_i(t') \right). \quad (105)$$

The only way we can have a four-dimensional, normalizable wave function is if  $\epsilon_i$  is positive for  $t \rightarrow \infty$  and negative for  $t \rightarrow -\infty$ . This explains how axial charge can be violated during tunneling. No fermion ever changes its chirality; all states simply move one level up or down. The axial charge comes, so to say, from the ‘‘bottom of the Dirac sea.’’

In QCD, the most important consequence of the anomaly is the fact that the would-be ninth Goldstone boson, the  $\eta'$ , is massive even in the chiral limit. The way the  $\eta'$  acquires its mass is also intimately related with instantons, and we shall come back to this topic a number of times during this review. Historically, the first attempt to understand the origin of the  $\eta'$  mass from the anomaly was based on anomalous Ward identities (Veneziano, 1979); see Sec. V.E. Saturating these Ward identities with hadronic resonances and using certain additional assumptions, one can derive the Witten-Veneziano relation (Veneziano, 1979; Witten, 1979b)

$$\chi_{\text{top}} = \int d^4x \langle Q(x) Q(0) \rangle = \frac{f_\pi^2}{2N_f} (m_\eta^2 + m_{\eta'}^2 - 2m_K^2). \quad (106)$$

In this relation, we have introduced an important new quantity, the topological susceptibility  $\chi_{\text{top}}$ , which measures fluctuations of the topological charge in the QCD vacuum. The combination of meson masses on the right-hand side corresponds to the part of the  $\eta'$  mass that is not due to the strange-quark mass.

There are several subtleties in connection with the Witten-Veneziano relation. In Sec. V.E, we shall show that in QCD with massless flavors the topological charge is screened and  $\chi_{\text{top}} = 0$ . This means that the quantity on the left-hand side of the Witten-Veneziano relation is

<sup>18</sup>We use Euclidean Dirac matrices that satisfy  $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$ . We also have  $\sigma_{\mu\nu} = i/2[\gamma_\mu, \gamma_\nu]$  and  $\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$ .

<sup>19</sup>This result is not an accident. Indeed, there is a mathematical theorem (the Atiyah-Singer index theorem) that requires that  $Q = n_L - n_R$  for every species of chiral fermion. In the case of instantons, this relation was proven by Schwarz (1977); see also the discussion by Coleman (1977).

the topological susceptibility in pure gauge theory<sup>20</sup> (without quarks). But in pure gauge theory, the  $\eta'$  correlation function is pathological (see Sec. VI.C.4), so the  $\eta'$  mass on the right-hand-side of the Witten-Veneziano relation has to be determined in full QCD. This means that the left-hand side and the right-hand side of the Witten-Veneziano relation are actually defined in different theories.<sup>21</sup> Nevertheless, the Witten-Veneziano relation provides a reasonable estimate of the quenched topological susceptibility (Sec. III.C.1), and effective Lagrangians that incorporate the Witten-Veneziano relation provide a good description of the pseudoscalar meson spectrum. We shall also show that the Witten-Veneziano relation can be derived within the instanton liquid model using the mean-field approximation (Sec. IV.G). An analog of the Witten-Veneziano relation in full QCD is described in Sec. V.E. A more detailed study of the  $\eta'$  correlation function in different instanton ensembles will be given in Sec. VI.C.4.

## 2. Tunneling amplitude in the presence of light fermions

The previous subsection was a small detour from the calculation of the tunneling amplitude. Now we return to our original problem and study the effect of light quarks on the tunneling rate. Quarks modify the weight in the Euclidean partition function by the fermionic determinant

$$\prod_f \det[\mathcal{D}(A_\mu) + m_f], \quad (107)$$

which depends on the gauge fields through the covariant derivative. Using the fact that nonzero eigenvalues come in pairs, we can write the determinant as

$$\det[\mathcal{D} + m] = m^\nu \prod_{\lambda > 0} (\lambda^2 + m^2), \quad (108)$$

where  $\nu$  is the number of zero modes. Note that the integration over fermions gives a determinant in the numerator. This means that (as  $m \rightarrow 0$ ) the fermion zero mode makes the tunneling amplitude vanish and individual instantons cannot exist.

We have already seen what the reason for this phenomenon is: During the tunneling event, the axial charge of the vacuum changes by two units, so instantons have to be accompanied by fermions. Indeed, it was pointed out by 't Hooft that the tunneling amplitude is nonzero in the presence of external quark sources, because zero modes in the denominator of the quark propagator may cancel against zero modes in the determinant. This is completely analogous to the situation in

the quantum-mechanical toy model of Sec. II.B. Nevertheless, there are important differences. In particular, we shall see that, in QCD, chiral symmetry breaking implies that isolated instantons can have a nonzero amplitude (Sec. IV.D).

Consider the fermion propagator in the instanton field

$$S(x, y) = \frac{\psi_0(x)\psi_0^+(y)}{im} + \sum_{\lambda \neq 0} \frac{\psi_\lambda(x)\psi_\lambda^+(y)}{\lambda + im}, \quad (109)$$

where we have written the zero-mode contribution separately. Suppose there are  $N_f$  light-quark flavors, so that the instanton amplitude is proportional to  $m^{N_f}$  (or, more generally, to  $\prod_f m_f$ ). Instead of the tunneling amplitude, let us calculate a  $2N_f$ -quark Green's function  $\langle \prod_f \bar{\psi}_f(x_f) \Gamma \psi_f(y_f) \rangle$ , containing one quark and antiquark of each flavor. Contracting all the quark fields, we obtain the Green's function by multiplying the tunneling amplitude by  $N_f$  fermion propagators. Every propagator has a zero-mode contribution with one power of the fermion mass in the denominator. As a result, the zero-mode contribution to the Green's function is finite in the chiral limit.<sup>22</sup>

The result can be written in terms of an effective Lagrangian ('t Hooft, 1976b; see Sec. IV.F, where we give a more detailed derivation). It is a nonlocal  $2N_f$ -fermion interaction, where the quarks are emitted or absorbed in zero-mode wave functions. In general, it has a fairly complicated structure, but under certain assumptions, it can be significantly simplified. First, if we limit ourselves to low momenta, the interaction is effectively local. Second, if instantons are uncorrelated, we can average over their orientation in color space. For SU(3) color and  $N_f=1$ , the result is (Shifman, Vainshtein, and Zakharov, 1980c)

$$\mathcal{L}_{N_f=1} = \int d\rho n_0(\rho) \left( m\rho - \frac{4}{3} \pi^2 \rho^3 \bar{q}_R q_L \right), \quad (110)$$

where  $n_0(\rho)$  is the tunneling rate without fermions. Note that the zero-mode contribution acts like a mass term. This is quite natural because, for  $N_f=1$ , there is only one chiral  $U(1)_A$  symmetry, which is anomalous. In contrast to the case  $N_f > 0$ , the anomaly can therefore generate a fermion mass term.

For  $N_f=2$ , the result is

$$\begin{aligned} \mathcal{L}_{N_f=2} = \int d\rho n_0(\rho) & \left[ \prod_f \left( m\rho - \frac{4}{3} \pi^2 \rho^3 \bar{q}_{f,R} q_{f,L} \right) \right. \\ & + \frac{3}{32} \left( \frac{4}{3} \pi^2 \rho^3 \right)^2 \left( \bar{u}_R \lambda^a u_L \bar{d}_R \lambda^a d_L \right. \\ & \left. \left. - \frac{3}{4} \bar{u}_R \sigma_{\mu\nu} \lambda^a u_L \bar{d}_R \sigma_{\mu\nu} \lambda^a d_L \right) \right], \quad (111) \end{aligned}$$

<sup>20</sup>It is usually argued that the Witten-Veneziano relation is derived in the large- $N_c$  approximation to QCD and that  $\chi_{\text{top}} = O(1)$  in this limit. That does not really solve the problem, however. In order to obtain a finite topological susceptibility, one has to set  $N_f=0$ , even if  $N_c \rightarrow \infty$ .

<sup>21</sup>This means that *a priori* it is not even defined how the two numbers should be compared.

<sup>22</sup>Note that Green's functions involving more than  $2N_f$  legs are not singular as  $m \rightarrow 0$ . The Pauli principle always ensures that no more than  $2N_f$  quarks can propagate in zero-mode states.

where the  $\lambda^a$  are color Gell-Mann matrices. One can easily check that the interaction is  $SU(2) \times SU(2)$  invariant, but  $U(1)_A$  is broken. This means that the 't Hooft Lagrangian provides another derivation of the  $U(1)_A$  anomaly. Furthermore, in Secs. III and IV we shall argue that the importance of this interaction goes far beyond the anomaly and that it explains the physics of chiral symmetry breaking and the spectra of light hadrons.

Finally, we need to include the effects of nonzero modes on the tunneling probability. One effect is that the coefficient in the beta function is changed to  $b = 11N_c/3 - 2N_f/3$ . In addition to that, there is an overall constant that was calculated by 't Hooft (1976b) and Carlitz and Creamer (1979a),

$$n(\rho) \sim 1.34(m\rho)^{N_f} [1 + N_f(m\rho)^2 \log(m\rho) + \dots], \quad (112)$$

where we have also specified the next-order correction in the quark mass. Note that at two-loop order we not only get the two-loop beta function in the running coupling but also one-loop anomalous dimensions in the quark masses.

We should emphasize that, in this section, we have studied the effect of fermions on the tunneling rate only for widely separated, individual instantons. But light fermions induce strong correlations between instantons, and the problem becomes very complicated. Many statements that are correct in the case of pure gauge theory, e.g., the fact that tunneling lowers the ground-state energy, are no longer obvious in the theory with quarks. But before we try to tackle these problems, we should like to review what is known phenomenologically about instantons in QCD.

### III. PHENOMENOLOGY OF INSTANTONS

#### A. How often does tunneling occur in the QCD vacuum?

In order to assess the importance of instantons in the QCD vacuum, we have to determine the total tunneling rate in QCD. Unfortunately, the standard semiclassical theory discussed in the last section is not able to answer this question. The problem is related to the presence of large-size instantons for which the action is of order one. The naive use of the one-loop running coupling leads to an infrared divergence in the semiclassical rate, which is a simple consequence of the Landau pole. Before we discuss any attempts to improve on the theoretical estimate, we should like to study phenomenological estimates of the instanton density in QCD.

The first attempt along these lines (Shifman, Vainshtein, and Zakharov, 1978) was based on information on properties of the QCD vacuum obtained from QCD sum rules. We cannot go into details of the sum-rule method, which is based on using dispersion theory to match experimental information with the operator product expansion (OPE) prediction for hadronic correlation functions. See the reviews of Reinders, Rubinstein, and Yazaki (1985), Narison (1989), Shifman (1992). The essential point is that the method provides

an estimate for the gluon condensate,<sup>23</sup>  $\langle 0 | (G_{\mu\nu}^a)^2 | 0 \rangle$ . From an analysis of the charmonium spectrum, Shifman *et al.* (1979) obtained

$$\langle 0 | (G_{\mu\nu}^a)^2 | 0 \rangle \approx 0.5 \text{ GeV}^4. \quad (113)$$

It is difficult to assess the accuracy of this number. The analysis has been repeated many times, including many more channels. Reinders, Rubinstein, and Yazaki (1985) agree with the value of Eq. (113) and quote an error of 25%. On the other hand, a recent analysis gives  $\langle 0 | (G_{\mu\nu}^a)^2 | 0 \rangle = (1.02 \pm 0.1) \text{ GeV}^4$ , about twice the original Shifman, Vainshtein, and Zakharov value (Narison, 1996).

The tunneling rate can now be estimated using the following simple idea. If the nonperturbative fields contributing to the gluon condensate are dominated by (weakly interacting) instantons, the condensate is simply proportional to the instantons density, because every single instanton contributes a finite amount  $\int d^4x (G_{\mu\nu}^a)^2 = 32\pi^2$ . Therefore the value of the gluon condensate provides an upper limit for the instanton density,

$$n = \frac{dN_{I+A}}{d^4x} \leq \frac{1}{32\pi^2} \langle (G_{\mu\nu}^a)^2 \rangle \approx 1 \text{ fm}^{-4}. \quad (114)$$

Another estimate of the instanton density can be obtained from the topological susceptibility. This quantity measures fluctuations of the topological charge in a four-volume  $V$ ,

$$\chi_{\text{top}} = \lim_{V \rightarrow \infty} \frac{\langle Q^2 \rangle}{V}. \quad (115)$$

On average, the topological charge vanishes,  $\langle Q \rangle = 0$ , but generally in a given configuration  $Q \neq 0$  (see Fig. 7). Topological fluctuations provide an important characteristic of the vacuum in pure gauge QCD. However, in the presence of massless quarks, the topological charge is screened, and  $\chi_{\text{top}} = 0$  (see Sec. V.E). The value of  $\chi_{\text{top}}$  in quenched QCD can be estimated using the Witten-Veneziano relation (106) to give

$$\chi_{\text{top}} = \frac{f_\pi^2}{2N_f} (m_\eta^2 + m_{\eta'}^2 - 2m_K^2) = (180 \text{ MeV})^4. \quad (116)$$

If one assumes that instantons and anti-instantons are uncorrelated, the topological susceptibility can be estimated as follows. The topological charge in some volume  $V$  is  $Q = N_I - N_A$ . For a system with Poissonian statistics, the fluctuations in the particle numbers are  $\Delta N_{I,A} \approx \sqrt{N_{I,A}}$ . This means that for a random system of

<sup>23</sup>Again, we use conventions appropriate for dealing with classical fields. In standard perturbative notations, the fields are rescaled by a factor of  $g$ , and the condensate is given by  $\langle 0 | (gG_{\mu\nu}^a)^2 | 0 \rangle$ .

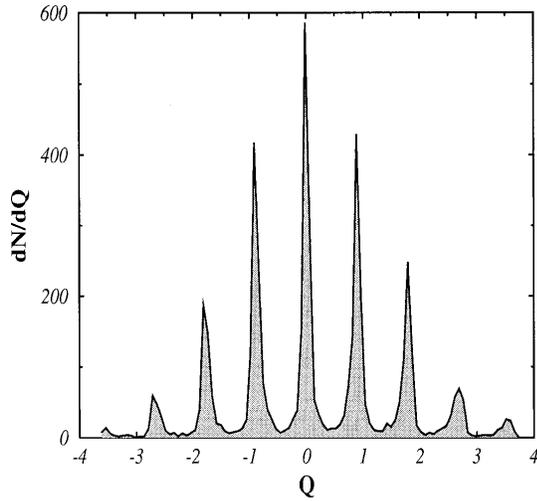


FIG. 7. Distribution of topological charges for an ensemble of 5000 thermalized configurations in pure gauge SU(3) at  $\beta=6.1$ , from Alles *et al.*, 1997.

instantons, we expect  $\chi_{\text{top}}=N/V$ . Using the phenomenological estimate from above, we again get  $(N/V) \approx 1 \text{ fm}^{-4}$ .

How reliable are these estimates? Both methods suffer from uncertainties that are hard to assess. For the gluon condensate, there is no systematic method of separating perturbative and nonperturbative contributions. The sum-rule method effectively determines contributions to the gluon condensate with momenta below a certain separation scale. This corresponds to instanton fluctuations above a given size, which in itself is not a problem, since the rate of small instantons can be determined perturbatively, but the value of the separation scale ( $\mu \sim 1 \text{ GeV}$ ) is not very well determined. In any case, the connection of the gluon condensate with instantons is indirect. Other fluctuations might very well play a role. The estimate of the instanton density from the (quenched) topological susceptibility relies on the assumption that instantons are distributed randomly. In the presence of light quarks, this assumption is certainly incorrect (that is why an extrapolation from quenched to real QCD is necessary).

### B. The typical instanton size and the instanton liquid mode

Next to the tunneling rate, the typical instanton size is the most important parameter characterizing the instanton ensemble. If instantons are too large, it does not make any sense to speak of individual tunneling events, and semiclassical theory is inapplicable. If instantons are too small, then semiclassical theory is good, but the tunneling rate is strongly suppressed. The first estimate of the typical instanton size was made by Shifman *et al.* (1978), based on the estimate of the tunneling rate given above.

If the total tunneling rate can be calculated from the semiclassical 't Hooft formula, we can ask up to what

critical size we have to integrate the rate in order to get the phenomenological instanton density<sup>24</sup>

$$\int_0^{\rho_{\text{max}}} d\rho n_0(\rho) = n_{\text{phen}}. \quad (117)$$

Using  $n_{\text{phen}}=1 \text{ fm}^{-4}$ , Shifman *et al.* concluded that  $\rho_{\text{max}} \approx 1 \text{ fm}$ . This is a very pessimistic result because it implies that the instanton action is not large, so the semiclassical approximation is useless. Also, if instantons are that large, they overlap strongly, and it makes no sense to speak of individual instantons.

There are two possible ways in which the semiclassical approximation can break down as the typical size becomes large. One possibility is that (higher-loop) perturbative fluctuations start to grow. We have discussed these effects in the double-well potential (see Sec. II.A.4), but in gauge theory, the two-loop [ $O(1/S_0)$ ] corrections to the semiclassical result are not yet known. Another possibility is that nonperturbative (multi-instanton etc.) effects become important. These effects can be estimated from the gluon condensate. The interaction of an instanton with an arbitrary, weak external field  $G_{\mu\nu}^a \text{ ext}$  is given by (see Sec. IV.A.1)

$$S_{\text{int}} = \frac{2\pi^2 \rho^2}{g^2} \bar{\eta}_{\mu\nu}^a U^{ab} G_{\mu\nu}^b \text{ ext}, \quad (118)$$

where  $U$  is the matrix that describes the instanton orientation in color space. This is a dipole interaction, so to first order it does not contribute to the average action. To second order in  $G_{\mu\nu}^a \text{ ext}$ , one has (Shifman *et al.*, 1980b)

$$n(\rho) = n_0(\rho) \left[ 1 + \frac{\pi^4 \rho^4}{2g^4} \langle (G_{\mu\nu}^a)^2 \rangle + \dots \right]. \quad (119)$$

From our knowledge of the gluon condensate, we can now estimate for what size nonperturbative effects become important. Using the Shifman, Vainshtein, and Zakharov value, we see that for  $\rho > 0.2 \text{ fm}$  the interaction with vacuum fields (of whatever origin) is not negligible. Unfortunately, for  $\rho < 0.2 \text{ fm}$  the total density of instantons is too small as compared to the phenomenological estimate.

However, it is important to note the sign of the correction. The nonperturbative contribution leads to a tunneling rate that grows even faster than the semiclassical rate. Accounting for higher-order effects by exponentiating the second-order contribution, Shuryak (1982a) suggested that the critical size be estimated from the modified condition

$$\int_0^{\rho_{\text{max}}} d\rho n_0(\rho) \exp \left[ \frac{\pi^4 \rho^4}{2g^4} \langle (G_{\mu\nu}^a)^2 \rangle \right] = n_{\text{phen}}. \quad (120)$$

<sup>24</sup>This procedure cannot be entirely consistent, since simply cutting off the size integration violates many exact relations such as the trace anomaly (see Sec. IV.C). Nevertheless, given all the other uncertainties, this method provides a reasonable first estimate.

Since the rate grows faster, the critical size is shifted to a smaller value,

$$\rho_{\max} \sim 1/3 \text{ fm.} \quad (121)$$

If the typical instanton is indeed small, we obtain a completely different perspective on the QCD vacuum:

- (1) Since the instanton size is significantly smaller than the typical separation  $R$  between instantons,  $\rho/R \sim 1/3$ , the vacuum is fairly dilute. The fraction of spacetime occupied by strong fields is only a few percent.
- (2) The fields inside the instanton are very strong,  $G_{\mu\nu} \gg \Lambda^2$ . This means that the semiclassical approximation is valid, and the typical action is large:

$$S_0 = 8\pi^2/g^2(\rho) \sim 10 - 15 \gg 1. \quad (122)$$

Higher-order corrections are proportional to  $1/S_0$  and presumably small.

- (3) Instantons retain their individuality and are not destroyed by interactions. From the dipole formula, one can estimate

$$|\delta S_{\text{int}}| \sim (2-3) \ll S_0. \quad (123)$$

- (1) Nevertheless, interactions are important for the structure of the instanton ensemble, since

$$\exp|\delta S_{\text{int}}| \sim 20 \gg 1. \quad (124)$$

This implies that interactions have a significant effect on correlations among instantons; the instanton ensemble in QCD is not a dilute gas, but an interacting liquid.

Improved estimates of the instanton size can be obtained from phenomenological applications of instantons. The average instanton size determines the structure of chiral symmetry breaking, in particular the values of the quark condensate, the pion mass, its decay constant, and its form factor. We shall discuss these observables in greater detail in the next sections.

In particular, the consequences of the vacuum structure advocated here were studied in the context of the “random-instanton liquid model.” The idea is to fix  $N/V = 1 \text{ fm}^{-4}$  and  $\rho = 1/3 \text{ fm}$  and add the assumption that the distribution of instanton positions, as well as color orientations, is completely random. This is not necessarily in contradiction with the observation (124) that interactions are important, as long as they do not induce strong correlations among instantons. The random model is sufficiently simple that one can study a large number of hadronic observables. The agreement with experimental results is quite impressive, thus providing support for the underlying parameters.

### C. Instantons on the lattice

#### 1. The topological charge and susceptibility

The most direct way to determine the parameters of the instanton liquid is provided by numerical simulations

on the lattice. Before we come to direct instanton searches, we should like to discuss the determination of the topological susceptibility, which requires measurements of the total topological charge inside a given volume. This has become a very technical subject, and we will not be able to go into much detail [see the nice, albeit somewhat dated, review by Kronfeld (1988)]. The standard techniques used to evaluate the topological charge are the (i) (naive) field-theoretical, (ii) geometric, and (iii) fermionic methods. Today, these methods are usually used in conjunction with various improvements, like cooling, blocking, or improved actions.

The field-theoretic method is based on the naive lattice discretization of the topological charge density  $G\tilde{G} \sim \epsilon_{\mu\nu\rho\sigma} \text{tr}[U_{\mu\nu}U_{\rho\sigma}]$ , where  $U_{\mu\nu}$  is the elementary plaquette in the  $\mu\nu$  plane. The method is simple to implement, but it has no topological meaning, and the naive topological charge  $Q$  is not even an integer. In addition to that, the topological susceptibility suffers from large renormalization effects and mixes with other operators, in particular, the unit operator and the gluon condensate (Campostrini, DiGiacomo, and Panagopoulos, 1988).

There are a number of “geometric” algorithms that ensure that  $Q$  has topological significance (Luescher, 1982; Voit, 1983; Phillips and Stone, 1986). This means that  $Q$  is always an integer and that the topological charge can be expressed as a surface integral. All these methods are based on fixing the gauge and using some interpolation procedure to reconstruct a smooth gauge potential from the discrete lattice data. For a finite lattice with the topology of a four-dimensional torus, the topology of the gauge fields resides in the transition functions that connect the gauge potential on the boundaries. The geometric method provides a well-defined topological charge for almost all gauge configurations. In the continuum, different topological sectors are separated by configurations with infinite action. On the lattice, however, different sectors are separated by exceptional finite-action configurations called dislocations. Although expected to be unimportant for sufficiently smooth fields, dislocations may spoil the continuum limit on the lattice (Pugh and Teper, 1989).

Fermionic methods for calculating the topological charge rely on the connection between instantons and fermionic zero modes. Exceptionally small eigenvalues of the Dirac operator on the lattice have been identified (Smit and Vink, 1987, 1988; Laursen, Smit, and Vink, 1990). Furthermore, it was demonstrated that the corresponding Dirac eigenvectors have the correct chirality and are spatially correlated with instantons. Fermionic methods are not sensitive to dislocations, but they suffer from problems connected with the difficulty of defining chiral fermions on the lattice. In particular, the (almost) zero modes connected with instantons are not exactly chiral, and the topological charge defined through a fermionic expectation value does suffer from renormalization (for both Wilson and staggered fermions). For this reason, fermionic methods have never been pursued very vigorously [see Vink (1988) for a rare exception].

Recently, some progress in constructing chiral fermions on the lattice has been made. See, for example, Kaplan (1992) and Narayanan and Neuberger (1995). These methods may provide improved measurements of the topological susceptibility (Narayanan and Vranas, 1997).

Since most of the difficulties with the field-theoretical and geometrical algorithms are related to fluctuations on very short scales, it is natural to supplement these algorithms with some sort of smoothing procedure. The best known method of this type is *cooling* (Hoek, 1986; Hoek, Teper, and Waterhouse, 1987). In the cooling method, one locally minimizes the action. This operation quickly eliminates short-range quantum fluctuations and eventually leads to a smooth configuration, corresponding to the classical content of the original field. It has been verified that these configurations are indeed dominated by instantons<sup>25</sup> (Hoek *et al.*, 1987; Chu *et al.*, 1994). It has also been checked that in the cooled configurations the field-theoretic definition of the topological charge agrees with the more sophisticated, geometrical methods (Alles, DiGiacomo, and Gianetti, 1990; Wiese, 1990). Unfortunately, the cooling method also suffers from systematic uncertainties. If the simplest Wilson action is used, instantons gradually shrink and finally fall through the lattice. Improved lattice actions can make instantons stable (de Forcrand, Perez, and Statamescu, 1995), but instanton anti-instanton pairs still annihilate during the cooling process.

The study of topological objects on the lattice is part of a larger effort to find improved or even “perfect” lattice actions and operators.<sup>26</sup> An example of such a method is the “inverse blocking” procedure considered by Hasenfratz, DeGrand, and Zhu (1996). From the field configuration on the original coarse lattice, one constructs a smoother configuration on a finer lattice by an approximate inverse renormalization-group transformation. The method has the advantage that it gives a larger action for dislocations than the standard Wilson action does, thus suppressing undesirable contributions to the topological charge. First tests of improved topological charges are very promising, correctly recovering instantons with sizes as small as  $0.8a$ . Significantly improved values for the topological susceptibility will hopefully be forthcoming.

Whatever method is used to define  $Q$  and measure the topological susceptibility on the lattice, one has to test the behavior of  $\chi_{\text{top}}$  as the lattice spacing is taken to zero,  $a \rightarrow 0$ . If the topological susceptibility is a physical quantity, it has to exhibit scaling towards the continuum limit. In the geometrical method,  $Q$  itself is not renormalized, and  $\chi_{\text{top}}$  is expected to show logarithmic scaling (if it were not for the contribution of dislocations). If the

<sup>25</sup>This “experimental” result also shows that all stable classical solutions of the Yang-Mills equations are of multi-instanton type.

<sup>26</sup>Classically improved (perfect) actions have no discretization errors of order  $a^n$  for some (all)  $n$ , where  $a$  is the lattice spacing.

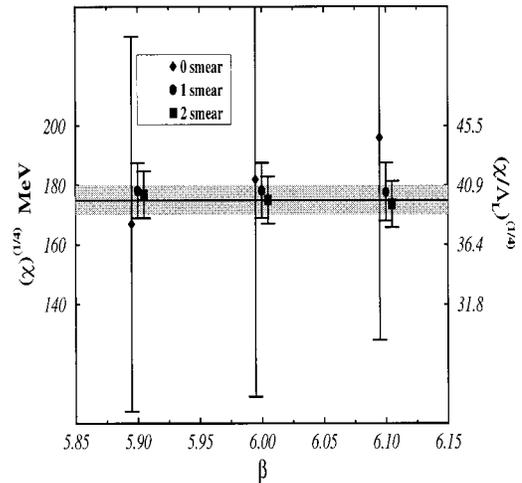


FIG. 8. Scaling test of the topological susceptibility in pure gauge SU(3), from Alles *et al.*, 1997. The improvement from  $Q_L^{(0)}$  to  $Q_L^{(2)}$  is clearly visible. Inside errors, the results are independent of the bare charge  $\beta = 6/g^2$ .

field-theoretic method is used,  $\chi_{\text{top}}$  mixes with the unit operator and the gluon condensate. Early studies of the scaling behavior of  $\chi_{\text{top}}$  can be found in Woit (1983), Ishikawa *et al.* (1983), and Fox *et al.* (1985). A more detailed study of scaling and the role of mixing effects [in pure gauge SU(2)] was recently carried out by Alles, Camprostrini, and DiGiacomo (1993). These authors use “heating” to investigate the effect of fluctuations on classical instanton configurations. As quantum fluctuations are turned on, one can study the renormalization of the topological charge. For comparison, if a configuration with no topology is heated, it is mainly sensitive to mixing. Alles *et al.* conclude that in their window,  $\beta = (2/g^2) = 2.45 - 2.525$ , everything scales correctly and  $\chi_{\text{top}} = 3.5(4) \times 10^5 \Lambda_L^4 \approx (195 \text{ MeV})^4$ , where we have used  $\Lambda_{LAT} = 8 \text{ MeV}$ . For comparison, the result from cooling is about 30% larger. This discrepancy gives a rough estimate of the uncertainty in the calculation.

An example of the use of an improved topological charge operator is shown in Figs. 7 and 8. Figure 7 shows the distribution of topological charges in pure gauge SU(3). As expected, the distribution is a Gaussian with zero mean. A scaling test of the topological susceptibility is shown in Fig. 8. The result clearly shows the reduction in the statistical error achieved using the improved operator and the quality of the scaling behavior. The topological susceptibility is  $\chi_{\text{top}} = [175(5) \text{ MeV}]^4$ . On the other hand, the geometric method (and preliminary results from inverse blocking) gives larger values, for example  $\chi_{\text{top}} = (260 \text{ MeV})^4$  in pure-gauge SU(3) simulations (Grandy and Gupta, 1994). We conclude that lattice determinations are consistent with the phenomenological value of  $\chi_{\text{top}}$ , but that the uncertainty is still rather large. Also, the scaling behavior of the topological susceptibility extracted from improved (perfect) operators or fermionic definitions still needs to be established in greater detail.

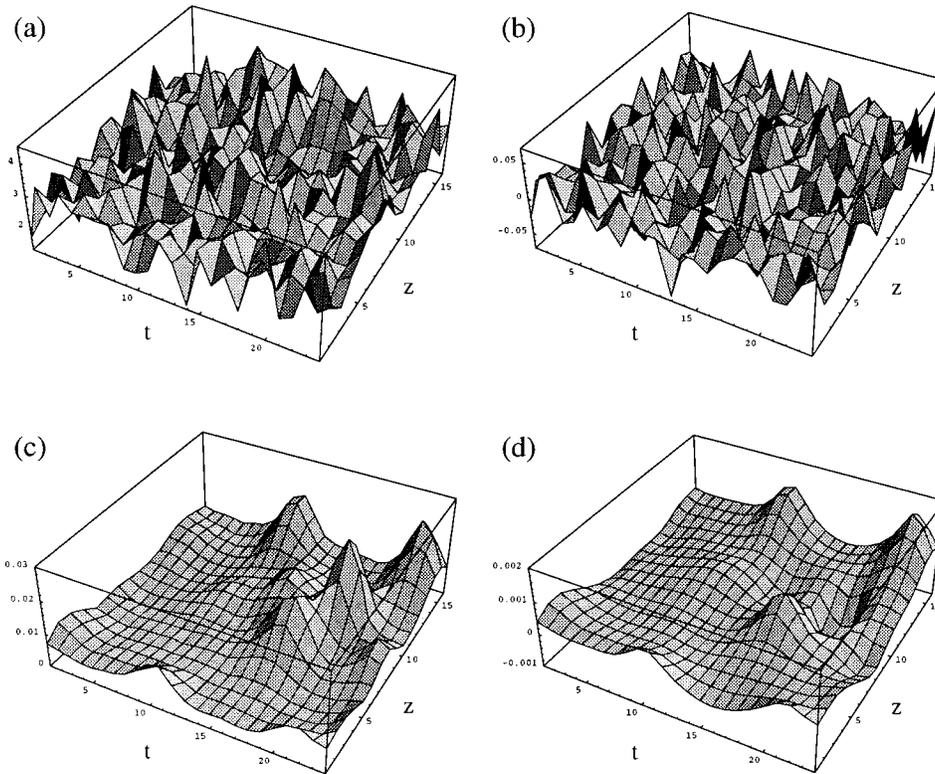


FIG. 9. A typical slice through a lattice configuration: (a), (b), before cooling; (c), (d), after 25 cooling sweeps, from Chu *et al.*, 1994. Figs. (a) and (c) show the fields strength  $G^2$ , while Figs. (b) and (d) show the topological charge density  $G\tilde{G}$ . The lattice units are  $a=0.16$  fm before and  $a=0.14$  fm after 25 cooling sweeps.

A by-product of the measurements of Alles *et al.* is a new determination of the gluon condensate in pure gauge SU(2),

$$\langle G^2 \rangle = (4\pi^2) \times 0.38(6) \times 10^8 \Lambda_L^4 \approx 1.5 \text{ GeV}^4. \quad (125)$$

This number is significantly larger than the Shifman, Vainshtein, and Zakharov value given in the last section, but it is consistent with other lattice data, for example those of Campostrini, DiGiacomo, and Gunduc (1989). However, when comparing the lattice data to the Shifman-Vainshtein-Zakharov estimate one should keep in mind that the two results are based on very different physical observables. While the Shifman-Vainshtein-Zakharov value is based on the OPE of a hadronic correlator, and the value of the separation scale is fairly well determined, lattice results are typically based on the value of the average plaquette, and the separation scale is not very well defined.

## 2. The instanton liquid on the lattice

The topological susceptibility provides only a global characterization of the instanton ensemble. In addition, we should like to identify individual instantons and study their properties. This is true in particular for theories with light quarks, where the total charge is suppressed because of screening effects (see Sec. V.E).

Most studies of instantons on the lattice are based on the cooling method. As already mentioned, cooling is not an ideal method for this purpose because instantons and anti-instantons annihilate during cooling. While this process does not affect the total charge, it does affect the density of instantons and anti-instantons. Ultimately, improved operators are certainly the method of choice.

Nevertheless, cooling has the advantage of providing smooth gauge configurations that are easily interpreted in terms of continuum fields.

A typical field configuration after the quantum noise has disappeared is shown in Fig. 9 (Chu *et al.*, 1994). The left panel shows the field strength on a slice through the lattice. One can clearly identify individual classical objects. The right panel shows the distribution of the topological charge, demonstrating that the classical configurations are indeed instantons and anti-instantons. Fixing physical units from the (quenched) rho-meson mass, Chu *et al.* conclude that the instanton density in pure gauge QCD is  $(1.3-1.6) \text{ fm}^{-4}$ . This number is indeed close to the estimate presented above.

The next question concerns the average size of an instanton. A qualitative confirmation of the instanton liquid value  $\rho \sim 1/3$  fm was first obtained by Polikarpov and Veselov (1988). More quantitative measurements (Chu *et al.*, 1994) are based on fitting of the topological charge correlation function after cooling. The fit is very good and gives an average size  $\rho = 0.35$  fm.

More detail is provided by measurements of the instanton size distribution. Lattice studies of this type were recently performed for pure gauge SU(2) by Michael and Spencer (1995) and de Forcrand, Perez, and Stamatescu (1997). The results obtained by Michael and Spencer on two different size lattices are shown in Fig. 10(a). The agreement between the two measurements is best for large instantons, while it is not so good for small  $\rho$ . That is of course exactly as expected; instantons of size  $\rho \sim a$  fall through the lattice during cooling. The most important result is the existence of a relatively sharp maximum in the size distribution together with a

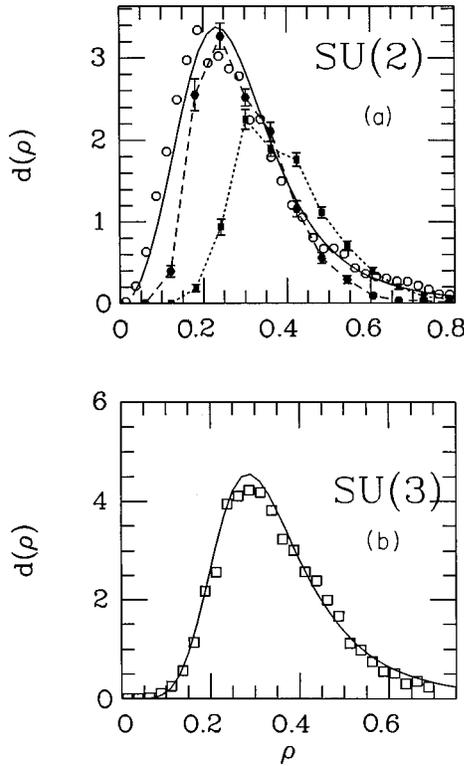


FIG. 10. Instanton size distribution in pure gauge SU(2), from Michael and Spencer (1995) and Shuryak (1995). The size  $\rho$  is given in fm, where lattice units have been fixed from the glueball mass  $m_{0^{++}}=1.7$  GeV:  $\blacksquare$ ,  $16^4$ ,  $4/g^2=2.4$ ;  $\bullet$ ,  $24^4$ ,  $4/g^2=2.5$ . The dotted and dashed lines simply serve to guide the eye. The open circles and squares come from an interacting instanton calculation, while the solid curve corresponds to the parametrization discussed in the text.

strong suppression of large-size instantons. The physical mechanism for this effect is still not adequately understood. In Fig. 10 we compare the lattice data with an instanton liquid calculation, in which large instantons are suppressed by a repulsive core in the instanton interaction (Shuryak, 1995).

In addition to the size distribution, Michael and Spencer (1995) also studied correlations among instantons. They found that, on average, unlike pairs were closer than like pairs,  $\langle R(\text{IA}) \rangle \approx 0.7 \langle R(\text{II}) \rangle$ . This clearly suggests an attractive interaction between instantons and anti-instantons. The distribution of pseudoparticle separations is shown in Fig. 11. There are very few IA-pairs with very small separation, because these pairs easily annihilate during cooling. Like pairs show an enhancement at small  $R$ , a feature that is not understood at present.

After identifying instantons on the lattice, the next step is to study the importance of instantons for physical observables. We shall discuss an example of this line of research in Sec. VI.E, where we present hadronic correlation functions in cooled configurations. Another example can be found in Thurner, Feurstein, and Markum (1997), who show that there is a strong correlation between the quark condensate and the location of instantons after cooling.

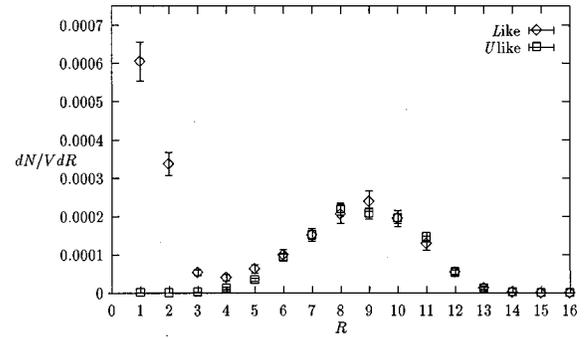


FIG. 11. The distribution of the separation of like and unlike instanton pairs in pure gauge SU(2), from Michael and Spencer, 1995. The distance to the closest neighbor (after cooling) is given in lattice units  $a=0.08$  fm.

### 3. The fate of large-size instantons and the beta function

We have seen that both phenomenology and the available lattice data suggest that instantons larger than  $\rho \approx 1/3$  fm are strongly suppressed in QCD. In Sec. II.C.4 we saw that this result could not be understood from the leading-order semiclassical formula. This leaves essentially three possibilities: The instanton distribution is regulated by higher-order quantum effects, by classical instanton interactions, or by the interaction of instantons with other classical objects (e.g., monopoles or strings).

The possible role of repulsive interactions between instantons will be discussed in Sec. IV (this is also what the open dots in Fig. 10 are based on). It is hard to speculate on the role of other classical fields, although we shall try to summarize some work in this direction in the next section. Here, we should like to discuss the possibility that the size distribution is regulated by quantum fluctuations (Shuryak, 1995). If this is the case, gross features of the size distribution can be studied by considering a single instanton rather than an interacting ensemble.

The Gell-Mann–Low beta function is defined as a derivative of the coupling constant  $g$  over the logarithm of the normalization scale  $\mu$  at which  $g$  is determined,

$$\beta(g) = \frac{\partial g}{\partial \log \mu} = -b \frac{g^3}{16\pi^2} - b' \frac{g^5}{(16\pi^2)^2} + \dots \quad (126)$$

In QCD with  $N_c$  colors and  $N_f$  light flavors, we have  $b = 11N_c/3 - 2N_f/3$  and  $b' = 34N_c^2/3 - 13N_c N_f/3 + N_f/N_c$ . Remember that the tunneling amplitude is  $n(\rho) \sim \rho^{-5} \exp[-(8\pi^2)/g^2(\rho)]$ . In the weak-coupling domain, one can use the one-loop running coupling and  $n(\rho) \sim \rho^{b-5} \Lambda^b$ . This means that the strong growth of the size distribution in QCD is related to the large value of  $b \approx 9$ .

For pedagogical reasons, we should like to start our discussion in the domain of the phase diagram where  $b$  is small and this dangerous phenomenon is absent. For  $N_c=3$  colors,  $b$  is zero if  $N_f=33/2 \approx 16$ . When  $b$  is small and positive, it turns out that the next coefficient  $b'$  is negative (Belavin and Migdal, 1974; Banks and Zaks, 1982), and therefore the beta function has a zero at

$$\frac{g_*^2}{16\pi^2} = -\frac{b}{b'}. \quad (127)$$

Since  $b$  is small, so is  $g_*$ , and the perturbative calculation is reliable. If the beta function has a zero, the coupling constant will first grow as one goes to larger distances but then stop running (“freeze”) as the critical value  $g_*$  is approached. In this case, the large-distance behavior of all correlation functions is governed by the infrared fixed point, and the correlators decay as fractional powers of the distance. In this domain the instanton contribution to the partition function is of the order  $\exp(-16\pi^2/g_*^2) \sim \exp(-|b'/b|)$ , so it is exponentially small if  $b$  is small.<sup>27</sup>

What happens if  $N_f$  is reduced, so that we move away from the  $b=0$  line? The infrared fixed point survives in some domain (the conformal region), but the value of the critical coupling grows. Eventually, non-perturbative effects (instantons in particular) grow exponentially, and the perturbative result (127) is unreliable. As we shall discuss in Sec. IX.D, existing lattice data suggest that the boundary of the conformal domain is around  $N_f=7$  for  $N_c=3$  (Iwasaki *et al.*, 1996).

There is no unique way to define the beta function in the nonperturbative regime. In our context, a preferred definition is provided by the value of instanton action  $S_{\text{inst}}=8\pi^2/g^2(\rho)$  as a function of the instanton size. This quantity can be studied directly on the lattice by heating (i.e., adding quantum fluctuations to) a smooth instanton of given size. This program is not yet implemented, but one can get some input from the measured instanton size distribution. If  $S_{\text{inst}} \gg 1$ , the semiclassical expression  $n(\rho) \sim \rho^{-5} \exp(-S_{\text{inst}})$  should be valid, and we can extract the effective charge from the measured size distribution.

We have already discussed lattice measurements of  $n(\rho)$  in Sec. III.B. These results are well reproduced using the semiclassical size distribution with a modified running coupling constant (Shuryak, 1995),

$$\frac{8\pi^2}{g^2(\rho)} = bL + \frac{b'}{b} \log L, \quad (128)$$

where (for  $N_f=0$ ) the coefficients are  $b=11N_c/3$ ,  $b'=17N_c^2/3$  as usual, but the logarithm is regularized according to

$$L = \frac{1}{p} \log \left[ \left( \frac{1}{\rho\Lambda} \right)^p + C^p \right]. \quad (129)$$

For small  $\rho$ , Eq. (128) reduces to the perturbative running coupling, but for large  $\rho$  the coupling stops running in a manner controlled by the two parameters  $C$  and  $p$ . A good description of the measured size distribution can

<sup>27</sup>Since  $N_f$  is large, the vacuum consists of a dilute gas of instanton/anti-instanton molecules, so the action is twice that for a single instanton.

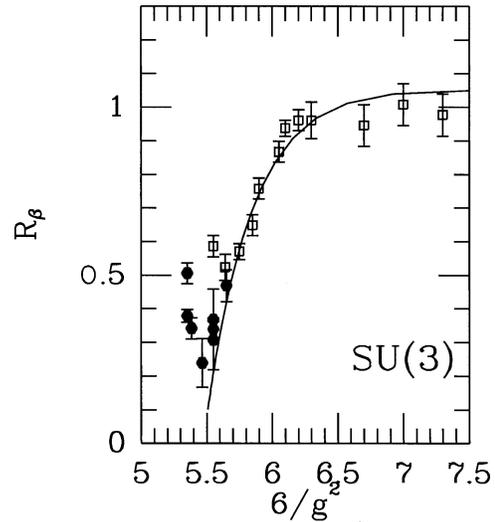


FIG. 12. Nonperturbative beta function from SU(3) lattice gauge theory with Wilson action:  $\square$ , Gupta, 1992;  $\bullet$ , Blum *et al.*, 1995. The solid line shows the fit discussed in the text.

be obtained with<sup>28</sup>  $\Lambda=0.66 \text{ fm}^{-1}$ ,  $p=3.5$ , and  $C=4.8$ , shown by the solid line in Fig. 10. In real QCD, the coupling cannot freeze because the theory certainly has no infrared fixed point. Nevertheless, in order to make the instanton density convergent, one does not need the beta function to vanish. It is sufficient that the coupling constant runs more slowly, with an effective  $b < 5$ .

There are some indications from lattice simulations that this is indeed the case, and that there is a consistent trend from  $N_f=0$  to  $N_f=16$ . These results are based on a definition of the nonperturbative beta function based on pre-asymptotic scaling of hadronic observables. Ideally, the lattice scale is determined by performing simulations at different couplings  $g$ , fixing the scale  $a$  from the asymptotic (perturbative) relation  $g(a)$ . Scaling behavior is established by studying many different hadronic observables. Although asymptotic scaling is often violated, a weaker form of scaling might still be present. In this case, the lattice scale  $a$  at a given coupling  $g$  is determined by measuring a hadronic observable in units of  $a$  and then fixing  $a$  to give the correct experimental value. This procedure makes sense as long as the result is universal (independent of the observable). Performing simulations at different  $g$ , one can determine the function  $g(a)$  and the beta function.

Lattice results for both pure gauge (Gupta, 1992; open points) and  $N_f=2$  (Blum *et al.*, 1995) SU(3) are shown in Fig. 12. In order to compare different theories, it is convenient to normalize the beta function to its asymptotic ( $g \rightarrow 0$ ) value. The ratio

<sup>28</sup>The agreement is even more spectacular in the  $O(3)$  model. In this case the instanton size distribution is measured over a wider range in  $\rho$  and shows a very nice  $n(\rho) \sim 1/\rho^3$  behavior, which is what one would expect if the coupling stops running.

$$R_\beta(g) = \frac{d(1/g^2)}{d \log(a)} \left( \frac{d(1/g^2)}{d \log(a)} \Big|_{g \rightarrow 0} \right)^{-1} \quad (130)$$

tends to 1 as  $g \rightarrow 0$  ( $6/g^2 \rightarrow \infty$ ). The two-loop correction is positive, so  $R_\beta \rightarrow 1$  from above. However, in the non-perturbative region the results show the opposite trend, with  $R_\beta$  dropping significantly below 1 around  $(6/g^2) \approx 6$ . This means that the coupling constant runs slower than the one-loop formula suggests. At somewhat smaller  $6/g^2$ ,  $R_\beta$  displays a rapid upward turn, which is known as the transition to the strong-coupling regime. This part of the data is clearly dominated by lattice artifacts. The significant reduction of the beta function (by about 50%) observed for intermediate values of  $g$  is precisely what is needed to explain the suppression of large-size instantons. Furthermore, the reduction of  $R_\beta$  is even larger for  $N_f=2$ . This might very well be a precursor of the infrared fixed point. At some critical  $N_f$  we expect  $R_\beta$  to touch zero. As  $N_f$  is further increased, the zero should move to weaker coupling (to the right) and reach infinity at  $N_f=33/2$ .

#### D. Instantons and confinement

After the discovery of instantons, it was hoped that instantons might help us to understand confinement in QCD. This hope was mainly inspired by Polyakov's proof that instantons lead to confinement in three-dimensional compact QED (Polyakov, 1987). However, there are important differences between three- and four-dimensional theories. In three dimensions, the field of an instanton looks like a magnetic monopole (with  $B \sim 1/r^2$ ), while in four dimensions it is a dipole field that falls off as  $1/r^4$ .

For a random ensemble of instantons, one can calculate the instanton contribution to the heavy-quark potential. In the dilute gas approximation, the result is determined by the Wilson loop in the field of an individual instanton (Callan, Dashen, and Gross, 1978b). The corresponding potential is  $V \sim x^2$  for small  $x$ , but tends to a constant at large distances. This result was confirmed by numerical simulations in the random ensemble (Shuryak, 1989), as well as the mean-field approximation (Diakonov, Petrov, and Poblitsa, 1989). The main instanton effect is a renormalization of the heavy-quark mass  $\delta M_Q = 50\text{--}70$  MeV. The force  $|dV/dx|$  peaks at  $x \approx 0.5$  fm, but even at this point it is almost an order of magnitude smaller than the string tension (Shuryak, 1989).

Later attempts to explain confinement in terms of instantons (or similar objects) fall roughly into three different categories: objects with fractional topological charge, strongly correlated instantons, and the effects of very large instantons.

Classical objects with fractional topological charge were first seriously considered by Callan *et al.* (1978a), who proposed a liquid consisting of instantons and merons. Merons are singular configurations with topological charge  $Q=1/2$  (Alfaro, Fubini, and Furlan, 1976). Basically, one can interpret merons as the result of split-

ting the dipole field of an instanton into two halves. This means that merons have long-range fields.

Another way to introduce fractional charge is by considering twisted boundary conditions ('t Hooft, 1981a). In this case, one finds fracton solutions (also known as 't Hooft fluxes) with topological charges quantized in units  $1/N_c$ . Gonzales-Arroyo and Montero (1996) suggested that confinement is produced by an ensemble of these objects glued together. In order to study this hypothesis, they measured the string tension in cooled and uncooled configurations with twisted boundary conditions. They found that fractionally charged objects could indeed be identified, and that their number roughly scaled with the string tension. Clearly, there are many problems that need to be understood, in particular, what the role of the boundary condition is and how regions with twisted boundary conditions can be glued together.

An important attempt at understanding confinement is based on the role of magnetic monopoles (see below). One can make monopolelike configuration (more precisely, dyonlike, since the fields are self-dual) from instantons by lining up identically oriented instantons.<sup>29</sup> Another possibility is a chain of strongly correlated instanton anti-instanton pairs, which might create an infinitely long monopole loop. Under normal circumstances, however, these objects have very small entropy, and they are not found in the simulations discussed in Sec. V.

The possible role of very large instantons was discussed by Diakonov and Petrov (1996). These authors propose that instantons can cause confinement if the size distribution behaves as  $n(\rho) \sim 1/\rho^3$ . This can be understood as follows. In Sec. IV.H, we shall show that the mass renormalization of a heavy quark due to instantons is  $\Delta M_Q \sim (N/V)\rho^3$ . For typical instanton radii this contribution is not very important, but if the size distribution has a  $1/\rho^3$  tail, then  $\Delta M_Q$  is linearly divergent, a possible signature of confinement. This is very intriguing, but again, there are a number of open questions. For one, the  $1/\rho^3$  distribution implies that the total volume occupied by instantons is infinite. Very large instantons would also introduce long-range correlations, which are inconsistent with the expected exponential decay of gluonic correlation functions.

Whatever the mechanism of confinement may turn out to be, it is clear that instantons should be affected by confinement in some way. One example of this line of reasoning is the idea that confinement might be the reason for the suppression of large-size instantons. A related, more phenomenological suggestion is that instantons provide a dynamic mechanism for bag formation (see Shuryak, 1978b; Callan, Dashen, and Gross, 1979). A lattice measurement of the suppression of instantons in the field of a static quark was recently performed by the Vienna group (see Faber *et al.*, 1995, and references therein). The measured distribution of the topological

<sup>29</sup>An example for this is the finite-temperature caloron solution; see Sec. VII.A.1.-

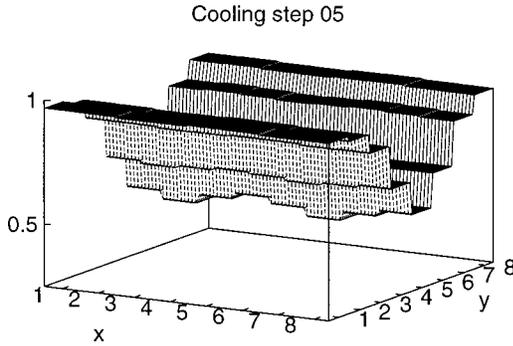


FIG. 13. Squared topological charge density around a static, heavy quark-antiquark pair in SU(3) with  $N_f=3$ , from Faber *et al.*, 1995. The results were obtained on an  $8^3 \times 4$  lattice at  $6/g^2=5.6$ . The topological charge was measured after five cooling steps.

charge is shown in Fig. 13. The instanton density around the static charge is indeed significantly suppressed, by about a factor of two. Since instantons give a vacuum energy density on the order of  $-1 \text{ GeV}/\text{fm}^3$ , this effect alone generates a significant difference in the nonperturbative energy density (bag constant) inside and outside of a hadron.

A very interesting picture of confinement in QCD is based on the idea that the QCD vacuum behaves like a dual superconductor, formed by the condensation of magnetic monopoles (Mandelstam, 1976). Although the details of this mechanism lead to many serious questions (DelDebbio *et al.*, 1997), there is some evidence in favor of this scenario from lattice simulations. There are no semiclassical monopole solutions in QCD, but monopoles can arise as singularities after imposing a gauge condition ('t Hooft, 1981b). The number and the location of monopole trajectories will then depend on the gauge choice. In practice, the so-called maximal Abelian gauge has turned out to be particularly useful (Kronfeld, Schierholz, and Wiese, 1987). The maximal Abelian gauge is specified by the condition that the norms of the off-diagonal components of the gauge fields be minimal, e.g.,  $\text{tr}[(A_\mu^1)^2 + (A_\mu^2)^2] = \min$  in SU(2). This leaves one U(1) degree of freedom unfixed, so in this preferred subgroup one can identify magnetic charges, study their trajectories, and evaluate their contribution to the (Abelian) string tension. The key observations are (a) that the Abelian string tension (in maximal Abelian gauge) is numerically close to the full non-Abelian string tension, and (b) that it is dominated by the longest monopole loops (Smit and van der Sijs, 1989; Suzuki, 1993).

We can not go into the details of these calculations, which would require a review in themselves, but only mention some ideas on how instantons and monopoles might be correlated. If monopoles are responsible for color confinement, then their paths should wind around the color flux tubes, just as electric-current lines wind around magnetic-flux tubes in ordinary superconductors. We have already mentioned that color flux tubes expel topological charges (see Fig. 13). This suggests that instantons and monopoles should be (anti?) correlated.

Lattice simulations that combine Abelian projection with the cooling technique or improved topological charges have indeed observed a strong positive correlation between monopoles and instantons (Feurstein, Markum, and Thurner, 1997). In order to understand this phenomenon in greater detail, a number of authors have studied monopole trajectories associated with individual instantons or instanton pairs (Chernodub and Gubarev, 1995; Hart and Teper, 1995; Suganuma *et al.*, 1995). It was observed that each instanton was associated with a monopole loop. For instanton anti-instanton pairs, these loops might either form two separate loops or one common loop, depending on the relative color orientation of the two instantons. Since then, it has been shown that the small loops associated with individual instantons do not correspond to global minima of the gauge-fixing functional and should be discarded (Brower, Orginos, and Tan, 1996). On the other hand, monopole trajectories connecting two or more instantons appear to be physical. The main physical question is then whether these monopole loops can percolate to form the long monopole loops responsible for confinement.

#### IV. TOWARDS A THEORY OF THE INSTANTON ENSEMBLE

##### A. The instanton interaction

###### 1. The gauge interaction

In the last section we argued that the density of instantons in the QCD vacuum is quite significant, implying that interactions among them are essential to an understanding of the instanton ensemble. In general, it is clear that field configurations containing both instantons and anti-instantons are not exact solutions of the equations of motion and that the action of an instanton/anti-instanton pair is not equal to twice the single instanton action. The interaction is defined by

$$S_{\text{int}} = S(A_\mu^{\text{IA}}) - 2S_0, \quad (131)$$

where  $A_\mu^{\text{IA}}$  is the gauge potential of the instanton/anti-instanton pair. Since  $A_\mu^{\text{IA}}$  is not an exact solution, there is some freedom in choosing an ansatz for the gauge field. This freedom corresponds to finding a convenient parametrization of the field configurations in the vicinity of an approximate saddle point. In general, we have to integrate over all field configurations anyway, but the ansatz determines the way we split coordinates into approximate zero modes and nonzero modes.

For well separated IA pairs, the fields are not strongly distorted and the interaction is well defined. For very close pairs, on the other hand, the fields are strongly modified and the interaction is not well determined. In addition to that, if the instanton and anti-instanton begin to annihilate, the gauge fields become perturbative and should not be included in semiclassical approximations. We comment on the present understanding of these questions below.

The interaction between instantons at large distances was derived by Callan, Dashen, and Gross (1978a; see also Förster, 1977). They began by studying the interaction of an instanton with a weak, slowly varying external field  $(G_{\mu\nu}^a)_{\text{ext}}$ . In order to ensure that the gauge field is localized, the instanton is put in the singular gauge. One then finds

$$S_{\text{int}} = -\frac{2\pi^2}{g^2} \rho^2 \bar{\eta}_{\mu\nu}^a (G_{\mu\nu}^a)_{\text{ext}}. \tag{132}$$

This result can be interpreted as the external field interacting with the color magnetic ‘‘dipole moment’’  $(2\pi^2/g^2)\rho^2 \bar{\eta}_{\mu\nu}^a$  of the instanton. Note that the interaction vanishes if the external field is self-dual. If the external field is taken to be an anti-instanton at distance  $R$ , the interaction is

$$S_{\text{int}} = +\frac{32\pi^2}{g^2} \rho_I^2 \rho_A^2 \bar{\eta}_{\mu\rho}^a \eta_{\nu\rho}^b R^{ab} \frac{\hat{R}_\mu \hat{R}_\nu}{R^4}, \tag{133}$$

where  $R^{ab}$  is the relative color orientation matrix  $R^{ab} = \frac{1}{2} \text{tr}(U \tau^a U^\dagger \tau^b)$  and  $\hat{R}$  is the unit vector connecting the centers of the instanton and anti-instanton. The dipole-dipole form of the interaction is quite general; the interaction of topological objects in other theories, such as skyrmions or  $O(3)$  instantons, is of similar type.

Instead of considering the dipole interaction as a classical phenomenon, we can also look at the interaction as a quantum effect, due to gluon exchanges between the instantons (Zakharov, 1992). The linearized interaction (132) corresponds to an instanton vertex that describes the emission of a gluon from the classical instanton field. The amplitude for the exchange of one gluon between an instanton and an anti-instanton is given by

$$\frac{4\pi^4}{g^2} \rho_1^2 \rho_2^2 R_I^{ab} R_A^{cd} \eta_{\mu\nu}^b \eta_{\alpha\beta}^d \langle G_{\mu\nu}^a(x) G_{\alpha\beta}^c(0) \rangle, \tag{134}$$

where  $\langle G_{\mu\nu}^a(x) G_{\alpha\beta}^c(0) \rangle$  is the free gauge-field propagator,

$$\langle G_{\mu\nu}^a(x) G_{\alpha\beta}^c(0) \rangle = \frac{2\delta^{ac}}{\pi^2 x^6} (g_{\nu\alpha} x_\mu x_\beta + g_{\mu\beta} x_\nu x_\alpha - g_{\nu\beta} x_\mu x_\alpha - g_{\mu\alpha} x_\nu x_\beta). \tag{135}$$

Inserting Eq. (135) into (134) gives the dipole interaction to first order. Summing all  $n$  gluon exchanges allows one to exponentiate the result and reproduce the full dipole interaction (133).

In order to study this interaction in greater detail, it is useful to introduce some additional notation. We can characterize the relative color orientation in terms of the four-vector  $u_\mu = (1/2i) \text{tr}(U \tau_\mu^+)$ , where  $\tau_\mu^+ = (\vec{\tau}, -i)$ . Note that for the gauge group  $SU(2)$ ,  $u_\mu$  is a real vector with  $u^2=1$ , whereas for  $SU(N>2)$ ,  $u_\mu$  is complex and  $|u|^2 \leq 1$ . Also note that, for  $SU(N>2)$ , the interaction is completely determined by the upper  $2 \times 2$  block of the  $SU(N)$  matrix  $U$ . We can define a relative color orientation angle  $\theta$  by

$$\cos^2 \theta = \frac{|u \cdot \hat{R}|^2}{|u|^2}. \tag{136}$$

In terms of this angle, the dipole interaction is given by

$$S_{\text{int}} = -\frac{32\pi^2}{g^2} \frac{\rho_1^2 \rho_2^2}{R^4} |u|^2 (1 - 4 \cos^2 \theta). \tag{137}$$

The orientational factor  $d = 1 - 4 \cos^2 \theta$  varies between 1 and  $-3$ . We also observe that the dipole interaction vanishes when we average over all angles  $\theta$ .

The dipole interaction is the leading part of the interaction at large distances. In order to determine the interaction at intermediate separation, we have to evaluate the total action for an IA pair. This is most easily done for the so-called sum ansatz

$$A_\mu^{\text{IA}} = A_\mu^I + A_\mu^A, \tag{138}$$

where again both the instanton and the anti-instanton are in the singular gauge. The action  $S = 1/4 \int d^4x G^2$  can easily be split into the free part  $2S_0$  and the interaction. Systematically expanding in  $1/R^2$ , one finds (Diakonov and Petrov, 1984)

$$S_{\text{int}}^{\text{IA}} = -\frac{8\pi^2}{g^2} \left\{ (|u|^2 - 4|u \cdot \hat{R}|^2) \times \left[ \frac{4\rho_1^2 \rho_2^2}{R^4} - \frac{15\rho_1^2 \rho_2^2 (\rho_1^2 + \rho_2^2)}{2R^6} \right] + |u|^2 \frac{9\rho_1^2 \rho_2^2 (\rho_1^2 + \rho_2^2)}{2R^6} + O(R^{-8}) \right\} \tag{139}$$

for the IA interaction and

$$S_{\text{int}}^{\text{II}} = \frac{8\pi^2}{g^2} (9 + 3|u|^2 - 4|\vec{u}|^2) \frac{\rho_1^2 \rho_2^2 (\rho_1^2 + \rho_2^2)}{2R^6} + O(R^{-8}) \tag{140}$$

for the instanton-instanton (II) interaction. Here,  $\vec{u}$  denotes the spatial components of the four-vector  $u_\mu$ . To first order, we find the dipole interaction in the IA channel and no interaction between two instantons. To next order, there is a repulsive interaction for both IA and II.

Clearly, it is important to understand to what extent this interaction is unique or if it depends on details of the underlying ansatz for the gauge potential. It was also realized that the simple sum ansatz leads to certain artifacts, for example, that the field strength at the centers of the two instantons becomes infinite and that there is an interaction between pseudoparticles of the same charge. One of us (Shuryak, 1988a) therefore proposed the ‘‘ratio’’ ansatz

$$A_\mu^a = \frac{2R_I^{ab} \bar{\eta}_{\mu\nu}^b \frac{\rho_I^2 (x-z_I)_\nu}{(x-z_I)^4} + 2R_A^{ab} \eta_{\mu\nu}^b \frac{\rho_A^2 (x-z_A)_\nu}{(x-z_A)^4}}{1 + \frac{\rho_I^2}{(x-z_I)^2} + \frac{\rho_A^2}{(x-z_A)^2}}, \tag{141}$$

whose form was inspired by 't Hooft's exact multi-instanton solution. This ansatz ensures that the field strength is regular everywhere and that (at least if they have the same orientation) there is no interaction between pseudoparticles of the same charge. Deriving an analytic expression for the interaction in the ratio ansatz

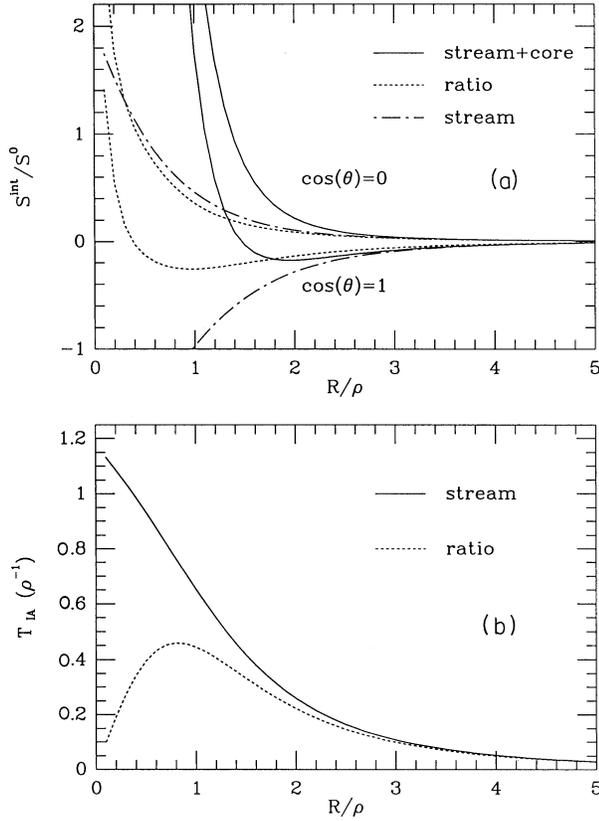


FIG. 14. Gauge field and fermion interaction for an instanton/anti-instanton pair: (a) the gluonic interaction in the streamline (dash-dotted curve) and ratio ansatz (short-dashed curve). The interaction is given in units of the single instanton action  $S_0$  for the most attractive ( $\cos\theta=1$ ) and most repulsive ( $\cos\theta=0$ ) orientations. The solid line shows the original streamline interaction supplemented by the core discussed in Sec. V.C. (b) Fermionic overlap matrix element in the streamline (solid curve) and ratio ansatz (dotted curve). The matrix elements are given in units of the geometric mean of the instanton radii.

is somewhat tedious. Instead, we have calculated the interaction numerically and parametrized the result; see Shuryak and Verbaarschot (1991) and Fig. 14. We observe that both the sum and the ratio ansatz lead to the dipole interaction at large distances, but the amount of repulsion at short distance is significantly reduced in the ratio ansatz.

Given these ambiguities, the question is whether one can find the best possible ansatz for the gauge field of an IA pair. This question was answered with the introduction of the streamline or valley method (Balitsky and Yung, 1986). The basic idea is to start from a well separated instanton/anti-instanton pair, which is an approximate solution of the equations of motion, and then let the system evolve under the force given by the variation of the action (Shuryak, 1988b). For a scalar field theory, the streamline equation is given by

$$f(\lambda) \frac{d\phi}{d\lambda} = \frac{\delta S}{\delta \phi}, \quad (142)$$

where  $\lambda$  is the collective variable corresponding to the

evolution of the pair and  $f(\lambda)$  is a function that depends on the parametrization of the streamline. The initial conditions are  $\phi(\infty) = \phi_c$  and  $\phi'(\infty) = \phi_0$ , where  $\phi_c$  is the classical solution corresponding to a well-separated instanton/anti-instanton pair and  $\phi_0$  is the translational zero mode.

In QCD, the streamline equations were solved by Verbaarschot (1991) using the conformal invariance of the classical equations of motion. Conformal invariance implies that an instanton/anti-instanton pair in the singular gauge at some distance  $R$  is equivalent to a regular-gauge instanton and a singular-gauge anti-instanton with the same center but different sizes. We do not want to go into details of the construction, but refer the reader to the original work. It turns out that the resulting gauge configurations are very close to the ansatz originally proposed by Yung (1988),

$$A_\mu^a = 2 \eta_{\mu\nu}^a x_\nu \frac{1}{x^2 + \rho^2 \lambda} + 2 R^{ab} \eta_{\mu\nu}^b x_\nu \frac{\rho^2}{\lambda} \frac{1}{x^2 (x^2 + \rho^2/\lambda)}, \quad (143)$$

where  $\rho = \sqrt{\rho_1 \rho_2}$  is the geometric mean of the two instanton radii and  $\lambda$  is the conformal parameter

$$\lambda = \frac{R^2 + \rho_1^2 + \rho_2^2}{2\rho_1\rho_2} + \left( \frac{(R^2 + \rho_1^2 + \rho_2^2)^2}{4\rho_1^2\rho_2^2} - 1 \right)^{1/2}. \quad (144)$$

Note that it is large not only if the distance  $R$  significantly exceeds the mean (geometric) size, but also if  $R$  is small and one instanton is much larger than another. (This latter situation is important when we consider suppression of large-size instantons.)

The interaction for this ansatz is given by (Verbaarschot, 1991).

$$S_{\text{IA}} = \frac{8\pi^2}{g^2} \frac{4}{(\lambda^2 - 1)^3} \{ -4(1 - \lambda^4 + 4\lambda^2 \log(\lambda)) \\ \times [ |u|^2 - 4|u \cdot \hat{R}|^2 ] + 2(1 - \lambda^2 + (1 + \lambda^2) \log(\lambda)) \\ \times [ (|u|^2 - 4|u \cdot \hat{R}|^2)^2 + |u|^4 + 2(u \cdot u^*)^2 ] \}, \quad (145)$$

which is also shown in Fig. 14. For the repulsive orientation, the interaction is similar to the ratio ansatz, and the average interaction is repulsive. For the most attractive orientation, however, the interaction approaches  $-2S_0$  at short distance, showing that the instanton and anti-instanton annihilate each other and the total action of the pair approaches zero.

It is interesting to note that the Yung ansatz, at least to order  $1/R^6$ , leads to the same interaction as the perturbative method (134) if carried out to higher order (Arnold and Mattis, 1991; Balitsky and Schäfer, 1993; Diakonov and Petrov, 1994). This problem is related to the calculation of instanton-induced processes at high energy (like multigluon production in QCD, or the baryon-number-violating cross section in electroweak theory). To leading order in  $E/M_{\text{sph}}$ , where  $E$  is the bombarding energy and  $M_{\text{sph}}$  the sphaleron barrier (see Sec. VIII.B), the cross section is given by

$$\sigma_{\text{tot}} = \text{disc} \int d^4 R e^{i p \cdot R} \int \int d\Omega_1 d\Omega_2 n(\rho_1) n(\rho_2) \times e^{-S_{\text{int}}(R, \Omega_{12})}, \tag{146}$$

where  $\Omega_i = (z_i, \rho_i, U_i)$  are the collective coordinates associated with the instanton and anti-instanton, and  $\text{disc } f(s) = (1/2i)[f(s+i\epsilon) - f(s-i\epsilon)]$  denotes the discontinuity of the amplitude after continuation to Minkowski space,  $s = p^2 > 0$ . Given the agreement of the streamline method and the perturbative calculation to leading and next-to-leading order, it has been suggested that the behavior of  $\sigma_{\text{tot}}$  at high energy be used to define the interaction  $S_{\text{int}}$  (Diakonov and Polyakov, 1993). The behavior of this quantity is still debated, but it is generally believed that the baryon-number-violating cross section does not reach the unitarity bound. In that case, the interaction would have to have some repulsion at short distance, unlike the streamline solution.

Another possible way to make sense of the short-distance part of the IA interaction in the streamline method is to use analytic continuation in the coupling constant, as discussed in Sec. II.A. Allowing  $g^2 \rightarrow -g^2$  gives a new saddle point in the complex  $g$  plane at which the IA interaction is repulsive and the semiclassical method is under control.

2. The fermionic interaction

In the presence of light fermions, instantons interact with each other not only through their gauge fields, but also through fermion exchanges. The basic idea in dealing with the fermionic interaction is that the Dirac spectrum can be split into quazero modes, linear combinations of the zero modes of the individual instantons, and nonzero modes. Here we shall focus on the interaction due to approximate zero modes. The interactions due to nonzero modes and the corrections due to interference between zero and nonzero modes were studied by Brown and Creamer (1978) and Lee and Bardeen (1979).

In the basis spanned by the zero modes, we can write the Dirac operator as

$$i\mathcal{D} = \begin{pmatrix} 0 & T_{IA} \\ T_{AI} & 0 \end{pmatrix}, \tag{147}$$

where we have introduced the overlap matrix elements  $T_{IA}$ ,

$$T_{IA} = \int d^4 x \psi_{0,I}^\dagger(x - z_I) i\mathcal{D} \psi_{0,A}(x - z_A). \tag{148}$$

Here,  $\psi_{0,I}$  is the fermionic zero mode (101). The matrix elements have the meaning of a hopping amplitude for a quark from one pseudoparticle to another. Indeed, the amplitude for an instanton to emit a quark is given by the amputated zero-mode wave function  $i\mathcal{D} \psi_{0,I}$ . This shows that the matrix element (148) can be written as two quark-instanton vertices connected by a propagator,  $\psi_{0,I}^\dagger i\mathcal{D} (i\mathcal{D})^{-1} i\mathcal{D} \psi_{0,A}$ . At large distance, the overlap matrix element decreases as  $T_{IA} \sim 1/R^3$ , which corresponds to the exchange of a massless quark. The deter-

minant of the matrix (147) can be interpreted as the sum of all closed diagrams consisting of (zero-mode) quark exchanges between pseudoparticles. In other words, the logarithm of the Dirac operator (147) is the contributions of the 't Hooft effective interaction to the vacuum energy.

Due to the chirality of the zero modes, the matrix elements  $T_{II}$  and  $T_{AA}$  between instantons of the same charge vanish. In the sum ansatz, we can use the equations of motion to replace the covariant derivative in (148) by an ordinary one. The dependence on the relative orientation is then given by  $T_{IA} = (u \cdot \hat{R}) f(R)$ . This means that, like the gluonic dipole interaction, the fermion overlap is maximal if  $\cos \theta = 1$ . The matrix element can be parametrized by

$$T_{IA} = i(u \cdot R) \frac{1}{\rho_I \rho_A} \frac{4.0}{[2.0 + R^2/(\rho_I \rho_A)]^2}. \tag{149}$$

A parametrization of the overlap matrix element for the streamline gauge configuration can be found in Shuryak and Verbaarschot (1992). The result is compared with the sum ansatz in Fig. 14 (for  $\cos \theta = 1$ ). We observe that the matrix elements are very similar at large distance but differ at short distance.

Using these results, we may write the contribution of an instanton/anti-instanton pair to the partition function as

$$Z_{IA} = V_4 \int d^4 z dU \exp[N_f \log |T_{IA}(U, z)|^2 - S_{\text{int}}(U, z)]. \tag{150}$$

Here,  $z$  is the distance between the centers, and  $U$  is the relative orientation of the pair. The fermionic part is attractive, while the bosonic part is either attractive or repulsive, depending on the orientation. If the interaction is repulsive, there is a real saddle point for the  $z$  integral, whereas for the attractive orientation there is only a saddle point in the complex  $z$  plane (as in Sec. II.B).

The calculation of the partition function (150) in the saddle-point approximation was recently attempted by Shuryak and Velkovsky (1997). They find that for a large number of flavors,  $N_f > 5$ , the ground-state energy oscillates as a function of  $N_f$ . The period of the oscillation is 4, and the real part of the energy shift vanishes for even  $N_f = 6, 8, \dots$ . The reason for these oscillations is exactly the same as in the case of supersymmetry quantum mechanics: the saddle point gives a complex contribution with a phase that is proportional to the number of flavors.

B. Instanton ensembles

In Sec. II.C.4 we studied the semiclassical theory of instantons, treating them as very rare (and therefore independent) tunneling events. However, as emphasized in Sec. III.A, in QCD instantons are not rare, so one cannot just exponentiate the results obtained for a single instanton. Before we study the instanton ensemble in

QCD, we should like to discuss a simple physical analogy. In this picture, we think of instantons as atoms and light quarks as valence electrons.

A system like this can exist in many different phases, e.g., as a gas, liquid, or solid. In theories with massless quarks, instantons have “unsaturated bonds” and cannot appear individually. Isolated instantons in the form of an atomic gas can only exist if there is a nonzero quark mass. The simplest “neutral” object is an instanton/anti-instanton (IA) molecule. Therefore, if the instanton density is low (for whatever reason—high temperature, a large Higgs expectation value, etc.) the system should be in a phase consisting of IA molecules. Below, we shall also argue that this is the case if the density is not particularly small, but the interactions are strong and favor the formation of molecules, for example, if the number of light quarks  $N_f$  exceeds some critical value. If the instanton ensemble is in the molecular phase, then quarks are bound to molecules and cannot propagate over large distances. This means that there are no eigenmodes with almost zero virtuality, the “conductivity” is zero, or chiral symmetry remains unbroken.

The liquid phase differs from the gas phase in many important respects. The density is determined by the interactions and cannot be arbitrarily small. A liquid has a surface tension, etc. As we shall see below, the instanton ensemble in QCD has all these properties. In QCD we also expect that chiral symmetry is spontaneously broken. This means that in the ground state there is a preferred direction in flavor space, characterized by the quark condensate  $\langle \bar{q}q \rangle$ . This preferred orientation can only be established if quarks form an infinite cluster. In terms of our analogy, this means that electrons are delocalized and the conductivity is nonzero. In the instanton liquid phase, quarks are delocalized because the instanton zero modes become collective.

At very high density, interactions are strong and the ensemble is expected to form a four-dimensional crystal. In this case, both Lorentz and gauge invariance would be broken spontaneously; clearly, this phase is very different from the QCD vacuum. Fortunately, however, explicit calculations (Diakonov and Petrov, 1984; Shuryak and Verbaarschot, 1990) show that the instanton liquid crystallizes only if the density is pushed to about two orders of magnitude larger than the phenomenological value. At physical densities, the crystalline phase is not favored: although the interaction is attractive, the entropy is too small as compared to the random liquid.<sup>30</sup> The electronic structure of a crystal consists of several bands. In our case, the Fermi surface lies between dif-

ferent bands, so the crystal is an insulator. In QCD, this means that chiral symmetry is not broken.

Another analogy with liquids concerns the question of density stabilization. In order for the system to saturate, we need a short-range repulsive force in addition to the attractive, long-range dipole interaction. We have already mentioned that the nature of the short-range interaction and the (possibly related) question of the fate of large instantons in QCD are not well understood. Lattice simulations indicate that large instantons are strongly suppressed, but for the moment we have to include this result in a phenomenological manner.

Let us summarize this qualitative discussion. Depending on the density, the instanton ensemble is expected to be in a gas, liquid, or solid phase. The phase boundaries will in general depend on the number of colors and flavors. Large  $N_c$  favors a crystalline phase, while large  $N_f$  favors a molecular gas. Neither case is phenomenologically acceptable as a description of real QCD, with two light and one intermediate mass flavor. We therefore expect (and will show below) that the instanton ensemble in QCD is in the liquid phase.

### C. The mean-field approximation

In order to study the structure of the instanton ensemble in a more quantitative fashion, we consider the partition function for a system of instantons in pure gauge theory,

$$Z = \frac{1}{N_+! N_-!} \prod_i^{N_+ + N_-} \int [d\Omega_i n(\rho_i)] \exp(-S_{\text{int}}). \quad (151)$$

Here,  $N_{\pm}$  are the numbers of instantons and anti-instantons,  $\Omega_i = (z_i, \rho_i, U_i)$  are the collective coordinates of the instanton  $i$ ,  $n(\rho)$  is the semiclassical instanton distribution function (93), and  $S_{\text{int}}$  is the bosonic instanton interaction. In general, the dynamics of a system of pseudoparticles governed by Eq. (151) is still quite complicated, so we have to rely on approximation schemes. There are a number of techniques well known from statistical mechanics that can be applied to the problem, for example, the mean-field approximation or the variational method. These methods are expected to be reliable as long correlations between individual instantons are weak.

The first such attempt was made by Callan, Dashen, and Gross (1978a). These authors included only the dipole interaction, which, as we noted above, vanishes on average and produces an attractive interaction to next order in the density. In this case, there is nothing to balance the growth  $n(\rho) \sim \rho^{b-5}$  of the semiclassical instanton distribution. In order to deal with this problem, Ilgenfritz and Müller-Preußker (1981) introduced an *ad hoc* “hard core” in the instanton interaction (see also Münster, 1982). The hard core automatically excludes large instantons and leads to a well-behaved partition function. It is important to note that one cannot simply cut the size integration at some  $\rho_c$ , but has to introduce the cutoff in a way that does not violate the conformal

<sup>30</sup>Diakonov and Petrov (1984) suggested that the instanton liquid crystallizes in the limit of a large number of colors,  $N_c \rightarrow \infty$ , because the interaction is proportional to the charge  $8\pi^2/g^2$ , which is of order  $N_c$ . As long as instantons do not disappear altogether (the action is of order 1), the interaction becomes increasingly important. However, little is known about the structure of large- $N_c$  QCD.

invariance of the classical equations of motion. This guarantees that the results do not spoil the renormalization properties of QCD, most notably the trace anomaly (see below). In practice, Ilgenfritz and Müller-Preußker chose

$$S_{\text{int}} = \begin{cases} \infty & |z_I - z_A| < (a\rho_I^2\rho_A^2)^{1/4} \\ 0 & |z_I - z_A| > (a\rho_I^2\rho_A^2)^{1/4}, \end{cases} \quad (152)$$

which leads to an excluded volume in the partition function, controlled by the dimensionless parameter  $a$ .

We do not go into the details of their results, but present the next development (Diakonov and Petrov, 1984). In this work, the arbitrary core is replaced by the interaction in the sum ansatz [see Eqs. (139) and Eq. (140)]. The partition function is evaluated using a trial distribution function. If we assume that correlations between instantons are not very important, then a good trial function is given by the product of single-instanton distributions  $\mu(\rho)$ ,

$$\begin{aligned} Z_1 &= \frac{1}{N_+!N_-!} \prod_i^{N_++N_-} \int d\Omega_i \mu(\rho_i) \\ &= \frac{1}{N_+!N_-!} (V\mu_0)^{N_++N_-}, \end{aligned} \quad (153)$$

where  $\mu_0 = \int d\rho \mu(\rho)$ . The distribution  $\mu(\rho)$  is determined from the variational principle  $\delta \log Z_1 / \delta \mu = 0$ . In quantum mechanics a variational wave function always provides an upper bound on the true ground-state energy. The analogous statement in statistical mechanics is known as Feynman's variational principle. Using convexity

$$Z = Z_1 \langle \exp[-(S - S_1)] \rangle \geq Z_1 \exp(-\langle S - S_1 \rangle), \quad (154)$$

where  $S_1$  is the variational action, one can see that the variational vacuum energy is always higher than the true one.

In our case, the single-instanton action is given by  $S_1 = \log[\mu(\rho)]$  while  $\langle S \rangle$  is the average action calculated from the variational distribution (153). Since the variational ansatz does not include any correlations, only the average interaction enters,

$$\langle S_{\text{int}} \rangle = \gamma^2 \rho_1^2 \rho_2^2, \quad \gamma^2 = \frac{27}{4} \frac{N_c}{N_c - 1} \pi^2 \quad (155)$$

for both IA and II pairs. Clearly, Eq. (155) is of the same form as the hard core (152) discussed above, only the dimensionless parameter  $\gamma^2$  is determined from the interaction in the sum ansatz. Applying the variational principle, one finds (Diakonov and Petrov, 1984)

$$\mu(\rho) = n(\rho) \exp\left[-\frac{\beta\gamma^2}{V} N \bar{\rho}^2 \rho^2\right], \quad (156)$$

where  $\beta = \beta(\bar{\rho})$  is the average instanton action and  $\bar{\rho}^2$  is the average size. We observe that the single-instanton distribution is cut off at large sizes by the average instanton repulsion. The average size  $\bar{\rho}^2$  is determined by the self-consistency condition  $\bar{\rho}^2 = \mu_0^{-1} \int d\rho \mu(\rho) \rho^2$ . The result is

$$\bar{\rho}^2 = \left(\frac{\nu V}{\beta\gamma^2 N}\right)^{1/2}, \quad \nu = \frac{b-4}{2}, \quad (157)$$

which determines the dimensionless diluteness of the ensemble,  $\rho^4(N/V) = \nu/(\beta\gamma^2)$ . Using the pure gauge beta function  $b=11$ ,  $\gamma^2 \approx 25$  from above and  $\beta \approx 15$ , we get  $\rho^4(N/V) = 0.01$ , even more dilute than phenomenology requires. The instanton density can be fixed from the second self-consistency requirement,  $(N/V) = 2\mu_0$  (the factor 2 comes from instantons and anti-instantons). We get

$$\frac{N}{V} = \Lambda_{PV}^4 [C_{N_c} \beta^{2N_c} \Gamma(\nu) (\beta\nu\gamma^2)^{-\nu/2}]^{2/(2+\nu)}, \quad (158)$$

where  $C_{N_c}$  is the prefactor in Eq. (93). The formula shows that  $\Lambda_{PV}$  is the only dimensionful parameter. The final results are

$$\begin{aligned} \chi_{\text{top}} &\approx \frac{N}{V} = (0.65\Lambda_{PV})^4, \quad (\bar{\rho}^2)^{1/2} = 0.47\Lambda_{PV}^{-1} \approx \frac{1}{3}R, \\ \beta = S_0 &\approx 15, \end{aligned} \quad (159)$$

consistent with the phenomenological values for  $\Lambda_{PV} \approx 300$  MeV. It is instructive to calculate the free energy as a function of the instanton density. Using  $F = -(1/V)\log Z$ , we have

$$F = \frac{N}{V} \left\{ \log\left(\frac{N}{2V\mu_0}\right) - \left(1 + \frac{\nu}{2}\right) \right\}. \quad (160)$$

The instanton density is determined by minimizing the free energy,  $\partial F / [\partial(N/V)] = 0$ . The vacuum energy density is given by the value of the free energy at the minimum,  $\epsilon = F_0$ . We find  $N/V = 2\mu_0$  as above and

$$\epsilon = -\frac{b}{4} \left(\frac{N}{V}\right). \quad (161)$$

Estimating the value of the gluon condensate in a dilute instanton gas,  $\langle G^2 \rangle = 32\pi^2(N/V)$ , we see that Eq. (161) is consistent with the trace anomaly. Note that, for non-interacting instantons (with the size integration regularized in some fashion), one would expect  $\epsilon \approx -(N/V)$ , which is inconsistent with the trace anomaly and shows the importance of respecting classical scale invariance.

The second derivative of the free energy with respect to the instanton density, the compressibility of the instanton liquid, is given by

$$\left. \frac{\partial^2 F}{\partial(N/V)^2} \right|_{n_0} = \frac{4}{b} \left(\frac{N}{V}\right), \quad (162)$$

where  $n_0$  is the equilibrium density. This observable is also determined by a low-energy theorem based on broken scale invariance (Novikov *et al.*, 1981),

$$\int d^4x \langle G^2(0)G^2(x) \rangle = (32\pi^2) \frac{4}{b} \langle G^2 \rangle. \quad (163)$$

Here, the left-hand side is given by an integral over the field-strength correlator, suitably regularized and with the constant disconnected term  $\langle G^2 \rangle^2$  subtracted. For a dilute system of instantons, the low-energy theorem gives

$$\langle N^2 \rangle - \langle N \rangle^2 = \frac{4}{b} \langle N \rangle. \quad (164)$$

Here,  $\langle N \rangle$  is the average number of instantons in a volume  $V$ . The result (164) shows that density fluctuations in the instanton liquid are not Poissonian. Using the general relation between fluctuations and the compressibility gives the result (162). This means that the form of the free energy near the minimum is determined by the renormalization properties of the theory. Therefore the functional form (160) is more general than the mean-field approximation used to derive it.

How reliable are the numerical results derived from the mean-field approximation? The accuracy of the mean-field approximation can be checked by doing statistical simulations of the full partition function. (We shall come to this approach in Sec. V.) Another question concerns the accuracy of the sum ansatz. This can be checked explicitly by calculating the induced current  $j_\mu = D_\nu G_{\mu\nu}^{\text{cl}}$  in the classical gauge configurations (Shuryak, 1985). This current measures the failure of the gauge potential to be a true saddle point. In the sum ansatz, the induced current gives a sizable contribution to the action, which means that this ansatz is not a good starting point for a self-consistent solution.

In principle, this problem is solved in the streamline method because, by construction,  $j_\mu$  is orthogonal to quantum fluctuations.<sup>31</sup> However, applying the variational method to the streamline configurations (Verbaarschot, 1991) is not satisfactory either, because the ensemble contains too many close pairs and too many large instantons.

In summary, phenomenology and the lattice seem to favor a fairly dilute instanton ensemble. This is well reproduced by the mean-field approximation based on the sum ansatz, but the results are not really self-consistent. How to generate a fully consistent ensemble in which large instantons are automatically suppressed remains an open problem. Nevertheless, as long as large instantons are excluded in a way that does not violate the symmetries of QCD, the results are likely to be independent of the precise mechanism that leads to the suppression of large instantons.

#### D. The quark condensate in the mean-field approximation

Proceeding from pure glue theory to QCD with light quarks, one has to deal with the much more complicated problem of quark-induced interactions. Not only does the fermion determinant induce a very nonlocal interaction, but the very presence of instantons cannot be understood in the single-instanton approximation. Indeed, as discussed in Sec. II.D, the semiclassical instanton density is proportional to the product of fermion masses and therefore vanishes in the chiral limit  $m \rightarrow 0$ . In the QCD

vacuum, however, chiral symmetry is spontaneously broken, and the quark condensate  $\langle \bar{q}q \rangle$  is nonzero. The quark condensate is the amplitude for a quark to flip its chirality, so we expect that the instanton density is controlled not by the current masses, but by the quark condensate, which does not vanish as  $m \rightarrow 0$ .

Given the importance of chiral symmetry breaking, we shall discuss this phenomenon on a number of different levels. In this section, we should like to give a simple qualitative derivation following (Shuryak, 1982b). Earlier works on chiral symmetry breaking by instantons are those of Caldi (1977) and Carlitz and Creamer (1979a, 1979b); see also the review of Diakonov (1995).

The simplest case is QCD with just one light flavor,  $N_f=1$ . In this theory, the only chiral symmetry is the axial  $U(1)_A$  symmetry, which is broken by the anomaly. This means that there is no spontaneous symmetry breaking, and the quark condensate appears at the level of a single instanton. The condensate is given by

$$\langle \bar{q}q \rangle = i \int d^4x \text{tr}[S(x,x)]. \quad (165)$$

In the chiral limit, nonzero modes do not contribute to the quark condensate. Using the zero-mode propagator  $S(x,y) = -\psi_0(x)\psi_0^\dagger(y)/(im)$ , we find that the contribution of a single instanton to the quark condensate is given by  $-1/m$ . Multiplying this result by the density of instantons, we have  $\langle \bar{q}q \rangle = -(N/V)/m$ . Since the instanton density is proportional to the quark mass  $m$ , the quark condensate is finite in the chiral limit.

The situation is different for  $N_f > 1$ . The theory still has an anomalous  $U(1)_A$  symmetry, which is broken by instantons. The corresponding order parameter  $\det_f(\bar{q}_{f,L}q_{f,R})$  (where  $f$  is the flavor index) appears already at the one-instanton level. But in addition to that, there is a chiral  $SU(N_f)_L \times SU(N_f)_R$  symmetry that is spontaneously broken to  $SU(N_f)_V$ . This effect cannot be understood on the level of a single instanton: the contribution to  $\langle \bar{q}q \rangle$  is still  $(N/V)/m$ , but the density of instantons is proportional to  $(N/V) \sim m^{N_f}$ .

Spontaneous symmetry breaking has to be a collective effect involving infinitely many instantons. This effect is most easily understood in the context of the mean-field method. For simplicity, we consider small-size instantons. Then the tunneling rate is controlled by the vacuum expectation value of the  $2N_f$ -fermion operator (111) in the 't Hooft effective Lagrangian. This vacuum expectation value can be estimated using the "vacuum dominance" (or factorization) approximation,

$$\langle \bar{\psi}\Gamma_1\psi\bar{\psi}\Gamma_2\psi \rangle = \frac{1}{N^2} (\text{Tr}[\Gamma_1]\text{Tr}[\Gamma_2] - \text{Tr}[\Gamma_1\Gamma_2]) \times \langle \bar{q}q \rangle^2, \quad (166)$$

where  $\Gamma_{1,2}$  is a spin, isospin, color matrix and  $N = 4N_fN_c$  is the corresponding degeneracy factor. Using this approximation, we find that the factor  $\prod_f m_f$  in the instanton density should be replaced by  $\prod_f m_f^*$ , where the effective quark mass is given by

<sup>31</sup>Except for the soft mode leading to the trivial gauge configuration, but the integration over this mode can be done explicitly.

$$m_f^* = m_f - \frac{2}{3} \pi^2 \rho^2 \langle \bar{q}_f q_f \rangle. \tag{167}$$

Thus, if chiral symmetry is broken, the instanton density is  $O((m^*)^{N_f})$ , finite in the chiral limit.

The next question is whether it is possible to make this estimate self-consistent and calculate the quark condensate from instantons. If we replace the current mass by the effective mass in the quark propagator,<sup>32</sup> the contribution of a single instanton to the quark condensate is given by  $1/m^*$  and, for a finite density of instantons, we expect

$$\langle \bar{q}q \rangle = - \frac{(N/V)}{m^*}. \tag{168}$$

This equation, taken together with Eq. (167), gives a self-consistent value for the quark condensate,

$$\langle \bar{q}q \rangle = - \frac{1}{\pi\rho} \sqrt{3N/2V}. \tag{169}$$

Using the phenomenological values  $(N/V) = 1 \text{ fm}^{-4}$  and  $\rho = 0.33 \text{ fm}$ , we get  $\langle \bar{q}q \rangle \approx -(215 \text{ MeV})^3$ , quite consistent with the experimental value  $\langle \bar{q}q \rangle \approx -(230 \text{ MeV})^3$ . The effective quark mass is given by  $m^* = \pi\rho(2/3)^{1/2}(N/V)^{1/2} \approx 170 \text{ MeV}$ . The self-consistent pair of equations (167), (168) has the general form of a gap equation. We shall provide a more formal derivation of the gap equation in Sec. IV.F.

### E. Dirac eigenvalues and chiral symmetry breaking

In this section we shall get a different and more microscopic perspective on the formation of the quark condensate. The main idea is to study the motion of quarks in a fixed gauge field and then average over all gauge-field configurations. This approach is quite natural from the point of view of the path integral (and lattice gauge theory). Since the integral over the quark fields can always be performed exactly, quark observables are determined by the exact quark propagator in a given gauge configuration, averaged over all gauge fields.

In a given gauge-field configuration, we can determine the spectrum of the Dirac operator  $i\mathcal{D} = [i\partial_\mu + A_\mu(x)]\gamma_\mu$ ,

$$i\mathcal{D}\psi_\lambda = \lambda\psi_\lambda, \tag{170}$$

where  $\psi_\lambda$  is an eigenstate with “virtuality”  $\lambda$ . In terms of the spectrum, the quark propagator  $S(x,y) = -\langle x|i\mathcal{D}^{-1}|y \rangle$  is given by

$$S(x,y) = - \sum_\lambda \frac{\psi_\lambda(x)\psi_\lambda^\dagger(y)}{\lambda + im}. \tag{171}$$

Using the fact that the eigenfunctions are normalized, we obtain the quark condensate,

$$i \int d^4x \text{tr}[S(x,x)] = -i \sum_\lambda \frac{1}{\lambda + im}. \tag{172}$$

We can immediately make a few important observations concerning the spectrum of the Dirac operator:

- (1) Since the Dirac operator is Hermitian, the eigenvalues  $\lambda$  are real. The inverse propagator  $(i\mathcal{D} + im)^{-1}$ , on the other hand, consists of a Hermitian and an anti-Hermitian piece.
- (2) For every nonzero eigenvalue  $\lambda$  with eigenvector  $\psi_\lambda$ , there is another eigenvalue  $-\lambda$  with eigenvector  $\gamma_5\psi_\lambda$ .
- (3) This implies that the fermion determinant is positive: combining the nonzero eigenvalues in pairs, we get

$$\prod_{\lambda \neq 0} (i\lambda - m) = \prod_{\lambda \geq 0} (i\lambda - m)(-i\lambda - m) = \prod_{\lambda \geq 0} (\lambda^2 + m^2). \tag{173}$$

- (4) Only zero modes can be unpaired. Since  $\gamma_5\psi_0 = \pm\psi_0$ , zero-mode wave functions have to have a definite chirality. We have already seen this in the case of instantons, where the Dirac operator has a left-handed zero mode.

Using the fact that nonzero eigenvalues are paired, we can write the trace of the quark propagator as

$$i \int d^4x \text{tr}[S(x,x)] = - \sum_{\lambda \geq 0} \frac{2m}{\lambda^2 + m^2}. \tag{174}$$

We have excluded zero modes since they do not contribute to the quark condensate in the limit  $m \rightarrow 0$  (for  $N_f > 1$ ). In order to determine the average quark condensate, we introduce the spectral density  $\rho(\nu) = [\sum_\lambda \delta(\nu - \lambda)]$ . We then have

$$\langle \bar{q}q \rangle = - \int_0^\infty d\lambda \rho(\lambda) \frac{2m}{\lambda^2 + m^2}. \tag{175}$$

This result shows that the order in which we take the chiral and thermodynamic limits is crucial. In a finite system, the integral is well behaved in the infrared, and the quark condensate vanishes in the chiral limit. This is consistent with the observation that there is no spontaneous symmetry breaking in a finite system. A finite spin system, for example, cannot be magnetized if there is no external field. If the energy barrier between states with different magnetization is finite and there is no external field that selects a preferred magnetization, the system will tunnel between these states and the average magnetization is zero. Only in the thermodynamic limit can the system develop a spontaneous magnetization.

However, if we take the thermodynamic limit first, we can have a finite density of eigenvalues arbitrarily close to zero. In this case, the  $\lambda$  integration is infrared divergent as  $m \rightarrow 0$ , and we get a finite quark condensate,

$$\langle \bar{q}q \rangle = - \pi\rho(\lambda=0), \tag{176}$$

a result known as the Banks-Casher relation (Banks and Casher, 1980). This expression shows that quark condensation is connected with quark states of arbitrarily small virtuality.

<sup>32</sup>We shall give a more detailed explanation for this approximation in Sec. VI.C.2.

Studying chiral symmetry breaking requires an understanding of quazero modes, the spectrum of the Dirac operator near  $\lambda=0$ . If there is only one instanton, the spectrum consists of a single zero mode plus a continuous spectrum of nonzero modes. If there is a finite density of instantons, the spectrum is complicated, even if the ensemble is very dilute. In the chiral limit, fluctuations of the topological charge are suppressed, so one can think of the system as containing as many instantons as anti-instantons. The zero modes are expected to mix, so that the eigenvalues spread over some range of virtualities  $\Delta\lambda$ . If chiral symmetry is broken, the natural value of the quark condensate is of order  $(N/V)/\Delta\lambda$ .

There is a useful analogy from solid-state physics. In condensed matter, atomic bound states may become delocalized and form a band. The material is a conductor if the Fermi surface lies inside a band. Such zones also exist in disordered systems like liquids, but in this case they do not have a sharp boundary.

In the basis spanned by the zero modes of the individual instantons the Dirac operator reduces to the matrix

$$i\mathcal{D} = \begin{pmatrix} 0 & T_{IA} \\ T_{AI} & 0 \end{pmatrix}, \quad (177)$$

already introduced in Sec. IV.A.2. The width of the zero-mode zone in the instanton liquid is governed by the off-diagonal matrix elements  $T_{IA}$  of the Dirac operator. The matrix elements depend on the relative color orientation of the pseudoparticles. If the interaction between instantons is weak, the matrix elements are distributed randomly with zero average, but their variance is nonzero. Averaging  $T_{IA}T_{IA}^*$  over the positions and orientations of a pair of pseudoparticles, one gets

$$\langle |T_{IA}|^2 \rangle = \frac{2\pi^2 N\rho^2}{3N_c V}. \quad (178)$$

If the matrix elements are distributed according to a Gaussian unitary ensemble,<sup>33</sup> the spectral density is a semicircle,

$$\rho(\lambda) = \frac{N}{\pi\sigma V} \left( 1 - \frac{\lambda^2}{4\sigma^2} \right)^{1/2}. \quad (179)$$

From the Casher-Banks formula, we then get the following result for the quark condensate:

$$\langle \bar{q}q \rangle = -\frac{1}{\pi\rho} \left( \frac{3N_c N}{2V} \right)^{1/2} \approx -(240 \text{ MeV})^3, \quad (180)$$

<sup>33</sup>In the original work (Diakonov and Petrov, 1985) from which these arguments are taken, it was assumed that the spectrum had a Gaussian shape, with the width given above. However, for a random matrix the correct result is a semicircle. In reality, the spectrum of the Dirac operator for a system of randomly distributed instantons is neither a semicircle nor a Gaussian; it has a nonanalytic peak at  $\lambda=0$  (Shuryak and Verbaarschot, 1990). This does not qualitatively change estimate for the quark condensate.

which has the same parametric dependence on  $(N/V)$  and  $\rho$  as the result in the previous section, only the coefficient is slightly different. In addition to that, we can identify the effective mass  $m^*$  introduced in the previous section with a weighted average of the eigenvalues,  $(m^*)^{-1} = N^{-1} \sum \lambda^{-1}$ .

It is very interesting to study the spectral density of the Dirac operator at small virtualities. Similar to the density of states near the Fermi surface in condensed-matter problems, this quantity controls the low-energy excitations of the system. If chiral symmetry is broken, the spectral density at  $\lambda=0$  is finite. In addition, chiral perturbation theory predicts the slope of the spectrum (Smilga and Stern, 1993),

$$\rho(\lambda) = -\frac{1}{\pi} \langle \bar{q}q \rangle + \frac{\langle \bar{q}q \rangle^2}{32\pi^2 f_\pi^4} \frac{N_f^2 - 4}{N_f} |\lambda| + \dots, \quad (181)$$

which is valid for  $N_f \geq 2$ . The second term is connected with the fact that for  $N_f > 2$  there is a Goldstone boson cut in the scalar-isovector ( $\delta$ -meson) correlator, while the decay  $\delta \rightarrow \pi\pi$  is not allowed for two flavors. The result implies that the spectrum is flat for  $N_f=2$  but has a cusp for  $N_f > 2$ .

## F. Effective interaction between quarks and the mean-field approximation

In this section we should like to discuss chiral symmetry breaking in terms of an effective, purely fermionic theory that describes the effective interaction between quarks generated by instantons (Diakonov and Petrov, 1986). For this purpose, we shall have to reverse the strategy used in the last section and integrate over the gauge field first. This will leave us with an effective theory of quarks that can be treated with standard many-body techniques. Using these methods allows us to study not only chiral symmetry breaking, but also the formation of quark-antiquark bound states in the instanton liquid.

For this purpose we rewrite the partition function of the instanton liquid

$$Z = \frac{1}{N_+! N_-!} \prod_i^{N_+ + N_-} \int [d\Omega_i n(\rho_i)] \times \exp(-S_{\text{int}}) \det(\mathcal{D} + m)^{N_f} \quad (182)$$

in terms of a fermionic effective action

$$Z = \int d\psi d\psi^\dagger \exp \left( \int d^4x \psi^\dagger (i\partial + im) \psi \right) \times \left\langle \prod_{I,f} (\Theta_{I^-} - im_f) \prod_{A,f} (\Theta_{A^-} - im_f) \right\rangle, \quad (183)$$

$$\Theta_{I,A} = \int d^4x (\psi^\dagger(x) i\partial \psi_{I,A}(x - z_{I,A}))$$

$$\times \int d^4y (\phi_{I,A}^\dagger(y-z_{I,A}) i \not{\partial} \psi(y)), \quad (184)$$

which describes quarks interacting via the 't Hooft vertices  $\Theta_{I,A}$ . The expectation value  $\langle \cdot \rangle$  corresponds to an average over the distribution of instanton collective coordinates. Formally, Eq. (183) can be derived by “fermionizing” the original action (see Nowak, 1991). In practice, it is easier to check the result by performing the integration over the quark fields and verifying that one recovers the fermion determinant in the zero-mode basis.

Here, however, we want to use a different strategy and exponentiate the 't Hooft vertices  $\Theta_{I,A}$  in order to derive the effective quark interaction. For this purpose we calculate the average in Eq. (183) with respect to the variational single-instanton distribution (156). There are no correlations, so only the average interaction induced by a single instanton enters. For simplicity, we only average over the position and color orientation and keep the average size  $\rho = \bar{\rho}$  fixed,

$$Y_\pm = \int d^4z \int dU \prod_f \Theta_{I,A}. \quad (185)$$

In order to exponentiate  $Y_\pm$ , we insert factors of unity  $\int d\Gamma_\pm \int d\beta_\pm / (2\pi) \exp[i\beta_\pm(Y_\pm - \Gamma_\pm)]$  and integrate over  $\Gamma_\pm$  using the saddle-point method to obtain

$$Z = \int d\psi d\psi^\dagger \exp\left(\int d^4x \psi^\dagger i \not{\partial} \psi\right) \int \frac{d\beta_\pm}{2\pi} \times \exp(i\beta_\pm Y_\pm) \exp\left[N_\pm \left(\log\left(\frac{N_\pm}{i\beta_\pm V}\right) - 1\right) + (+ \leftrightarrow -)\right],$$

where we have neglected the current quark mass. In this partition function, the saddle-point parameters  $\beta_\pm$  play the role of an activity for instantons and anti-instantons.

### 1. The gap equation for $N_f=1$

The form of the saddle-point equations for  $\beta_\pm$  depends on the number of flavors. The simplest case is  $N_f=1$ , where the Grassmann integration is quadratic. The average over the 't Hooft vertex is most easily performed in momentum space:

$$Y_\pm = \int \frac{d^4k}{(2\pi)^4} \int dU \times \psi^\dagger(k) [\not{k} \phi_{I,A}(k) \phi_{I,A}^\dagger(k) \not{k}] \psi(k), \quad (186)$$

where  $\phi(k)$  is the Fourier transform of the zero-mode profile (see Appendix A.2). Performing the average over the color orientation, we get

$$Y_\pm = \int \frac{d^4k}{(2\pi)^4} \frac{1}{N_c} k^2 \varphi'^2(k) \psi^\dagger(k) \gamma_\pm \psi(k), \quad (187)$$

where  $\gamma_\pm = (1 \pm \gamma_5)/2$  and  $\varphi'(k)$  is defined in the Appendix. Clearly, the saddle-point equations are symmetric in  $\beta_\pm$ , so that the average interaction is given by

$Y_+ + Y_-$ , which acts like a mass term. This can be seen explicitly by first performing the Grassmann integration

$$Z = \int \frac{d\beta_\pm}{2\pi} \exp\left[N_\pm \left(\log\left(\frac{N_\pm}{i\beta_\pm V}\right) - 1\right) + N_c V \int \frac{d^4k}{(2\pi)^4} \text{tr} \log\left(\not{k} + \gamma_\pm \beta_\pm \frac{k^2 \varphi'^2(k)}{N_c}\right)\right] \quad (188)$$

and then doing the saddle-point integral. Varying with respect to  $\beta = \beta_\pm$  gives the gap equation (Diakonov and Petrov, 1986)

$$\int \frac{d^4k}{(2\pi)^4} \frac{M^2(k)}{k^2 + M^2(k)} = \frac{N}{4N_c V}, \quad (189)$$

where  $M(k) = \beta k^2 \varphi'^2(k)/N_c$  is the momentum-dependent effective quark mass. The gap equation determines the effective constituent mass  $M(0)$  in terms of the instanton density  $N/V$ . For the parameters (159), the effective mass is  $M \approx 350$  MeV. We can expand the gap equation in the instanton density (Pobylitsa, 1989). For small  $N/V$ , one finds  $M(0) \sim \rho(N/2VN_c)^{1/2}$ , which parametrically behaves like the effective mass  $m^*$  introduced above. Note that a dynamic mass is generated for arbitrarily small values of the instanton density. This is expected for  $N_f=1$ , since there is no spontaneous symmetry breaking and the effective mass is generated by the anomaly at the level of an individual instanton.

### 2. The effective interaction for two or more flavors

In the context of QCD, the more interesting case is the one of two or more flavors. For  $N_f=2$ , the effective 't Hooft vertex is a four-fermion interaction

$$Y_\pm = \left[ \prod_{i=1,4} \int \frac{d^4k_i}{(2\pi)^4} k_i \varphi'(k_i) \right] \times (2\pi)^4 \delta^4\left(\sum_i k_i\right) \frac{1}{4(N_c^2 - 1)} \times \left( \frac{2N_c - 1}{2N_c} (\psi^\dagger \gamma_\pm \tau_a^- \psi)^2 + \frac{1}{8N_c} (\psi^\dagger \gamma_\pm \sigma_{\mu\nu} \tau_a^- \psi)^2 \right), \quad (190)$$

where  $\tau_a^- = (\vec{\tau}, i)$  is an isospin matrix and we have suppressed the momentum labels on the quark fields. In the long-wavelength limit  $k \rightarrow 0$ , the 't Hooft vertex (190) corresponds to a local four-quark interaction<sup>34</sup>

$$\mathcal{L} = \beta (2\pi\rho)^4 \frac{1}{4(N_c^2 - 1)} \left( \frac{2N_c - 1}{2N_c} [(\psi^\dagger \tau_a^- \psi)^2] \right)$$

<sup>34</sup>The structure of this interaction is identical to that one given in Eq. (111), as one can check using Fierz identities. The only new ingredient is that the overall constant  $\beta$  is determined self-consistently from a gap equation.

$$+(\psi^\dagger \gamma_5 \tau_a^- \psi)^2] + \frac{1}{4N_c} (\psi^\dagger \sigma_{\mu\nu} \tau_a^- \psi)^2). \quad (191)$$

This Lagrangian is of the type first studied by Nambu and Jona-Lasinio (1961). It is widely used as a model for chiral symmetry breaking and as an effective description for low-energy chiral dynamics (see the reviews of Vogl and Weise, 1991; Klevansky, 1992; Hatsuda and Kunihiro, 1994).

Unlike the Nambu–Jona-Lasinio model, however, the interaction has a natural cutoff parameter  $\Lambda \sim \rho^{-1}$ , and the coupling constants in Eq. (191) are determined in terms of a physical parameter, the instanton density  $(N/V)$ . The interaction is attractive for quark-antiquark pairs with the quantum numbers of the  $\pi$  and  $\sigma$  meson. If the interaction is sufficiently strong, the vacuum is rearranged, quarks condense, and a light (Goldstone) pion is formed. The interaction is repulsive in the pseudoscalar-isoscalar [the SU(2) singlet  $\eta'$ ] and scalar-isovector  $\delta$  channel, showing the effect of the  $U(1)_A$  anomaly. Note that, to first order in the instanton density, there is no interaction in the vector  $\rho, \omega, a_1, f_1$  channels. We shall explore the consequences of this interaction in much greater detail below.

In the case of two (or more) flavors the Grassmann integration cannot be done exactly, since the effective action is more than quadratic in the fermion fields. Instead, we perform the integration over the quark fields in mean-field approximation. This procedure is consistent with the approximations used to derive the effective interaction (185). The mean-field approximation is most easily derived by decomposing fermion bilinears into a constant and a fluctuating part. The integral over the fluctuations is quadratic and can be done exactly. Technically, this can be achieved by introducing auxiliary scalar fields  $L_a, R_a$  into the path integral<sup>35</sup> and then shifting the integration variables in order to linearize the interaction. Using this method, the four-fermion interaction becomes

$$(\psi^\dagger \tau_a^- \gamma_- \psi)^2 \rightarrow 2(\psi^\dagger \tau_a^- \gamma_- \psi) L_a - L_a L_a, \quad (192)$$

$$(\psi^\dagger \tau_a^- \gamma_+ \psi)^2 \rightarrow 2(\psi^\dagger \tau_a^- \gamma_+ \psi) R_a - R_a R_a. \quad (193)$$

In the mean-field approximation, the  $L_a, R_a$  integration can be done using the saddle-point method. Since isospin and parity are not broken, only  $\sigma = L_4 = R_4$  can have a nonzero value. At the saddle point, the free energy  $F = -(1/V) \log Z$  is given by

$$F = 4N_c \int \frac{d^4k}{(2\pi)^4} \log[k^2 + \beta\sigma k^2 \varphi'^2(k)] - 2 \frac{2N_c(N_c^2 - 1)}{2N_c - 1} \beta\sigma^2 - \frac{N}{V} \log\left(\frac{\beta V}{N}\right). \quad (194)$$

<sup>35</sup>In the mean-field approximation, we do not need to introduce auxiliary fields  $T_{\mu\nu}$  in order to linearize the tensor part of the interaction, since  $T_{\mu\nu}$  cannot have a vacuum expectation value.

Varying with respect to  $\beta\sigma$  gives the same gap equation as in the  $N_f=1$  case, now with  $M(k) = \beta\sigma k^2 \varphi'^2(k)$ . We also find  $(N/V) = 2f\sigma^2\beta$  where  $f = 2N_c(N_c^2 - 1)/(2N_c - 1)$ . Expanding everything in  $(N/V)$ , one can show that  $M(0) \sim (N/V)^{1/2}$ ,  $\sigma \sim (N/V)^{1/2}$ , and  $\beta \sim \text{const}$ .

The fact that the gap equation is independent of  $N_f$  is a consequence of the mean-field approximation. It implies that even for  $N_f=2$  chiral symmetry is spontaneously broken for arbitrarily small values of the instanton density. As we shall see in the next section, this is not correct if the full fermion determinant is included. If the instanton density is too small, the instanton ensemble forms a molecular gas, and chiral symmetry is unbroken. However, as we shall show in Sec. V, the mean-field approximation is quite useful for physically interesting values of the instanton density.

The quark condensate is given by

$$\langle \bar{q}q \rangle = -4N_c \int \frac{d^4k}{(2\pi)^4} \frac{M(k)}{M^2(k) + k^2}. \quad (195)$$

Solving the gap equation numerically, we get  $\langle \bar{q}q \rangle \approx -(255 \text{ MeV})^3$ . It is easy to check that  $\langle \bar{q}q \rangle \sim (N/V)^{1/2} \rho^{-1}$ , in agreement with the results obtained in Secs. IV.D and IV.E. The relation (195) was first derived by Diakonov and Petrov using somewhat different techniques see (Sec. VI.B.3).

The procedure for three flavors is very similar, so we do not need to go into detail here. Let us simply quote the effective Lagrangian for  $N_f=3$  (Nowak, Verbaarschot, and Zahed, 1989a),

$$\begin{aligned} \mathcal{L} = & \beta(2\pi\rho)^6 \frac{1}{6N_c(N_c^2 - 1)} \epsilon_{f_1 f_2 f_3} \epsilon_{g_1 g_2 g_3} \\ & \times \left( \frac{2N_c + 1}{2N_c + 4} (\psi_{f_1}^\dagger \gamma_+ \psi_{g_1}) (\psi_{f_2}^\dagger \gamma_+ \psi_{g_2}) (\psi_{f_3}^\dagger \gamma_+ \psi_{g_3}) \right. \\ & + \frac{3}{8(N_c + 2)} (\psi_{f_1}^\dagger \gamma_+ \psi_{g_1}) (\psi_{f_2}^\dagger \gamma_+ \sigma_{\mu\nu} \psi_{g_2}) \\ & \left. \times (\psi_{f_3}^\dagger \gamma_+ \sigma_{\mu\nu} \psi_{g_3}) + (+ \leftrightarrow -) \right), \quad (196) \end{aligned}$$

which was first derived in slightly different form by Shifman *et al.* (1980c). So far, we have neglected the current quark mass dependence and considered the SU( $N_f$ ) symmetric limit. Real QCD is intermediate between the  $N_f=2$  and  $N_f=3$  cases. Flavor mixing in the instanton liquid with realistic values of quark masses was studied by Nowak *et al.* (1989a) to which we refer the reader for more details.

Before we discuss the spectrum of hadronic excitations let us briefly summarize the last three subsections. A random system of instantons leads to spontaneous chiral symmetry breaking. If the system is not only random, but also sufficiently dilute, this phenomenon is most easily studied using the mean-field approximation. We have presented the mean-field approximation in three slightly different versions: one using a schematic model in Sec. IV.D, one using random matrix arguments in Sec. IV.E, and one using an effective quark model in

this section. The results are consistent and illustrate the phenomenon of chiral symmetry breaking in complementary ways. Of course, the underlying assumptions of randomness and diluteness have to be checked. We shall come back to this problem in Sec. V.

### G. Bosonization and the spectrum of pseudoscalar mesons

We have seen that the 't Hooft interaction leads to the spontaneous breakdown of chiral symmetry and generates a strong interaction in the pseudoscalar meson channels. We shall discuss mesonic correlation functions in great detail in Sec. VI and now consider only the consequences for the spectrum of pseudoscalar mesons. The pseudoscalar spectrum is most easily obtained by bosonizing the effective action. In order to describe the  $U(1)_A$  anomaly and the scalar  $\sigma$  channel correctly, one has to allow for fluctuations of the number of instantons. The fluctuation properties of the instanton ensemble can be described by the following "coarse-grained" partition function (Nowak, Verbaarschot, and Zahed, 1989b):

$$S_{\text{eff}} = \frac{b}{4} \int d^4z [n^+(z) + n^-(z)] \left[ \log \left( \frac{n^+(z) + n^-(z)}{n_0} \right) - 1 \right] \\ + \frac{1}{2n_0} \int d^4z [n^+(z) - n^-(z)]^2 \\ + \int d^4z [\psi^\dagger(i\partial + im)\psi - n^+(z)\bar{\Theta}_I(z) \\ - n^-(z)\bar{\Theta}_A(z)], \quad (197)$$

where  $n^\pm(z)$  is the local density of instantons/anti-instantons and  $\bar{\Theta}_{I,A}(z)$  is the 't Hooft interaction (184) averaged over the instanton orientation. This partition function reproduces the low-energy theorem (164) as well as the relation  $\chi_{\text{top}} = (N/V)$  expected for a dilute system in the quenched approximation. In addition to that, the divergence of the flavor-singlet axial current is given by  $\partial_\mu j_\mu^5 = 2N_f [n^+(z) - n^-(z)]$ , consistent with the axial  $U(1)_A$  anomaly.

Again, the partition function can be bosonized by introducing auxiliary fields, linearizing the fermion interaction, and performing the integration over the quarks. In addition to that, we expand the fermion determinant in derivatives of the meson fields in order to recover kinetic terms for the meson fields. This procedure gives the standard nonlinear  $\sigma$ -model Lagrangian. To leading order in the quark masses, the pion and kaon masses satisfy the Gell-Mann, Oakes, Renner relations

$$f_\pi^2 m_\pi^2 = -2m \langle \bar{q}q \rangle, \quad (198)$$

$$f_K^2 m_K^2 = -(m + m_s) \langle \bar{q}q \rangle, \quad (199)$$

with the pion-decay constant

$$f_\pi^2 = 4N_c \int \frac{d^4k}{(2\pi)^4} \frac{M^2(k)}{[k^2 + M^2(k)]^2} \approx (100 \text{ MeV})^2. \quad (200)$$

To this order in the quark masses,  $f_K^2 = f_\pi^2$ . The mass matrix in the flavor singlet and octet sector is more complicated. One finds

$$V = \frac{1}{2} \left( \frac{4}{3} m_K^2 - \frac{1}{3} m_\pi^2 \right) \eta_8^2 + \frac{1}{2} \left( \frac{2}{3} m_K^2 + \frac{1}{3} m_\pi^2 \right) \eta_0^2 \\ + \frac{1}{2} \frac{4\sqrt{2}}{3} (m_\pi^2 - m_K^2) \eta_0 \eta_8 + \frac{N_f}{f_\pi^2} \left( \frac{N}{V} \right) \eta_0^2. \quad (201)$$

The last term gives the anomalous contribution to the  $\eta'$  mass. It agrees with the effective Lagrangian originally proposed by Veneziano (1979) and leads to the Witten-Veneziano relation [with  $\chi_{\text{top}} \approx (N/V)$ ]

$$f_\pi^2 (m_{\eta'}^2 + m_\eta^2 - 2m_K^2) = 2N_f (N/V). \quad (202)$$

Diagonalizing the mass matrix for  $m = 5 \text{ MeV}$  and  $m_s = 120 \text{ MeV}$ , we find  $m_\eta = 527 \text{ MeV}$ ,  $m_{\eta'} = 1172 \text{ MeV}$ , and a mixing angle  $\theta = -11.5^\circ$ . The  $\eta'$  mass is somewhat too heavy, but given the crude assumptions, the result is certainly not bad. One should also note that the result corresponds to an ensemble of uncorrelated instantons. In full QCD, however, the topological susceptibility is zero and correlations between instantons have to be present (see Sec. V.E).

### H. Spin-dependent interactions induced by instantons

The instanton-induced effective interaction between light quarks produces large spin-dependent effects. In this section, we wish to compare these effects with other spin-dependent interactions in QCD and study the effect of instantons on spin-dependent forces in heavy-quark systems. In QCD, the simplest source of spin-dependent effects is the hyperfine interaction generated by the one-gluon exchange potential

$$V_{ij}^{\text{OGE}} = -\frac{\alpha_s}{m_i m_j} \frac{\pi}{6} (\lambda_i^a \lambda_j^a) (\vec{\sigma}_i \vec{\sigma}_j) \delta^3(\vec{r}). \quad (203)$$

This interaction has at least two phenomenologically important features: (a) The  $(\vec{\sigma}_i \vec{\sigma}_j) (\lambda_i^a \lambda_j^a)$  term is twice as large in mesons as it is in baryons, and (b) the ratio of spin-dependent forces in strange and nonstrange system is controlled by the inverse (constituent) quark mass.

For comparison, the nonrelativistic limit of the 't Hooft effective interaction is

$$V_{ij}^{\text{inst}} = -\frac{\pi^2 \rho^2}{6} \frac{(m^*)^2}{m_i^* m_j^*} \left( 1 + \frac{3}{32} (1 + 3 \vec{\sigma}_i \vec{\sigma}_j) \lambda_i^a \lambda_j^a \right) \\ \times \left( \frac{1 - \tau_i^a \tau_j^a}{2} \right) \delta^3(\vec{r}). \quad (204)$$

The spin-dependent part of  $V^{\text{inst}}$  clearly shares the attractive features mentioned above. The dependence on the effective mass comes from having to close two of the external legs in the three-flavor interaction (196). However, there are important differences in the flavor dependence of the one-gluon exchange and instanton interactions. In particular, there is no 't Hooft interaction between quarks of the same flavor ( $uu$ ,  $dd$ , or  $ss$ ). Nev-

ertheless, as shown by Shuryak and Rosner (1989), the potential provides a description of spin splittings in the octet and decuplet that is as good as the one-gluon exchange. The instanton-induced potential has two additional advantages over the one-gluon-exchange potential (Dorokhov *et al.*, 1992). First, we do not have to use an uncomfortably large value of the strong-coupling constant  $\alpha_s$ , and the instanton potential does not have a (phenomenologically unwanted) large spin-orbit part.

In addition to that, instantons provide genuine three-body forces. These forces act only in  $uds$  singlet states, like the flavor singlet  $\Lambda$  [usually identified with the  $\Lambda(1405)$ ] or the hypothetical dilambda (H-dibaryon).<sup>36</sup>

Another interesting question concerns instanton-induced forces between heavy quarks (Callan *et al.*, 1978b). For heavy quarks, the dominant part of the interaction is due to nonzero modes, which we have completely neglected in the discussion above. These effects can be studied using the propagator of an infinitely heavy quark,

$$S(x) = \frac{1 + \gamma_4}{2} \delta^3(\vec{r}) \Theta(\tau) P \exp\left(i \int A_\mu dx_\mu\right), \quad (205)$$

in the field of an instanton. Here,  $P$  denotes a path-ordered integral, and we have eliminated the mass of the heavy quark using a Foldy-Wouthuysen transformation. The phase accumulated by a heavy quark in the field of a single instanton (in singular gauge) is

$$\begin{aligned} U(\vec{r}) &= P \exp\left(i \int_{-\infty}^{\infty} A_4 dx_4\right) \\ &= \cos[F(r)] + i\vec{r} \cdot \vec{\tau} \sin[F(r)], \quad \text{with} \\ F(r) &= \pi \left(1 - \frac{r}{\sqrt{r^2 + \rho^2}}\right), \end{aligned} \quad (206)$$

where  $\vec{r}$  is the spatial distance between the instanton and the heavy quark, and  $r = |\vec{r}|$ . From this result, we can determine the mass renormalization of the heavy quark due to the interaction with instantons in the dilute-gas approximation (Callan *et al.*, 1978b; Diakonov *et al.*, 1989; Chernyshev, Nowak, and Zahed, 1996):

$$\Delta M_Q = \frac{16\pi}{N_c} \left(\frac{N}{V}\right) \rho^3 \times 0.552 \approx 70 \text{ MeV}. \quad (207)$$

In a similar fashion, one can determine the spin-independent part of the heavy-quark potential (for color singlet  $\bar{q}q$  systems),

$$\begin{aligned} V_{QQ}(\vec{r}_i, \vec{r}_j) &= \int d\rho n(\rho) \int d^3r \frac{1}{3} \text{Tr}[U(\vec{r}_i - \vec{r}) \\ &\quad \times U^\dagger(\vec{r}_j - \vec{r}) - 1]. \end{aligned} \quad (208)$$

The potential (208) rises quadratically at short distance but levels off at a value of  $2\Delta M_Q$  for  $r > 5\rho$ . This is a

<sup>36</sup>Takeuchi and Oka (1991) found that instanton-induced forces made the H unbound, a quite welcome conclusion, since so far it has eluded all searches.

reflection of the fact that dilute instantons do not confine. Also, the magnitude of the potential is on the order of 100 MeV, too small to be of much importance in charmonium or bottomonium systems. The spin-dependent part of the heavy-quark potential is

$$V_{QQ}^{ss}(\vec{r}_i, \vec{r}_j) = -\frac{1}{4M_i M_j} (\vec{\sigma}_i \cdot \vec{\nabla}_i) (\vec{\sigma}_j \cdot \vec{\nabla}_j) V(\vec{r}_i - \vec{r}_j), \quad (209)$$

and it, too, is too small to be important phenomenologically. More important is the instanton-induced interaction in heavy-light systems. This problem was studied in some detail by Chernyshev *et al.* (1996). The effective potential between the heavy and the light quark is given by

$$\begin{aligned} V_{qQ}(\vec{r}) &= \frac{\Delta M_Q m_q^*}{2(N/V)N_c} \left[ \left(1 + \frac{\lambda_Q^a \lambda_q^a}{4}\right) \right. \\ &\quad \left. - \frac{\Delta M_Q^{\text{spin}}}{\Delta M_Q} \vec{\sigma}_q \cdot \vec{\sigma}_Q \frac{\lambda_Q^a \lambda_q^a}{4} \right] \delta^3(\vec{r}), \end{aligned} \quad (210)$$

where  $\Delta M_Q$  is the mass renormalization of the heavy quark and

$$\Delta M_Q^{\text{spin}} = \frac{16\pi}{N_c} \left(\frac{N}{V}\right) \frac{\rho^2}{M_Q} \times 0.193 \quad (211)$$

controls the hyperfine interaction. This interaction gives very reasonable results for spin splittings in heavy-light mesons. In addition to that, instantons generate many-body forces which might be important in heavy-light baryons. We conclude that instantons can account for spin-dependent forces required in light-quark spectroscopy without the need for large hyperfine interactions. Instanton-induced interactions are not very important in heavy-quark systems, but may play a role in heavy-light systems.

## V. THE INTERACTING INSTANTON LIQUID

### A. Numerical simulations

In the last section we discussed an analytic approach to the statistical mechanics of the instanton ensemble based on the mean-field approximation. This approach provides important insights into the structure of the instanton ensemble and the qualitative dependence on the interaction. However, the method ignores correlations among instantons, which are important for a number of phenomena, such as topological charge screening (Sec. V.E), chiral symmetry restoration (Sec. VII.B), and hadronic correlation functions (Sec. VI).

In order to go beyond the mean-field approximation and study the instanton liquid with the 't Hooft interaction included to all orders, we have performed numerical simulations of the interacting instanton liquid (Shuryak, 1988a; Shuryak and Verbaarschot, 1990; Nowak *et al.*, 1989a; Schäfer and Shuryak, 1996a). In these simulations, we make use of the fact that the quantum-field-theory problem is analogous to the statistical, mechanics of a system of pseudoparticles in four

dimensions. The distribution of the  $4N_c N$  collective coordinates associated with a system of  $N$  pseudoparticles can be studied using standard Monte Carlo techniques (e.g., the Metropolis algorithm), originally developed for simulations of statistical systems containing atoms or molecules.

These simulations have a number of similarities to lattice simulations of QCD [see the textbooks of Creutz (1983), Rothe (1992), and Montvay and Münster (1995)]. As in lattice gauge theory, we consider systems in a finite four-dimensional volume, subject to periodic boundary conditions. This means that both approaches share finite-size problems, especially the difficulty of working with realistic quark masses. Both methods are also formulated in Euclidean space, which means that it is difficult to extract real-time (in particular nonequilibrium) information. However, in contrast to the lattice, spacetime is continuous, so we have no problems with chiral fermions. Furthermore, the number of degrees is drastically reduced, and meaningful (unquenched) simulations of the instanton ensemble can be performed in a few minutes on an average workstation. Finally, using the analogy with an interacting liquid, it is easier to develop an intuitive understanding of the numerical simulations.

The instanton ensemble is defined by the partition function

$$Z = \sum_{N_+, N_-} \frac{1}{N_+! N_-!} \int \prod_i^{N_+ + N_-} [d\Omega_i n(\rho_i)] \times \exp(-S_{\text{int}}) \prod_f^{N_f} \det(\mathcal{D} + m_f), \quad (212)$$

describing a system of pseudoparticles interacting via the bosonic action and the fermionic determinant. Here,  $d\Omega_i = dU_i d^4 z_i d\rho_i$  is the measure in the space of collective coordinates (color orientation, position, and size) associated with a single instanton, and  $n(\rho)$  is the single-instanton density [Eq. (93)].

The gauge interaction between instantons is approximated by a sum of pure two-body interactions  $S_{\text{int}} = \frac{1}{2} \sum_{I \neq J} S_{\text{int}}(\Omega_{IJ})$ . Genuine three-body effects in the instanton interaction are not important as long as the ensemble is reasonably dilute. This part of the interaction is fairly easy to deal with. The computational effort is similar to that for a statistical system with long-range forces.

The fermion determinant, however, introduces nonlocal interactions among many instantons. Changing the coordinates of a single instanton requires the calculation of the full  $N$ -instanton determinant, not just  $N$  two-body interactions. Evaluating the determinant exactly is a quite formidable problem. In practice we proceed as in Sec. IV.F and factorize the determinant into a low- and a high-momentum part,

$$\det(\mathcal{D} + m_f) = \left( \prod_i^{N_+ + N_-} 1.34 \rho_i \right) \det(T_{\text{IA}} + m_f), \quad (213)$$

where the first factor, the high-momentum part, is the product of contributions from individual instantons calculated in the Gaussian approximation, whereas the low-momentum part associated with the fermionic zero modes of individual instantons is calculated exactly.  $T_{\text{IA}}$  is the  $N_+ \times N_-$  matrix of overlap matrix elements introduced in Sec. IV.A.2. As emphasized before, this determinant contains the 't Hooft interaction to all orders.

In practice, it is simpler to study the instanton ensemble for a fixed particle number  $N = N_+ + N_-$ . This means that instead of the grand-canonical partition function (212), we consider the canonical ensemble

$$Z_N = \frac{1}{N_+! N_-!} \prod_i^{N_+ + N_-} [d\Omega_i n(\rho_i)] \times \exp(-S_{\text{int}}) \prod_f^{N_f} \det(\mathcal{D} + m_f) \quad (214)$$

for different densities and determine the ground state by minimizing the free energy  $F = -1/V \log Z_N$  of the system. Furthermore, we shall consider only ensembles with no net topology. The two constraints  $N/V = \text{const}$  and  $Q = N_+ - N_- = 0$  do not affect the results in the thermodynamic limit. The only exceptions are of course fluctuations of  $Q$  and  $N$ , the topological susceptibility and the compressibility of the instanton liquid, respectively. In order to study these quantities, one has to consider appropriately chosen subsystems (see Sec. V.E).

In order to simulate the partition function (214), we generate a sequence  $\{\Omega_i\}_j$  ( $i = 1, \dots, N$ ;  $j = 1, \dots, N_{\text{conf}}$ ) of configurations according to the weight function  $p(\{\Omega_i\}) \sim \exp(-S)$ , where

$$S = - \sum_{i=1}^{N_+ + N_-} \log[n(\rho_i)] + S_{\text{int}} + \text{tr} \log(\mathcal{D} + m_f) \quad (215)$$

is the total action of the configuration. This is most easily accomplished using the Metropolis algorithm: a new configuration is generated using some microreversible procedure  $\{\Omega_i\}_j \rightarrow \{\Omega_i\}_{j+1}$ . The configuration is always accepted if the new action is smaller than the old one, and it is accepted with the probability  $\exp(-\Delta S)$  if the new action is larger. Alternatively, one can generate the ensemble using other techniques, e.g., the Langevin (Nowak *et al.*, 1989a), heat bath, or microcanonical methods.

## B. The free energy of the instanton ensemble

Using the sequence of configurations generated by the Metropolis algorithm, it is straightforward to determine expectation values by averaging measurements in many configurations,

$$\langle \mathcal{O} \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mathcal{O}(\{\Omega_i\}_j). \quad (216)$$

This is how the quark and gluon condensates, as well as the hadronic correlation functions discussed in this and the following section, have been determined. However,

more work is required to determine the partition function, which gives the overall normalization of the instanton distribution. The knowledge of the partition function is necessary in order to calculate the free energy and the thermodynamics of the system. In practice, the partition function is most easily evaluated using the thermodynamic integration method (Kirkwood, 1931). For this purpose we write the total action as

$$S(\alpha) = S_0 + \alpha \Delta S, \quad (217)$$

which interpolates between a solvable action  $S_0$  and the full action  $S(\alpha=1) = S_0 + \Delta S$ . If the partition function for the system governed by the action  $S_0$  is known, the full partition function can be determined from

$$\log Z(\alpha=1) = \log Z(\alpha=0) - \int_0^1 d\alpha' \langle 0 | \Delta S | 0 \rangle_{\alpha'}, \quad (218)$$

where the expectation value  $\langle 0 | \cdot | 0 \rangle_{\alpha}$  depends on the coupling constant  $\alpha$ . The obvious choice for decomposing the action of the instanton liquid would be to use the single-instanton action,  $S_0 = \sum_i \log[n(\rho_i)]$ , but this does not work since the  $\rho$  integration in the free partition function is not convergent. We therefore consider the following decomposition:

$$S(\alpha) = \sum_{i=1}^{N_+ + N_-} \left( -\log[n(\rho_i)] + (1-\alpha) \nu \frac{\rho_i^2}{\bar{\rho}^2} \right) + \alpha (S_{\text{int}} + \text{tr} \log(\mathcal{D} + m_f)), \quad (219)$$

where  $\nu = (b-4)/2$  and  $\bar{\rho}^2$  is the average size squared of the instantons with the full interaction included. The  $\rho_i^2$  term serves to regularize the size integration for  $\alpha=0$ . It does not affect the final result for  $\alpha=1$ . The specific form of this term is irrelevant; our choice here is motivated by the fact that  $S(\alpha=0)$  gives a single-instanton distribution with the correct average size  $\bar{\rho}^2$ . The  $\alpha=0$  partition function corresponds to the variational single-instanton distribution

$$Z_0 = \frac{1}{N_+! N_-!} (V \mu_0)^{N_+ + N_-}, \quad \mu_0 = \int_0^\infty d\rho n(\rho) \exp\left(-\nu \frac{\rho^2}{\bar{\rho}^2}\right), \quad (220)$$

where  $\mu_0$  is the normalization of the one-body distribution. The full partition function obtained from integrating over the coupling  $\alpha$  is

$$\log Z = \log(Z_0) + N \int_0^1 d\alpha' \times \langle 0 | \nu \frac{\rho^2}{\bar{\rho}^2} - \frac{1}{N} [S_{\text{int}} + \text{tr} \log(\mathcal{D} + m_f)] | 0 \rangle_{\alpha'}, \quad (221)$$

where  $N = N_+ + N_-$ . The free-energy density is finally given by  $F = -(1/V) \log Z$ , where  $V$  is the four-volume of the system. The pressure and the energy density are related to  $F$  by  $p = -F$  and  $\epsilon = T dp/dT - p$ .

### C. The instanton ensemble

If correlations among instantons are important, the variational (or mean-field) partition function  $Z_0$  is not expected to provide an accurate estimate for the partition function. This is certainly the case in the presence of light fermions, in particular at finite temperature. In this section we want to present numerical results obtained from simulations of the instanton liquid at zero temperature.

As discussed in Sec. IV.A.1, a general problem in the interacting instanton model is the treatment of very close instanton anti-instanton pairs. In practice, we have decided to deal with this difficulty by introducing a phenomenological short-range repulsive core,

$$S_{\text{core}} = \frac{8\pi^2 A}{g^2} \frac{1}{\lambda^4} |u|^2, \quad \lambda = \frac{R^2 + \rho_I^2 + \rho_A^2}{2\rho_I \rho_A} + \left( \frac{(R^2 + \rho_I^2 + \rho_A^2)^2}{4\rho_I^2 \rho_A^2} - 1 \right)^{1/2}, \quad (222)$$

into the streamline interaction. Here,  $\lambda$  is the conformal parameter (144) and  $A$  controls the strength of the core. This parameter essentially governs the dimensionless diluteness  $f = \rho^4(N/V)$  of the ensemble. The second parameter of the instanton liquid is the scale  $\Lambda_{\text{QCD}}$  in the instanton size distribution, which fixes the absolute units.

We have defined the scale parameter by fixing the instanton density to be  $N/V = 1 \text{ fm}^{-4}$ . This means that, in our units, the average distance between instantons is 1 fm by definition. Alternatively, one can proceed as in lattice gauge simulations and use an observable such as the  $\rho$  meson mass to set the scale. Using  $N/V$  is very convenient and, as we shall see in the next section, using the  $\rho$  or nucleon mass would not make much of a difference. We use the same scale-setting procedure for all QCD-like theories, independent of  $N_c$  and  $N_f$ . This provides a simple prescription for comparing dimensional quantities in theories with different matter content.

The remaining free parameter is the (dimensionless) strength of the core  $A$ , which determines the (dimensionless) diluteness of the ensemble. In Schäfer and Shuryak (1996a), we chose  $A = 128$ , which gives  $f = \bar{\rho}^4(N/V) = 0.12$  and  $\bar{\rho} = 0.43 \text{ fm}$ . As a result, the ensemble is not quite as dilute as phenomenology seems to demand [ $(N/V) = 1 \text{ fm}^{-4}$  and  $\bar{\rho} = 0.33 \text{ fm}$ ] but comparable to the lattice result  $(N/V) = (1.4-1.6) \text{ fm}^{-4}$  and  $\bar{\rho} = 0.35 \text{ fm}$  (Chu *et al.*, 1994). The average instanton action is  $S \approx 6.4$ , while the average interaction is  $S_{\text{int}}/N \approx 1.0$ , showing that the system is still semiclassical and that interactions among instantons are important, but not dominant.

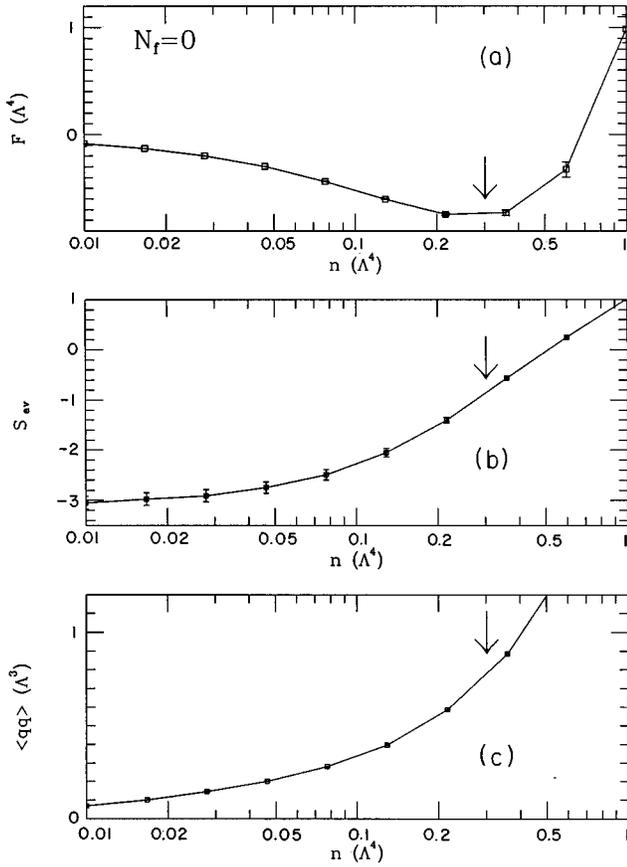


FIG. 15. Free energy, average instanton action, and quark condensate as a function of the instanton density in the pure gauge theory, from Schäfer and Shuryak, 1996a. All dimensional quantities are given in units of the scale parameter  $\Lambda_{\text{QCD}}$ .

Detailed simulations of the instanton ensemble in QCD are discussed by Schäfer and Shuryak (1996a). As an example, we show the free energy versus the instanton density in pure gauge theory (without fermions) in Fig. 15. At small density the free energy is roughly proportional to the density, but at larger densities repulsive interactions become important, leading to a well-defined minimum. We also show the average action per instanton as a function of density. The average action controls the probability  $\exp(-S)$  to find an instanton, but has no minimum in the range of densities studied. This shows that the minimum in the free energy is a compromise between maximum entropy and minimum action.

Fixing the units such that  $N/V=1 \text{ fm}^{-4}$ , we have  $\Lambda=270 \text{ MeV}$ , and the vacuum energy density generated by instantons is  $\epsilon=-526 \text{ MeV/fm}^3$ . We have already stressed that the vacuum energy is related to the gluon condensate by the trace anomaly. Estimating the gluon condensate from the instanton density, we have  $\epsilon=-b/4(N/V)=-565 \text{ MeV/fm}^3$ , which is in good agreement with the direct determination of the energy density. Not only the depth of the free energy, but also its curvature (the instanton compressibility) is fixed from the low-energy theorem (162). The compressibility determined from Fig. 15 is  $3.2(N/V)^{-1}$ , which can be compared with  $2.75(N/V)^{-1}$  from the low-energy theorem.

At the minimum of the free energy we can also determine the quark condensate [see Fig. 15(c)]. In quenched QCD, we have  $\langle \bar{q}q \rangle = -(251 \text{ MeV})^3$ , while a similar simulation in full QCD gives  $\langle \bar{q}q \rangle = -(216 \text{ MeV})^3$ , in good agreement with the phenomenological value.

#### D. Dirac spectra

We have already emphasized that the distribution of eigenvalues of the Dirac operator  $i\mathcal{D}\psi_\lambda = \lambda\psi_\lambda$  is of great interest for many phenomena in QCD. In this section, we wish to study the spectral density of  $i\mathcal{D}$  in the instanton liquid for different numbers of flavors. Before we present the results, let us review a few general arguments. First, since the weight function contains the fermion determinant  $\det_f(i\mathcal{D}) = (\prod_i \lambda_i)^{N_f}$ , it is clear that small eigenvalues will be suppressed if the number of flavors is increased. This can also be seen from the Smilga-Stern result [Eq. (181)]. For  $N_f=2$  the spectral density at  $\lambda=0$  is flat, while for  $N_f>2$  the slope of spectrum is positive.

In general, one might therefore expect that there is a critical number of flavors (smaller than  $N_f=17$ , where asymptotic freedom is lost) for which chiral symmetry is restored. There are a number of arguments that this indeed happens in non-Abelian gauge theories (see Sec. IX.D). Let us only mention the simplest one here. The number of quark degrees of freedom is  $N_q=4N_cN_f$ , while, if chiral symmetry is broken, the number of low-energy degrees of freedom (“pions”) is  $N_\pi=N_f^2-1$ . If chiral symmetry is still broken (for  $N_f>12$ , this leads to the unusual situation that the effective number of degrees of freedom at low energy is larger than the number of elementary degrees of freedom at high energy. In this case it is hard to see how one could ever have a transition to a phase of weakly interacting quarks and gluons, since the pressure of the low-temperature phase is always bigger.

In Fig. 16 we show the spectrum of the Dirac operator in the instanton liquid for  $N_f=0,1,2,3$  light flavors (Verbaarschot, 1994a). Clearly, the results are qualitatively consistent with the Smilga-Stern theorem<sup>37</sup> for  $N_f \geq 2$ . In addition to that, the trend continues for  $N_f < 2$ , where the result is not applicable. We also note that for  $N_f=3$  (massless) flavors, a gap starts to open up in the spectrum. In order to check whether this gap indicates chiral symmetry restoration in the infinite-volume limit, one has to investigate finite-size scaling. The problem was studied in more detail by Schäfer and Shuryak (1996a), who concluded that chiral symmetry was restored in the instanton liquid between  $N_f=4$  and  $N_f=5$ . Another interesting problem is the dependence on the dynamic quark mass in the chirally restored phase  $N_f > N_f^{\text{crit}}$ . If the quark mass is increased, the influence of the fermion determinant is reduced, and eventually

<sup>37</sup>Numerically, the slope in the  $N_f=3$  spectrum appears to be too large, but it is not clear how small  $\lambda$  has to be for the theorem to be applicable.

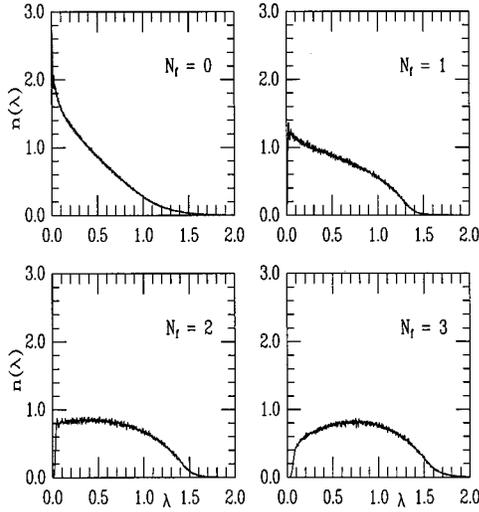


FIG. 16. Spectrum of the Dirac operator for different values of the number of flavors  $N_f$ , from Verbaarschot, 1994b. The eigenvalue is given in units of the scale parameter  $\Lambda_{\text{QCD}}$  and the distribution function is normalized to one.

“spontaneous” symmetry breaking is recovered. As a consequence, QCD has an interesting phase structure as a function of the number of flavors and their masses, even at zero temperature.

### E. Screening of the topological charge

Another interesting phenomenon associated with dynamical quarks is topological charge screening. This effect is connected with properties of the  $\eta'$  meson, strong  $CP$  violation, and the structure of QCD at finite  $\theta$  angle.

Topological charge screening can be studied in a number of complementary ways. Historically, it was first discussed on the basis of Ward identities associated with the anomalous  $U(1)_A$  symmetry (Veneziano, 1979). Let us consider the flavor singlet correlation function  $\Pi_{\mu\nu} = \langle \bar{q} \gamma_\mu \gamma_5 q(x) \bar{q} \gamma_\nu \gamma_5 q(0) \rangle$ . Taking two derivatives and using the anomaly relation (95), we can derive the low-energy theorem

$$\chi_{\text{top}} = \int d^4x \langle Q(x) Q(0) \rangle = -\frac{m \langle \bar{q} q \rangle}{2N_f} + \frac{m^2}{4N_f^2} \int d^4x \langle \bar{q} \gamma_5 q(x) \bar{q} \gamma_5 q(0) \rangle. \quad (223)$$

Since the correlation function on the right-hand side does not have any massless poles in the chiral limit, the topological susceptibility  $\chi_{\text{top}} \sim m$  as  $m \rightarrow 0$ . More generally,  $\chi_{\text{top}}$  vanishes if there is at least one massless quark flavor.

Alternatively, we can use the fact that the topological susceptibility is the second derivative of the vacuum energy with respect to the  $\theta$  angle. Writing the QCD partition function as a sum over all topological sectors and extracting the zero modes from the fermion determinant, we have

$$Z = \sum_\nu e^{i\theta\nu} \int_\nu dA_\mu e^{-S} \det_f(\mathcal{D} + m_f) = \sum_\nu (e^{i\theta} \det \mathcal{M})^\nu \int_\nu dA_\mu e^{-S} \prod'_{n,f} (\lambda_n^2 + m_f^2). \quad (224)$$

Here,  $\nu$  is the winding number of the configuration,  $\mathcal{M}$  is the mass matrix, and  $\Pi'$  denotes the product of all eigenvalues with the zero modes excluded. The result shows that the partition function depends on  $\theta$  only through the combination  $e^{i\theta} \det \mathcal{M}$ , so the vacuum energy is independent of  $\theta$  if one of the quark masses vanishes.

The fact that  $\chi_{\text{top}}$  vanishes implies that fluctuations in the topological charge are suppressed, so instantons, and anti-instantons have to be correlated. Every instanton is surrounded by a cloud of anti-instantons which completely screens its topological charge, analogous to Debye screening in ordinary plasmas. This fact can be seen most easily by using the bosonized effective Lagrangian (201). For simplicity we consider instantons to be point-like, but, in contrast to the procedure in Sec. IV.G, we do allow the positions of the pseudoparticles to be correlated. The partition function is given by Nowak *et al.*, 1989b; Kikuchi and Wudka, 1992; Dowrick and McDougall, 1993; and Shuryak and Verbaarschot, 1995 as

$$Z = \sum_{N_+, N_-} \frac{\mu_0^{N_+ + N_-}}{N_+! N_-!} \prod_i d^4z_i \exp(-S_{\text{eff}}), \quad (225)$$

where  $\mu_0$  is the single-instanton normalization (153) and the effective action is given by

$$S_{\text{eff}} = i \int d^4x \frac{\sqrt{2N_f}}{f_\pi} \eta_0 Q + \int d^4x \mathcal{L}(\eta_0, \eta_8). \quad (226)$$

Here, the topological charge density is  $Q(x) = \Sigma Q_i \delta(x - z_i)$ , and  $\mathcal{L}(\eta_0, \eta_8)$  is the nonanomalous part of the pseudoscalar meson Lagrangian with the mass terms given in Eq. (201). We can perform the sum in Eq. (225) and, keeping only the quadratic terms, integrate out the meson fields and determine the topological charge correlator in this model. The result is

$$\langle Q(x) Q(0) \rangle = \left( \frac{N}{V} \right) \left\{ \delta^4(x) - \frac{2N_f N}{f_\pi^2 V} [\cos^2(\phi) D(m_{\eta'}, x) + \sin^2(\phi) D(m_{\eta, x})] \right\}, \quad (227)$$

where  $D(m, x) = m/(4\pi^2 x) K_1(mx)$  is the (Euclidean) propagator of a scalar particle and  $\phi$  is the  $\eta - \eta'$  mixing angle. The correlator (227) has an obvious physical interpretation. The local terms is the contribution from a single instanton located at the center, while the second term is the contribution from the screening cloud. One can easily check that the integral of the correlator is of order  $m_\pi^2$ , so  $\chi_{\text{top}} \sim m$  in the chiral limit. We also observe that the screening length is given by the mass of the  $\eta'$ .

Detailed numerical studies of topological charge screening in the interacting instanton model were performed by Shuryak and Verbaarschot (1995). The authors verified that complete screening took place if one of the quark masses went to zero and that the screening length was consistent with the  $\eta'$  mass. They also addressed the question of how the  $\eta'$  mass could be extracted from topological charge fluctuations. The main idea was not to study the limiting value of  $\langle Q^2 \rangle / V$  for large volumes but to determine its dependence on  $V$  for small volumes  $V < 1 \text{ fm}^4$ . In this case, one has to worry about possible surface effects. It is therefore best to consider the topological charge in a segment  $H(l_4) = l_4 \times L^3$  of the torus  $L^4$  (a hypercube with periodic boundary conditions). This construction ensures that the surface area of  $H(l_4)$  is independent of its volume. Using the effective meson action introduced above, we expect (in the chiral limit)

$$K_P(l_4) \equiv \langle Q(l_4)^2 \rangle = L^3 \left( \frac{N}{V} \right) \frac{1}{m_{\eta'}} (1 - e^{-m_{\eta'} l_4}). \quad (228)$$

Numerical results for  $K_P(l_4)$  are shown in Fig. 17. The solid line shows the result for a random system of instantons with a finite topological susceptibility  $\chi_{\text{top}} \simeq (N/V)$ , and the dashed curve is a fit using the parametrization (228). Again, we clearly observe topological charge screening. Furthermore, the  $\eta'$  mass extracted from the fit is  $m_{\eta'} = 756 \text{ MeV}$  (for  $N_f = 2$ ), quite consistent with what one would expect. The figure also shows the behavior of the scalar correlation function, related to the compressibility of the instanton liquid. The instanton number  $N = N_+ + N_-$  is of course not screened, but the fluctuations in  $N$  are reduced by a factor  $4/b$  due to the interactions [see Eq. (162)]. For a more detailed analysis of the correlation function, see Shuryak and Verbaarschot (1995).

We conclude that the topological charge in the instanton liquid is completely screened in the chiral limit. The  $\eta'$  mass is not determined by the topological susceptibility, but by fluctuations of the charge in small subvolumes.

## VI. HADRONIC CORRELATION FUNCTIONS

### A. Definitions and generalities

In a relativistic field theory, current correlation functions determine the spectrum of hadronic resonances. In addition to that, hadronic correlation functions provide a bridge between hadronic phenomenology, the underlying quark-gluon structure, and the structure of the QCD vacuum. The available theoretical and phenomenological information was recently reviewed by Shuryak (1993), so we give only a brief overview here.

In the following, we consider hadronic point-to-point correlation functions,

$$\Pi_h(x) = \langle 0 | j_h(x) j_h(0) | 0 \rangle. \quad (229)$$

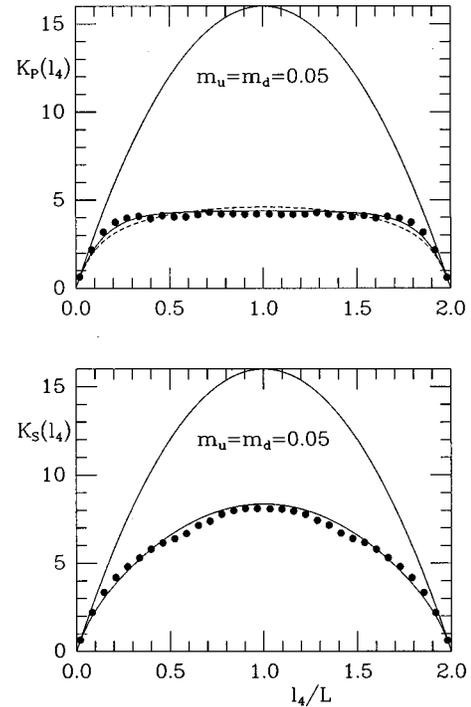


FIG. 17. Pseudoscalar and scalar correlators  $K_{P,S}(l_4)$  as a function of the length  $l_4$  of the subvolume  $l_4 \times L^3$ , from Shuryak and Verbaarschot, 1995. Screening implies that the correlator depends only on the surface, not on the volume of the torus. This means that in the presence of screening, the correlator goes to a constant. The results were obtained for  $N_c = 3$  and  $m_u = m_d = 10 \text{ MeV}$  and  $m_s = 150 \text{ MeV}$ . The upper solid line corresponds to a random system of instantons, while the other solid line shows the parametrization discussed in the text (the dashed line in the upper panel shows a slightly more sophisticated parametrization). Note the qualitative difference between the data for topological (upper panel) and number fluctuations (lower panel). The topological charge correlator is flat, corresponding to charge screening, while the number fluctuations are only reduced in size as compared to the random ensemble.

Here,  $j_h(x)$  is a local operator with the quantum numbers of a hadronic state  $h$ . We shall concentrate on mesonic and baryonic currents of the type

$$j_{mes}(x) = \delta^{ab} \bar{\psi}^a(x) \Gamma \psi^b(x), \quad (230)$$

$$j_{bar}(x) = \epsilon^{abc} [\psi^a T(x) C \Gamma \psi^b(x)] \Gamma' \psi^c(x). \quad (231)$$

Here,  $a, b, c$  are color indices and  $\Gamma, \Gamma'$  are isospin and Dirac matrices. At zero temperature, we shall focus exclusively on correlators for spacelike (or Euclidean) separation  $\tau = \sqrt{-x^2}$ . The reason is that spacelike correlators are exponentially suppressed rather than oscillatory at large distance. In addition, most theoretical approaches, like the operator product expansion (OPE), lattice calculations, or the instanton model deal with Euclidean correlators.

Hadronic correlation functions are completely determined by the spectrum (and the coupling constants) of the physical excitations with the quantum numbers of the current  $j_h$ . For a scalar correlation function, we have

the standard dispersion relation

$$\begin{aligned} \Pi(Q^2) = & \frac{(-Q^2)^n}{\pi} \int ds \frac{\text{Im } \Pi(s)}{s^n(s+Q^2)} \\ & + a_0 + a_1 Q^2 + \dots, \end{aligned} \quad (232)$$

where  $Q^2 = -q^2$  is the Euclidean momentum transfer and we have indicated possible subtraction constants  $a_i$ . The spectral function  $\rho(s) = (1/\pi)\text{Im } \Pi(s)$  can be expressed in terms of physical states,

$$\begin{aligned} \rho(s = -q^2) = & (2\pi)^3 \sum_n \delta^4(q - q_n) \langle 0 | j_h(0) | n \rangle \\ & \times \langle n | j_h^\dagger(0) | 0 \rangle, \end{aligned} \quad (233)$$

where  $|n\rangle$  is a complete set of hadronic states. Correlation functions with nonzero spin can be decomposed into Lorentz-covariant tensors and scalar functions. Fourier-transforming Eq. (232) gives a spectral representation of the coordinate-space correlation function

$$\Pi(\tau) = \int ds \rho(s) D(\sqrt{s}, \tau). \quad (234)$$

Here,  $D(m, \tau)$  is the Euclidean propagator of a scalar particle with mass  $m$ ,

$$D(m, \tau) = \frac{m}{4\pi^2 \tau} K_1(m\tau). \quad (235)$$

Note that, except for possible contact terms, subtraction constants do not affect the coordinate-space correlator. For large arguments, the correlation function decays exponentially,  $\Pi(\tau) \sim \exp(-m\tau)$ , where the decay is governed by the lowest pole in the spectral function. This is the basis of hadronic spectroscopy on the lattice. In practice, however, lattice simulations often rely on more complicated sources in order to improve the suppression of excited states. While useful in spectroscopy, these correlation functions mix short- and long-distance effects and are not as interesting theoretically.

Correlation functions of currents built from quark fields only (like the meson and baryon currents introduced above) can be expressed in terms of the full quark propagator. For an isovector meson current  $j_{I=1} = \bar{u}\Gamma d$  (where  $\Gamma$  is only a Dirac matrix), the correlator is given by the ‘‘one-loop’’ term

$$\Pi_{I=1}(x) = \langle \text{Tr}[S^{ab}(0,x)\Gamma S^{ba}(x,0)\Gamma] \rangle. \quad (236)$$

The averaging is performed over all gauge configurations, with the weight function  $\det(\mathcal{D} + m)\exp(-S)$ . Note that the quark propagator is not translation invariant before the vacuum average is performed, so the propagator depends on both arguments. Also note that while Eq. (236) has the appearance of a one-loop (perturbative) graph, it includes arbitrarily complicated, multi-

loop, gluon exchanges as well as nonperturbative effects. All of these effects are hidden in the vacuum average. Correlators of isosinglet meson currents  $j_{I=0} = (1/\sqrt{2})(\bar{u}\Gamma u + \bar{d}\Gamma d)$  receive an additional two-loop, or disconnected, contribution

$$\begin{aligned} \Pi_{I=0}(x) = & \langle \text{Tr}[S^{ab}(0,x)\Gamma S^{ba}(x,0)\Gamma] \rangle \\ & - 2\langle \text{Tr}[S^{aa}(0,0)\Gamma] \text{Tr}[S^{bb}(x,x)\Gamma] \rangle. \end{aligned} \quad (237)$$

In an analogous fashion, baryon correlators can be expressed as vacuum averages of three-quark propagators.

At short distance, asymptotic freedom implies that the correlation functions are determined by free-quark propagation. The free-quark propagator is given by

$$S_0(x) = \frac{i}{2\pi^2} \frac{\gamma \cdot x}{x^4}. \quad (238)$$

This means that mesonic and baryonic correlation functions at short distance behave as  $\Pi_{\text{mes}} \sim 1/x^6$  and  $\Pi_{\text{bar}} \sim 1/x^9$ , respectively. Deviations from asymptotic freedom at intermediate distances can be studied using the operator product expansion. The basic idea (Wilson, 1969) is to expand the product of currents in Eq. (229) into a series of coefficient functions  $c_n(x)$  multiplied by local operators  $\mathcal{O}_n(0)$

$$\Pi(x) = \sum_n c_n(x) \langle \mathcal{O}_n(0) \rangle. \quad (239)$$

From dimensional considerations it is clear that the most singular contributions correspond to operators of the lowest possible dimension. Ordinary perturbative contributions are contained in the coefficient of the unit operator. The leading nonperturbative corrections are controlled by the quark and gluon condensates of dimension three and four.

In practice, the OPE for a given correlation function is most easily determined using the short-distance expansion of the propagator in external, slowly varying quark and gluon fields (Novikov, *et al.*, 1985b),<sup>38</sup>

$$\begin{aligned} S_f^{ab}(x) = & \frac{i\delta^{ab}}{2\pi^2} \frac{\gamma \cdot x}{x^4} - \frac{\delta^{ab}}{4\pi^2} \frac{m_f}{x^2} + q^a(0)\bar{q}^b(0) \\ & - \frac{i}{32\pi^2} (\lambda^k)^{ab} G_{\alpha\beta}^k \frac{\gamma \cdot x \sigma_{\alpha\beta} + \sigma_{\alpha\beta} \gamma \cdot x}{x^2} + \dots. \end{aligned} \quad (240)$$

The corrections to the free propagator have an obvious interpretation in terms of the interaction of the quark with the external quark and gluon fields. There is an enormous literature about QCD sum rules based on the OPE. See the reviews of Reinders *et al.* (1985), Narison

<sup>38</sup>This expression was derived in the Fock-Schwinger gauge  $x_\mu A_\mu^a = 0$ . The resulting hadronic correlation functions are of course gauge invariant.

(1989), and Shifman (1992). The general idea is easily explained. If there is a window in which the OPE [Eq. (239)] has reasonable accuracy and the spectral representation [Eq. (234)] is dominated by the ground state, one can match the two expressions in order to extract ground-state properties. In general, the two requirements are in conflict with each other, so the existence of a sum-rule window has to be established in each individual case.

The main sources of phenomenological information about the correlation functions are summarized as follows (Shuryak, 1993):

- (1) Ideally, the spectral function is determined from an experimentally measured cross section using the optical theorem. This is the case, for example, in the vector-isovector (rho-meson) channel, where the necessary input is provided by the ratio

$$R(s) = \frac{\sigma(e^+e^- \rightarrow (I=1 \text{ hadrons}))}{\sigma(e^+e^- \rightarrow \mu^+\mu^-)}, \quad (241)$$

where  $s$  is the invariant mass of the lepton pair. Similarly, in the axial-vector ( $a_1$  meson channel) the spectral function below the  $\tau$  mass can be determined from the hadronic decay width of the  $\tau$  lepton  $\Gamma(\tau \rightarrow \nu_\tau + \text{hadrons})$ .

- (2) In some cases, the coupling constants of a few resonances can be extracted indirectly, for example using low-energy theorems. In this way, the approximate shape of the pseudoscalar  $\pi, K, \eta, \eta'$  and some glueball correlators can be determined.
- (3) Ultimately, the best source of information about hadronic correlation functions is the lattice. At present most lattice calculations use complicated nonlocal sources. Exceptions can be found in Chu *et al.* (1993a, 1993b) and Leinweber (1995a, 1995b). So far, all results have been obtained in the quenched approximation.

In general, given the fundamental nature of hadronic correlators, all models of hadronic structure or the QCD vacuum should be tested against the available information on the correlators. We shall discuss some of these models as we go along.

## B. The quark propagator in the instanton liquid

As we have emphasized above, the complete information about mesonic and baryonic correlation functions is encoded in the quark propagator in a given gauge-field configuration. Interactions among quarks are represented by the failure of expectation values to factorize, e.g.,  $\langle S(\tau)^2 \rangle \neq \langle S(\tau) \rangle^2$ . In the following, we shall construct the quark propagator in the instanton ensemble, starting from the propagator in the background field of a single instanton.

From the quark propagator, we calculate the ensemble-averaged meson and baryon correlation func-

tions. However, it is also interesting to study the vacuum expectation value of the propagator<sup>39</sup>

$$\langle S(x) \rangle = S_S(x) + \gamma \cdot x S_V(x). \quad (242)$$

From the definition of the quark condensate, we have  $\langle \bar{q}q \rangle = S_S(0)$ , which means that the scalar component of the quark propagator provides an order parameter for chiral symmetry breaking. To obtain more information, we can define a gauge-invariant propagator by adding a Wilson line,

$$S_{\text{inv}}(x) = \left\langle \psi(x) P \exp \left( \int_0^x A_\mu(x') dx'_\mu \right) \bar{\psi}(0) \right\rangle. \quad (243)$$

This object has a direct physical interpretation because it describes the propagation of a light quark coupled to an infinitely heavy, static source (Shuryak, 1982a; Shuryak and Verbaarschot, 1993b; Chernyshev *et al.*, 1996). It therefore determines the spectrum of heavy-light mesons (with the mass of the heavy quark subtracted) in the limit where the mass of the heavy quark goes to infinity.

### 1. The propagator in the field of a single instanton

The propagators of massless scalar bosons, gauge fields, and fermions in the background field of a single instanton can be determined analytically<sup>40</sup> (Brown *et al.*, 1978; Levine and Yaffe, 1979). We do not go into details of the construction, which is quite technical, but only provide the main results.

We have already seen that the quark propagator in the field of an instanton is ill behaved because of the presence of a zero mode. Of course, the zero mode does not cause any harm, since it is compensated by a zero in the tunneling probability. The remaining non-zero-mode part of the propagator satisfies the equation

$$i \not{D} S^{nz}(x, y) = \delta(x - y) - \psi_0(x) \psi_0^\dagger(y), \quad (244)$$

which ensures that all modes in  $S^{nz}$  are orthogonal to the zero mode. Formally, this equation is solved by

$$S^{nz}(x, y) = \vec{D}_x \Delta(x, y) \left( \frac{1 + \gamma_5}{2} \right) + \Delta(x, y) \vec{D}_y \left( \frac{1 - \gamma_5}{2} \right), \quad (245)$$

where  $\Delta(x, y)$  is the propagator of a scalar quark in the fundamental representation,  $-D^2 \Delta(x, y) = \delta(x, y)$ .

<sup>39</sup>The quark propagator is of course not a gauge-invariant object. Here, we imply that a gauge has been chosen or the propagator is multiplied by a gauge string. Also note that, before averaging, the quark propagator has a more general Dirac structure,  $S(x) = E + P \gamma_5 + V_\mu \gamma_\mu + A_\mu \gamma_\mu \gamma_5 + T_{\mu\nu} \sigma_{\mu\nu}$ . This decomposition, together with positivity, is the basis of a number of exact results about correlation functions (Weingarten, 1983; Vafa and Witten, 1984).

<sup>40</sup>The result is easily generalized to 't Hooft's exact multi-instanton solution, but much more effort is required to construct the quark propagator in the most general (Atiyah, Hitchin, Drinfeld, and Manin, 1977) instanton background (Corrigan, Goddard, and Templeton, 1979).

Equation (245) is easily checked using the methods we employed in order to construct the zero-mode solution; see Eq. (102). The scalar propagator does not have any zero modes, so it can be constructed using standard techniques. The result (in singular gauge) is (Brown *et al.*, 1978)

$$\Delta(x,y) = \frac{1}{4\pi^2(x-y)^2} \frac{1}{\sqrt{1+\rho^2/x^2}} \frac{1}{\sqrt{1+\rho^2/y^2}} \times \left( 1 + \frac{\rho^2 \tau^- \cdot x \tau^+ \cdot y}{x^2 y^2} \right). \quad (246)$$

For an instanton located at  $z$ , one has to make the obvious replacements  $x \rightarrow (x-z)$  and  $y \rightarrow (y-z)$ . The propagator in the field of an anti-instanton is obtained by interchanging  $\tau^+$  and  $\tau^-$ . If the instanton is rotated by the color matrix  $R^{ab}$ , then  $\tau^\pm$  has to be replaced by  $(R^{ab} \tau^b, \mp i)$ .

Using the result for the scalar quark propagator and the representation (245) of the spinor propagator introduced above, the non-zero-mode propagator is given by

$$S^{nz}(x,y) = \frac{1}{\sqrt{1+\rho^2/x^2}} \frac{1}{\sqrt{1+\rho^2/y^2}} \left[ S_0(x,y) \times \left( 1 + \frac{\rho^2 \tau^- \cdot x \tau^+ \cdot y}{x^2 y^2} \right) - D_0(x,y) \frac{\rho^2}{x^2 y^2} \times \left( \frac{\tau^- \cdot x \tau^+ \cdot y \tau^- \cdot (x-y) \tau^+ \cdot y}{\rho^2 + x^2} \gamma_+ + \frac{\tau^- \cdot x \tau^+ \cdot (x-y) \tau^- \cdot y \tau^+ \cdot y}{\rho^2 + x^2} \gamma_- \right) \right], \quad (247)$$

where  $\gamma_\pm = (1 \pm \gamma_5)/2$ . The propagator can be generalized to arbitrary instanton positions and color orientations in the same way as the scalar quark propagator discussed above.

At short distances, as well as far away from the instanton, the propagator reduces to the free one. At intermediate distances, the propagator is modified due to gluon exchanges with the instanton field,

$$S^{nz}(x,y) = -\frac{\gamma \cdot (x-y)}{2\pi^2(x-y)^4} - \frac{1}{16\pi^2(x-y)^2} \times (x-y)_\mu \gamma_\nu \gamma_5 \tilde{G}_{\mu\nu} + \dots \quad (248)$$

This result is consistent with the OPE of the quark propagator in a general background field [see Eq. (240)]. It is interesting to note that all the remaining terms are regular as  $(x-y)^2 \rightarrow 0$ . This has important consequences for the OPE of hadronic correlators in a general, self-dual background field (Dubovikov and Smilga, 1981).

Finally, we need the quark propagator in the instanton field for small but nonvanishing quark mass. Expanding the quark propagator for small  $m$ , we get

$$S(x,y) = \frac{1}{i\not{D} + im} = \frac{\psi_0(x)\psi_0^\dagger(y)}{m} + S^{nz}(x,y) + m\Delta(x,y) + \dots \quad (249)$$

## 2. The propagator in the instanton ensemble

In this section we extend the results of the last section to the more general case of an ensemble consisting of many pseudoparticles. The quark propagator in an arbitrary gauge field can always be expanded as

$$S = S_0 + S_0 \mathcal{A} S_0 + S_0 \mathcal{A} S_0 \mathcal{A} S_0 + \dots, \quad (250)$$

where the individual terms have an obvious interpretation as arising from multiple gluon exchanges with the background field. If the gauge field is a sum of instanton contributions,  $A_\mu = \sum_I A_{I\mu}$ , then Eq. (250) becomes

$$S = S_0 + \sum_I S_0 \mathcal{A}_I S_0 + \sum_{I,J} S_0 \mathcal{A}_I S_0 \mathcal{A}_J S_0 + \dots \quad (251)$$

$$= S_0 + \sum_I (S_I - S_0) + \sum_{I \neq J} (S_I - S_0) S_0^{-1} (S_J - S_0) + \sum_{I \neq J, J \neq K} (S_I - S_0) S_0^{-1} (S_J - S_0) S_0^{-1} (S_K - S_0) + \dots \quad (252)$$

Here,  $I, J, K, \dots$  refer to both instantons and anti-instantons. In the second line, we have resummed the contributions corresponding to an individual instanton.  $S_I$  refers to the sum of zero and non-zero-mode components. At large distances from the center of the instanton,  $S_I$  approaches the free propagator  $S_0$ . Thus Eq. (252) has a nice physical interpretation: Quarks propagate by jumping from one instanton to the other. If  $|x-z_I| \ll \rho_I$ , and  $|y-z_I| \ll \rho_I$  for all  $I$ , the free propagator dominates. At large distances, terms involving more and more instantons become important.

In the QCD ground state, chiral symmetry is broken. The presence of a condensate implies that quarks can propagate over large distances. Therefore we cannot expect that truncating the series (252) will provide a useful approximation to the propagator at low momenta. Furthermore, we know that spontaneous symmetry breaking is related to small eigenvalues of the Dirac operator. A good approximation to the propagator is obtained by assuming that  $(S_I - S_0)$  is dominated by fermion zero modes,

$$(S_I - S_0)(x,y) \approx \frac{\psi_I(x)\psi_I^\dagger(y)}{im}. \quad (253)$$

In this case, the expansion (252) becomes

$$\begin{aligned}
S(x,y) &\approx S_0(x,y) \\
&+ \sum_I \frac{\psi_I(x)\psi_I^\dagger(y)}{im} + \sum_{I \neq J} \frac{\psi_I(x)}{im} \\
&\times \left( \int d^4r \psi_I^\dagger(r)(-i\not{p}-im)\psi_J(r) \right) \frac{\psi_J^\dagger(y)}{im} + \dots, \quad (254)
\end{aligned}$$

which contains the overlap integrals  $T_{IJ}$  defined in Eq. (148). This expansion can easily be summed to give

$$\begin{aligned}
S(x,y) &\approx S_0(x,y) \\
&+ \sum_{I,J} \psi_I(x) \frac{1}{T_{IJ} + imD_{IJ} - im\delta_{IJ}} \psi_J^\dagger(y). \quad (255)
\end{aligned}$$

Here,  $D_{IJ} = \int d^4r \psi_I^\dagger(r)\psi_J(r) - \delta_{IJ}$  arises from the restriction  $I \neq J$  in the expansion (252). The quantity  $mD_{IJ}$  is small in both the chiral expansion and the packing fraction of the instanton liquid and will be neglected in what follows. Comparing the resummed propagator (255) with the single-instanton propagator (253) shows the importance of chiral symmetry breaking. While Eq. (253) is proportional to  $1/m$ , the diagonal part of the full propagator is proportional to  $(T^{-1})_{II} = 1/m^*$ .

The result (255) can also be derived by inverting the Dirac operator in the basis spanned by the zero modes of the individual instantons:

$$S(x,y) \approx S_0(x,y) + \sum_{I,J} |I\rangle\langle I| \frac{1}{i\not{D} + im} |J\rangle\langle J|. \quad (256)$$

The equivalence of Eqs. (255) and (256) can be easily seen using the fact that, in the sum ansatz, the derivative in the overlap matrix element  $T_{IJ}$  can be replaced by a covariant derivative.

The propagator (255) can be calculated either numerically or using the mean-field approximation introduced in Sec. IV.F. We shall discuss the mean-field propagator in the following section. For our numerical calculations, we have improved the zero-mode propagator by adding the contributions from nonzero modes to first order in the expansion (252). The result is

$$\begin{aligned}
S(x,y) &= S_0(x,y) + S^{ZMZ}(x,y) \\
&+ \sum_I [S_I^{NZM}(x,y) - S_0(x,y)]. \quad (257)
\end{aligned}$$

How accurate is this propagator? We have seen that the propagator agrees with the general OPE result at short distance. We also know that it accounts for chiral symmetry breaking and spontaneous mass generation at large distances. In addition to that, we have performed a number of checks on the correlation functions that are sensitive to the degree to which (257) satisfies the equations of motion, for example, by testing whether the vector correlator is transverse (the vector current is conserved).

### 3. The propagator in the mean-field approximation

In order to understand the propagation of quarks in the “zero-mode zone” it is very instructive to construct

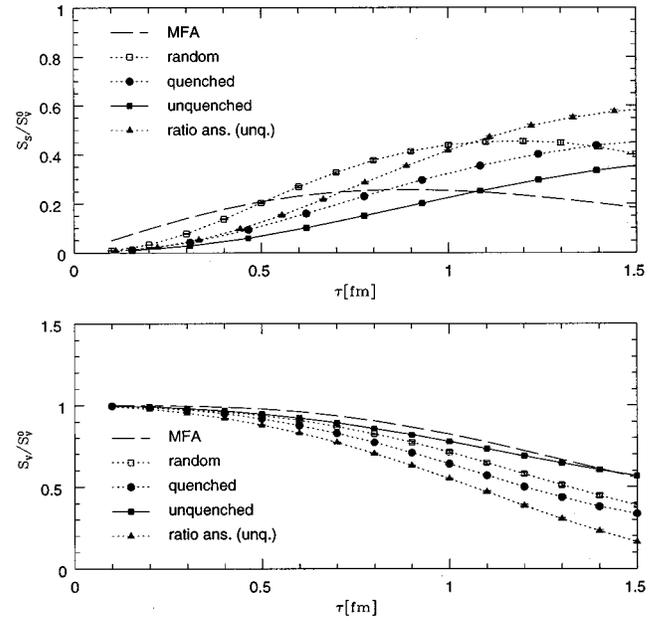


FIG. 18. Scalar and vector components of the quark propagator, normalized to the free vector propagator. The dashed lines show the result of the mean-field approximation, while the data points were obtained in different instanton ensembles.

the propagator in the mean-field approximation. The mean-field propagator can be obtained in several ways. Most easily, we can read off the propagator directly from the effective partition function (188). We find

$$S(p) = \frac{\not{p} + iM(p)}{p^2 + M^2(p)} \quad (258)$$

with the momentum-dependent effective quark mass

$$M(p) = \frac{\beta}{2N_c} \frac{N}{V} p^4 \varphi'^2(p). \quad (259)$$

Here,  $\beta$  is the solution of the gap equation (189). Originally, the result (258) was obtained by Diakonov and Petrov from the Dyson-Schwinger equation for the quark propagator in the large- $N_c$  limit (Diakonov and Petrov, 1986). At small momenta, chiral symmetry breaking generates an effective mass  $M(0) = (\beta/2N_c)(N/V)(2\pi\rho)^2$ . The quark condensate was already given in Eq. (195). At large momenta, we have  $M(p) \sim 1/p^6$  and constituent quarks become free current quarks.

For comparison with our numerical results, it is useful to determine the mean-field propagator in coordinate space. Fourier-transforming the result (258) gives

$$S_V(x) = \frac{1}{4\pi^2 x} \int dp \frac{p^4}{p^2 + M^2(p)} J_2(px), \quad (260)$$

$$S_S(x) = \frac{1}{4\pi^2 x} \int dp \frac{p^3 M(p)}{p^2 + M^2(p)} J_1(px). \quad (261)$$

The result is shown in Fig. 18. The scalar and vector components of the propagator are normalized to the free propagator. The vector component of the propagator is exponentially suppressed at large distance, show-

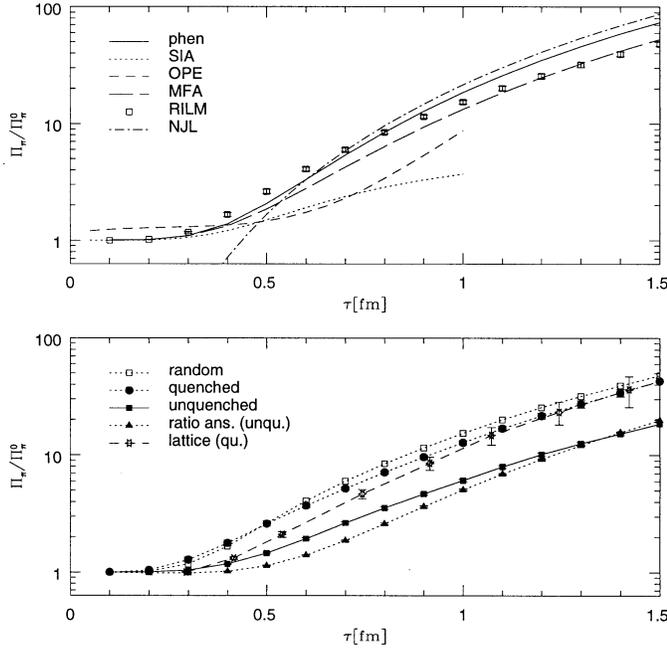


FIG. 19. Pion correlation function in various approximations and instanton ensembles. Upper panel: solid curve, the phenomenological expectation; short-dashed curve, the operator product expansion; dotted curve, the single instanton; long-dashed curve, the mean-field approximation;  $\square$ , data in the random instanton ensemble. Lower panel: comparison of different instanton ensembles;  $\square$ , random;  $\bullet$ , quenched;  $\blacksquare$ , unquenched;  $\blacktriangle$ , ratio ansatz;  $\star$ , lattice; crown, interacting.

ing the formation of a constituent mass. The scalar component again shows the breaking of chiral symmetry. At short distances, the propagator is consistent with the coordinate-dependent quark mass  $m = \pi^2/3 \cdot x^2 \langle \bar{\psi}\psi \rangle$  inferred from the OPE equation (240). At large distances the exponential decay is governed by the constituent mass  $M(0)$ .

For comparison, we also show the quark propagator in different instanton ensembles. The result is very similar to the mean-field approximation; the differences are mainly due to different values of the quark condensate. The quark propagator is not very sensitive to correlations in the instanton liquid. In Shuryak and Verbaarschot (1993b), and Chernyshev *et al.* (1996), the quark propagator was also used to study heavy-light mesons in the  $1/M_Q$  expansion. The results are very encouraging, and we refer the reader to the original literature for details.

### C. Mesonic correlators

#### 1. General results and the operator product expansion

A large number of mesonic correlation functions have been studied in the instanton model, and clearly this is not the place to list all of them. Instead, we have decided to discuss three examples that are illustrative of the techniques and the most important effects. We shall consider the  $\pi$ ,  $\rho$ , and  $\eta'$  channels, related to the currents

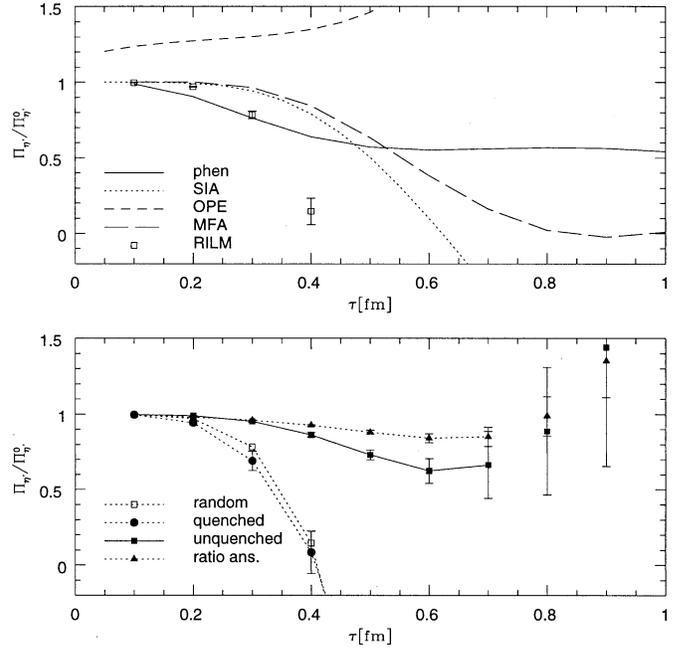


FIG. 20. Eta-prime meson correlation functions. The various curves and data sets are labeled as in Fig. 19.

$$j_\pi = \bar{q} \tau^a \gamma_5 q, \quad j_\rho = \bar{q} \frac{\tau^a}{2} \gamma_\mu q, \quad j_{\eta_{ns}} = \bar{q} \gamma_5 q, \quad (262)$$

where  $q = (u, d)$ . Here, we consider only the nonstrange  $\eta'$  and refer the reader to Schäfer (1996), and Shuryak and Verbaarschot (1993a) for a discussion of SU(3) flavor symmetry breaking and  $\eta - \eta'$  mixing. The three channels discussed here are instructive because the instanton-induced interaction is attractive for the pion channel, repulsive for the  $\eta'$ , and (to first order in the instanton density) does not affect the  $\rho$  meson. The  $\pi$ ,  $\rho$ , and  $\eta'$  are therefore representative of a much larger set of correlation functions. In addition, these three mesons have special physical significance. The pion is the lightest hadron, connected with the spontaneous breakdown of  $SU(N_f)_L \times SU(N_f)_R$  chiral symmetry. The  $\eta'$  is surprisingly heavy, a fact related to the anomalous  $U(1)_A$  symmetry. The  $\rho$  meson, finally, is the lightest non-Goldstone particle and the first  $\pi\pi$  resonance.

Phenomenological predictions for the correlation functions are shown in Figs. 19–21 (Shuryak, 1993). All correlators are normalized to the free ones,  $R(\tau) = \Pi_\Gamma(\tau)/\Pi_\Gamma^0(\tau)$ , where  $\Pi_\Gamma^0(\tau) = \text{Tr}[\Gamma S_0(\tau)\Gamma S_0(-\tau)]$ . At short distances, asymptotic freedom implies that this ratio approaches one. At large distances, the correlators are exponential and  $R$  is small. At intermediate distances,  $R$  depends on the quark-quark interaction in that channel. If  $R > 1$ , we shall refer to the correlator as attractive, while  $R < 1$  implies repulsive interactions. The normalized pion correlation function  $R_\pi$  is significantly larger than one, showing a strongly attractive interaction and a light bound state. The rho-meson correlator is close to one out to fairly large distances  $x \approx 1.5$  fm, a phenomenon referred to as “superduality” by Shuryak (1993). The  $\eta'$  channel is strongly repulsive,

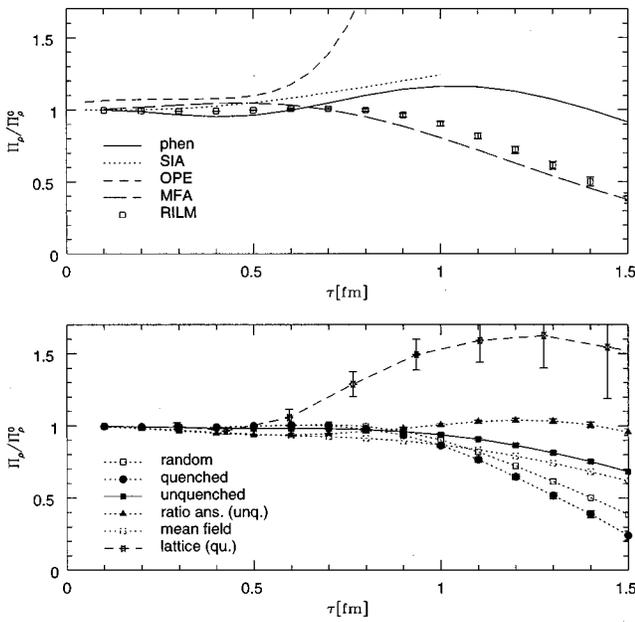


FIG. 21. Rho-meson correlation functions. The various curves and data sets are labeled as in Fig. 19. The dashed squares show the noninteracting part of the rho meson correlator in the interacting ensemble.

showing an enhancement only at intermediate distances due to mixing with the  $\eta$ .

Theoretical information about the short-distance behavior of the correlation functions comes from the operator product expansion (Shifman *et al.*, 1979). For the  $\pi$  and  $\rho$  channels, we have

$$\begin{aligned} \Pi_\pi^{\text{OPE}}(x) = & \Pi_\pi^0(x) \left( 1 + \frac{11}{3} \frac{\alpha_s(x)}{\pi} - \frac{\pi^2}{3} m \langle \bar{q}q \rangle x^4 \right. \\ & + \frac{1}{384} \langle G^2 \rangle x^4 + \frac{11\pi^3}{81} \alpha_s(x) \\ & \left. \times \langle \bar{\psi}\psi \rangle^2 \log(x^2) x^6 + \dots \right), \end{aligned} \quad (263)$$

$$\begin{aligned} \Pi_\rho^{\text{OPE}}(x) = & \Pi_\rho^0(x) \left( 1 + \frac{\alpha_s(x)}{\pi} - \frac{\pi^2}{4} m \langle \bar{q}q \rangle x^4 \right. \\ & - \frac{1}{384} \langle G^2 \rangle x^4 + \frac{7\pi^3}{81} \alpha_s(x) \\ & \left. \times \langle \bar{\psi}\psi \rangle^2 \log(x^2) x^6 + \dots \right), \end{aligned} \quad (264)$$

where we have restricted ourselves to operators of dimension up to six and the leading-order perturbative corrections. We have also used the factorization hypothesis for the expectation value of four-quark operators. Note that, to this order, the nonstrange eta-prime, and pion correlation functions are identical. This result shows that QCD sum rules cannot account for the  $U(1)_A$  anomaly.

The OPE predictions [Eqs. (263) and (264)] are also shown in Figs. 19–21. The leading quark and gluon

power corrections in the  $\pi$  channel are both attractive, so the OPE prediction has the correct tendency, but underpredicts the rise for  $x > 0.25$  fm. In the  $\rho$ -meson channel, the leading corrections have a tendency to cancel each other, in agreement with the superduality phenomenon mentioned above.

## 2. The single-instanton approximation

After this prelude we come to the instanton model. We start by considering instanton contributions to the short-distance behavior of correlation functions in the single-instanton approximation<sup>41</sup> (Shuryak, 1983). The main idea is that if  $x - y$  is small compared to the typical instanton separation  $R$ , we expect that the contribution from the instanton  $I = I_*$  closest to the points  $x$  and  $y$  will dominate over all others. For the propagator in the zero-mode zone, this implies

$$\begin{aligned} S(x, y) = & \sum_{IJ} \psi_I(x) \left( \frac{1}{T + im} \right)_{IJ} \psi_J^\dagger(y) \\ \approx & \psi_{I_*}(x) \left( \frac{1}{T + im} \right)_{I_* I_*} \psi_{I_*}^\dagger(y) \approx \frac{\psi_{I_*}(x) \psi_{I_*}^\dagger(y)}{m^*}, \end{aligned} \quad (265)$$

where we have approximated the diagonal matrix element by its average,  $(T + im)_{I_* I_*}^{-1} \approx N^{-1} \sum_I (T + im)_{II}^{-1}$ , and introduced the effective mass  $m^*$  defined in Sec. IV.E,  $(m^*)^{-1} = N^{-1} \sum \lambda^{-1}$ . In the following we shall use the mean-field estimate  $m^* = \pi \rho (2n/3)^{1/2}$ . As a result, the propagator in the single-instanton approximation looks like the zero-mode propagator of a single instanton, but for a particle with an effective mass  $m^*$ .

The  $\pi$  and  $\eta'$  correlators receive zero-mode contributions. In the single-instanton approximation, we find (Shuryak, 1983)

$$\begin{aligned} \Pi_{\pi, \eta'}^{\text{SIA}}(x) = & \pm \int d\rho n(\rho) \frac{6\rho^4}{\pi^2} \frac{1}{(m^*)^2} \frac{\partial^2}{\partial(x^2)^2} \\ & \times \left\{ \frac{4\xi^2}{x^4} \left( \frac{\xi^2}{1 - \xi^2} + \frac{\xi}{2} \log \frac{1 + \xi}{1 - \xi} \right) \right\}, \end{aligned} \quad (266)$$

where  $\xi^2 = x^2/(x^2 + 4\rho^2)$ . There is also a non-zero-mode contribution to these correlation functions. It was calculated by Shuryak (1989), but numerically it is not very important.

We show the result [Eq. (266)] in Fig. 19. For simplicity, we have chosen  $n(\rho) = n_0 \delta(\rho - \rho_0)$  with the standard parameters  $n_0 = 1 \text{ fm}^{-4}$  and  $\rho_0 = 0.33 \text{ fm}$ . The pion correlator is similar to the OPE prediction at short distances,  $x \leq 0.25 \text{ fm}$ , but follows the phenomenological

<sup>41</sup>We should like to distinguish this method from the dilute-gas approximation. In the dilute-gas approximation, we systematically expand the correlation functions in terms of the one- (two-, three-, etc.) instanton contribution. In the presence of light fermions (for  $N_f > 1$ ), however, this method is useless because there is no zero-mode contribution to chirality-violating operators from any finite number of instantons.

result out to larger distances,  $x \approx 0.4$  fm. In particular, we find that even at very short distances,  $x \approx 0.3$  fm, the regular contributions to the correlator coming from instanton zero modes are larger than the singular contributions included in the OPE. The fact that nonperturbative corrections not accounted for in the OPE are particularly large in spin-zero channels ( $\pi$ ,  $\eta'$ ,  $\sigma$ , and scalar glueballs) was first emphasized by Novikov *et al.* (1981). For the  $\eta'$ , the instanton contribution is strongly repulsive. This means that, in contrast to the OPE, the single-instanton approximation at least qualitatively accounts for the  $U(1)_A$  anomaly. The single-instanton approximation was extended to the full pseudoscalar nonet by Shuryak (1983). It was shown that the substitution  $m^* \rightarrow m^* + m_s$  gives a good description of SU(3) flavor symmetry breaking and the  $\pi, K, \eta$  correlation functions.

In the  $\rho$ -meson correlator, zero modes cannot contribute since the chiralities do not match. Nonvanishing contributions come from the non-zero-mode propagator (247) and from interference between the zero-mode part and the leading mass correction in Eq. (249):

$$\begin{aligned} \Pi_\rho^{\text{SIA}}(x, y) = & \text{Tr}[\gamma_\mu S^{nz}(x, y) \gamma_\mu S^{nz}(y, x)] \\ & + 2 \text{Tr}[\gamma_\mu \psi_0(x) \psi_0^\dagger(y) \gamma_\mu \Delta(y, x)]. \end{aligned} \quad (267)$$

The latter term survives even in the chiral limit because the factor  $m$  in the mass correction is cancelled by the  $1/m$  from the zero mode. Also note that the result corresponds to the standard dilute-gas approximation, so true multi-instanton effects are not included. After averaging over the instanton coordinates, we find<sup>42</sup> (Andrei and Gross, 1978)

$$\begin{aligned} \Pi_\rho^{\text{SIA}}(x) = & \Pi_\rho^0 + \int d\rho n(\rho) \frac{12}{\pi^2} \frac{\rho^4}{x^2} \frac{\partial}{\partial(x^2)} \\ & \times \left\{ \frac{\xi}{x^2} \log \frac{1+\xi}{1-\xi} \right\}. \end{aligned} \quad (268)$$

The result is also shown in Fig. 21. As in the OPE, the correlator is attractive at intermediate distances. The correlation function does not go down at larger distances, since the dilute-gas approximation does not account for a dynamically generated mass. It is very instructive to compare the result to the OPE in greater detail. Expanding Eq. (268), we get

$$\Pi_\rho^{\text{SIA}}(x) = \Pi_\rho^0(x) \left( 1 + \frac{\pi^2 x^4}{6} \int d\rho n(\rho) \right). \quad (269)$$

This agrees exactly with the OPE, provided we use the average values of the operators in the dilute-gas approximation,

$$\begin{aligned} \langle G^2 \rangle = & 32\pi^2 \int d\rho n(\rho), \text{ and} \\ m\langle \bar{q}q \rangle = & - \int d\rho n(\rho). \end{aligned} \quad (270)$$

Note that the value of  $m\langle \bar{q}q \rangle$  is ‘‘anomalously’’ large in the dilute-gas limit. This means that the contribution from dimension-four operators is attractive, in contradiction with the OPE prediction based on the canonical values of the condensates.

An interesting observation is the fact that Eq. (269) is the only singular term in the dilute-gas approximation correlation function. In fact, the OPE of *any* mesonic correlator in *any* self-dual field contains only dimension-four operators (Dubovikov and Smilga, 1981). This means that for all higher-order operators, either the Wilson coefficient vanishes [as it does, for example, for the triple gluon condensate  $\langle f^{abc} G_{\mu\nu}^a G_{\nu\rho}^b G_{\rho\mu}^c \rangle$ ] or the matrix elements of various operators of the same dimension cancel each other.<sup>43</sup> This is a very remarkable result because it helps to explain the success of QCD sum rules based on the OPE in many channels. In the instanton model, the gluon fields are very inhomogeneous, so one would expect the OPE to fail for  $x > \rho$ . The Dubovikov-Smilga result shows that quarks can propagate through very strong gauge fields (as long as they are self-dual) without suffering strong interactions.

### 3. The random-phase approximation

The single-instanton approximation clearly improves on the short-distance behavior of the  $\pi, \eta'$  correlation functions as compared to the OPE. However, in order to describe a true pion bound state one has to resum the attractive interaction generated by the 't Hooft vertex. This is most easily accomplished using the random-phase approximation (RPA), which corresponds to iterating the average 't Hooft vertex in the  $s$  channel (see Fig. 2). The solution of the Bethe-Salpeter equation can be written as (Diakonov and Petrov, 1986; Hutter, 1995; Kacir, Prakash, and Zahed, 1996)

$$\Pi_\pi^{\text{RPA}}(x) = \Pi_\pi^{\text{MFA}}(x) + \Pi_\pi^{\text{int}}, \quad (271)$$

$$\Pi_\rho^{\text{RPA}}(x) = \Pi_\rho^{\text{MFA}}(x).$$

Here,  $\Pi_\Gamma^{\text{MFA}}$  denotes the mean-field (noninteracting) part of the correlation functions

$$\Pi_\Gamma^{\text{MFA}}(x) = \text{Tr}[\bar{S}(x) \Gamma \bar{S}(-x) \Gamma], \quad (272)$$

where  $\bar{S}(x)$  is the mean-field propagator discussed in Sec. VI.B.3. In the  $\rho$ -meson channel, the 't Hooft vertex vanishes and the correlator is given by the mean-field contribution only. The interacting part of the  $\pi, \eta'$  correlation functions is given by

$$\Pi_{\pi, \eta'}^{\text{int}}(x) = \int d^4 q e^{iq \cdot x} \Gamma_5(q) \frac{\pm 1}{1 \mp C_5(q)} \Gamma_5(q), \quad (273)$$

where the elementary loop function  $C_5$  and the vertex function  $\Gamma_5$  are given by

<sup>43</sup>Here, we do not consider radiative corrections like  $\alpha_s \langle \bar{\psi} \psi \rangle^2$  because we evaluate the OPE in a fixed (classical) background field without taking into account radiative corrections to the background field.

<sup>42</sup>There is a mistake by an overall factor of 3/2 in the original work.

$$C_5(q) = 4N_c \left( \frac{V}{N} \right) \int \frac{d^4p}{(2\pi)^4} \frac{M_1 M_2 (M_1 M_2 - p_1 \cdot p_2)}{(M_1^2 + p_1^2)(M_2^2 + p_2^2)}, \quad (274)$$

$$\Gamma_5(q) = 4 \int \frac{d^4p}{(2\pi)^4} \frac{\sqrt{M_1 M_2} (M_1 M_2 - p_1 \cdot p_2)}{(m_1^2 + p_1^2)(M_2^2 + p_2^2)}. \quad (275)$$

Here,  $p_1 = p + q/2$ ,  $p_2 = p - q/2$ , and  $M_{1,2} = M(p_{1,2})$  are the momentum-dependent effective quark masses. We have already stressed that in the long-wavelength limit, the effective interaction between quarks is of the Nambu–Jona-Lasinio type. Indeed, the pion correlator in the random-phase approximation to the Nambu–Jona-Lasinio model is given by

$$\Pi_\pi^{\text{NJL}}(x) = \int d^4q e^{iq \cdot x} \frac{J_5(q)}{1 - G J_5(q)}, \quad (276)$$

where  $J_5(q)$  is the pseudoscalar loop function

$$J_5(q) = 4N_c \int^\Lambda \frac{d^4p}{(2\pi)^4} \frac{(M^2 - p_1 \cdot p_2)}{(M^2 + p_1^2)(M^2 + p_2^2)}, \quad (277)$$

$G$  is the four-fermion coupling constant and the (momentum-independent) constituent mass  $M$  is the solution of the Nambu–Jona-Lasinio gap equation. The loop function (277) is divergent and is regularized using a cutoff  $\Lambda$ .

Clearly, the RPA correlation functions (273) and (276) are very similar; the only difference is that in the instanton liquid both the quark mass and the vertex are momentum dependent. The momentum dependence of the vertex ensures that no regulator is required. The loop integral is automatically cut off at  $\Lambda \sim \rho^{-1}$ . We also note that the momentum dependence of the effective interaction at the mean-field level is very simple; the vertex function is completely separable.

The mean-field correlation functions are shown in Figs. 19–21. The coordinate-space correlators are easily determined by Fourier-transforming the result (273). We also show the Nambu–Jona-Lasinio result for the pion correlation function. In this case, the correlation function at short distances is not very meaningful, since the cutoff  $\Lambda$  eliminates all contributions to the spectral function above that scale.

The pion correlator nicely reproduces the phenomenological result. The  $\eta'$  correlation function has the correct tendency at short distances but becomes unphysical for  $x > 0.5$  fm. The result in the  $\rho$ -meson channel is also close to phenomenology, despite the fact that it corresponds to two unbound constituent quarks.

#### 4. The interacting instanton liquid

What is the quality and the range of validity of the random-phase approximation? The RPA is usually motivated by the large- $N_c$  approximation, and the dimensionless parameter that controls the expansion is  $\rho^4(N/V)/(4N_c)$ . This parameter is indeed very small, but in practice there are additional parameters that determine the size of corrections to the RPA.

First of all,  $\rho^4(N/V)$  is a useful expansion parameter only if the instanton liquid is random and the role of correlations is small. As discussed in Sec. IV.B, if the density is very small, the instanton liquid is in a molecular phase, while it is in a crystalline phase if the density is large. Clearly, the RPA is expected to fail in both of these limits. In general, the RPA corresponds to using linearized equations for the fluctuations around a mean-field solution. In our case, low-lying meson states are collective fluctuations of the chiral order parameter. For isovector scalar mesons, the RPA is expected to be good because the scalar mean field (the condensate) is large and the masses are light. For isosinglet mesons, the fluctuations are much larger, and the RPA is likely to be less useful.

In the following we shall therefore discuss results from numerical calculations of hadronic correlators in the instanton liquid. These calculations go beyond the RPA in two ways: (i) the propagator includes genuine many-instanton effects and non-zero-mode contributions; and (ii) the ensemble is determined using the full (fermionic and bosonic) weight function, so it includes correlations among instantons. In addition, we shall also consider baryonic correlators and three-point functions that are difficult to handle in the RPA.

We discuss correlation function in three different ensembles, the random, the quenched, and the fully interacting instanton ensembles. In the random model, the underlying ensemble is the same as in the mean-field approximation, only the propagator is more sophisticated. In the quenched approximation, the ensemble includes correlations due to the bosonic action, while the fully interacting ensemble also includes correlations induced by the fermion determinant. In order to check the dependence of the results on the instanton interaction, we study correlation functions in two different, unquenched ensembles, one based on the streamline interaction (with a short-range core) and one based on the ratio ansatz interaction. The bulk parameters of these ensembles are compared in Table II. We note that the ratio ansatz ensemble is denser than the streamline ensemble.

We are interested not only in the behavior of the correlation functions, but also in numerical results for the ground-state masses and coupling constants. For this purpose we have fitted the correlators using a simple “pole-plus-continuum” model for the spectral functions. In the case of the pion, this leads to the following parametrization of the correlation function:

$$\Pi_\pi(x) = \lambda_\pi^2 D(m_\pi, x) + \frac{3}{8\pi^2} \int_{s_0}^\infty ds s D(\sqrt{s}, x), \quad (278)$$

where  $\lambda_\pi$  is the pion coupling constant defined in Table III and  $s_0$  is the continuum threshold. Physically,  $s_0$  roughly represents the position of the first excited state. Resolving higher resonances requires high-quality data and more sophisticated techniques. The model spectral function used here is quite popular in connection with QCD sum rules. It provides a surprisingly good descrip-

TABLE II. Bulk parameters density  $n=N/V$ , average size  $\bar{\rho}$ , diluteness  $\bar{\rho}^4(N/V)$ , and quark condensate  $\langle\bar{q}q\rangle$  in the different instanton ensembles. RILM=random instanton liquid model. We also give the value of the Pauli-Villars scale parameter  $\Lambda^4$  that corresponds to our choice of units,  $n \equiv 1 \text{ fm}^{-4}$ .

	Unquenched	Quenched	RILM	Ratio ansatz (unquenched)
$N/V$	$0.174\Lambda^4$	$0.303\Lambda^4$	$1.0 \text{ fm}^4$	$0.659\Lambda^4$
$\bar{\rho}$	$0.64\Lambda^{-1}$ (0.42 fm)	$0.58\Lambda^{-1}$ (0.43 fm)	$0.33 \text{ fm}$	$0.66\Lambda^{-1}$ (0.59 fm)
$\bar{\rho}^4(N/V)$	0.029	0.034	0.012	0.125
$\langle\bar{q}q\rangle$	$0.359\Lambda^3$ (219 MeV) <sup>3</sup>	$0.825\Lambda^3$ (253 MeV) <sup>3</sup>	$(264 \text{ MeV})^3$	$0.882\Lambda^3$ (213 MeV) <sup>3</sup>
$\Lambda$	306 MeV	270 MeV	-	222 MeV

tion of all measured correlation functions, not only in the instanton model but also on the lattice (Chu *et al.*, 1993b; Leinweber, 1995a).

Correlation functions in the different instanton ensembles were calculated by Shuryak and Verbaarschot (1993a), Schäfer *et al.* (1994), and Schäfer and Shuryak (1996a), to whose articles we refer the reader for more details. The results are shown in Figs. 19–21 and summarized in Table IV. The pion correlation functions in the different ensembles are qualitatively very similar. The differences are mostly due to different values of the quark condensate (and the physical quark mass) in the different ensembles. Using the Gell-Mann, Oaks, Renner relation, one can extrapolate the pion mass to the physical value of the quark masses (see Table IV). The results are consistent with the experimental value in the streamline ensemble (both quenched and unquenched) but clearly too small in the ratio ansatz ensemble. This is a reflection of the fact that the ratio ansatz ensemble is not sufficiently dilute.

In Fig. 21 we also show the results in the  $\rho$  channel. The  $\rho$ -meson correlator is not affected by instanton zero modes to first order in the instanton density. The results in the different ensembles are fairly similar to each other

and all fall somewhat short of the phenomenological result at intermediate distances  $x \approx 1 \text{ fm}$ . We have determined the  $\rho$ -meson mass and coupling constant from a fit similar to Eq. (278). The results are given in Table IV. The  $\rho$ -meson mass is somewhat too heavy in the random and quenched ensembles but in good agreement with the experimental value  $m_\rho = 770 \text{ MeV}$  in the unquenched ensemble.

Since there are no interactions in the  $\rho$ -meson channel to first order in the instanton density, it is important to study whether the instanton liquid provides any significant binding. In the instanton model, there is no confinement, and  $m_\rho$  is close to the two- (constituent) quark threshold. In QCD, the  $\rho$  meson is also not a true bound state but a resonance in the  $2\pi$  continuum. To determine whether the continuum contribution in the instanton liquid is predominantly from two-pion or two-quark states would require the determination of the corresponding three-point functions, which has not yet been done. Instead, we have compared the full correlation function with the noninteracting (mean-field) correlator (272), where we use the average (constituent-quark) propagator determined in the same ensemble (see Fig. 21). This comparison provides a measure of the strength

TABLE III. Definition of various currents and hadronic matrix elements referred to in this work.

Channel	Current	Matrix element	Experimental value
$\pi$	$j_\pi^a = \bar{q} \gamma_5 \tau^a q$	$\langle 0   j_\pi^a   \pi^b \rangle = \delta^{ab} \lambda_\pi$	$\lambda_\pi \approx (480 \text{ MeV})^3$
	$j_{\mu 5}^a = \bar{q} \gamma_\mu \gamma_5 \frac{\tau^a}{2} q$	$\langle 0   j_{\mu 5}^a   \pi^b \rangle = \delta^{ab} q_\mu f_\pi$	$f_\pi = 93 \text{ MeV}$
$\delta$	$j_\delta^a = \bar{q} \tau^a q$	$\langle 0   j_\delta^a   \delta^b \rangle = \delta^{ab} \lambda_\delta$	
$\sigma$	$j_\sigma = \bar{q} q$	$\langle 0   j_\sigma   \sigma \rangle = \lambda_\sigma$	
$\eta_{ns}$	$j_{\eta_{ns}} = \bar{q} \gamma_5 q$	$\langle 0   j_{\eta_{ns}}   \eta_{ns} \rangle = \lambda_{\eta_{ns}}$	
$\rho$	$J_\mu^a = \bar{q} \gamma_\mu \frac{\tau^a}{2} q$	$\langle 0   j_{\mu}^a   \rho^b \rangle = \delta^{ab} \epsilon_\mu \frac{m_\rho^2}{g_\rho}$	$g_\rho = 5.3$
$a_1$	$j_{\mu 5}^a = \bar{q} \gamma_\mu \gamma_5 \frac{\tau^a}{2} q$	$\langle 0   j_{\mu 5}^a   a_1^b \rangle = \delta^{ab} \epsilon_\mu \frac{m_{a_1}^2}{g_{a_1}}$	$g_{a_1} = 9.1$
$N$	$\eta_1 = \epsilon^{abc} (u^a C \gamma_\mu u^b) \gamma_5 \gamma_\mu d^c$	$\langle 0   \eta_1   N(p, s) \rangle = \lambda_1^N u(p, s)$	
$N$	$\eta_2 = \epsilon^{abc} (u^a C \sigma_{\mu\nu} u^b) \gamma_5 \sigma_{\mu\nu} d^c$	$\langle 0   \eta_2   N(p, s) \rangle = \lambda_2^N u(p, s)$	
$\Delta$	$\eta_\mu = \epsilon^{abc} (u^a C \gamma_\mu u^b) u^c$	$\langle 0   \eta_\mu   N(p, s) \rangle = \lambda^\Delta u_\mu(p, s)$	

TABLE IV. Meson parameters in the different instanton ensembles. RILM= random instanton liquid model. All quantities are given in units of GeV. The current quark mass is  $m_u = m_d = 0.1\Lambda$ . Except for the pion mass, no attempt has been made to extrapolate the parameters to physical values of the quark mass.

	Unquenched	Quenched	RILM	Ratio ansatz (unquenched)
$m_\pi$	0.265	0.268	0.284	0.128
$m_\pi$ (extr.)	0.117	0.126	0.155	0.067
$\lambda_\pi$	0.214	0.268	0.369	0.156
$f_\pi$	0.071	0.091	0.091	0.183
$m_\rho$	0.795	0.951	1.000	0.654
$g_\rho$	6.491	6.006	6.130	5.827
$m_{a_1}$	1.265	1.479	1.353	1.624
$g_{a_1}$	7.582	6.908	7.816	6.668
$m_\sigma$	0.579	0.631	0.865	0.450
$m_\delta$	2.049	3.353	4.032	1.110
$m_{\eta_{ns}}$	1.570	3.195	3.683	0.520

of the interaction. We observe that there is an attractive interaction generated in the interacting liquid due to correlated instanton/anti-instanton pairs [see the discussion in Sec. VII.B.2, in particular, Eq. (332)]. This is consistent with the fact that the interaction is considerably smaller in the random ensemble. In the random model, the strength of the interaction grows as the ensemble becomes more dense. However, the interaction in the full ensemble is significantly larger than in the random model at the same diluteness. Therefore most of the interaction is due to dynamically generated pairs.

The situation is drastically different in the  $\eta'$  channel. Among the  $\sim 40$  correlation functions calculated in the random ensemble, only the  $\eta'$  and the isovector-scalar  $\delta$  discussed in the next section, are completely unacceptable. The correlation function decreases very rapidly and becomes negative at  $x \sim 0.4$  fm. This behavior is incompatible with the positivity of the spectral function. The interaction in the random ensemble is too repulsive, and the model “overexplains” the  $U(1)_A$  anomaly.

The results in the unquenched ensembles (closed and open points) significantly improve the situation. This is related to dynamic correlations between instantons and anti-instantons (topological charge screening). The single-instanton contribution is repulsive, but the contribution from pairs is attractive (Schäfer *et al.*, 1995). Only if correlations among instantons and anti-instantons are sufficiently strong will the correlators be prevented from becoming negative. Quantitatively, the  $\delta$  and  $\eta_{ns}$  masses in the streamline ensemble are still heavier than their experimental values. In the ratio ansatz, on the other hand, the correlation functions even show an enhancement at distances on the order of 1 fm, and the fitted masses are too light. This shows that the  $\eta'$  channel is very sensitive to the strength of correlations among instantons.

In summary, pion properties are mostly sensitive to

global properties of the instanton ensemble, in particular its diluteness. Good phenomenology demands  $\bar{\rho}^4 n \approx 0.03$ , as originally suggested by Shuryak (1982c). The properties of the  $\rho$  meson are essentially independent of the diluteness but show sensitivity to IA correlations. These correlations become crucial in the  $\eta'$  channel.

Finally, we compare the correlation functions in the instanton liquid to lattice measurements reported by Chu *et al.* (1993a, 1993b; see Sec. VI.E). These correlation functions were measured in quenched QCD, so they should be compared to the random or quenched instanton ensembles. The results agree very well in the pion channel, while the lattice correlation function in the rho-meson channel is somewhat more attractive than the correlator in the instanton liquid.

## 5. Other mesonic correlation functions

After discussing the  $\pi, \rho, \eta'$  in some detail we comment only briefly on other correlation functions. The remaining scalar states are the isoscalar  $\sigma$  and the isovector  $\delta$  (the  $f_0$  and  $a_0$  according to the notation of the Particle Data Group). The sigma correlator has a disconnected contribution, which is proportional to  $\langle \bar{q}q \rangle^2$  at large distances. In order to determine the lowest resonance in this channel, the constant contribution has to be subtracted, which makes it difficult to obtain reliable results. Nevertheless, we find that the instanton liquid favors a (presumably broad) resonance around 500–600 MeV. The isovector channel is in many ways similar to the  $\eta'$ . In the random ensemble, the interaction is too repulsive, and the correlator becomes unphysical. This problem is solved in the interacting ensemble, but the  $\delta$  is still very heavy,  $m_\delta > 1$  GeV.

The remaining nonstrange vectors are the  $a_1$ ,  $\omega$ , and  $f_1$ . The  $a_1$  mixes with the pion, which allows a determination of the pion decay constant  $f_\pi$  (as does a direct measurement of the  $\pi - a_1$  mixing correlator). In the instanton liquid, disconnected contributions in the vector channels are small. This is consistent with the fact that the  $\rho$  and the  $\omega$ , as well as the  $a_1$  and the  $f_1$ , are almost degenerate.

Finally, we can also include strange quarks. SU(3) flavor breaking in the 't Hooft interaction nicely accounts for the masses of the  $K$  and the  $\eta$ . More difficult is a correct description of  $\eta - \eta'$  mixing, which can only be achieved in the full ensemble. The random ensemble also has a problem with the mass splittings among the vectors  $\rho$ ,  $K^*$ , and  $\phi$  (Shuryak and Verbaarschot, 1993a). This is related to the fact that flavor symmetry breaking in the random ensemble is so strong that the strange and nonstrange constituent-quark masses are almost degenerate. This problem is improved (but not fully solved) in the interacting ensemble.

## D. Baryonic correlation functions

After discussing quark-antiquark systems in the last section, we now proceed to three-quark (baryon) channels. As emphasized by Shuryak and Rosner (1989), the existence of a strongly attractive interaction in the pseu-

TABLE V. Definition of nucleon and delta correlation functions.

Correlator	Definition	Correlator	Definition
$\Pi_1^N(x)$	$\langle \text{tr}(\eta_1(x) \bar{\eta}_1(0)) \rangle$	$\Pi_1^\Delta(x)$	$\langle \text{tr}(\eta_\mu(x) \bar{\eta}_\mu(0)) \rangle$
$\Pi_2^N(x)$	$\langle \text{tr}(\gamma \cdot \hat{x} \eta_1(x) \bar{\eta}_1(0)) \rangle$	$\Pi_2^\Delta(x)$	$\langle \text{tr}(\gamma \cdot \hat{x} \eta_\mu(x) \bar{\eta}_\mu(0)) \rangle$
$\Pi_3^N(x)$	$\langle \text{tr}(\eta_2(x) \bar{\eta}_2(0)) \rangle$	$\Pi_3^\Delta(x)$	$\langle \text{tr}(\hat{x}_\mu \hat{x}_\nu \eta_\mu(x) \bar{\eta}_\nu(0)) \rangle$
$\Pi_4^N(x)$	$\langle \text{tr}(\gamma \cdot \hat{x} \eta_2(x) \bar{\eta}_2(0)) \rangle$	$\Pi_4^\Delta(x)$	$\langle \text{tr}(\gamma \cdot \hat{x} \hat{x}_\mu \hat{x}_\nu \eta_\mu(x) \bar{\eta}_\nu(0)) \rangle$
$\Pi_5^N(x)$	$\langle \text{tr}(\eta_1(x) \bar{\eta}_2(0)) \rangle$		
$\Pi_6^N(x)$	$\langle \text{tr}(\gamma \cdot \hat{x} \eta_1(x) \bar{\eta}_2(0)) \rangle$		

doscalar quark-antiquark (pion) channel also implies an attractive interaction in the scalar quark-quark (diquark) channel. This interaction is phenomenologically very desirable because it explains not only why the spin-1/2 nucleon is lighter than the spin-3/2 delta but also why lambda is lighter than sigma.

### 1. Nucleon correlation functions

The proton current can be constructed by coupling a  $d$  quark to a  $uu$  diquark. The diquark has the structure  $\epsilon_{abc} u_b C \Gamma u_c$ , which requires that the matrix  $C\Gamma$  be symmetric. This condition is satisfied for the  $V$  and  $T$  gamma matrix structures. The two possible currents (with no derivatives and the minimum number of quark fields) with positive parity and spin 1/2 are given by (Ioffe, 1981)

$$\begin{aligned} \eta_1 &= \epsilon_{abc} (u^a C \gamma_\mu u^b) \gamma_5 \gamma_\mu d^c, \\ \eta_2 &= \epsilon_{abc} (u^a C \sigma_{\mu\nu} u^b) \gamma_5 \sigma_{\mu\nu} d^c. \end{aligned} \quad (279)$$

It is sometimes useful to rewrite these currents in terms of scalar and pseudoscalar diquarks,

$$\eta_{1,2} = (2,4) \{ \epsilon_{abc} (u^a C d^b) \gamma_5 u^c \mp \epsilon_{abc} (u^a C \gamma_5 d^b) u^c \}. \quad (280)$$

Nucleon correlation functions are defined by  $\Pi_{\alpha\beta}^N(x) = \langle \eta_\alpha(0) \bar{\eta}_\beta(x) \rangle$ , where  $\alpha, \beta$  are the Dirac indices of the nucleon currents. In total, there are six different nucleon correlators: the diagonal  $\eta_1 \bar{\eta}_1$ , diagonal  $\eta_2 \bar{\eta}_2$ , and off-diagonal  $\eta_1 \bar{\eta}_2$  correlators, each contracted with either the identity or  $\gamma \cdot x$  (see Table V). Let us focus on the first two of these correlation functions. For more detail, we refer the reader to Schäfer *et al.* (1994) and references therein. The OPE predicts (Belyaev and Ioffe, 1982)

$$\frac{\Pi_1^N}{\Pi_2^{N0}} = \frac{\pi^2}{12} |\langle \bar{q}q \rangle| \tau^3, \quad (281)$$

$$\frac{\Pi_2^N}{\Pi_2^{N0}} = 1 + \frac{1}{768} \langle G^2 \rangle \tau^4 + \frac{\pi^4}{72} |\langle \bar{q}q \rangle|^2 \tau^6. \quad (282)$$

The vector components of the diagonal correlators receive perturbative quark-loop contributions, which are

dominant at short distances. The scalar components of the diagonal correlators, as well as the off-diagonal correlation functions, are sensitive to chiral symmetry breaking, and the OPE starts at order  $\langle \bar{q}q \rangle$  or higher. Single-instanton corrections to the correlation functions were calculated by Dorokhov and Kochelev (1990), and Forkel and Banerjee (1993).<sup>44</sup> Instantons introduce additional, regular contributions in the scalar channel and violate the factorization assumption for the four-quark condensates. As in the pion case, both of these effects increase the attraction already seen in the OPE.

The correlation function  $\Pi_2^N$  in the interacting ensemble is shown in Fig. 22. There is a significant enhancement over the perturbative contribution which corresponds to a tightly bound nucleon state with a large coupling constant. Numerically, we find<sup>45</sup>  $m_N = 1.019$  GeV (see Table VI). In the random ensemble, we have measured the nucleon mass at smaller quark masses and found  $m_N = 0.960 \pm 0.30$  GeV. The nucleon mass is fairly insensitive to the instanton ensemble. However, the strength of the correlation function depends on the instanton ensemble. This is reflected by the value of the nucleon coupling constant, which is smaller in the interacting model.

Figure 22 also shows the nucleon correlation function measured in a quenched lattice simulation (Chu *et al.*, 1993b). The agreement with the instanton liquid results is quite impressive, especially given the fact that, before the lattice calculations were performed, there was no phenomenological information on the value of the nucleon coupling constant and the behavior of the correlation function at intermediate and large distances.

The fitted position of the threshold is  $E_0 \approx 1.8$  GeV, larger than the mass of the first nucleon resonance, the Roper  $N^*(1440)$ , and above the  $\pi\Delta$  threshold  $E_0 = 1.37$  GeV. This might indicate that the coupling of the nucleon current to the Roper resonance is small. In the case of the  $\pi\Delta$  continuum, this can be checked directly using the phenomenologically known coupling constants. The large value of the threshold energy also implies that there is little strength in the (unphysical) three-quark continuum. The fact that the nucleon is

<sup>44</sup>The latter paper corrects a few mistakes in the original work by Dorokhov and Kochelev.

<sup>45</sup>Note that this value corresponds to a relatively large current quark mass,  $m = 30$  MeV.-

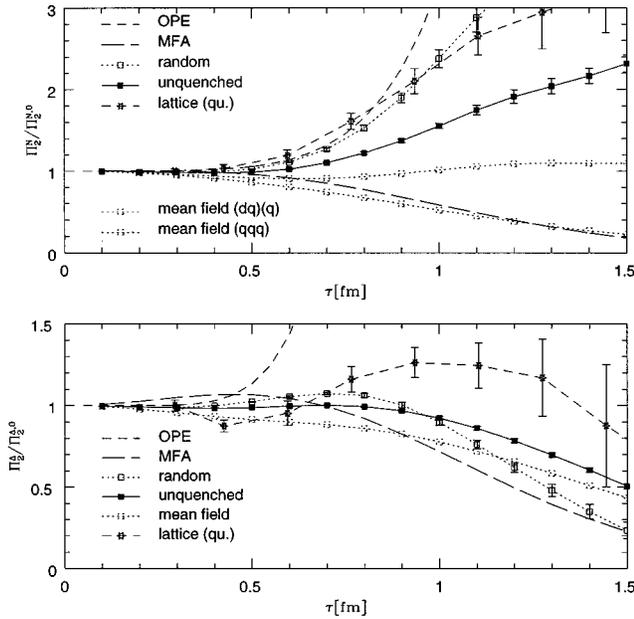


FIG. 22. Nucleon and delta correlation functions  $\Pi_2^N$  and  $\Pi_2^\Delta$ . Curves labeled as in Fig. 19.

deeply bound can also be demonstrated by comparing the full nucleon correlation function with that of three noninteracting quarks (see Fig. 22). The full correlator is significantly larger than the noninteracting (mean-field) result, indicating the presence of a strong, attractive interaction.

Some of this attraction is due to the scalar diquark content of the nucleon current. This raises the question whether the nucleon (in our model) is a strongly bound diquark very loosely coupled to a third quark. In order to check this, we have decomposed the nucleon correlation function into quark and diquark components. Using the mean-field approximation, that means treating the nucleon as a noninteracting quark-diquark system, for which we get the correlation function labeled  $(dq)(q)$  in Fig. 22. We observe that the quark-diquark model explains some of the attraction seen in  $\Pi_2^N$  but falls short of the numerical results. This means that, while diquarks may play some role in making the nucleon bound, there are substantial interactions in the quark-diquark system. Another hint of the qualitative role of diquarks is provided by the values of the nucleon coupling constants  $\lambda_N^{1,2}$ . Using Eq. (280), we can translate these results into the coupling constants  $\lambda_N^{s,p}$  of nucleon currents built from scalar or pseudoscalar diquarks. We find that the coupling to the scalar diquark current  $\eta_s = \epsilon_{abc}(u^a C \gamma_5 d^b) u^c$  is an order of magnitude larger than the coupling to the pseudoscalar current  $\eta_p = \epsilon_{abc}(u^a C d^b) \gamma_5 u^c$ . This is in agreement with the idea that the scalar diquark channel is very attractive and that these configurations play an important role in the nucleon wave function.

## 2. Delta correlation functions

In the case of the delta resonance, there exists only one independent current, given by (for the  $\Delta^{++}$ )

TABLE VI. Nucleon and delta parameters in the different instanton ensembles. RILM=random instanton liquid model. All quantities are given in units of GeV. The current quark mass is  $m_u = m_d = 0.1\Lambda$ .

	Unquenched	Quenched	RILM	Ratio ansatz (unquenched)
$m_N$	1.019	1.013	1.040	0.983
$\lambda_N^1$	0.026	0.029	0.037	0.021
$\lambda_N^2$	0.061	0.074	0.093	0.048
$m_\Delta$	1.428	1.628	1.584	1.372
$\lambda_\Delta$	0.027	0.040	0.036	0.026

$$\eta_\mu^\Delta = \epsilon_{abc}(u^a C \gamma_\mu u^b) u^c. \quad (283)$$

However, the spin structure of the correlator  $\Pi_{\mu\nu;\alpha\beta}^\Delta(x) = \langle \eta_{\mu\alpha}^\Delta(0) \bar{\eta}_{\nu\beta}^\Delta(x) \rangle$  is much richer. In general, there are ten independent tensor structures, but the Rarita-Schwinger constraint,  $\gamma^\mu \eta_\mu^\Delta = 0$ , reduces this number to four; see Table V. The OPE predicts

$$\frac{\Pi_1^\Delta}{\Pi_2^{\Delta 0}} = 4 \frac{\pi^2}{12} |\langle \bar{q}q \rangle| \tau^3, \quad (284)$$

$$\frac{\Pi_2^\Delta}{\Pi_2^{\Delta 0}} = 1 - \frac{25}{18} \frac{1}{768} \langle G^2 \rangle \tau^4 + 6 \frac{\pi^4}{72} |\langle \bar{q}q \rangle|^2 \tau^6, \quad (285)$$

which implies that power corrections, in particular those due to the quark condensate, are much larger for the delta than for the nucleon. On the other hand, nonperturbative effects due to instantons are much smaller. The reason is that, while there are large, direct instanton contributions in the nucleon, there are none in the delta.

The delta correlation function in the instanton liquid is shown in Fig. 22. The result is qualitatively different from that for the nucleon channel: the correlator at intermediate distance,  $x \approx 1$  fm, is significantly smaller and close to perturbation theory. This is in agreement with the results of the lattice calculation (Chu *et al.*, 1993b). Note that, again, this is a quenched result, which should be compared to the predictions of the random instanton model.

The mass of the delta resonance is too large in the random model, but closer to experiment in the unquenched ensemble. Note that just as for the nucleon, part of this discrepancy is due to the value of the current mass. Nevertheless, the delta-nucleon mass splitting in the unquenched ensemble is  $m_\Delta - m_N = 409$  MeV, still too large as compared to the experimental value 297 MeV. Just as for the  $\rho$  meson, there is no interaction in the delta channel to first order in the instanton density. However, if we compare the correlation function with the mean-field approximation based on the full propagator (see Fig. 22), we find evidence for substantial attraction between the quarks. Again, more detailed checks, for example, concerning the coupling to the  $\pi N$  continuum, are necessary.

## E. Correlation functions on the lattice

The study of hadronic (point-to-point) correlation functions on the lattice was pioneered by the MIT group

(Chu *et al.*, 1993a, 1993b), who measured correlation functions of the  $\pi$ ,  $\delta$ ,  $\rho$ ,  $a_1$ ,  $N$ , and  $\Delta$  in quenched QCD. The correlation functions were calculated on a  $16^3 \times 24$  lattice at  $6/g^2 = 5.7$ , corresponding to a lattice spacing of  $a \approx 0.17$  fm. A more detailed investigation of baryonic correlation functions on the lattice was subsequently carried out by Leinweber (1995a, 1995b). We have already shown some of the results of the MIT group in Figs. 19–22. The correlators were measured for distances up to  $\sim 1.5$  fm. Using the parametrization introduced above, they extracted ground-state masses and coupling constants and found good agreement with phenomenological results. What is even more important, they found the full correlation functions to agree with the predictions of the instanton liquid, even in channels (like the nucleon and delta) where no phenomenological information was available.

In order to check this result in greater detail, they also studied the behavior of the correlation functions under cooling (Chu *et al.*, 1994). The cooling procedure was monitored by studying a number of gluonic observables, like the total action, the topological charge, and the Wilson loop. From these observables, the authors concluded that the configurations were dominated by interacting instantons after  $\sim 25$  cooling sweeps. Instanton/anti-instanton pairs were continually lost during cooling, and after  $\sim 50$  sweeps the topological charge fluctuations were consistent with a dilute gas. The characteristics of the instanton liquid have already been discussed in Sec. III.C.2. After 50 sweeps the action was reduced by a factor  $\sim 300$ , while the string tension (measured from  $7 \times 4$  Wilson loops) had dropped by a factor of 6.

The behavior of the pion and nucleon correlation functions under cooling is shown in Fig. 23. The behavior of the  $\rho$  and  $\Delta$  correlators is quite similar. During the cooling process, the scale was readjusted by keeping the nucleon mass fixed. This introduced only a small uncertainty; the change in scale was  $\sim 16\%$ . We observe that the correlation functions are *stable under cooling*. They agree almost within error bars. This can also be seen from the extracted masses and coupling constants. While  $m_N$  and  $m_\pi$  are stable by definition,  $m_\rho$  and  $g_\rho$  change by less than 2%,  $\lambda_\pi$  by 7%, and  $\lambda_N$  by 1%. Only the delta mass is too small after cooling; it changes by 27%.

### F. Gluonic correlation functions

One of the most interesting problems in hadronic spectroscopy is whether one can identify glueballs, bound states of pure glue, among the spectrum of observed hadrons. This question has two aspects. In pure glue theory, stable glueball states exist, and they have been studied for a number of years in lattice simulations. In full QCD, glueballs mix with quark states, making it difficult to identify glueball candidates unambiguously.

Even in pure gauge theory, lattice simulations still require large numerical efforts. Nevertheless, a few results appear to be firmly established (Weingarten, 1994): (i) The lightest glueball is the scalar  $0^{++}$ , with a mass in the 1.5–1.8 GeV range. (ii) The tensor glueball is signifi-

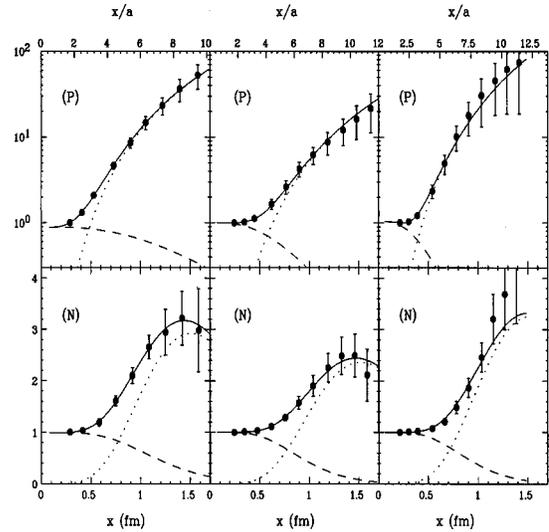


FIG. 23. Behavior of pion ( $P$ ) and nucleon ( $N$ ) correlation functions under cooling, from Chu *et al.*, 1994: Left panels, results in the original ensemble; center panels, after 25 cooling sweeps; right panels, after 50 cooling sweeps. The solid lines show fits to the data based on a pole-plus-continuum model for the spectral function. The dotted and dashed lines show the individual contributions from the pole and the continuum parts.

cantly heavier,  $m_{2^{++}}/m_{0^{++}} \approx 1.4$ , and the pseudoscalar is heavier still,  $m_{0^{-+}}/m_{0^{++}} = 1.5\text{--}1.8$  (Bali *et al.*, 1993). (iii) The scalar glueball is much smaller than other glueballs. The size of the scalar is  $r_{0^{++}} \approx 0.2$  fm, while  $r_{2^{++}} \approx 0.8$  fm (de Forcrand and Liu, 1992). For comparison, a similar measurement for the  $\pi$  and  $\rho$  mesons gives 0.32 fm and 0.45 fm, indicating that spin-dependent forces are stronger between gluons than between quarks.

Gluonic currents with the quantum numbers of the lowest glueball states are the field strength squared ( $S = 0^{++}$ ), the topological charge ( $P = 0^{-+}$ ), and the energy-momentum tensors ( $T = 2^{++}$ ):

$$j_S = (G_{\mu\nu}^a)^2, \quad j_P = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} G_{\mu\nu}^a G_{\rho\sigma}^a, \quad j_T = \frac{1}{4} (G_{\mu\nu}^a)^2 - G_{0\alpha}^a G_{0\alpha}^a. \quad (286)$$

The short-distance behavior of the corresponding correlation functions is determined by the OPE (Novikov *et al.*, 1980): to give

$$\Pi_{S,P}(x) = \Pi_{S,P}^0 \left( 1 \pm \frac{\pi^2}{192g^2} \langle f^{abc} G_{\mu\nu}^a G_{\nu\beta}^b G_{\beta\mu}^c \rangle x^6 + \dots \right) \quad (287)$$

$$\Pi_T(x) = \Pi_T^0 \left( 1 + \frac{25\pi^2}{9216g^2} \langle 2\mathcal{O}_1 - \mathcal{O}_2 \rangle \log(x^2) x^8 + \dots \right), \quad (288)$$

where we have defined the operators  $\mathcal{O}_1 = (f^{abc} G_{\mu\alpha}^b G_{\nu\alpha}^c)^2$ , and  $\mathcal{O}_2 = (f^{abc} G_{\mu\nu}^b G_{\alpha\beta}^c)^2$ , and the free correlation functions are given by

$$\Pi_{S,P}(x) = (\pm) \frac{384g^4}{\pi^4 x^8}, \quad \Pi_T(x) = \frac{24g^4}{\pi^4 x^8}. \quad (289)$$

Power corrections in the glueball channels are remarkably small. The leading-order power correction  $O(\langle G^2 \rangle/x^4)$  vanishes,<sup>46</sup> while radiative corrections of the form  $\alpha_s \log(x^2) \langle G^2 \rangle/x^4$  [not included in Eq. (287)], or higher-order power corrections like  $\langle f^{abc} G_{\mu\nu}^a G_{\nu\rho}^b G_{\rho\mu}^c \rangle/x^2$ , are very small.

On the other hand, there is an important low-energy theorem that controls the large-distance behavior of the scalar correlation function (Novikov *et al.*, 1979),

$$\int d^4x \Pi_S(x) = \frac{128\pi^2}{b} \langle G^2 \rangle, \quad (290)$$

where  $b$  denotes the first coefficient of the beta function. In order to make the integral well defined, we have to subtract the constant term  $\sim \langle G^2 \rangle^2$  as well as singular (perturbative) contributions to the correlation function. Analogously, the integral over the pseudoscalar correlation functions is given by the topological susceptibility  $\int d^4x \Pi_P(x) = \chi_{\text{top}}$ . In pure gauge theory  $\chi_{\text{top}} \approx (32\pi^2) \langle G^2 \rangle$ , while in unquenched QCD  $\chi_{\text{top}} = O(m)$  (see Sec. V.E). These low-energy theorems indicate the presence of rather large nonperturbative corrections in the scalar glueball channels. This can be seen as follows: We can incorporate the low-energy theorem into the sum rules by using a subtracted dispersion relation

$$\frac{\Pi(Q^2) - \Pi(0)}{Q^2} = \frac{1}{\pi} \int ds \frac{\text{Im} \Pi(s)}{s(s+Q^2)}. \quad (291)$$

In this case, the subtraction constant acts like a power correction. In practice, however, the subtraction constant totally dominates ordinary power corrections. For example, using pole dominance, the scalar glueball coupling  $\lambda_S = \langle 0 | j_S | 0^{++} \rangle$  is completely determined by the subtraction,  $\lambda_S^2/m_S^2 \approx (128\pi^2/b) \langle G^2 \rangle$ .

For this reason, we expect instantons to give a large contribution to scalar glueball correlation functions. Expanding the gluon operators around the classical fields, we have

$$\begin{aligned} \Pi_S(x, y) &= \langle 0 | G^{\text{cl}}(x) G^{\text{cl}}(y) | 0 \rangle + \langle 0 | G_{\mu\nu}^{a, \text{cl}}(x) \\ &\quad \times [D_\mu^x D_\alpha^y D_{\nu\beta}(x, y)]^{ab} G_{\alpha\beta}^{b, \text{cl}}(y) | 0 \rangle + \dots, \end{aligned} \quad (292)$$

where  $D_{\mu\nu}^{ab}(x, y)$  is the gluon propagator in the classical background field. If we insert the classical field of an instanton, we find (Novikov *et al.*, 1979; Shuryak, 1982c; Schäfer and Shuryak, 1995a)

$$\begin{aligned} \Pi_{S,P}^{\text{SIA}}(x) &= \int d\rho n(\rho) \frac{8192\pi^2}{\rho^4} \frac{\partial^3}{\partial(x^2)^3} \\ &\quad \times \left\{ \frac{\xi^6}{x^6} \left( \frac{10 - 6\xi^2}{(1 - \xi^2)^2} + \frac{3}{\xi} \log \frac{1 + \xi}{1 - \xi} \right) \right\}, \end{aligned} \quad (293)$$

where  $\xi$  is defined as in Eq. (266). There is no classical contribution in the tensor channel, since the stress ten-

<sup>46</sup>There is a  $\langle G^2 \rangle \delta^4(x)$  contact term in the scalar glueball correlators which, depending on the choice of sum rule, may enter momentum space correlation functions.

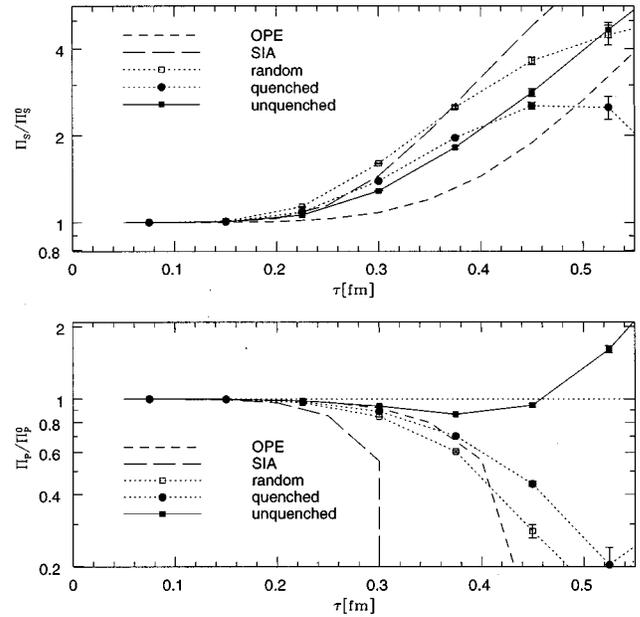


FIG. 24. Scalar and pseudoscalar glueball correlation functions. Curves labeled as in Fig. 19.

sor in the self-dual field of an instanton is zero. Note that the perturbative contributions in the scalar and pseudoscalar channels have opposite sign, while the classical contribution has the same sign. We therefore find, to first order in the instanton density, the three scenarios discussed in Sec. VI.C: attraction in the scalar channel, repulsion in the pseudoscalar, and no effect in the tensor channel. The single-instanton prediction is compared with the OPE in Fig. 24. We can clearly see that classical fields are much more important than the power corrections.

Quantum corrections to this result can be calculated from the second term in Eq. (292), using the gluon propagator in the instanton field (L. S. Brown *et al.*, 1978; Levine and Yaffe, 1979). The singular contributions correspond to the OPE in the instanton field. There is an analog of the Dubovikov-Smilga result for glueball correlators: In a general self-dual background field, there are no power corrections to the tensor correlator (Novikov *et al.*, 1980). This is consistent with the result (288), since the combination  $\langle 2\mathcal{O}_1 - \mathcal{O}_2 \rangle$  vanishes in a self-dual field. Moreover, the sum of the scalar and pseudoscalar glueball correlators does not receive any power corrections, while the difference does, starting at  $O(G^3)$ .

Numerical calculations of glueball correlators in different instanton ensembles were performed by Schäfer and Shuryak (1995a). At short distances, the results were consistent with the single-instanton approximation. At larger distances, the scalar correlator was modified due to the presence of the gluon condensate. This means that (like the  $\sigma$  meson), the correlator has to be subtracted, and the determination of the mass is difficult. In the pure gauge theory we find  $m_{0^{++}} \approx 1.5 \text{ GeV}$  and  $\lambda_{0^{++}} = 16 \pm 2 \text{ GeV}^3$ . While the mass is consistent with QCD sum-rule predictions, the coupling is much larger

than would be expected from calculations that do not enforce the low-energy theorem (Narison, 1984; Bagan and Steele, 1990).

In the pseudoscalar channel the correlator is very repulsive, and there is no clear indication of a glueball state. In the full theory (with quarks) the correlator is modified due to topological charge screening. The non-perturbative correction changes sign and a light (on the glueball mass scale) state, the  $\eta'$ , appears. Nonperturbative corrections in the tensor channel are very small. Isolated instantons and anti-instantons have a vanishing energy-momentum tensor, so the result is entirely due to interactions.

In Schäfer and Shuryak (1995a) we also measured glueball wave functions. The most important result is that the scalar glueball is indeed small,  $r_{0^{++}}=0.2$  fm, while the tensor is much bigger,  $r_{2^{++}}=0.6$  fm. The size of the scalar is determined by the size of an instanton, whereas in the case of the tensor the scale is set by the average distance between instantons. This number is comparable to the confinement scale, so the tensor wave function is probably not very reliable. On the other hand, the scalar is much smaller than the confinement scale, so the wave function of the  $0^{++}$  glueball may provide an indication of the importance of instantons in pure gauge theory.

### G. Hadronic structure and $n$ -point correlators

So far, we have focused on two-point correlation in the instanton liquid. However, in order to study hadronic properties like decay widths, form, structure functions, etc., we have to calculate  $n$ -point correlators. There is no systematic study of these objects in the literature. In the following, we discuss two exploratory attempts and point out a number of phenomenologically interesting questions.

Both of our examples are related to the question of hadronic sizes. Hadronic wave functions (or Bethe-Salpeter amplitudes) are defined by three-point correlators of the type

$$\begin{aligned} \Pi_{\pi}(x,y) &= \langle 0 | T(\bar{d}(x) P e^{i \int_x^{x+y} A(x') dx'} \gamma_5 u(x+y) \\ &\quad \times \bar{d}(0) \gamma_5 u(0)) | 0 \rangle \\ &\xrightarrow{x \rightarrow \infty} \psi(y) e^{-m_{\pi} x}. \end{aligned} \quad (294)$$

These wave functions are not directly accessible to experiment, but they have been studied in a number of lattice gauge simulations, both at zero (Velikson and Weingarten, 1985; Chu, Lissia, and Negele, 1991; Hecht and DeGrand, 1992; Gupta, Daniel, and Grandy, 1993) and at finite temperature (Bernard *et al.*, 1992; Schramm and Chu, 1993). In the single-instanton approximation, we find that light states (like the pion or the nucleon) receive direct instanton contributions, so their size is controlled by the typical instanton size. Particles like the  $\rho$  or the  $\Delta$ , on the other hand, are not sensitive to direct instantons and are therefore less bound and larger in size.

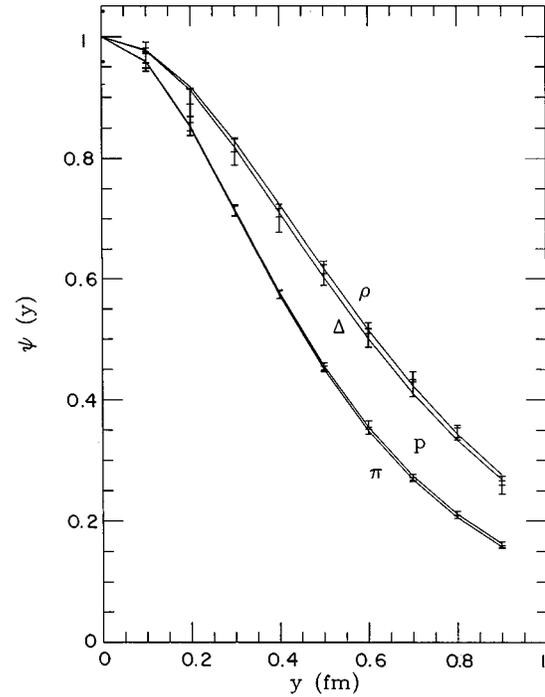


FIG. 25. Hadronic wave functions of the pion, rho meson, proton, and delta resonances in the random instanton ensemble.

This can be seen from Fig. 25, where we show results obtained in the random instanton ensemble (Schäfer and Shuryak, 1995a). In particular, we observe that the pion and the proton, as well as the  $\rho$  and the  $\Delta$ , have essentially the same size. In the case of the pion and the proton, this is in agreement with lattice data reported by Chu *et al.* (1991). These authors also find that the  $\rho$  meson is significantly larger (they did not study the  $\Delta$ ).

A more detailed comparison with the lattice data reveals some of the limitations of the instanton model. Lattice wave functions are linear at the origin, not quadratic as in the instanton liquid. Presumably, this is due to the lack of a perturbative Coulomb interaction between the quarks in the instanton model. Also, lattice wave functions decay faster at large distances, which might be related to the absence of a string potential in the instanton liquid.

Let us now focus on another three-point function that is more directly related to experiment. The pion electromagnetic form factor is determined by the correlation function

$$\Gamma_{\mu}(x,y) = \langle j_5^+(-x/2) j_{\mu}^3(y) j_5^-(x/2) \rangle, \quad (295)$$

where  $j_5^{\pm}$  are the pseudoscalar currents of the charged pion and  $j_{\mu}^3$  is the third component of the isovector-vector current. In this case, the correlator is completely determined by the triangle diagram. In the context of QCD sum rules, the pion form factor is usually analyzed using a three-point function built from axial-vector currents (because the pseudoscalar sum rule is known to be unreliable). In the single-instanton approximation, the problem was recently studied by Forkel and Nielsen (1995), where it was shown that, when direct instantons

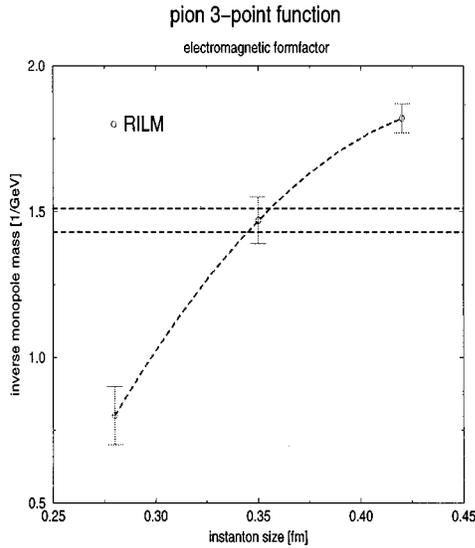


FIG. 26. The size of the pion, determined from the inverse monopole mass in the pion electromagnetic form, as a function of the instanton size in the random instanton ensemble, from Blotz and Shuryak, 1997. The horizontal lines show the experimental uncertainty.

are included (as in Sec. VI.C.2), the pion form factor can also be extracted from the pseudoscalar correlator (295). Using the standard instanton liquid parameters, these authors calculated the pion form factor for  $Q^2 \sim 1$  GeV, obtaining good agreement with experimental data.

Recently, the three-point function (295) in the instanton liquid was also calculated numerically (Blotz and Shuryak, 1997). In this case, one can go to arbitrarily large distances and determine the pion form factor at small momentum. In this regime, the pion form factor has a simple monopole shape,  $F_\pi(Q^2) = M^2/(Q^2 + M^2)$ , with a characteristic mass close to the rho-meson mass. However, at intermediate momenta,  $Q^2 \sim 1$ , the pion form factor is less sensitive to the meson cloud and largely determined by the quark-antiquark interaction inside the pion. In the instanton model, the interaction is dominated by single-instanton effects, and the range is controlled by the instanton size. This can be seen from Fig. 26, where we show the fitted monopole mass as a function of the instanton size. Clearly, the two are strongly correlated, and the instanton model reproduces the experimental value of the monopole mass if the mean instanton size is  $\rho \approx 0.35$  fm.

Finally, let us mention a few experimental observables that might be interesting in connection with instanton effects in hadronic structure. These are observables that are sensitive to the essential features of the 't Hooft interaction, the strong correlation between quark and gluon polarization and flavor mixing. The first is the flavor singlet axial coupling constant of the nucleon  $g_A^0$ , which is a measure of the quark contribution to the nucleon spin. This quantity can be determined in polarized, deep-inelastic scattering and was found to be unex-

pectedly small,  $g_A^0 \approx 0.35$ . The quark model suggests  $g_A^0 \approx 1$ , a discrepancy which is known as the “proton spin crisis.”

Since  $g_A^0$  is defined by a matrix element of the flavor singlet current, one might suspect that the problem is somehow related to the anomaly. In fact, if it were not for the anomaly, one could not transfer polarization from quarks to gluons and  $g_A^0$  could not evolve. All this suggests that instantons might be crucial in understanding the proton spin structure. While some attempts in this direction have been made (Forte and Shuryak, 1991; Dorokhov, Kochelev, and Zubov, 1993), a realistic calculation is still lacking.

Another question concerns the flavor structure of the unpolarized quark sea in the nucleon. From measurements of the Gottfried sum rule, the quark sea is known to be flavor asymmetric; there are more  $d$  than  $u$  quarks in the proton sea. From a perturbative point of view this is puzzling because gluons are flavor blind, so a radiatively generated sea is flavor symmetric. However, as pointed out by Dorokhov and Kochelev (1993), quark pairs produced by instantons are flavor asymmetric. For example, a valence  $u$  quark generates  $\bar{d}d, \bar{s}s$  sea, but no  $\bar{u}u$  pairs. Since a proton has two valence  $u$  quarks, this gives the right sign of the observed asymmetry.

## VII. INSTANTONS AT FINITE TEMPERATURE

### A. Introduction

#### 1. Finite-temperature field theory and the caloron solution

In the previous sections we have shown that the instanton liquid model provides a mechanism for chiral symmetry breaking and describes a large number of hadronic correlation functions. Clearly, it is of interest to generalize the model to finite temperature and/or density. This will allow us to study the behavior of hadrons in matter, the mechanism of the chiral phase transition, and possible nonperturbative effects in the high-temperature/density phase. Extending the methods described in the last three sections to nonzero temperature is fairly straightforward. In Euclidean space, finite temperature corresponds to periodic boundary conditions on the fields. Basically, all we have to do is replace all the gauge potentials and fermionic wave functions by their  $T \neq 0$  periodic counterparts. Nevertheless, doing so leads to a number of interesting and very nontrivial phenomena, which we describe in detail below. Extending the instanton model to finite density is more difficult. In Euclidean space a finite chemical potential corresponds to a complex weight in the functional integral, so applying standard methods from statistical mechanics is less straightforward. While some work on the subject has been done (Carvalho, 1980; Chemtob, 1981; Shuryak, 1982d; Abrikosov, 1983), many questions remain to be understood.

Before we study the instanton liquid at  $T \neq 0$ , we should like to give a very brief introduction to finite-temperature field theory in Euclidean space (Shuryak,

1980; Gross, Pisarski, and Yaffe, 1981; McLerran, 1986; Kapusta, 1989). The basic object is the partition function

$$Z = \text{Tr}(e^{-\beta H}), \tag{296}$$

where  $\beta=1/T$  is the inverse temperature,  $H$  is the Hamiltonian, and the trace is performed over all physical states. In Euclidean space, the partition function can be written as a functional integral,

$$Z = \int_{\text{per}} DA_\mu \int_{\text{aper}} D\bar{\psi} D\psi \exp\left(-\int_0^\beta d\tau \int d^3x \mathcal{L}\right), \tag{297}$$

where the gauge fields and fermions are subject to periodic/antiperiodic boundary conditions

$$A_\mu(\vec{r}, \beta) = A_\mu(\vec{r}, 0), \tag{298}$$

$$\psi(\vec{r}, \beta) = -\psi(\vec{r}, 0), \quad \bar{\psi}(\vec{r}, \beta) = -\bar{\psi}(\vec{r}, 0). \tag{299}$$

The boundary conditions imply that the fields can be expanded in a Fourier series  $\phi(\vec{r}, \tau) = e^{i\omega_n \tau} \phi_n(\vec{r})$ , where  $\omega_n = 2\pi n T$  and  $\omega_n = (2n+1)\pi T$  are the Matsubara frequencies for boson and fermions.

As an example, it is instructive to consider the free propagator of a massless fermion at finite temperature. Summing over all Matsubara frequencies we get

$$S_T(r, \tau) = \frac{i}{4\pi^2} \gamma \cdot \partial \sum_n \frac{(-1)^n}{r^2 + (\tau - n\beta)^2}. \tag{300}$$

The sum can easily be performed, and in the spatial direction one finds

$$S_T(r, 0) = \frac{i\vec{\gamma} \cdot \vec{r}}{2\pi^2 r^4 z} \exp(-z) \times \frac{(z+1) + (z-1)\exp(-2z)}{[1 + \exp(-2z)]^2}, \tag{301}$$

where  $z = \pi r T$ . This result shows that, at finite temperature, the propagation of massless fermions in the spatial direction is exponentially suppressed. The screening mass  $m = \pi T$  is the lowest Matsubara frequency for fermions at finite temperature.<sup>47</sup> The energy  $\pi T$  acts like a (chiral) mass term for spacelike propagation. For bosons, on the other hand, the wave functions are periodic, the lowest Matsubara frequency is zero, and the propagator is not screened. The propagator in the temporal direction is given by

$$S_T(0, \tau) = \frac{i\gamma_4}{2\pi^2 \tau^3} \frac{y^3}{2} \frac{1 + \cos^2(y)}{\sin^3(y)}, \tag{302}$$

with  $y = \pi \tau T$ . Clearly, there is no suppression factor for propagation in the temporal direction.

<sup>47</sup>Let us mention another explanation for this fact which does not use the Matsubara formalism. In the real-time formalism,  $S_T(r, 0)$  is the spatial Fourier transform of  $[1/2 - n_f(E_k)]$ , where  $n_f$  is the Fermi-Dirac distribution. For small momenta,  $k < T$ , the two terms cancel and there is no power-law decay. In other words, the  $1/r^3$  behavior disappears because the contributions from real and virtual fermions cancel each other.

Periodic instanton configurations can be constructed by lining up zero-temperature instantons along the imaginary time direction with the period  $\beta$ . Using 't Hooft's multi-instanton solution (A7), the explicit expression for the gauge field is given by Harrington and Shepard (1978)

$$A_\mu^a = \bar{\eta}_{\mu\nu}^a \Pi(x) \partial_\nu \Pi^{-1}(x), \tag{303}$$

where

$$\Pi(x) = 1 + \frac{\pi\rho^2}{\beta r} \frac{\sinh(2\pi r/\beta)}{\cosh(2\pi r/\beta) - \cos(2\pi\tau/\beta)}. \tag{304}$$

Here,  $\rho$  denotes the size of the instanton. The solution (303) is sometimes referred to as the caloron. It has topological charge  $Q=1$  and action  $S=8\pi^2/g^2$  independent of temperature.<sup>48</sup> A caloron with  $Q=-1$  can be constructed by making the replacement  $\bar{\eta}_{\mu\nu}^a \rightarrow \eta_{\mu\nu}^a$ . Of course, as  $T \rightarrow 0$  the caloron field reduces to the field of an instanton. In the high-temperature limit  $T\rho \gg 1$ , however, the field looks very different (Gross *et al.*, 1981). In this case, the caloron develops a core of size  $O(\beta)$  where the fields are very strong,  $G_{\mu\nu} \sim O(\beta^2)$ . In the intermediate regime,  $O(\beta) < r < O(\rho^2/\beta)$ , the caloron looks like a ( $T$ -independent) dyon with unit electric and magnetic charges,

$$E_i^a = B_i^a \approx \frac{\hat{r}^a \hat{r}^i}{r^2}. \tag{305}$$

In the far region,  $r > O(\rho^2/\beta)$ , the caloron resembles a three-dimensional dipole field,  $E_i^a = B_i^a \sim O(1/r^3)$ .

At finite temperature, tunneling between degenerate classical vacua is related to the anomaly in exactly the same way as it is at  $T=0$ . This means that, during tunneling, the fermion vacuum is rearranged and the Dirac operator in the caloron field has a normalizable left-handed zero mode. This zero mode can be constructed from the zero modes of the exact  $n$ -instanton solution (Grossman, 1977). The result is

$$\psi_i^a = \frac{1}{2\sqrt{2}\pi\rho} \sqrt{\Pi(x)} \partial_\mu \left( \frac{\Phi(x)}{\Pi(x)} \right) \left( \frac{1 - \gamma_5}{2} \gamma_\mu \right)_{ij} \epsilon_{aj}, \tag{306}$$

where  $\Phi(x) = (\Pi(x) - 1) \cos(\pi\tau/\beta) / \cosh(\pi r/\beta)$ . Note that the zero-mode wave function also shows an exponential decay  $\exp(-\pi r T)$  in the spatial direction, despite the fact the eigenvalue of the Dirac operator is exactly zero. This will have important consequences for instanton interactions at nonzero temperature.

## 2. Instantons at high temperature

At finite temperature, just as at  $T=0$ , the instanton density is controlled by fluctuations around the classical

<sup>48</sup>The topological classification of smooth gauge fields at finite temperature is more complicated than at  $T=0$  (Gross *et al.*, 1981). The situation simplifies in the absence of magnetic charges. In this case topological charge is quantized just as it is at zero temperature.

caloron solution. In the high-temperature plasma phase, the gluoelectric fields in the caloron are Debye screened, so we expect that instantons are strongly suppressed at high temperature (Shuryak, 1978b). The perturbative Debye mass in the quark-gluon plasma is (Shuryak, 1978a)

$$m_D^2 = (N_c/3 + N_f/6)g^2 T^2. \quad (307)$$

Normal  $O(1)$  electric fields are screened at distances  $1/(gT)$ , while the stronger  $O(1/g)$  nonperturbative fields of the instantons should be screened for sizes  $\rho > T^{-1}$ . An explicit calculation of the quantum fluctuations around the caloron was performed by Pisarski and Yaffe (1980). Their result is

$$n(\rho, T) = n(\rho, T=0) \times \exp\left(-\frac{1}{3}(2N_c + N_f)(\pi\rho T)^2 - B(\lambda)\right)$$

with

$$B(\lambda) = \left(1 + \frac{N_c}{6} - \frac{N_f}{6}\right) \times \left[-\log\left(1 + \frac{\lambda^2}{3}\right) + \frac{0.15}{(1 + 0.15\lambda^{-3/2})^8}\right] \quad (308)$$

where  $\lambda = \pi\rho T$ . As expected, large instantons,  $\rho \gg 1/T$ , are exponentially suppressed. This means that the instanton contribution to physical quantities like the energy density (or pressure, etc.) is of the order

$$\epsilon(T) \sim \int_0^{1/T} \frac{d\rho}{\rho^5} (\rho\Lambda)^b \sim T^4 (\Lambda/T)^b. \quad (309)$$

At high temperature, this is small compared to the energy density of an ideal gas,  $\epsilon(T)_{SB} \sim T^4$ .

It was emphasized by Shuryak and Velkovsky (1994) that, although the Pisarski-Yaffe result contains only one dimensionless parameter  $\lambda$ , its applicability is controlled by two separate conditions:

$$\rho \ll 1/\Lambda, \quad T \gg \Lambda. \quad (310)$$

The first condition ensures the validity of the semiclassical approximation, while the second justifies the perturbative treatment of the heat bath. In order to illustrate this point, we should like to discuss the derivation of the semiclassical result [Eq. (308)] in somewhat greater detail. Our first point is that the finite-temperature correction to the instanton density can be split into two parts of different physical origin. For this purpose, let us consider the determinant of a scalar field in the fundamental representation.<sup>49</sup> The temperature-dependent part of the one-loop effective action

$$\log \det\left(\frac{-D^2}{-\partial^2}\right)\Bigg|_T = \log \det\left(\frac{-D^2}{-\partial^2}\right)\Bigg|_{T=0} + \delta \quad (311)$$

<sup>49</sup>The quark and gluon (non-zero-mode) determinants can be reduced to the determinant of a scalar field in the fundamental and adjoint representation (Gross *et al.*, 1981).

can be split into two pieces,  $\delta = \delta_1 + \delta_2$ , where

$$\delta_1 = \text{Tr}_T \log\left(\frac{-D^2(A(\rho))}{-\partial^2}\right) - \text{Tr} \log\left(\frac{-D^2(A(\rho))}{-\partial^2}\right), \quad (312)$$

$$\delta_2 = \text{Tr}_T \log\left(\frac{-D^2(A(\rho, T))}{-\partial^2}\right) - \text{Tr}_T \log\left(\frac{-D^2(A(\rho))}{-\partial^2}\right). \quad (313)$$

Here  $\text{Tr}_T$  includes an integration over  $R^3 \times [0, \beta]$ , and  $A(\rho, T)$  and  $A(\rho)$  are the gauge potentials of the caloron and the instanton, respectively. The two terms  $\delta_1$  and  $\delta_2$  correspond to the two terms in the exponent in the semiclassical result (308).

It was shown by Shuryak (1982d) that the physical origin of the first term is scattering of particles in the heat bath on the instanton field. The forward scattering amplitude  $T(p, p)$  of a scalar quark can be calculated using the standard Lehmann-Symanzik-Zimmermann reduction formula

$$\text{Tr}(T(p, p)) = \int d^4x d^4y e^{ip \cdot (x-y)} \text{Tr}(\partial_x^2 \Delta(x, y) \partial_y^2), \quad (314)$$

where  $\Delta(x, y)$  is the scalar quark propagator introduced in Sec. VI.B.1. By subtracting the trace of the free propagator and going to the physical pole  $p^2=0$ , one gets a very simple answer,

$$\text{Tr}(T(p, p)) = -4\pi^2 p^2. \quad (315)$$

Since the result is just a constant, there is no problem with analytic continuation to Minkowski space. Integrating the result over a thermal distribution, we get

$$\begin{aligned} \delta_1 &= \int \frac{d^3p}{(2\pi)^3} \frac{1}{2p(\exp(p/T)-1)} \text{Tr}(T(p, p)) \\ &= -\frac{1}{6}(\pi\rho T)^2 = -\frac{1}{6}\lambda^2. \end{aligned} \quad (316)$$

The constants appearing in the result are easily interpreted:  $\rho^2$  comes from the scattering amplitude, while the temperature dependence enters via the integral over the heat bath. Note also that the scattering amplitude has the same origin (and the same dependence on  $N_c, N_f$ ) as the Debye mass.

Formally, the validity of this perturbative calculation requires that  $g(T) \ll 1$ , but in QCD this criterion is satisfied only at extremely high temperatures. We would argue, however, that the accuracy of the calculation is controlled by the same effects that determine the validity of the perturbative result for the Debye mass. Available lattice data (Irback *et al.*, 1991) suggest that screening sets in right after the phase transition, and that the perturbative prediction for the Debye mass works to about 10% accuracy for  $T > 3T_c \approx 500$  MeV. Near the critical temperature one expects that the Debye mass becomes small; screening disappears together with the plasma. Below  $T_c$  the instanton density should be deter-

mined by the scattering of hadrons, not quarks and gluons, on the instanton. We shall return to this point in the next section.

The second term in the finite-temperature effective action [Eq. (313)] has a different physical origin. It is determined by quantum corrections to the colored current (Brown and Creamer, 1978), multiplied by the  $T$ -dependent variation of the instanton field, the difference between the caloron and the instanton. For  $\lambda$  small, the correction is given by

$$\delta_2 = -\frac{1}{36}\lambda^2 + O(\lambda^3). \quad (317)$$

The result has the same sign and the same dependence on the parameters as  $\delta_1$ , but a smaller coefficient. Thus finite- $T$  effects not only lead to the appearance of the usual (perturbative) heat bath, they also modify the strong  $O(1/g)$  classical gauge field of the instanton. In Euclidean space, we can think of this effect as arising from the interaction of the instanton with its periodic “mirror images” along the imaginary-time direction. However, below  $T_c$ , gluon correlators are exponentially suppressed since glueball states are very heavy. Therefore we expect that for  $T < T_c$  the instanton field is not modified by the boundary conditions. Up to effects of the order  $O[\exp(-M/T)]$ , where  $M$  is the mass of the lowest glueball, there is no instanton suppression due to the change in the classical field below  $T_c$ .

### 3. Instantons at low temperature

The behavior of the instanton density at low temperature was studied by Shuryak and Velkovsky (1994). At temperatures well below the phase transition the heat bath consists of weakly interacting pions. The interaction of the pions in the heat bath with instantons can be determined from the effective Lagrangian (111). For two flavors, this Lagrangian is a product of certain four-fermion operators and the semiclassical single-instanton density. In the last section, we argued that the semiclassical instanton density is not modified at small temperatures. The temperature dependence is then determined by the  $T$  dependence of the vacuum expectation value of the four-fermion operators. For small  $T$ , the situation can be simplified even further because the wavelength of the pions in the heat bath is large, and the four-fermion operators can be considered local.

The temperature dependence of the expectation value of a general four-fermion operator  $\langle (\bar{q}Aq)(\bar{q}Bq) \rangle$  can be determined using soft-pion techniques (Gerber and Leutwyler, 1989). To order  $T^2/f_\pi^2$  the result is determined by the matrix element of the operator between one-pion states, integrated over a thermal pion distribution. The one-pion matrix element can be calculated using the reduction formula and current-algebra commutators. These methods allow us to prove the general formula (Eletsy, 1993)

$$\langle (\bar{q}Aq)(\bar{q}Bq) \rangle_T = \langle (\bar{q}Aq)(\bar{q}Bq) \rangle_0 - \frac{T^2}{96f_\pi^2} \langle (\bar{q}\{\Gamma_5^a, \{\Gamma_5^a, A\}\}q)(\bar{q}Bq) \rangle_0 \quad (318)$$

$$- \frac{T^2}{96f_\pi^2} \langle (\bar{q}Aq)(\bar{q}\{\Gamma_5^a, \{\Gamma_5^a, B\}\}q) \rangle_0 - \frac{T^2}{48f_\pi^2} \langle (\bar{q}\{\Gamma_5^a, A\}q)(\bar{q}\{\Gamma_5^a, B\}q) \rangle_0, \quad (319)$$

where  $A, B$  are arbitrary flavor-spin-color matrices and  $\Gamma_5^a = \tau^a \gamma_5$ . Using this result, we obtain the instanton density at low temperature

$$n(\rho, T) = n(\rho) \left( \frac{4}{3} \pi^2 \rho^3 \right)^2 \left[ \langle K_1 \rangle_0 \frac{1}{4} \left( 1 - \frac{T^2}{6f_\pi^2} \right) - \langle K_2 \rangle_0 \frac{1}{12} \left( 1 + \frac{T^2}{6f_\pi^2} \right) \right], \quad (320)$$

where we have defined the two operators ( $\tau$  are isospin matrices)

$$K_1 = \bar{q}_L q_L \bar{q}_L q_L + \frac{3}{32} \bar{q}_L \lambda^a q_L \bar{q}_L \lambda^a q_L - \frac{9}{128} \bar{q}_L \sigma_{\mu\nu} \lambda^a q_L \bar{q}_L \sigma_{\mu\nu} \lambda^a q_L, \quad (321)$$

$$K_2 = \bar{q}_L \tau^i q_L \bar{q}_L \tau^i q_L + \frac{3}{32} \bar{q}_L \tau^i \lambda^a q_L \bar{q}_L \tau^i \lambda^a q_L - \frac{9}{128} \bar{q}_L \sigma_{\mu\nu} \tau^i \lambda^a q_L \bar{q}_L \sigma_{\mu\nu} \tau^i \lambda^a q_L. \quad (322)$$

Although the vacuum expectation values of these two operators are unknown, it is clear that (barring unexpected cancellations) the  $T$  dependence should be rather weak, most likely inside the range  $n = n_0(1 \pm T^2/(6f_\pi^2))$ . Furthermore, if one estimates the expectation values using the factorization assumption (166), the  $T$  dependence cancels to order  $T^2/f_\pi^2$ .

Summarizing the last two sections, we conclude that the instanton density is expected to be essentially constant below the phase transition, but exponentially suppressed at large temperature. We shall explore the physical consequences of this result in the remainder of

this section. Numerical evidence for our conclusions that comes from lattice simulations will be presented in Sec. VII.D.

#### 4. Instanton interaction at finite temperature

In order to study an interacting system of instantons at finite temperature, we need to determine the bosonic and fermionic interaction between instantons at  $T \neq 0$ . This subject was studied by Diakonov and Mirlin (1988) and later, in much greater detail, by Shuryak and Verbaarschot (1991), whom we follow here. As for  $T=0$ , the gluonic interactions between an instanton and an anti-instanton is defined as  $S_{\text{int}} = S[A_{\text{IA}}] - 2S_0$ , where we have to specify an ansatz  $A_{\text{IA}}$  for the gauge potential of the IA configuration. The main difficulty at  $T \neq 0$  is that Lorenz invariance is lost, and the interaction depends on more variables: the spatial and temporal separation  $r, \tau$  of the instantons, the sizes  $\rho_I, \rho_A$ , the relative orientation matrix  $U$ , and the temperature  $T$ . There are also more invariant structures that can be constructed from the relative orientation vector  $u_\mu$ . In the case of color SU(2), we have

$$\begin{aligned} S_{\text{IA}} = & s_0 + s_1(u \cdot \hat{R})^2 + s_2(u \cdot \hat{R})^4 \\ & + s_3(u^2 - (u \cdot \hat{R})^2 - u_4^2) \\ & + s_4(u^2 - (u \cdot \hat{R})^2 - u_4^2)^2 \\ & + s_5(u \cdot \hat{R})^2(u^2 - (u \cdot \hat{R})^2 - u_4^2), \end{aligned} \quad (323)$$

where  $s_i = s_i(r, \tau, \rho_I, \rho_A, \beta)$ . Because of the reduced symmetry of the problem, it is difficult to implement the streamline method. Instead, we shall study the interaction in the ratio ansatz. The gauge-field potential is given by a straightforward generalization of Eq. (141). Except for certain limits, the interaction cannot be obtained analytically. We therefore give a parametrization of the numerical results obtained by Shuryak and Verbaarschot (1991):

$$\begin{aligned} S_{\text{IA}} = & \frac{8\pi^2}{g^2} \left\{ \frac{4.0}{(r^2 + 2.0)^2} \frac{\beta^2}{\beta^2 + 5.21} |u|^2 \right. \\ & - \left[ \frac{1.66}{(1 + 1.68r^2)^3} + \frac{0.72 \log(r^2)}{(1 + 0.42r^2)^4} \right] \frac{\beta^2}{\beta^2 + 0.75} |u|^2 \\ & + \left[ -\frac{16.0}{(r^2 + 2.0)^2} + \frac{2.73}{(1 + 0.33r^2)^3} \right] \\ & \times \frac{\beta^2}{\beta^2 + 0.24 + 11.50r^2/(1 + 1.14r^2)} |u \cdot \hat{R}|^2 \\ & + 0.36 \log \left( 1 + \frac{\beta}{r} \right) \frac{1}{(1 + 0.013r^2)^4} \frac{1}{\beta^2 + 1.73} \\ & \left. \times (|u|^2 - |u \cdot \hat{R}|^2 - |u_4|^2) \right\}. \end{aligned} \quad (324)$$

This parametrization is shown in Fig. 27. We observe that the qualitative form of the interaction does not

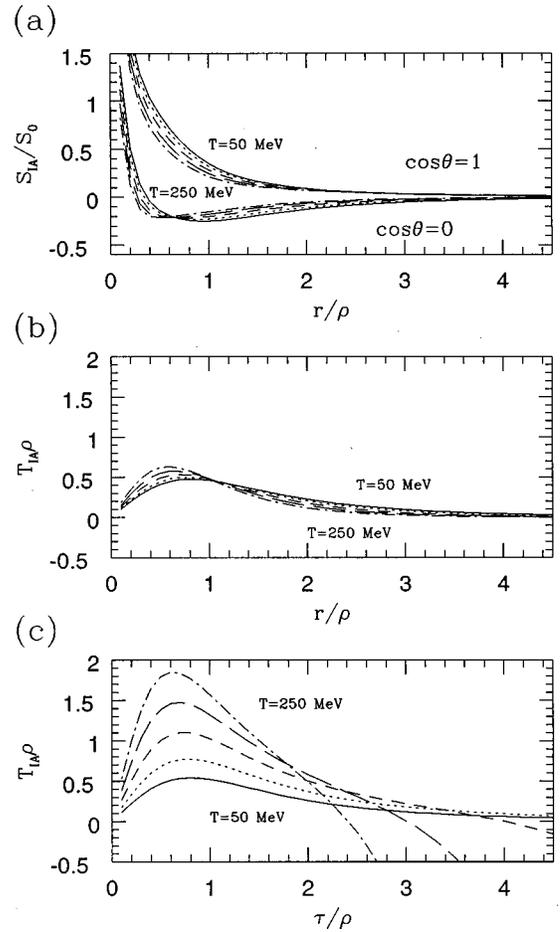


FIG. 27. Instanton/anti-instanton interaction at finite temperature: (a), the pure gauge interaction in units of  $S_0$  as a function of the spatial separation in units of  $\rho$ . The temperature is given in MeV for  $\rho=0.33$  fm. (b) the fermionic overlap matrix element  $T_{\text{IA}}$  for spatial separation  $r$ ; (c) same element for temporal separation  $\tau$ .

change, but it becomes more short range as the temperature increases. This is a consequence of the core in the instanton gauge field discussed above. At temperatures  $T < 1/(3\rho)$  the interaction is essentially isotropic. As we shall see below, this is not the case for the fermionic matrix elements.

The fermion determinant is calculated from the overlap matrix elements  $T_{\text{IA}}$  with the finite-temperature, zero-mode wave functions. Again, the orientational dependence is more complicated at  $T \neq 0$ . We have

$$T_{\text{IA}} = u_A f_1 + \vec{u} \cdot \hat{r} f_2 \quad (325)$$

where  $f_i = f_i(r, \tau, \rho_I, \rho_A, \beta)$ . The asymptotic form of  $T_{\text{IA}}$  for large temperatures  $\beta \rightarrow 0, R \rightarrow \infty$  can be determined analytically. The result is (Khoze and Yung, 1991; Shuryak and Verbaarschot, 1991)

$$f_1^{as} = i \frac{\pi^2}{\beta} \sin\left(\frac{\pi\tau}{\beta}\right) \exp\left(-\frac{\pi r}{\beta}\right), \quad (326)$$

$$f_2^{as} = i \frac{\pi^2}{\beta} \cos\left(\frac{\pi\tau}{\beta}\right) \exp\left(-\frac{\pi r}{\beta}\right). \quad (327)$$

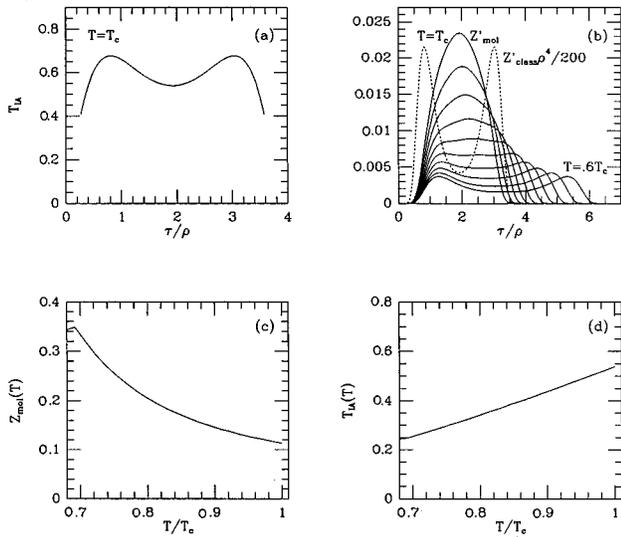


FIG. 28. Mean-field results at finite temperature, from Velkovsky and Shuryak, 1996. (a) Fermionic overlap matrix element  $T_{IA}(\tau, r=0, T=T_c)$  as a function of the distance  $\tau$  between the centers of the instanton and anti-instanton in the time direction; (b) Same for the partially integrated partition function  $Z'_{mol} = \int d^{10}\Omega T_{IA}^{2N_f} e^{-S_{int}}$  ( $Z$  is not yet integrated over Matsubara time) at different temperatures  $T = (0.6-1.0)T_c$ , in steps of  $\Delta T = 0.04T_c$ . The dashed line shows the behavior of  $T_{IA}$ . (c) Fermionic overlap matrix element after integration over Matsubara time; (d) partition function after integration over Matsubara time.

A parametrization of the full result is shown in Fig. 27 (Schäfer and Shuryak, 1996b). The essential features of the result can be seen from the asymptotic form [Eqs. (326) and (327)]: At large temperature,  $T_{IA}$  is very anisotropic. The matrix element is exponentially suppressed in the spatial direction and periodic along the temporal axis. The exponential decay in the spatial direction is governed by the lowest Matsubara frequency  $\pi T$ . The qualitative behavior of  $T_{IA}$  has important consequences for the structure of the instanton liquid at  $T \neq 0$ . In particular, the overlap matrix element favors the formation of instanton/anti-instanton molecules oriented along the time direction.

These configurations were studied in greater detail by Velkovsky and Shuryak (1996). These authors calculate the IA partition function (150) from the bosonic and fermionic interaction at finite temperature. The integration over the collective coordinates was performed as follows: The point  $r=0$  (same position in space) and the most attractive relative orientation ( $U=1, \cos \theta=1$ ) are maxima of the partition function, so one can directly use the saddle-point method for 10 integrals out of 11 (three over relative spatial distance between the centers and over seven relative orientation angles). The remaining integral over the temporal distance  $\tau$  is more complicated and has to be done numerically.

In Fig. 28(a) we show the dependence of the overlap matrix element  $T_{IA}$  on  $\tau$  for  $T=T_c, r=0$ , and  $U=1$ . Even at  $T_c, T_{IA}$  does not have a maximum at the symmetric point  $\tau=1/(2T)$ , but a minimum. However, when

one includes the bosonic interaction and the pre-exponential factor from taking into account fluctuations around the saddle point, the result looks different. The partition function after integrating over 10 of the 11 collective coordinates is shown in Fig. 28(b) for temperatures in the range  $(0.6-1.0)T_c$ . We observe that there is a maximum in the partition function at the symmetric point  $\tau=1/(2T)$  if the temperature is bigger than  $T_{molec} \approx 0.2\rho \approx 120$  MeV. The temperature dependence of the partition function integrated over all variables is shown in Figs. 28(c) and 28(d).

This means that there is a qualitative difference between the status of instanton/anti-instanton molecules at low and high ( $T > T_{molec}$ ) temperatures. At low temperatures, saddle points in the instanton/anti-instanton separation exist only if the collective coordinates are analytically continued into the complex plane<sup>50</sup> (as in Sec. II.A). At high temperatures there is a real saddle point in the middle of the Matsubara box [ $\tau=1/(2T)$ ], which gives a large contribution to the free energy. It is important that this happens close to the chiral phase transition. In fact, we shall argue that the phase transition is caused by the formation of these molecules.

## B. Chiral symmetry restoration

### 1. Introduction to QCD phase transitions

Before we come to a detailed discussion of the instanton liquid at finite temperature, we should like to give a brief summary of what is known about the phase structure of QCD at finite temperature. It is generally believed that at high temperature (or density) QCD undergoes a phase transition from hadronic matter to the quark-gluon plasma. In the plasma phase, color charges are screened (Shuryak, 1978a) rather than confined, and chiral symmetry is restored. At sufficiently high temperature, perturbation theory should be applicable, and the physical excitations are quarks and gluons. In this case, the thermodynamics of the plasma are governed by the Stefan-Boltzmann law, just like ordinary blackbody radiation.

This basic scenario has been confirmed by a large number of lattice simulations. As an example, we show the equation of state for  $N_f=2$  (Kogut-Susskind) QCD in Fig. 29 (Blum *et al.*, 1995). The transition temperature is  $T_c(N_f=2) \approx 150$  MeV. The energy density and pressure are small below the phase transition, but the energy density rises quickly to its perturbative (Stefan-Boltzmann) value. The pressure, on the other hand, lags behind and remains small up to  $T \approx 2T_c$ .

In Fig. 29 the energy density and pressure are measured with respect to the perturbative vacuum. However, we have repeatedly emphasized that in QCD there is a nonperturbative vacuum energy density (the bag pressure) even at  $T=0$ . In order to compare a noninter-

<sup>50</sup>The two maxima at  $\tau \approx \rho$  and  $\tau \approx 3\rho$  in Fig. 28(a) are not really physical; they are related to the presence of a core in the ratio ansatz interaction.

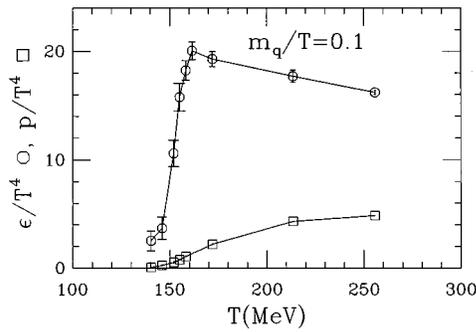


FIG. 29. Equation of state for  $N_f=2$  QCD with Kogut-Susskind fermions, from Blum *et al.*, 1995.

acting gas of quarks and gluons with the hadronic phase, we have to shift the pressure in the high- $T$  phase by this amount,  $p_{\text{QGP}} \rightarrow p_{\text{QGP}} - B$ . Since the pressure from thermal hadrons below  $T_c$  is small, a rough estimate of the transition temperature can be obtained from the requirement that the (shifted) pressure in the plasma phase be positive,  $p_{\text{QGP}} > 0$ . From this inequality we expect the plasma phase to be favored for  $T > T_c = [(90B)/(N_d \pi^2)]^{1/4}$ , where  $N_d$  is the effective number of degrees of freedom in the quark-gluon-plasma phase. For  $N_f=2$  we have  $N_d=37$  and  $T_c \approx 180$  MeV.

Lattice simulations indicate that in going from quenched QCD to QCD with two light flavors the transition temperature drops by almost a factor of two, from  $T_c(N_f=0) \approx 260$  MeV to  $T_c(N_f=2) \approx 150$  MeV. The number of degrees of freedom in the plasma phase, on the other hand, increases only from 16 to 37 (and  $T_c$  does not vary much with  $N_d$ ,  $T_c \sim N_d^{-1/4}$ ). This implies that there are significant differences between the pure gauge and unquenched phase transitions.

In particular, the low transition temperature observed for  $N_f=2,3$  suggests that the energy scales are quite different. We have already emphasized that the bag pressure is directly related to the gluon condensate (Shuryak, 1978b). This relation was studied in greater detail by Deng (1989), Adami, Hatsuda, and Zahed (1991), and Koch and Brown (1993). From the canonical energy-momentum tensor and the trace anomaly, the gluonic contributions to the energy density and pressure are related to the electric- and magnetic-field strengths,

$$\epsilon = \frac{1}{2} \langle B^2 - E^2 \rangle - \frac{g^2 b}{128 \pi^2} \langle E^2 + B^2 \rangle, \quad (328)$$

$$p = \frac{1}{6} \langle B^2 - E^2 \rangle + \frac{g^2 b}{128 \pi^2} \langle E^2 + B^2 \rangle. \quad (329)$$

Using the available lattice data one finds that in pure gauge theory the gluon condensate essentially disappears in the high-temperature phase, while in full QCD ( $N_f=2,3$ ) about half of the condensate remains. G. Brown has referred to this part of the gluon condensate as the “hard glue” or “epoxy.” It plays an important role in keeping the pressure positive despite the low transition temperature.

More general information on the nature of the phase

transition is provided by universality arguments. For this purpose, we have to identify an order parameter that characterizes the transition. In QCD, there are two limits in which this can be achieved. For infinitely heavy quarks, the Polyakov line (Polyakov, 1978) provides an order parameter for the deconfinement transition, while for massless quarks the fermion condensate is an order parameter for spontaneous chiral symmetry breaking. The Polyakov line is defined by

$$\langle L(\vec{x}) \rangle = \left\langle P \exp \left( i \int_0^{\beta} d\tau A_0(\vec{x}, \tau) \right) \right\rangle, \quad (330)$$

which can be interpreted as the propagator of a heavy quark. The free energy of a single static quark (minus the free energy of the vacuum) is given by  $F_q - F_0 = -T \log(\langle L(\vec{x}) \rangle)$ . For  $\langle L \rangle = 0$ , the free energy of an isolated quark is infinite, and the theory is in the confining phase. For  $\langle L \rangle \neq 0$  the free energy is finite, and quarks are screened rather than confined.

The Polyakov line has a nontrivial symmetry. Under gauge transformations in the center of the gauge group,  $Z_{N_c} \subset \text{SU}(N_c)$ , local observables are invariant but the Polyakov line picks up a phase  $L \rightarrow zL$  with  $z \in Z_{N_c}$ . The deconfinement transition was therefore related to spontaneous breakdown of the  $Z_{N_c}$  center symmetry. Following Landau, long wave excitations near the phase transition should be governed by an effective  $Z_{N_c}$  symmetric Lagrangian for the Polyakov line (Svetitsky and Yaffe, 1982). Since long-range properties are determined by the lowest Matsubara modes, the effective action is defined in three dimensions. For  $N_c=2$ , this means that the transition is in the same universality class as the  $d=3$  Ising model, which is known to have a second-order phase transition. For  $N_c \geq 3$ , on the other hand, we expect the transition to be first order.<sup>51</sup> These expectations are supported by lattice results.

For massless quarks, chiral symmetry is exact, and the quark condensate  $\langle \bar{q}q \rangle$  provides a natural order parameter. The symmetry of the order parameter is determined by the transformation properties of the matrix  $U_{ij} = \langle \bar{q}_i q_j \rangle$ . For  $N_f=2$  flavors, this symmetry is given by  $\text{SU}(2) \times \text{SU}(2) = O(4)$ , so the transition is governed by the effective Lagrangian for a four-dimensional vector field in three dimension (Pisarski and Wilczek, 1984; Wilczek, 1992; Rajagopal and Wilczek, 1993). The  $O(4)$  Heisenberg magnet is known to have a second-order phase transition, and the critical indices have been determined from both numerical simulations and the  $\epsilon$  expansion. For  $N_f \geq 3$ , however, the phase transition is predicted to be first order.

From these arguments, one expects the schematic phase structure of QCD in the  $m_{ud} - m_s$  (with  $m_{ud} \equiv m_u = m_d$ ) plane to look as shown in Fig. 30 (Brown *et al.*, 1990; Iwasaki *et al.*, 1996). The upper right-hand corner corresponds to the first-order pure gauge phase

<sup>51</sup>Under certain assumptions, the  $N_c > 3$  phase transition can be second order; see Pisarski and Tytgat (1997).

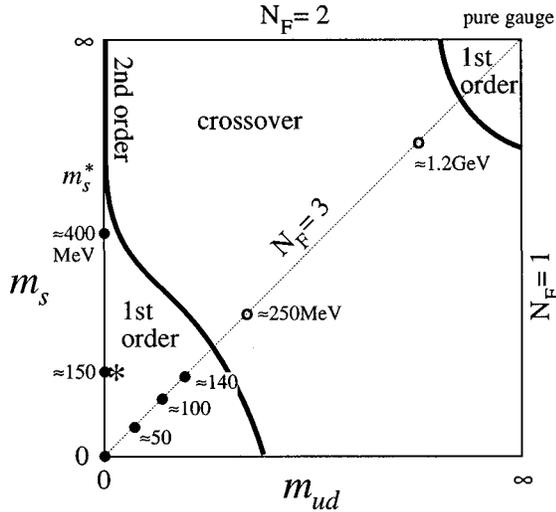


FIG. 30. Schematic phase diagram of  $N_f=3$  QCD with dynamic quark masses  $m_{ud} \equiv m_u = m_d$  and  $m_s$ , from Iwasaki *et al.*, 1996: ●, simulations (with Wilson fermions) that found a first-order transition; ○, simulations that found a smooth crossover. The (approximate) location of QCD with realistic masses is shown by a star.

transition. Presumably, this first-order transition extends to lower quark masses before it ends in a line of second-order phase transitions. The first-order  $N_f=3$  chiral phase transition is located in the lower left corner and continues in the mass plane before it ends in another line of second-order phase transitions. At the left edge there is a tricritical point where this line meets the line of  $N_f=2$  second-order phase transitions extending down from the upper left corner.

Simulations suggest that the gap between the first-order chiral and deconfinement transitions is very wide, extending from  $m \equiv m_u = m_d = m_s \approx 0.2$  GeV to  $m \approx 0.8$  GeV. This is in line with the arguments given above, indicating that there are important differences between the phase transitions in pure gauge and full QCD. Nevertheless, one should not take this distinction too literally. In the presence of light quarks, there is no deconfinement phase transition in a strict, mathematical sense. From a practical point of view, however, deconfinement plays an important role in the chiral transition. In particular, the equation of state shows that the jump in energy density is dominated by the release of 37 quark and gluon degrees of freedom.

There are many important questions related to the phase diagram that still have to be resolved. First of all, we have to determine the location of real QCD (with two light, one intermediate-mass, and several heavy flavors) on this phase diagram. While results using staggered fermions (Brown *et al.*, 1990) seem to suggest that QCD lies outside the range of first-order chiral phase transitions and shows only a smooth crossover, simulations using Wilson fermions (Iwasaki *et al.*, 1991) with realistic masses find a first-order transition.

A more general problem is to verify the structure of the phase diagram and to check the values of the critical indices near second-order transitions. While earlier stud-

ies confirmed (within errors) the  $O(4)$  values of the critical indices for the  $N_f=2$  chiral transition with both staggered (Karsch and Laermann, 1994) and Wilson fermions (Iwasaki *et al.*, 1996), more recent work concludes that  $1/\delta$  (defined as  $\langle \bar{q}q \rangle_{T_c} \sim m^{1/\delta}$ ) is consistent with zero (Ukawa, 1997). If this result is correct, it would imply that there is no  $N_f=2$  second-order phase transition, and the only second-order line in Fig. 30 corresponds to the boundary of the first-order region, with standard Ising indices.

Are there any possible effects that could upset the expected  $O(4)$  scenario for the  $N_f=2$  transition? One possibility discussed in the literature is related to the fate of the  $U(1)_A$  symmetry at  $T_c$ . At zero temperature the axial  $U(1)_A$  symmetry is explicitly broken by the chiral anomaly and instantons. If instantons are strongly suppressed (Pisarski and Wilczek, 1984) or rearranged (Shuryak, 1994) at  $T_c$ , then the  $U(1)_A$  symmetry might effectively be restored at the phase transition. In this case, the masses of the  $U(1)_A$  partners of the  $\pi, \sigma$ , the  $\delta, \eta'$ , could become sufficiently small that fluctuations of these modes would affect the universality arguments given above. In particular, if there were eight rather than four light modes at the phase transition, the transition would be expected to be first order. Whether this is the correct interpretation of the lattice data remains unclear at the moment. We shall return to the question of  $U(1)_A$  breaking at finite temperature in Sec. VII.C.3 below. Let us only note that the simulations are very difficult and that there are several possible artifacts. For example, there could be problems with the extrapolation to small masses. Also, first-order transitions need not have an order parameter, and it is difficult to distinguish very weak first-order transitions from second-order transitions.

Completing this brief review of general arguments and lattice results on the chiral phase transition, let us also comment on some of the theoretical approaches. Just as perturbative QCD describes the thermodynamics of the plasma phase at very high temperature, effective chiral Lagrangians provide a very powerful tool at low temperature. In particular, chiral perturbation theory predicts the temperature dependence of the quark and gluon condensates as well as of the masses and coupling constants of light hadrons (Gerber and Leutwyler, 1989). These results are expressed as an expansion in  $T^2/f_\pi^2$ . It is difficult to determine what the range of validity of these predictions is, but the approach certainly has to fail as one approaches the phase transition.

The phase transition has also been studied in various effective models, for example, in the linear sigma model, the chiral quark model, or the Nambu–Jona-Lasinio model (Hatsuda and Kunihiro, 1985; Bernard and Meissner, 1988; Klimt, Lutz, and Weise, 1990). In the Nambu–Jona-Lasinio model, the chiral transition connects the low-temperature phase of massive constituent quarks with the high-temperature phase of massless quarks (but of course not gluons). The mechanism for the transition is similar to that for a transition from a superconductor to a normal metal: The energy gain due

to pairing disappears once a sufficient number of positive-energy states is excited. We have seen that, at zero temperature, the instanton liquid leads to a picture of chiral symmetry breaking which closely resembles the Nambu–Jona-Lasinio model. Nevertheless, we shall argue that the mechanism for the phase transition is quite different. In the instanton liquid, it is the ensemble itself that is rearranged at  $T_c$ . This means not only that we have nonzero occupation numbers, but also that the effective interaction itself changes.

Finally, a number of authors have extended random matrix models to finite temperature and density (Jackson and Verbaarschot, 1996; Novak, Papp, and Zahed, 1996; Wettig, Schäfer, and Weidenmüller, 1996). It is important to distinguish these models from random matrix methods at  $T, \mu = 0$ . In this case, there is evidence that certain observables,<sup>52</sup> like scaled correlations between eigenvalues of the Dirac operator, are universal and can be described in terms of suitably chosen random matrix ensembles. The effects of nonzero temperature and density, on the other hand, are included in a rather schematic fashion, by putting terms like  $(\pi T + i\mu)$  into the Dirac operator.<sup>53</sup> This procedure is certainly not universal. From the point of view of the instanton model, the entries of the random matrix correspond to matrix elements of the Dirac operator in the zero-mode basis. If we include the effects of a nonzero  $\mu, T$  in the schematic form mentioned above, we assume that at  $\mu, T \neq 0$  there are no zero modes in the spectrum. But this is not true. We have explicitly constructed the zero mode for the caloron configuration in Sec. VII.A.1, for  $\mu \neq 0$  [see Carvalho (1980), and Abrikosov (1983)]. As we shall see below, in the instanton model the phase transition is not caused by a constant contribution to the overlap matrix elements, but by specific correlations in the ensemble.

## 2. The instanton liquid at finite temperature and IA molecules

In Sec. VII.A.3, we argued that the instanton density remains roughly constant below the phase transition. This means that the chiral phase transition has to be caused by a rearrangement of the instanton ensemble. Furthermore, we have shown that the gluonic interaction between instantons remains qualitatively unchanged even at fairly high temperatures. This suggests that fermionic interactions between instantons drive the phase transition (Ilgenfritz and Shuryak, 1994; Schäfer *et al.*, 1995; Schäfer and Shuryak, 1996a).

The mechanism for this transition is most easily understood by considering the fermion determinant for one instanton/anti-instanton pair (Schäfer *et al.*, 1995). Using the asymptotic form of the overlap matrix elements specified above, we have

<sup>52</sup>On the other hand, macroscopic observables, like the average level density (and the quark condensate) are not expected to be universal.

<sup>53</sup>See Wettig *et al.* (1996) for an attempt to model the results of the instanton liquid in terms of a random matrix ensemble.

$$\det(\mathcal{D}) \sim \left| \frac{\sin(\pi T \tau)}{\cosh(\pi T r)} \right|^{2N_f}. \quad (331)$$

This expression is maximal for  $r=0$  and  $\tau=1/(2T)$ , which is the most symmetric orientation of the instanton/anti-instanton pair on the Matsubara torus. Since the fermion determinant controls the probability of the configuration, we expect polarized molecules to become important at finite temperature. The effect should be largest when the IA pairs exactly fit onto the torus, i.e.,  $4\rho \approx \beta$ . Using the zero-temperature value  $\rho \approx 0.33$  fm, we get  $T \approx 150$  MeV, close to the expected transition temperature for two flavors.

In general, the formation of molecules is controlled by the competition between minimum action, which favors correlations, and maximum entropy, which favors randomness. Determining the exact composition of the instanton liquid as well as the transition temperature requires the calculation of the full partition function, including the fermion-induced correlations. We shall do this using a schematic model in Sec. VII.B.3 and using numerical simulations in Sec. VII.B.4.

Before we come to this, we should like to study what physical effects are caused by the presence of molecules. Qualitatively, it is clear why the formation of molecules leads to chiral symmetry restoration. If instantons are bound into pairs, then quarks mostly propagate from one instanton to the corresponding anti-instanton and back. In addition to that, quarks propagate mostly along the imaginary-time direction, so all eigenstates are well localized in space and no quark condensate is formed. Another way to see this is by noting that the Dirac operator essentially decomposes into  $2 \times 2$  blocks corresponding to the instanton/anti-instanton pairs. This means that the eigenvalues will be concentrated around some typical  $\pm \lambda$  determined by the average size of the pair, so the density of eigenvalues near  $\lambda=0$  vanishes. We have studied the eigenvalue distribution in a schematic model of random and correlated instantons in Schäfer *et al.* (1995), and a random matrix model of the transition based on these ideas was discussed by Wettig *et al.* (1996).

The effect of molecules on the effective interaction between quarks at high temperature was studied by Schäfer *et al.* (1995), using methods similar to those introduced in Sec. IV.F. In order to determine the interaction in a quark-antiquark (meson) channel with given quantum numbers, it is convenient to calculate both the direct and the exchange terms and Fierz-rearrange the exchange term into an effective direct interaction. The resulting Fierz symmetric Lagrangian reads (Schäfer *et al.*, 1995)

$$\begin{aligned} \mathcal{L}_{\text{mol sym}} = G \left\{ \frac{2}{N_c^2} [(\bar{\psi} \tau^a \psi)^2 - (\bar{\psi} \tau^a \gamma_5 \psi)^2] \right. \\ \left. - \frac{1}{2N_c^2} [(\bar{\psi} \tau^a \gamma_\mu \psi)^2 + (\bar{\psi} \tau^a \gamma_\mu \gamma_5 \psi)^2] \right. \\ \left. + \frac{2}{N_c^2} (\bar{\psi} \gamma_\mu \gamma_5 \psi)^2 \right\} + \mathcal{L}_8, \quad (332) \end{aligned}$$

with the coupling constant

$$G = \int n(\rho_1, \rho_2) d\rho_1 d\rho_2 \frac{1}{8T_{IA}^2} (2\pi\rho_1)^2 (2\pi\rho_2)^2. \quad (333)$$

Here,  $\mathcal{L}_8$  is the color-octet part of the interaction and  $\tau^a$  is a four-vector with components  $(\vec{\tau}, 1)$ . Also,  $n(\rho_1, \rho_2)$  is the tunneling probability for the IA pair and  $T_{IA}$  the corresponding overlap matrix element. The effective Lagrangian (332) was determined by averaging over all possible molecule orientations, with the relative color orientation  $\cos(\theta)=1$  fixed. Near the phase transition, molecules are polarized in the temporal direction, Lorentz invariance is broken, and vector interactions are modified according to  $(\bar{\psi}\gamma_\mu\Gamma\psi)^2 \rightarrow 4(\bar{\psi}\gamma_0\Gamma\psi)^2$ .

Like the zero-temperature effective Lagrangian (111), the interaction (332) is  $SU(2) \times SU(2)$  symmetric. Since molecules are topologically neutral, the interaction is also  $U(1)_A$  symmetric. This does not mean that  $U(1)_A$  symmetry is restored in the molecular vacuum. Even a very small  $O(m^{N_f})$  fraction of random instantons will still lead to  $U(1)_A$  breaking effects of order  $O(1)$  (see Sec. VII.C.3). If chiral symmetry is restored, the effective interaction (332) is attractive not only in the pion channel, but also in the other scalar-pseudoscalar channels  $\sigma$ ,  $\delta$ , and  $\eta'$ . Furthermore, unlike the 't Hooft interaction, the effective interaction in the molecular vacuum also includes an attractive interaction in the vector and axial-vector channels. If molecules are unpolarized, the corresponding coupling constant is a factor of 4 smaller than the scalar coupling. If they are fully polarized, only the longitudinal vector components are affected. In fact, the coupling constant is equal to the scalar coupling. A more detailed study of the quark interaction in the molecular vacuum was performed by Schäfer *et al.* (1995) and Schäfer and Shuryak (1995b), in which hadronic correlation functions in the spatial and temporal direction were calculated in the schematic model mentioned above. We shall discuss the results below.

### 3. Mean-field description at finite temperature

In the following two sections we wish to study the statistical mechanics of the instanton liquid at finite temperature. This is necessary not only in order to study thermodynamic properties of the system, but also to determine the correct ensemble for calculations of hadronic correlation functions. We first extend the mean-field calculation of Sec. IV.F to finite temperature. In the next section we shall study the problem numerically, using the methods introduced in Sec. V.B.

For pure gauge theory, the variational method was extended to finite temperature by Diakonov and Mirlin (1988) and Kanki (1988). The gluonic interaction between instantons changes very little with temperature, so we shall ignore this effect. In this case, the only difference as compared to zero temperature is the appearance of the perturbative suppression factor (308) (for  $T > T_c$ , although Diakonov and Mirlin used it for all  $T$ ).

Since the interaction is unchanged, so is the form of the single-instanton distribution,

$$\mu(\rho) = n(\rho, T) \exp\left[\frac{-\beta\gamma^2 NT}{V_3} \bar{\rho}^2 \rho^2\right], \quad (334)$$

where  $n(\rho, T)$  is the semiclassical result (308) and the four-dimensional volume is given by  $V = V_3/T$ . The  $T$  dependence of the distribution functions modifies the self-consistency equations for  $\bar{\rho}^2$  and  $N/V$ . Following Diakonov and Mirlin (1988), we can expand the coefficient  $B(\lambda)$  in (308) to order  $T^2$ :  $n(\rho, T) \approx \exp[-\frac{1}{3}(\frac{11}{6}N_c - 1)(\pi\rho T)^2] n(\rho, T=0)$ . In this case, the self-consistency condition for the average size is given by

$$\bar{\rho}^2 = \nu \left[ \frac{1}{3} \left( \frac{11}{6} N_c - 1 \right) \pi^2 T^2 + \frac{N\beta\gamma^2}{V_3} T \bar{\rho}^2 \right]^{-1}. \quad (335)$$

For  $T=0$ , this gives  $\bar{\rho}^2 = [(\nu V)/(\beta\gamma^2 N)]^{1/2}$  as before, while for large  $T > (\pi\rho)^{-1}$  we have  $\bar{\rho}^2 \sim 1/T^2$ . The instanton density follows from the self-consistency equation for  $\mu_0$ . For large  $T$  we have

$$N/V = C_{N_c} \beta^{2N_c} \Lambda^4 \Gamma(\nu) \left[ \frac{1}{3} \left( \frac{11}{6} N_c - 1 \right) \pi^2 \frac{T^2}{\Lambda^2} \right]^{-\nu}, \quad (336)$$

so that  $N/V \sim 1/T^{b-4}$ , which is what one would expect from simply cutting the size integration at  $\rho \approx 1/T$ .

The situation is somewhat more interesting if one extends the variational method to full QCD (Ilgenfritz and Shuryak, 1989; Nowak, Verbaarschot, and Zahed, 1989c). In this case, an additional temperature dependence enters through the  $T$  dependence of the average fermionic overlap matrix elements. More importantly, the average determinant depends on the instanton size,

$$\det(\mathcal{D}) = \prod_I (\rho m_{\det}),$$

$$m_{\det} = \rho^{3/2} \left[ \frac{1}{2} I(T) \int d\rho n(\rho) \rho \right]^{1/2}, \quad (337)$$

where  $I(T)$  is the angle and distance-averaged overlap matrix element  $T_{IA}$ . The additional  $\rho$  dependence modifies the instanton distribution (334) and introduces an additional nonlinearity into the self-consistency equation. As a result, the instanton density at large  $T$  depends crucially on the number of flavors. For  $N_f=0,1$ , the density drops smoothly with  $N/V \sim 1/T^{2a}$  and  $a = (b-4+2N_f)/(2-N_f)$  for large  $T$ . For  $N_f=2$ , the instanton density goes to zero continuously at the critical temperature  $T_c$ , whereas for  $N_f > 2$  the density goes to zero discontinuously at  $T_c$ . This behavior can be understood from the form of the gap equation for the quark condensate. We have  $\langle \bar{q}q \rangle \sim \text{const} \langle \bar{q}q \rangle^{N_f-1}$ , which, for  $N_f > 2$ , cannot have a solution for arbitrarily small  $\langle \bar{q}q \rangle$ .

In the mean-field approximation, the instanton ensemble remains random at all temperatures. This of course implies that instantons cannot exist above  $T_c$ . We have already argued that this is not correct, and that instantons can be present above  $T_c$ , provided they are

bound in molecules (or larger clusters). In order to include this effect in a variational calculation, Ilgenfritz and Shuryak introduced a ‘‘cocktail’’ model (Ilgenfritz and Shuryak, 1989; 1994), in which the instanton ensemble consists of a random and a molecular component. The composition of the instanton liquid is determined by minimizing the free energy with respect to the two concentrations.

As above, the distribution functions for random instantons is given by

$$\mu(\rho) = n(\rho) \exp[-\kappa \rho^2 (\overline{\rho_a^2} n_a + 2\overline{\rho_m^2} n_m)] (m_{\det} \rho)^{N_f}, \quad (338)$$

where  $n_{a,m}$  are the densities of the random and molecular contributions and  $\overline{\rho_{a,m}^2}$  are the corresponding average radii. The parameter  $\kappa = \beta \gamma^2$  characterizes the average repulsion between instantons. The distribution of instantons bound in molecules is given by

$$\begin{aligned} \mu(\rho_1, \rho_2) &= n(\rho_1) n(\rho_2) \\ &\times \exp[-\kappa(\rho_1^2 + \rho_2^2)(\overline{\rho_a^2} n_a + 2\overline{\rho_m^2} n_m)] \\ &\times \langle (T_{IA} T_{IA}^*)^{N_f} \rangle, \end{aligned} \quad (339)$$

where  $I_m(N_f, T) \equiv \langle (T_{IA} T_{IA}^*)^{N_f} \rangle$  is the average determinant for an instanton/anti-instanton pair, with the relative orientation  $\cos \theta = 1$  fixed. Summing the contributions from both the random and the molecular components, we find that the self-consistency condition for the instanton size becomes

$$\frac{\overline{\rho_m^2}}{\rho_a^2} = \frac{\alpha}{\beta}, \quad \frac{1}{\kappa} = \frac{2(\overline{\rho_a^2})^2 n_a}{\beta} + \frac{4(\overline{\rho_m^2})^2 n_m}{\alpha}, \quad (340)$$

where  $\alpha = b/2 - 1$  and  $\beta = b/2 + 3N_f/4 - 2$ . Using this result, one can eliminate the radii and determine the normalizations:

$$\mu_{0,m} = \frac{A}{[n_a + (2\alpha/\beta)n_m]^\alpha}, \quad A = \frac{I_m(N_f, T) C^2 \Gamma^2(\alpha)}{(4\kappa\beta)^\alpha}, \quad (341)$$

$$\begin{aligned} \mu_{0,a} &= \frac{B n_a^{N_f/2}}{[n_a + (2\alpha/\beta)n_m]^{\beta/2 + N_f/8}}, \\ B &= \frac{C\Gamma(\beta)}{(2\kappa)^\beta} \left( \frac{I(T)}{2} \right)^{N_f/2} \left( \frac{\beta}{\kappa} \right)^{N_f/8 - \beta/2}. \end{aligned} \quad (342)$$

Finally, the free energy is given by

$$\begin{aligned} F &= -\frac{1}{V_4} \log Z = \frac{N_a}{V_4} \log \left( \frac{e\mu_{0,a} V_4}{N_a} \right) \\ &+ \frac{N_m}{V_4} \log \left( \frac{e\mu_{0,m} V_4}{N_m} \right), \end{aligned} \quad (343)$$

and the instanton density is determined by minimizing  $F$  with respect to the densities  $n_{a,m}$  of random and correlated instantons. The resulting free energy determines the instanton contribution to the pressure  $p = -F$  and

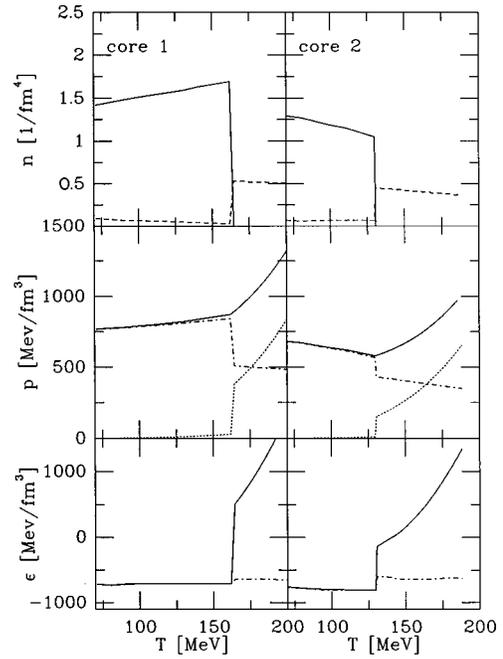


FIG. 31. Chiral restoration phase transitions for two massless flavors and two different core parameters  $R_c$ . The upper panel shows the  $T$  dependence of the densities: solid curve,  $n_a$ ; dotted curve,  $n_m$ . In the middle panel the  $T$  dependence is shown: solid curve, the total pressure  $p$ ; dotted curve, the contributions of the pion gas/quark-gluon plasma; dash-dotted curve, contribution of instantons. The lower panel shows the energy density (solid curve), which is modified by the instanton contribution (dash-dotted curve).

the energy density  $\epsilon = -p + T(\partial p/\partial T)$ . In order to provide a more realistic description of the thermodynamics of the chiral phase transition, Ilgenfritz and Shuryak added the free energy of a noninteracting pion gas in the broken phase and a quark-gluon plasma in the symmetric phase.

Minimizing the free energy gives a set of rather cumbersome equations. It is clear that in general there will be two phases, a low-temperature phase containing a mixture of molecules and random instantons, and a chirally restored high-temperature phase consisting of molecules only. The density of random instantons is suppressed as the temperature increases because the average overlap matrix element decreases. Molecules, on the other hand, are favored at high  $T$ , because the overlap for the most attractive orientation increases. Both of these results are simple consequences of the anisotropy of the fermion wave functions.

Numerical results for  $N_f = 2$  are shown in Fig. 31. In practice, the average molecular determinant  $I_m(N_f, T)$  was calculated by introducing a core into the IA interaction. In order to assess the uncertainty involved, we show the results for two different cores,  $R_c = \rho$  and  $2\rho$ . The overall normalization was fixed such that  $N/V = 1.4 \text{ fm}^{-4}$  at  $T = 0$ . Figure 31 shows that the random component dominates the broken phase and that the density of instantons is only weakly dependent on  $T$  below the phase transition. The number of molecules is small for

$T < T_c$  but jumps up at the transition. The total instanton density above  $T_c$ ,  $(N/V) = 2n_m$  turns out to be comparable to that at  $T=0$ .

The importance of the molecular component above  $T_c$  can be seen from the temperature dependence of the pressure. For  $T = (1-2)T_c$ , the contribution of molecules (dash-dotted line) is crucial in order to keep the pressure positive. This is the same phenomenon we already mentioned in our discussion of the lattice data. If the transition temperature is as low as  $T_c = 150$  MeV, then the contribution of quarks and gluons is not sufficient to match the  $T=0$  bag pressure. The lower panel in Fig. 31 shows the behavior of the energy density. The jump in the energy density at  $T_c$  is  $\Delta \epsilon \approx (0.5 - 1.0)$  GeV/fm<sup>3</sup>, depending on the size of the core. Although most of the latent heat is due to the liberation of quarks and gluons, a significant part is generated by molecules.

#### 4. Phase transitions in the interacting instanton model

In this section, we go beyond this schematic model and study the phase transition using numerical simulations of the interacting instanton liquid (Schäfer and Shuryak, 1996a). This means that we do not have to make any assumptions about the nature of the important configurations (molecules, larger clusters,...), nor do we have to rely on variational estimates to determine their relative importance. Neither, are we limited to a simple two-phase picture with a first-order transition.

In Fig. 32 we show the instanton density, free energy, and quark condensate for the physically relevant case of two light and one intermediate-mass flavor. In the ratio ansatz the instanton density at zero temperature is given by  $N/V = 0.69\Lambda^4$ . Taking the density to be  $1 \text{ fm}^{-4}$  at  $T=0$  fixes the scale parameter  $\Lambda = 222$  MeV and determines the absolute units. The temperature dependence of the instanton density<sup>54</sup> is shown in Fig. 32(a). It shows a slight increase at small temperatures, starts to drop around 115 MeV, and becomes very small for  $T > 175$  MeV. The free energy closely follows the behavior of the instanton density. This means that the instanton-induced pressure first increases slightly, but then drops and eventually vanishes at high temperature. This behavior is expected for a system of instantons, but if all fluctuations are included, the pressure should always increase as a function of the temperature.

The temperature dependence of the quark condensate is shown in Fig. 32(c). At temperatures below  $T = 100$  MeV it is practically temperature independent. Above that,  $\langle \bar{q}q \rangle$  starts to drop and becomes very small above the critical temperature  $T \approx 140$  MeV. Note that at this point the instanton density is  $N/V = 0.6 \text{ fm}^{-4}$ , slightly more than half the zero-temperature value. This

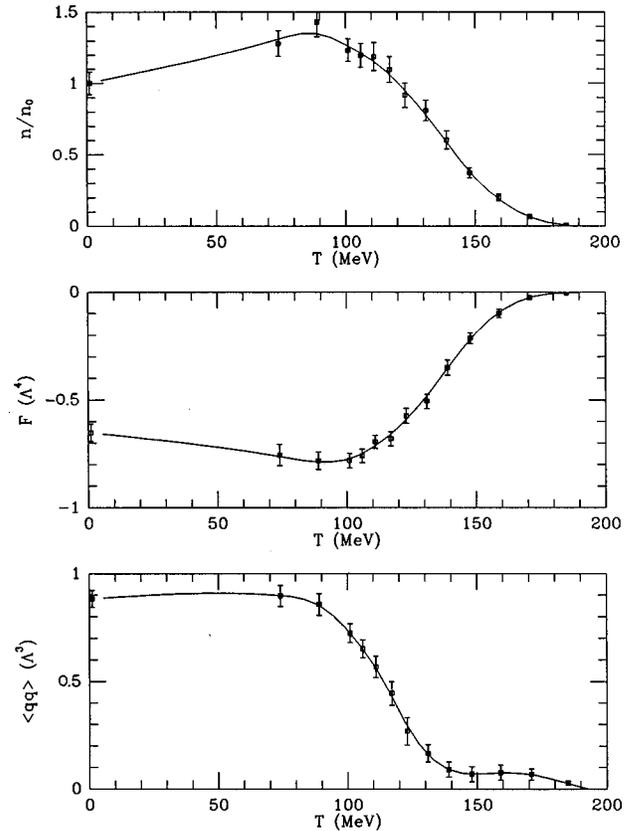


FIG. 32. Instanton density, free energy, and quark condensate as a function of the temperature in the instanton liquid with two light and one intermediate mass flavor, from Schäfer and Shuryak, 1996a. The instanton density is given in units of the zero-temperature value  $n_0 = 1 \text{ fm}^{-4}$ , while the free energy and the quark condensate are given in units of the Pauli-Villars scale parameter,  $\Lambda = 222$  MeV.

means that the phase transition is indeed caused by a transition within the instanton liquid, not by the disappearance of instantons. This point is illustrated in Fig. 33, which shows projections of the instanton liquid at  $T = 74$  MeV and  $T = 158$  MeV, below and above the phase transition. The figures are projections of a cube  $V = (3.0\Lambda^{-1})^3 \times T^{-1}$  into the 3-4 plane. The positions of instantons and anti-instantons are denoted by  $+/-$  symbols. The lines connecting them indicate the strength of the fermionic overlap (“hopping”) matrix elements. Below the phase transition, there is no clear pattern. Instantons are unpaired, part of molecules or larger clusters. Above the phase transition, the ensemble is dominated by polarized instanton/anti-instanton molecules.

Figure 34 shows the spectrum of the Dirac operator. Below the phase transition it has the familiar flat shape near the origin and extrapolates to a nonzero density of eigenvalues at  $\lambda = 0$ . Near the phase transition the eigenvalue density appears to extrapolate to 0 as  $\lambda \rightarrow 0$ , but there is a spike in the eigenvalue density at  $\lambda = 0$ . This spike contains the contribution from unpaired instantons. In the high-temperature phase, we expect the density of random instantons to be  $O(m^{N_f})$ , giving a con-

<sup>54</sup>The instanton density is of course sensitive to our assumptions concerning the role of the finite-temperature suppression factor. In practice, we have chosen a functional form that interpolates between no suppression below  $T_c$  and full suppression above  $T_c$ , with a width  $\Delta T = 0.3T_c$ .

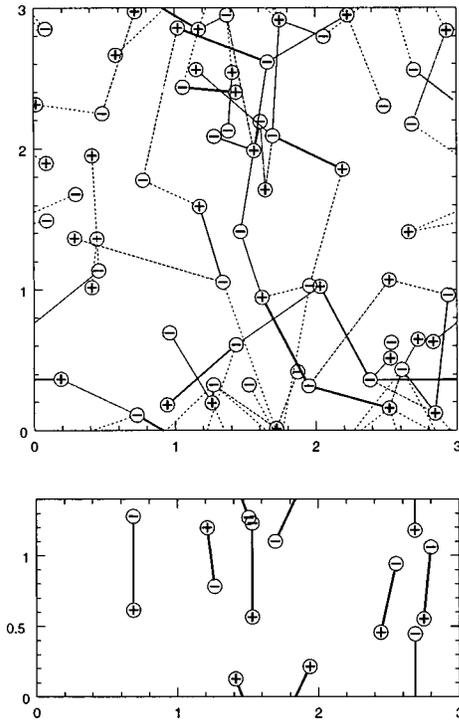


FIG. 33. Typical instanton ensembles for  $T=75$  and  $158$  MeV, from Schäfer and Shuryak, 1996a. The plots show projections of a four-dimensional  $(3.0\Lambda^{-1})^3 \times T^{-1}$  box into the 3-4 plane. The positions of instantons and anti-instanton are indicated by + and - symbols: Dashed lines, fermionic overlap matrix elements  $T_{IA}\Lambda > 0.40$ ; solid lines, matrix elements  $T_{IA}\Lambda > 0.56$ ; heavy solid lines, matrix elements  $T_{IA}\Lambda > 0.64$ .

tribution of the form  $\rho(\lambda) \sim m^{N_f} \delta(\lambda)$  to the Dirac spectrum. These eigenvalues do not contribute to the quark condensate in the chiral limit, but they are important for  $U(1)_A$ -violating observables.

The nature of the phase transition for different numbers of flavors and values of the quark masses was studied by Schäfer and Shuryak (1996a). The order of the phase transition is most easily determined by studying order-parameter fluctuations near  $T_c$ . For a first-order transition one expects two (meta) stable phases. The system tunnels back and forth between the two phases, with tunneling events signaled by sudden jumps in the order parameter. Near a second-order phase transition, on the other hand, the order parameter shows large fluctuations. These fluctuations can be studied in greater detail by measuring the scaling behavior of the mean-square fluctuations (the scalar susceptibility) with the current mass and temperature. Universality makes definite predictions for the corresponding critical exponents. The main conclusion in Schäfer and Shuryak (1996a) was that the transition in QCD is consistent with a nearby ( $N_f=2$ ) second-order phase transition with  $O(4)$  critical indices. For three flavors with physical masses, the transition is either very weakly first order or just a rapid crossover. As the number of flavors is increased, the transition becomes more strongly first order.

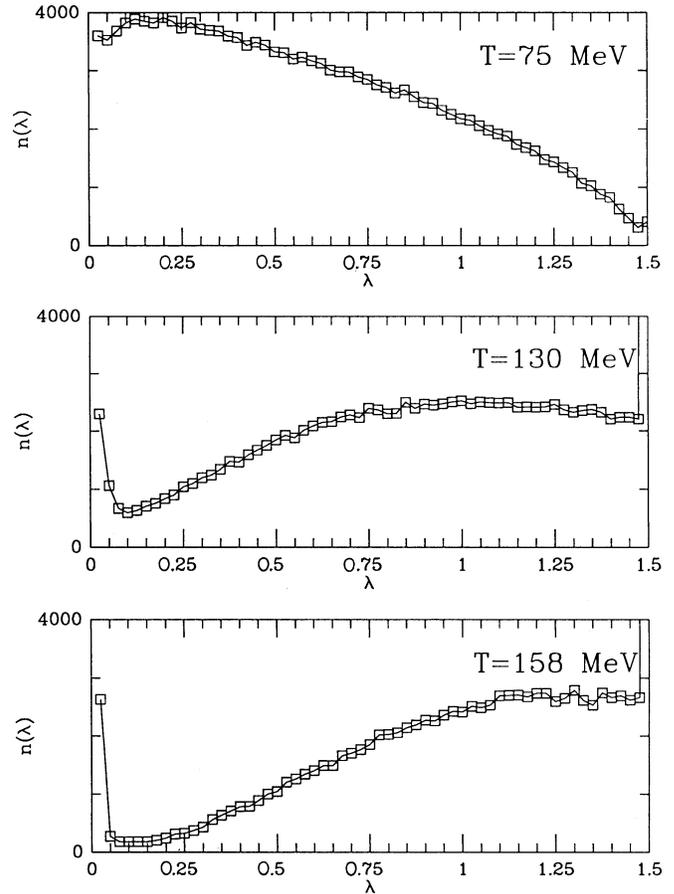


FIG. 34. Spectrum of the Dirac operator for different temperatures  $T=75$ ,  $130$ , and  $158$  MeV, from Schäfer and Shuryak, 1996a. Eigenvalues are given in units of the scale parameter. The normalization of the spectra is arbitrary (but identical).

The results of simulations with<sup>55</sup>  $N_f=2,3,5$  flavors with equal masses are summarized in the phase diagram 35 (Schäfer and Shuryak, 1996a). For  $N_f=2$  there is a second-order phase transition which turns into a line of first-order transitions in the  $m$ - $T$  plane for  $N_f>2$ . If the system is in the chirally restored phase ( $T>T_c$ ) at  $m=0$ , we find a discontinuity in the chiral order parameter if the mass is increased beyond some critical value. Qualitatively, the reason for this behavior is clear. While increasing the temperature increases the role of correlations caused by fermion determinant, increasing the quark mass has the opposite effect. We also observe that increasing the number of flavors lowers the transition temperature. Again, increasing the number of flavors means that the determinant is raised to a higher power, so fermion-induced correlations become stronger. For  $N_f=5$  we find that the transition temperature drops to zero and the instanton liquid has a chirally symmetric

<sup>55</sup>The case  $N_f=4$  is omitted because the contribution of small-size (semiclassical) instantons to the quark condensate is very small and the precise location of the phase boundary hard to determine.

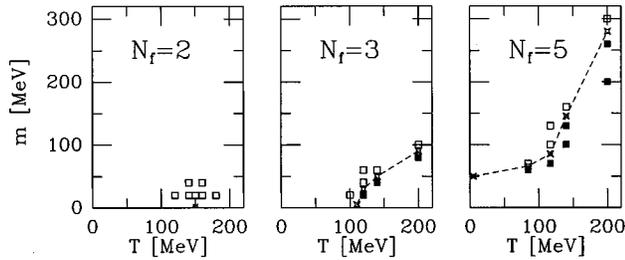


FIG. 35. Phase diagram of the instanton liquid in the  $T$ - $m$  plane for different numbers of flavors. The open and closed squares show points on the phase diagram where we have performed simulations. For  $N_f=2$ , the open squares mark points where we found large fluctuations of the order parameter, indicative of a nearby second-order phase transition (marked by a star). In the cases  $N_f=3$  and 5, open and closed squares mark points in the chirally broken and restored phases, respectively. The (approximate) location of the discontinuity is shown by the dashed line.

ground state, provided the dynamic quark mass is less than some critical value.

Studying the instanton ensemble more closely shows that, in this case, all instantons are bound into molecules. The molecular vacuum at  $T=0$  and large  $N_f$  has a somewhat different theoretical status from that of the molecular vacuum for small  $N_f$  and large  $T$ . In the high-temperature phase, large instantons are suppressed and long-range interactions are screened. This is not the case at  $T=0$  and  $N_f$  large, where these effects may contribute to chiral symmetry breaking (see the discussion in Sec. IX.D).

Unfortunately, little is known about QCD with different numbers of flavors from lattice simulations. There are some results on the phase structure of large- $N_f$  QCD that we shall discuss in Sec. IX.D. In addition to that, there are some recent data by the Columbia group (Chen and Mawhinney, 1997) on the hadron spectrum for  $N_f=4$ . The most important result is that chiral symmetry-breaking effects were found to be drastically smaller than those for  $N_f=0,2$ . In particular, the mass splittings between chiral partners such as  $\pi$ - $\sigma$ ,  $\rho$ - $a_1$ ,  $N(\frac{1}{2}^+) - N(\frac{1}{2}^-)$ , were found to be four to five times smaller. This agrees well with what was found in the interacting instanton model, but more work in this direction is certainly needed.

### C. Hadronic correlation functions at finite temperature

#### 1. Introduction

Studying the behavior of hadronic correlation functions at finite temperature is of great interest in connection with possible modifications of hadronic properties in hot hadronic matter. In addition to that, the structure of correlation functions at intermediate distances directly reflects on changes in the interaction between quarks and gluons. There is very little phenomenological information on this subject, but the problem has been studied extensively in the context of QCD sum rules

(Bochkarev and Shaposhnikov, 1986; Eletsii and Ioffe, 1993; Hatsuda, Koike, and Lee, 1993). At finite temperatures, however, the predictive power of QCD sum rules is very limited, because additional assumptions about the temperature dependence of the condensates and the shape of the spectrum are needed. There is an extensive literature on spacelike screening masses on the lattice, but only very limited information on temporal, point-to-point correlation functions (Boyd *et al.*, 1994).

At finite temperature, the heat bath breaks Lorentz invariance, and correlation functions in the spatial and temporal direction are independent. In addition, mesonic and baryonic correlation functions have to obey periodic or antiperiodic boundary conditions, respectively, in the temporal direction. In the case of spacelike correlators, one can still go to large  $x$  and filter out the lowest exponents known as screening masses. While these states are of theoretical interest and have been studied in a number of lattice calculations, they do not correspond to poles of the spectral function in energy. In order to look for real bound states, one has to study temporal correlation functions. However, at finite temperatures the periodic boundary conditions restrict the useful range of temporal correlators to the interval  $\tau < 1/(2T)$  (about 0.6 fm at  $T = T_c$ ). This means that there is no direct procedure for extracting information about the ground state. The underlying physical reason is clear: at finite temperatures excitations are always present. Below we explore how much can be learned from temporal correlation functions in the interacting instanton liquid. In the next section we also present the corresponding screening masses.

#### 2. Temporal correlation functions

Finite-temperature correlation functions in the temporal direction are shown in Fig. 36. These correlators were obtained from simulations of the interacting instanton liquid (Schäfer and Shuryak, 1995b). Correlators in the random-phase approximation were studied by Velkovsky and Shuryak (1996). The results, shown in Fig. 36, are normalized to the corresponding noninteracting correlators, calculated from the free  $T \neq 0$  propagator (300). Figures 36(a) and 36(b) show the pion and sigma correlators for different temperatures below (open symbols) and above (closed symbols) the chiral phase transition. The normalized  $\pi$  and  $\sigma$  correlators are larger than one at all temperatures, implying that there is an attractive interaction even above  $T_c$ . In particular, the value of the pion correlator at  $\tau=0.6$  fm, which is not directly affected by the periodic boundary conditions, is essentially temperature independent. This suggests that there is a strong  $(\pi, \sigma)$ -like mode present even above  $T_c$ . Schäfer and Shuryak (1996b) tried to determine the properties of this mode from a simple fit to the measured correlation function, similar to the  $T=0$  parametrization (278). Above  $T_c$ , the mass of the  $\pi$ -like mode is expected to grow, but the precise value is hard to determine. The coupling constant is  $\lambda_\pi \approx 1 \text{ fm}^{-2}$  at  $T = 170 \text{ MeV}$ , as compared to  $\lambda_\pi \approx 3 \text{ fm}^{-2}$  at  $T=0$ .

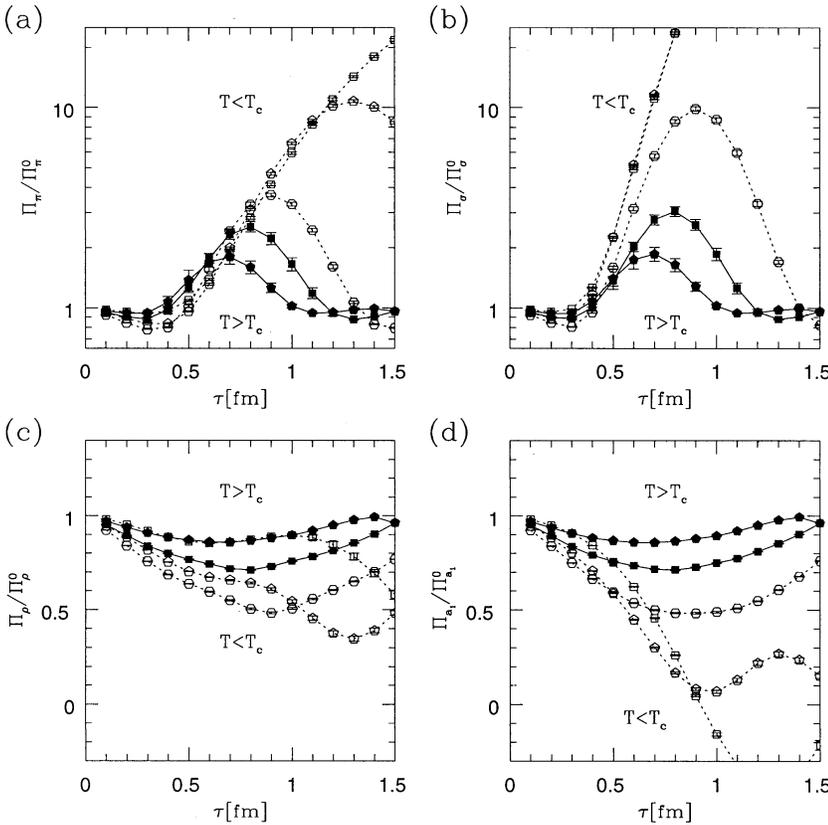


FIG. 36. Temporal correlation functions at  $T \neq 0$ , normalized to free thermal correlators: (a) the pseudoscalar (pion) correlator; (b) the isoscalar scalar  $\sigma$ ; (c) the isovector vector ( $\rho$ ); and (d) the axial vector ( $a_1$ ). Correlators in the chirally symmetric phase ( $T \geq T_c$ ) are shown as solid points, below  $T_c$  as open points.  $\Delta$ ,  $T = 0.43T_c$ ;  $\square$ ,  $T = 0.60T_c$ ;  $\circ$ ,  $T = 0.86T_c$ ;  $\blacktriangle$ ,  $T = 1.00T_c$ ;  $\blacksquare$ ,  $T = 1.13T_c$ .

The  $T$  dependence of the  $\sigma$  correlator is more pronounced because there is a disconnected contribution which tends to the square of the quark condensate at large distance. Above  $T_c$ , chiral symmetry is restored and the  $\sigma$  and  $\pi$  correlation functions are equal up to small corrections due to the current quark masses.

Vector and axial-vector correlation functions are shown in Figs. 36(c) and 36(d). At low  $T$  the two are very different, while above  $T_c$  they become indistinguishable, again in accordance with chiral symmetry restoration. In the vector channel, the changes in the correlation function indicate the “melting” of the resonance contribution. At the lowest temperature, there is a small enhancement in the correlation function at  $\tau \approx 1$  fm, indicating the presence of a bound state separated from the two-quark (or two-pion) continuum. However, this signal disappears at  $T \approx 100$  MeV, implying that the  $\rho$  meson coupling to the local current becomes small.<sup>56</sup> This is consistent with the idea that hadrons “swell” in hot and dense matter. At low temperature, the dominant effect is mixing between the vector and axial-vector channels (Dey, Eletsky, and Ioffe, 1990). In particular, there is a pion contribution to the vector correlator at finite  $T$ , which is most easily observed in the longitudinal vector channel  $\Pi_{44}^V(\tau)$  [in Fig. 36 we show the trace  $\Pi_{\mu\mu}^V(\tau)$  of the vector correlator].

<sup>56</sup>Note, however, that in the instanton model, there is no confinement and the amount of binding in the  $\rho$ -meson channel is presumably small. In full QCD, the  $\rho$  resonance might therefore be more stable as the temperature increases.

Schäfer and Shuryak (1995b) also studied baryon correlation functions at finite temperatures. Because the different nucleon and delta correlation functions have different transformation properties under the chiral  $SU(2) \times SU(2)$  and  $U(1)_A$  symmetries, one can distinguish different modes of symmetry restoration. The main possibilities are (i) that all resonances simply disappear, (ii) that all states become massless as  $T \rightarrow T_c$ , or (iii) that above  $T_c$  all states occur in parity doublets. In the nucleon channel, we find clear evidence for the survival of a massive nucleon mode. There is also support for the presence of a degenerate parity partner above  $T_c$ , so the data seem to favor the third possibility.

In summary, we find that correlation functions in strongly attractive, chiral even channels are remarkably stable as a function of temperature, despite the fact that the quark condensate disappears. There is evidence for the survival of  $(\pi, \sigma)$ -like modes even above  $T_c$ . These modes are bound by the effective quark interaction generated by polarized instanton molecules [see Eq. (332)]. In channels that do not receive large nonperturbative contributions at  $T = 0$  (like the  $\rho$ ,  $a_1$ , and  $\Delta$ ), the resonances disappear, and the correlators can be described in terms of the free-quark continuum (possibly with a residual “chiral” mass).

### 3. $U(1)_A$ breaking

So far, we have not discussed correlation functions related to the  $U(1)_A$  anomaly, in particular the  $\eta$  and  $\eta'$  channels. A number of authors have considered the possibility that the  $U(1)_A$  symmetry is at least partially re-

stored near the chiral phase transition (Shuryak, 1994; Huang and Wang, 1996; Kapusta, Kharzeev, and McLerran, 1996). Given the fact that the  $\eta' - \pi$  mass difference is larger than all other meson mass splittings, any tendency towards  $U(1)_A$  restoration is expected to lead to rather dramatic observable effects. Possible signatures of a hot hadronic phase with partially restored  $U(1)_A$  symmetry that might be produced in heavy-ion collisions are anomalous values of the  $\eta/\pi$  and  $\eta'/\pi$  ratios, as well as an enhancement in the number of low-mass dilepton pairs from the Dalitz decay of the  $\eta'$ .

In general, we know that isolated instantons disappear above the chiral phase transition. In the presence of external sources, however, isolated tunneling events can still occur (see Sec. II.D.2). The density of random instantons above  $T_c$  is expected to be  $O(m^N)$ , but the contribution of isolated instantons to the expectation value of the 't Hooft operator  $\det_f(\bar{\psi}_L \psi_R)$  (and other  $U(1)_A$ -violating operators) is of order  $O(1)$  (Evans, Hsu, and Schwetz, 1996; Lee and Hatsuda, 1996). The presence of isolated zero modes in the spectrum of the Dirac operator above  $T_c$  can be seen explicitly in Fig. 34. The problem was studied in greater detail by Schäfer (1996), where it was concluded that the number of (almost) zero modes above  $T_c$  scales correctly with the dynamic quark mass and the volume.

A number of groups have measured  $U(1)_A$ -violating observables at finite temperature on the lattice. Most work focuses on the susceptibility  $\chi_\pi - \chi_\delta$  (Chandrasekharan, 1995; Boyd, 1996; Bernard *et al.*, 1997). Above  $T_c$ , when chiral symmetry is restored, this quantity is a measure of  $U(1)_A$  violation. Most of the published results conclude that  $U(1)_A$  remains broken, although recent results by the Columbia group have questioned that conclusion<sup>57</sup> (Christ, 1996). In any case, one should keep in mind that all results involve extrapolations to  $m=0$ , and that both instanton and lattice simulations suffer from certain artifacts in this limit.

Phenomenological aspects of the  $U(1)_A$  anomaly at finite temperature are usually discussed in terms of the effective Lagrangian (Pisarski and Wilczek, 1984),

$$\mathcal{L} = \frac{1}{2} \text{Tr}((\partial_\mu \Phi)(\partial_\mu \Phi^\dagger)) - \text{Tr}(\mathcal{M}(\Phi + \Phi^\dagger)) + V(\Phi \Phi^\dagger) + c(\det \Phi + \det \Phi^\dagger), \quad (344)$$

where  $\Phi$  is a meson field in the (3,3) representation of  $U(3) \times U(3)$ ,  $V(\Phi \Phi^\dagger)$  is a  $U(3) \times U(3)$  symmetric potential (usually taken to be quartic),  $\mathcal{M}$  is a mass matrix, and  $c$  controls the strength of the  $U(1)_A$  breaking interaction. If the coupling is chosen as  $c = \chi_{\text{top}} / (12f_\pi^3)$ , the

<sup>57</sup>In addition, it is not clear whether lattice simulations observe the peak in the spectrum at  $\lambda=0$ , which is due to isolated instantons. The Columbia group has measured the valence mass dependence of the quark condensate, which is a folded version of the Dirac spectrum (Chandrasekharan, 1995). The result looks very smooth, with no hint of an enhancement at small virtuality.

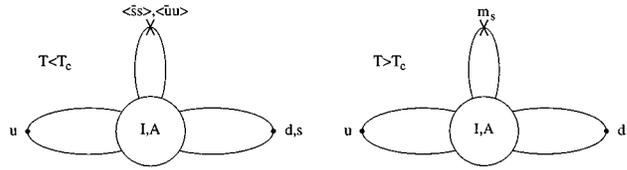


FIG. 37. Leading contributions to flavor mixing in the  $\eta - \eta'$  system below and above the chiral phase transition.

effective Lagrangian reproduces the Witten-Veneziano relation  $f_\pi^2 m_{\eta'}^2 = \chi_{\text{top}}$ . In a quenched ensemble, we can further identify  $\chi_{\text{top}} \approx (N/V)$ . The temperature dependence of  $c$  is usually estimated from the semiclassical tunneling amplitude  $n(\rho) \sim \exp[-(8/3)(\pi\rho T)^2]$ . As a result, the strength of the anomaly is reduced by a factor  $\sim 5$  near  $T_c$ . If the anomaly becomes weaker, the eigenstates are determined by the mass matrix. In that case, the mixing angle is close to the ideal value  $\theta = -54.7^\circ$  and the nonstrange  $\eta_{\text{NS}}$  is almost degenerate with the pion.

There are several points in this line of argument that are not entirely correct. The strength of the 't Hooft term is not controlled by the topological susceptibility ( $\chi_{\text{top}}=0$  in massless QCD),  $\chi_{\text{top}}$  is not proportional to the instanton density (for the same reason), and, at least below  $T_c$ , the semiclassical estimate for the instanton density is not applicable. Only above  $T_c$  do we expect instantons to be suppressed. However, chiral symmetry restoration affects the structure of flavor mixing in the  $\eta - \eta'$  system (see Fig. 37). The mixing between the strange and nonstrange eta is controlled by the light-quark condensate, so  $\eta_{\text{NS}}$  and  $\eta_S$  do not mix above  $T_c$ . As a result, the mixing angle is not close to zero, as it is at  $T=0$ , but close to the ideal value. Furthermore, the anomaly can only affect the nonstrange  $\eta$ , not the strange one. Therefore, if the anomaly is sufficiently strong, the  $\eta_{\text{NS}}$  will be *heavier* than the  $\eta_S$ .

This phenomenon can also be understood from the effective Lagrangian (344). The determinant is third order in the fields, so it only contributes to mass terms if some of the scalar fields have a vacuum expectation value. Above  $T_c$ , only the strange scalar has a vacuum expectation value, so only light flavors mix, and only the  $\eta_{\text{NS}}$  receives a contribution to its mass from the anomaly. This effect was studied more quantitatively by Schäfer (1996). Singlet and octet, as well as strange and nonstrange, eta correlation functions in the instanton liquid are shown in Fig. 38. Below  $T_c$  the singlet correlator is strongly repulsive, while the octet correlator shows some attraction at larger distance. The off-diagonal correlator is small and positive, corresponding to a negative mixing angle. The strange and nonstrange eta correlation functions are very similar, which is a sign for strong flavor mixing. This is also seen directly from the off-diagonal correlator between  $\eta_S$  and  $\eta_{\text{NS}}$ .

Above  $T_c$ , the picture changes. The off-diagonal singlet-octet correlator changes sign, and its value at intermediate distances  $\tau \approx 0.5$  fm is significantly larger. The strange and nonstrange eta correlators are very dif-

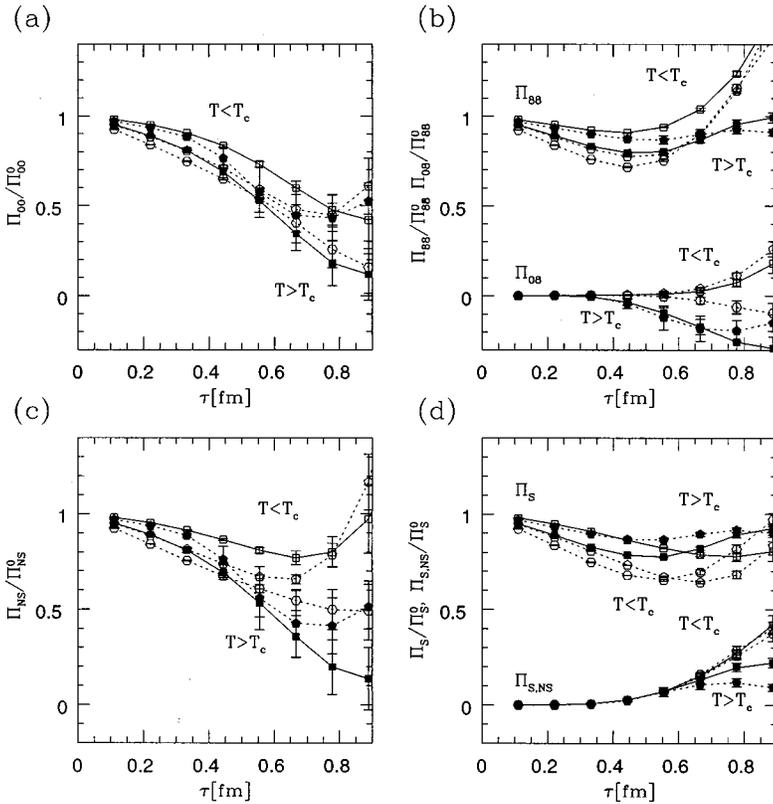


FIG. 38. Eta-meson correlation functions at  $T \neq 0$ , normalized to free thermal correlators: (a) the flavor singlet; (b) the octet and off-diagonal singlet-octet; (c) the nonstrange eta correlation function; (d) the strange and off-diagonal strange-nonstrange eta correlation functions. The correlators are labeled as in Fig. 36.

ferent from each other. The nonstrange correlation function is very repulsive, but no repulsion is seen in the strange channel. This clearly supports the scenario presented above. Near  $T_c$  the eigenstates are essentially the strange and nonstrange components of the  $\eta$ , with the  $\eta_S$  being the lighter of the two states. This picture is not realized completely;  $\Pi_{S,NS}$  does not vanish, and the singlet eta is still somewhat more repulsive than the octet eta correlation function. This is due to the fact that the light-quark mass does not vanish. In particular, in this simulation, the ratio  $(m_u + m_d)/(2m_s) = 1/7$ , which is about three times larger than the physical mass ratio.

It is difficult to provide a quantitative analysis of temporal correlation functions in the vicinity of the phase transition. At high temperatures the temporal direction in a Euclidean box becomes short, and there is no unique way to separate out the contribution from excited states. Nevertheless, under some simplifying assumptions one can try to translate the correlation functions shown in Fig. 38 into definite predictions concerning the masses of the  $\eta$  and  $\eta'$ . For definiteness, we use ideal mixing above  $T_c$  and fix the threshold for the perturbative continuum at 1 GeV. In this case, the masses of the strange and nonstrange components of the  $\eta$  at  $T = 126$  MeV are given by  $m_{\eta_S} = (0.420 \pm 0.120)$  GeV and  $m_{\eta_8} = (1.250 \pm 0.400)$  GeV.

#### 4. Screening masses

Correlation functions in the spatial direction can be studied at arbitrarily large distances, even at finite temperature. This means that (in contrast to the temporal

correlators) the spectrum of the lowest states can be determined with good accuracy. Although it is not directly related to the spectrum of physical excitations, the structure of spacelike screening masses is of theoretical interest and has been investigated in a number of lattice (Tar and Kogut, 1987; Gocksch, 1991) and theoretical (Eletsky and Ioffe, 1988; Hansson and Zahed, 1992; Koch *et al.*, 1992; Hansson, Sporre, and Zahed, 1994) works.

At finite temperatures, the antiperiodic boundary conditions in the temporal direction require the lowest Matsubara frequency for fermions to be  $\pi T$ . This energy acts like a mass term for propagation in the spatial direction, so quarks effectively become massive. At asymptotically large temperatures, quarks propagate only in the lowest Matsubara mode, and the theory undergoes dimensional reduction (Appelquist and Pisarski, 1981). The spectrum of spacelike screening states is then determined by a three-dimensional theory of quarks with chiral mass  $\pi T$ , interacting via the three-dimensional Coulomb law and the nonvanishing spacelike string tension (Borgs, 1985; Manoussakis and Polonyi, 1987).

Dimensional reduction at large  $T$  predicts almost degenerate multiplets of mesons and baryons with screening masses close to  $2\pi T$  and  $3\pi T$ . The splittings of mesons and baryons with different spins can be understood in terms of the nonrelativistic spin-spin interaction. The resulting pattern of screening states is in qualitative agreement with lattice results even at moderate temperatures  $T \approx 1.5T_c$ . The most notable exception is the pion, whose screening mass is significantly below  $2\pi T$ .

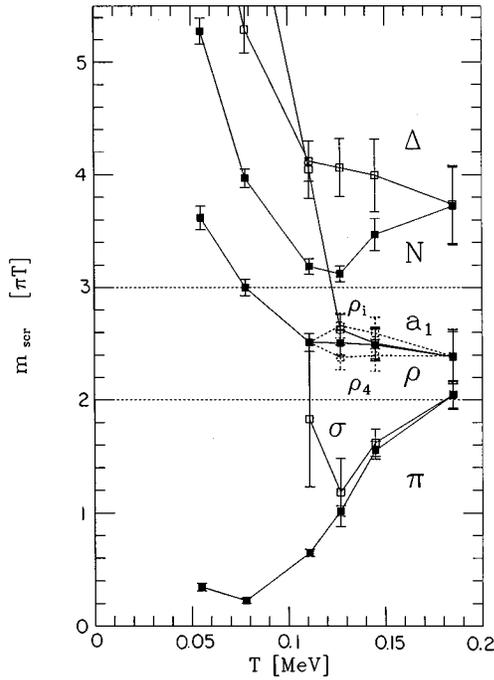


FIG. 39. Spectrum of spacelike screening masses in the instanton liquid as a function of the temperature, from Schäfer and Shuryak, 1996b. The masses are given in units of the lowest fermionic Matsubara frequency  $\pi T$ . The dotted lines correspond to the screening masses  $m=2\pi T$  and  $3\pi T$  for mesons and baryons in the limit when quarks are noninteracting.

Screening masses in the instanton liquid are summarized in Fig. 39. First of all, the screening masses clearly show the restoration of chiral symmetry as  $T \rightarrow T_c$ : chiral partners like the  $\pi$  and  $\sigma$  or the  $\rho$  and  $a_1$  become degenerate. Furthermore, the mesonic screening masses are close to  $2\pi T$  above  $T_c$ , while the baryonic ones are close to  $3\pi T$ , as expected. Most of the screening masses are slightly higher than  $n\pi T$ , consistent with a residual chiral quark mass on the order of 120–140 MeV. The most striking observation is the strong deviation from this pattern seen in the scalar channels  $\pi$  and  $\sigma$ , with screening masses significantly below  $2\pi T$  near the chiral phase transition. The effect disappears around  $T \approx 1.5T_c$ . We also find that the nucleon-delta splitting does not vanish at the phase transition, but decreases smoothly.

In summary, the pattern of screening masses seen in the instanton liquid is in agreement with the results of lattice calculations.<sup>58</sup> In particular, the attractive interaction provided by instanton molecules accounts for the fact that the  $\pi, \sigma$  screening masses are much smaller than  $2\pi T$  near  $T_c$ .

<sup>58</sup>There is one exception, which concerns the screening masses in the longitudinal  $\Pi_{44}$  and transverse  $\Pi_{ii}$  vector channels. In agreement with dimensional reduction, we find  $m_{\rho_i} > m_{\rho_4}$ , while lattice results reported by Tar and Kogut (1987) find the opposite pattern.

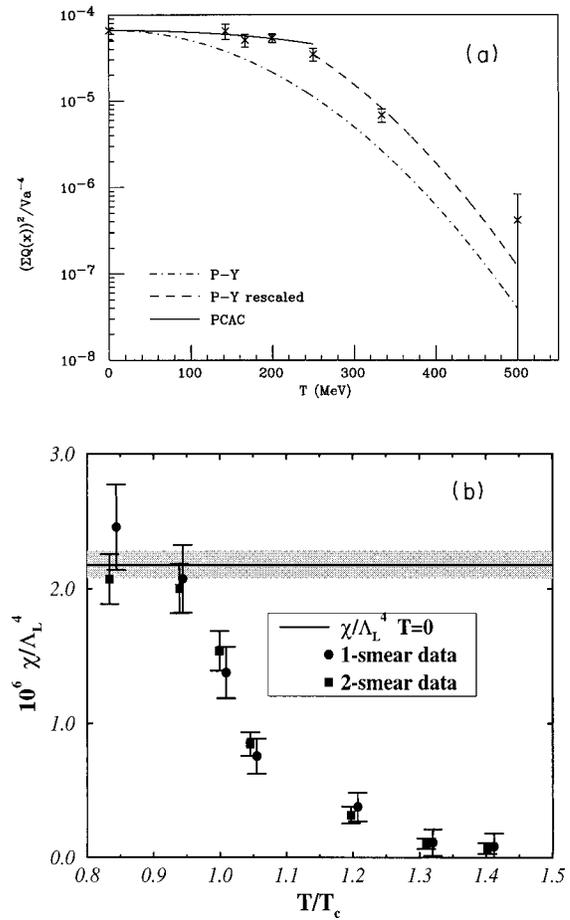


FIG. 40. Lattice measurements of the temperature dependence of the topological susceptibility in pure gauge SU(3): (a) from Chu and Schramm, 1995; (b) from Alles *et al.*, 1997. The dash-dotted line labeled P-Y in (a) corresponds to the semiclassical (Pisarski-Yaffe) result [Eq. (308)], while the dashed curve shows the rescaled formula discussed in the text. The solid line labeled PCAC is not very relevant for pure gauge theory. In (b), the circles and squares correspond to two different topological charge operators.

#### D. Instantons at finite temperature: Lattice studies

Very few lattice simulations have focused on the role of instantons at finite temperature, and all the results that have been reported were obtained in the quenched approximation (Teper, 1986; Hoek *et al.*, 1987; Chu and Schramm, 1995; Ilgenfritz, Müller-Preussker, and Meggiolaro, 1995). In this section we shall concentrate on results obtained by the cooling method, in particular, the work of Chu and Schramm (1995). In this work, the temperature was varied by changing the number of time slices  $N_\tau=2,4,\dots,16$ , while the spatial extent of the lattice and the coupling constant  $\beta=6$  were kept fixed.

The topological susceptibility was calculated from topological charge fluctuations  $\langle Q^2 \rangle / V$  in the cooled configurations. In the quenched theory, correlations between instantons are not very important, and the topological susceptibility provides a good estimate of the instanton density. Figure 40(a) shows their results as a function of  $T$ . At  $T=0$ ,  $\chi_{\text{top}} \approx (180 \text{ MeV})^4$ , in agree-

ment with the phenomenological value. The topological susceptibility is almost temperature independent below the critical temperature ( $T_c \approx 260$  MeV in quenched QCD), but drops very fast above  $T_c$ . The temperature dependence of  $\chi_{\text{top}}$  above  $T_c$  is consistent with the Debye screening suppression factor (308), but with a shifted temperature,  $T^2 \rightarrow (T^2 - T_c^2)$ . Clearly, these results support the arguments presented in Sec. VII.A.3. The conclusions of Chu and Schramm are consistent with results reported by Ilgenfritz *et al.* (1995) and Alles, D'Elia, and DiGiacomo (1996). We show the results of Alles *et al.* in Fig. 40(b).

Chu and Schramm also extracted the average instanton size from the correlation function of the topological charge density. Below  $T_c$ , they found  $\rho = 0.33$  fm, independent of temperature, while at  $T = 334$  MeV, they got a smaller value,  $\rho = 0.26$  fm. This result is in good agreement with the Debye screening dependence discussed above.

Finally, they considered instanton contributions to the pressure and spacelike hadronic wave functions. They found that instantons contributed roughly 15% of the pressure at  $T = 334$  MeV and 5% at  $T = 500$  MeV. While spacelike hadronic wave functions were dominated by instantons at  $T = 0$  (see Sec. VI.G), this was not true at  $T > T_c$ . This is consistent with the idea that spacelike wave functions above  $T_c$  are determined by the spacelike string tension (Koch *et al.*, 1992), which disappears during cooling.

Clearly, studies with dynamic fermions are of great interest. Some preliminary results have been obtained by Ilgenfritz *et al.* (private communication). Using a ‘‘gentle cooling’’ algorithm with only a few cooling iterations in order to prevent instantons and anti-instantons from annihilating each other, they found evidence for an anticorrelation of topological charges in the time direction above  $T_c$ . This would be the first direct lattice evidence for the formation of polarized instanton molecules in the chirally symmetric phase.

## VIII. INSTANTONS IN RELATED THEORIES

### A. Two-dimensional theories

Although two-dimensional theories should logically be placed between the simplest quantum-mechanical systems and Yang-Mills theories, we have postponed their discussion up to now in order not to disrupt the main line of the review. Nevertheless, topological objects play an important role in many low-dimensional models. We do not want to give an exhaustive survey of these theories, but have selected two examples, the  $O(2)$  and  $O(3)$  models, which, in our opinion, provide a few interesting lessons for QCD. As far as other theories are concerned, we refer the reader to the extensive literature, in particular on the Schwinger model (Smilga, 1994a; Steele, Subramanian, and Zahed, 1995) and two-dimensional QCD with fundamental or adjoint fermions (Smilga, 1994b).

### 1. The $O(2)$ sigma model

The  $O(2)$  model is also known as the  $d=2$  Heisenberg magnet or the  $XY$  model. It describes a two-dimensional spin vector  $\vec{S}$  governed by the Hamiltonian

$$\frac{E}{T} = \frac{1}{2t} \int d^2x (\partial_\mu \vec{S})^2, \quad (345)$$

together with the constraint  $\vec{S}^2 = 1$ . Here,  $t = T/J$  is a dimensionless parameter and  $J$  the coupling constant. In this section we follow the more traditional language of statistical mechanics rather than Euclidean quantum field theory. This means that the coordinates are  $x, y$  and the weight factor in the functional integral is  $\exp(-E/T)$ . Of course, one can always switch to field theory language by replacing the energy by the action and the temperature by the coupling constant  $g^2$ .

The statistical sum is Gaussian except for the constraint. We can make this more explicit by parametrizing the two-dimensional spin vector  $\vec{S}$  in the form  $S_1 = \cos \theta$ ,  $S_2 = \sin \theta$ . In this case, the energy is given by  $E/T = (1/2t) \int d^2x (\partial_\mu \theta)^2$ , which would describe a non-interacting scalar field if it were not for the fact that  $\theta$  is a periodic variable. It is often useful to define the theory directly on the lattice. The partition function is given by

$$Z = \int \left( \prod_x \frac{d\theta_x}{2\pi} \right) \exp \left( \beta \sum_{x, \hat{e}_\mu} [\cos(\theta_x - \theta_{x+\hat{e}_\mu}) - 1] \right), \quad (346)$$

which is automatically periodic in  $\theta$ . The lattice spacing  $a$  provides an ultraviolet cutoff. We should also note that the  $O(2)$  model has a number of physical applications. First of all, the model obviously describes a two-dimensional magnet. In addition, fluctuations of the order parameter in superconducting films of liquid  $^4\text{He}$  and the dynamics of dislocations in the melting of two-dimensional crystals are governed by effective  $O(2)$  models.

A two-dimensional theory with a continuous symmetry cannot have an ordered phase at nonzero temperatures. This means that, under ordinary circumstances, two-dimensional models cannot have a phase transition at finite temperatures (Mermin and Wagner, 1966). The  $O(2)$  model is special because it has a phase transition at  $T_c \approx (\pi J)/2 > 0$  (although, in agreement with the Mermin-Wagner theorem, the transition is not characterized by a local order parameter). Both the low- and the high-temperature phase are disordered, but the functional form of the spin-correlation function  $K(x) = \langle \vec{S}(x) \vec{S}(0) \rangle$  changes. In some sense, the whole region  $T < T_c$  is critical because the correlation function exhibits a power-law decay (Beresinsky, 1971). For  $T > T_c$  the spin correlator decays exponentially, and the theory has a mass gap.

The mechanism of this phase transition was clarified in the seminal paper by Koesterlitz and Thouless (1973; see also the review of Kogut, 1979). Let us start with the low-temperature phase. In terms of the angle variable  $\theta$ , the spin-correlation function is given by

$$K(x) = N^{-1} \int d\theta(x) e^{i\theta(x)} e^{-i\theta(0)} e^{-1/(2t) \int d^2x (\partial_\mu \theta)^2}. \quad (347)$$

At low temperatures, the system is dominated by spin waves, and we expect that fluctuations in  $\theta$  are small. In this case we can ignore the periodic character of  $\theta$  for the moment. Using the propagator for the  $\theta$  field,  $G(r) = -1/(2\pi) \log(r/a)$ , we have

$$K(r) = \exp[tG(r)] \sim r^{-t/2\pi}. \quad (348)$$

The correlator shows a power law, with a temperature-dependent exponent  $\eta = T/(2\pi J)$ .

In order to understand the phase transition in the  $O(2)$  model, we have to go beyond Gaussian fluctuations and include topological objects. These objects can be classified by a winding number

$$q = \frac{1}{2\pi} \oint d\vec{x} \cdot \vec{\nabla} \theta. \quad (349)$$

Solutions with  $q = \pm 1$  are called (anti) vortices. A solution with  $q = n$  is given by  $\theta = n\alpha$  with  $\alpha = \arctan(y/x)$ . The energy of a vortex is

$$\frac{E(q=1)}{T} = \frac{\pi}{t} \int \frac{dr}{r} \left( \frac{\partial \theta}{\partial \alpha} \right)^2 = \frac{\pi n^2}{t} \log(R/a), \quad (350)$$

where  $R$  is the IR cutoff. Since the energy is logarithmically divergent, one might think that vortex configurations are irrelevant. In fact, they are crucial for the dynamics of the phase transition.

The reason is that it is not the energy, but the free energy  $F = E - TS$ , which is relevant for the statistical sum. The entropy of an isolated vortex is essentially the logarithm of all possible vortex positions, given by  $S = \log[(R/a)^2]$ . For temperatures  $T > T_c \approx (\pi J)/2$ , entropy dominates over energy, and vortices are important. The presence of vortices implies that the system is even more disordered, and the spin correlator decays exponentially.

What happens to the vortex gas below  $T_c$ ? Although isolated vortices have infinite energy, vortex-antivortex pairs (molecules) have finite energy  $E \approx \pi/(2t) [\log(R/a) + \text{const}]$ , where  $R$  is the size of the molecule. At low temperatures, molecules are strongly suppressed, but as the temperature increases they become more copious. Above the critical temperature, molecules are ionized and a vortex plasma is formed.

There is yet another way to look at this transition. We can decompose any field configuration into the contribution of vortices and a smooth field. Using this decomposition, one can see that the  $O(2)$  sigma model is equivalent to a two-dimensional Coulomb gas. The Kosterlitz-Thouless transition describes the transition from a system of dipoles to an ionized plasma. The correlation length of the spin system is nothing but the Debye screening length in the Coulomb plasma. At the transition point, the screening length has an essential singularity (Kosterlitz and Thouless, 1973)

$$\xi \sim \exp\left(\frac{\text{const}}{|T - T_c|^{1/2}}\right) \quad (351)$$

rather than the power-law divergence observed in ordinary phase transitions.<sup>59</sup>

There is an obvious lesson we should like to draw from this: the two phases of the  $O(2)$  model resemble the two phases found in QCD (Sec. VII.B.4). In both cases the system is described by an (approximately) random ensemble of topological objects in one phase and by bound pairs of pseudoparticles with opposite charge in the other. The only difference is that the high- and low-temperature phases are interchanged. This is essentially due to the fact that temperature has a different meaning in both problems. In the  $O(2)$  model, instantons are excited at high temperature, while in QCD they are suppressed.

## 2. The $O(3)$ sigma model

The  $O(2)$  model can be generalized to a  $d=2$  spin system with a three-dimensional (or, more generally,  $N$ -dimensional) spin vector  $\vec{S}$  of unit length. Unlike the  $O(2)$  model, the  $O(3)$  model is not just a free field theory for a periodic variable. In fact, in many ways the model resembles QCD much more than the  $O(2)$  model does. First of all, for  $N > 2$  the rotation group is non-Abelian, and spin waves [with  $(N-1)$  polarizations] interact. In order to treat the model perturbatively, it is useful to decompose the vector field in the form  $\vec{S} = (\sigma, \vec{\pi})$ , where  $\vec{\pi}$  is an  $(N-1)$ -dimensional vector field and  $\sigma = \sqrt{1 - \vec{\pi}^2}$ . The perturbative analysis of the  $O(N)$  model gives the beta function (Polyakov, 1975)

$$\beta(t) = -\frac{N-2}{2\pi} t^2 + O(t^3), \quad (352)$$

which shows that just like QCD, the  $O(N)$  model is asymptotically free for  $N > 2$ . In the language of statistical mechanics, the beta function describes the evolution of the effective temperature as a function of the scale. A negative beta function then implies that if the temperature is low at the atomic scale  $a$ , the effective temperature grows as one goes to larger scales. Eventually, the reduced temperature is  $t = O(1)$  ( $T$  is comparable to  $J$ ), fluctuations are large, and long-range order is destroyed. Unlike the  $N=2$  model, the  $O(3)$  model has only one critical point,  $t=0$ , and correlation functions decay exponentially for all temperatures  $T > 0$ .

Furthermore, like non-Abelian gauge theory, the  $O(3)$  model has classical instanton solutions. The topology is provided by the fact that field configurations are maps from two-dimensional spacetime (compactified to a

<sup>59</sup>The interested reader can find further details in review talks on spin models given at the annual Lattice meetings. Using cluster algorithms to fight critical slowdown, one can obtain very accurate data. In Wolff (1989), the correlation length reaches about 70 lattice units, confirming Eq. (351). Nevertheless, not all Kosterlitz-Thouless results are reproduced: e.g., the value of index  $\eta$  (defined by  $K(r) \sim 1/r^\eta$  as  $T \rightarrow T_c$ ) is not 1/4 but noticeably larger.

sphere) into another sphere which describes the orientation of  $\vec{S}$ . The winding number is defined by

$$q = \frac{1}{8\pi} \int d^2x \epsilon_{\mu\nu} \vec{S} \cdot (\partial_\mu \vec{S} \times \partial_\nu \vec{S}). \quad (353)$$

Similar to QCD, solutions with integer winding number and energy  $E = 4\pi J|q|$  can be found from the self-duality equation

$$\partial_\mu \vec{S} = \pm \epsilon_{\mu\nu} \vec{S} \times \partial_\nu \vec{S}. \quad (354)$$

A solution with  $q=1$  can be constructed by means of the stereographic projection

$$\vec{S} = \left( \frac{2\rho x_1}{x^2 + \rho^2}, \frac{2\rho x_2}{x^2 + \rho^2}, \frac{x^2 - \rho^2}{x^2 + \rho^2} \right). \quad (355)$$

At large distances all spins point up, in the center the spin points down, and at intermediate distances  $x \sim \rho$  all spins are horizontal, pointing away from the center. Like QCD, the theory is scale invariant on the classical level, and the energy of an instanton is independent of the size  $\rho$ . Instantons in the  $O(2)$  model are sometimes called Skyrmons, in analogy with the static (three-dimensional) solitons introduced by Skyrme in the  $d=4$   $O(4)$  model (Skyrme, 1961).

The next logical step is the analog of the 't Hooft calculation of fluctuations around the classical instanton solution. The result of the semiclassical calculation is (Polyakov, 1975)

$$dN_{\text{inst}} \sim \frac{d^2x d\rho}{\rho^3} \exp(-E/T) \sim \frac{d^2x d\rho}{\rho}, \quad (356)$$

which is divergent both for large and small radii.<sup>60</sup> The result for large  $\rho$  is of course not reliable, since it is based on the one-loop beta function.

Because the  $O(3)$  model shows so many similarities with QCD, it is natural to ask whether one can learn anything of relevance for QCD. Indeed, the  $O(3)$  model has been widely used as a testing ground for new methods in lattice gauge theory. The numerical results provide strong support for the renormalization-group analysis. For example, Caracciolo, Edwards, and Sokal (1995) studied the  $O(3)$  model at correlation lengths as large as  $\xi/a \sim 10^5$ . The results agree with state-of-the-art theoretical predictions (based on the three-loop beta function and an overall constant determined from the Bethe ansatz) with a very impressive accuracy, on the order of a few percent.

Unfortunately, studies of instantons in the  $O(3)$  model have not produced any significant insights. In particular, there are no indications that small-size (semiclassical) instantons play any role in the dynamics of the

theory. Because of the divergence in the instanton density, the topological susceptibility in the  $O(3)$  has no continuum limit. This conclusion is supported by lattice simulations (Michael and Spencer, 1994). The simulations also indicate that, for large-size instantons, the size distribution is  $dN \sim d\rho/\rho^{-3}$ . This power differs from the semiclassical result, but it agrees with the  $\rho$  dependence coming from the Jacobian alone, without the running coupling in the action. This result supports the idea of a frozen coupling constant discussed in Sec. III.C.3.

## B. Instantons in electroweak theory

In the context of this review, we cannot provide a detailed discussion of instantons and baryon-number violation in electroweak theory. Nevertheless, we briefly touch on this subject because electroweak theory provides an interesting theoretical laboratory. The coupling constant is small, the instanton action is large, and the semiclassical approximation is under control. Unfortunately, this means that under ordinary conditions tunneling events are too rare to be of physical importance. Interesting questions arise when one tries to increase the tunneling rate, e.g., by studying scattering processes with collision energies close to the barrier height, or processes in the vicinity of the electroweak phase transition. Another interesting problem is what happens if we consider the Higgs expectation value to be a free parameter. When the Higgs vacuum expectation value is lowered, electroweak theory becomes a strongly interacting theory, and we encounter many of the problems we have to deal with in QCD.

Electroweak theory is an  $SU(2)_L \times U(1)$  chiral gauge theory coupled to a Higgs doublet. For simplicity we shall neglect the  $U(1)$  interactions in the following, i.e., set the Weinberg angle to zero. The most important difference, as compared to QCD, is the fact that gauge invariance is spontaneously broken. In the ground state, the Higgs field acquires an expectation value, which gives masses to the  $W$  bosons as well as to the quarks and charged leptons. If the Higgs field has a nonzero vacuum expectation value the instanton is, strictly speaking, not a solution of the equations of motion. Nevertheless, it is clear that if  $\rho \ll v^{-1}$  (where  $v$  is the Higgs vacuum expectation value), the gauge fields are much stronger than the Higgs field, and there should be an approximate instanton solution.<sup>61</sup> In the central region  $x < m_W^{-1}, m_H^{-1}$ , the solution can be found by keeping the instanton gauge field fixed and solving the equations of motion for the Higgs field. The result is

$$\phi = \frac{x^2}{x^2 + \rho^2} U \left( \begin{array}{c} 0 \\ v/\sqrt{2} \end{array} \right), \quad (357)$$

<sup>60</sup>This result marks the first nonperturbative ultraviolet divergence ever discovered. It is similar to the divergence of the density of instanton molecules for large  $N_f$  discussed in Sec. IX.D.

<sup>61</sup>This notion can be made more precise using the constrained instanton solution (Affleck, 1981). This technique is similar to the construction that defines the streamline solution for an instanton anti-instanton pair.

where  $U$  is the color orientation matrix of the instanton. The result shows that the instanton makes a hole of size  $\rho$  in the Higgs condensate. Scale invariance is lost, and the instanton action depends on the Higgs vacuum expectation value ('t Hooft, 1976b)

$$S = \frac{8\pi^2}{g^2} + 2\pi^2\rho^2v^2. \quad (358)$$

The tunneling rate is  $p \sim \exp(-S)$ , and large instantons with  $\rho \gg v^{-1}$  are strongly suppressed.

The loss of scale invariance also implies that, in electroweak theory, the height of the barrier separating different topological vacua can be determined. There is a static solution with winding number 1/2, corresponding to the top of the barrier, called the sphaleron (Klinkhammer and Manton, 1984). The sphaleron energy is  $E_{\text{sph}} \approx 4m_W/\alpha_W \approx 10$  TeV.

Electroweak instantons also have fermionic zero modes, and as usual the presence of these zero modes is connected with the axial anomaly. Since only left-handed fermions participate in weak interactions, neither vector nor axial-vector currents are conserved. The 't Hooft vertex contains all twelve weak doublets,

$$\begin{aligned} &(\nu_e, e), \quad (\nu_\mu, \mu), \quad (\nu_\tau, \tau), \quad 3*(u, d), \\ &3*(c, s), \quad 3*(t, b), \end{aligned} \quad (359)$$

where the factors of three come from color. Each doublet provides one fermionic zero mode, the flavor depending on the isospin orientation of the instanton. The 't Hooft vertex violates both baryon and lepton number. These processes are quite spectacular because all families have to be involved, for example,

$$u + d \rightarrow \bar{d} + \bar{s} + 2\bar{c} + 3\bar{t} + e^+ + \mu^+ + \tau^+, \quad (360)$$

$$u + d \rightarrow \bar{u} + 2\bar{s} + \bar{c} + \bar{t} + 2\bar{b} + \nu_e + \nu_\mu + \tau^+. \quad (361)$$

Note that  $\Delta B = \Delta L = -3$ , so  $B + L$  is violated, but  $B - L$  is conserved. Unfortunately, the probability of such an event is tiny, proportional to the square of the tunneling amplitude  $P \sim \exp(-16\pi^2/g_w^2) \sim 10^{-169}$  ('t Hooft, 1976a), many orders of magnitude smaller than any known radioactive decay.

Many authors have discussed the possibility of increasing the tunneling rate by studying processes near the electroweak phase transition (Kuzmin, Rubakov, and Shaposhnikov, 1985) or scattering processes involving energies close to the sphaleron barrier  $E \approx E_{\text{sph}}$ . Since  $E_{\text{sph}} \approx 10$  TeV, this energy would have been accessible at the SSC and will possibly be within reach at the LHC. The latter idea became attractive when it was realized that associated multi-Higgs and  $W$  production increases the cross section (Espinosa, 1990; Ringwald, 1990). On general grounds one expects

$$\sigma_{\Delta(B+L)} \sim \exp\left[-\frac{4\pi}{\alpha_W} F\left(\frac{E}{E_{\text{sph}}}\right)\right], \quad (362)$$

where  $F(\epsilon)$  is called the ‘‘holy grail’’ function. At low energy,  $F(0)=1$  and baryon-number violation is strongly suppressed. The first correction is  $F(\epsilon)=1 - \frac{9}{8}\epsilon^{4/3} + O(\epsilon^2)$ , indicating that the cross section rises with energy. Clearly, the problem is the behavior of  $F(\epsilon)$  near  $\epsilon=1$ . Most authors now seem to agree that  $F(\epsilon)$  will not drop below  $F \approx 1/2$ , implying that baryon-number violation in  $pp$  collisions will remain unobservable (Maggiore and Shifman, 1992a, 1992b; Veneziano, 1992; Zakharov, 1992; Diakonov and Petrov, 1994). The question is of interest also for QCD because the holy grail function at  $\epsilon=1$  is related to the instanton/anti-instanton interaction at short distance. In particular, taking into account unitarity in the multi- $W$  (or multigluon) production process corresponds to an effective instanton/anti-instanton repulsion; see Sec. IV.A.1.

The other question of interest for QCD is what happens if the Higgs vacuum expectation value  $v$  is gradually reduced. As  $v$  becomes smaller the theory moves from the weak-coupling to the strong-coupling regime. The vacuum structure changes from a very dilute system of IA molecules to a more dense (and more interesting) nonperturbative vacuum. Depending on the number of light fermions, one should eventually reach a confining, QCD-like phase. In this phase, leptons are composite, but the low-energy effective action is probably similar to the one in the Higgs phase (Abbott and Farhi, 1981; Claudson, Farhi, and Jaffe, 1986). Unfortunately, the importance of nonperturbative effects has never been studied.

Let us comment only on one element that is absent in QCD, the scalar-induced interaction between instantons. Although the scalar interaction is of order  $O(1)$  and therefore suppressed with respect to the gauge interaction  $O(g^{-2})$ , it is long range if the Higgs mass is small. For  $m_H^{-1} \gg R \gg \rho$  the IA interaction is (Yung, 1988)

$$S_{\text{Higgs}} = 4\pi^2 a^2 \rho^2 \left[ 1 + \frac{\rho^2}{R^2} [2(u \cdot \hat{R})^2 - 1] \right] + O\left(\frac{1}{R^4}\right). \quad (363)$$

Unlike the gluonic dipole interaction, it does not vanish if averaged over all orientations,  $\langle (u \cdot \hat{R})^2 \rangle = 1/4$ . This means that the scalar interaction can provide coherent attraction for distances  $Rm_H < 1$ , which is of the order  $v^2 \rho^4 n/m_H^2$  where  $n$  is the instanton density. This is large if the Higgs mass is small.

Another unusual feature of Yung interaction (363) is that it is repulsive for  $u \cdot \hat{R} = 1$  (which is the most attractive orientation for the dipole interaction). This would suggest that, for a light Higgs mass, there is no small- $R$  problem. This question was studied by Velkovsky and Shuryak (1993). For the complete Yung ansatz (which is a good approximation to the full streamline solution) the approximate result (363) is valid only for  $R > 10\rho$ , while for smaller separation the dependence on  $u \cdot \hat{R}$  is reversed.

### C. Supersymmetric QCD

Supersymmetry (SUSY) is a powerful theoretical concept, which is of great interest in constructing field theories beyond the standard model. In addition, supersymmetric field theories provide a very useful theoretical laboratory and have been responsible for a number of important advances in our understanding of the ground state of strongly coupled field theories. We have already seen one example of the usefulness of supersymmetry in isolating instanton effects in the context of SUSY quantum mechanics, in Sec. II.B. In that case, SUSY implies that perturbative contributions to the vacuum energy vanish and allows a precise definition of the instanton/anti-instanton contribution. In the following, we shall see that supersymmetry can be used in very much the same way in the context of gauge-field theories.

In general, one should keep in mind that SUSY theories are just ordinary field theories with a very specific matter content and certain relations between different coupling constants. Eventually, we hope to understand non-Abelian gauge theories for all possible matter sectors. In particular, we want to know how the structure of the theory changes as one goes from QCD to supersymmetric generalizations, where many exact statements about instantons and the vacuum structure are known. Deriving these results often requires special techniques that go beyond the scope of this review. For details of the supersymmetric instanton calculus we refer the reader to the extensive review by Amati *et al.* (1988). Nevertheless, we have tried to include a number of interesting results and explain them in standard language.

As theoretical laboratories, SUSY theories have several advantages over ordinary field theories. We have already mentioned one of them: Nonrenormalization theorems imply that many quantities do not receive perturbative contributions, so instanton effects are more easily identified. In addition to that, SUSY gauge theories usually have many degenerate classical vacua. These degeneracies cannot be lifted to any order in perturbation theory, and instantons often play an important role in determining the ground state. In most cases, the classical vacua are characterized by scalar field vacuum expectation values. If the scalar field vacuum expectation value is large, one can perform reliable semiclassical calculations. Decreasing the scalar vacuum expectation value, one moves towards strong coupling, and the dynamics of the theory is nontrivial. Nevertheless, supersymmetry restricts the functional dependence of the effective potential on the scalar vacuum expectation value (and other parameters, like masses or coupling constants), so that instanton calculations can often be continued into the strong-coupling domain.

Ultimately, we should like to understand the behavior of SUSY QCD as we introduce soft supersymmetry breaking terms and send the masses of the gluinos and squarks to infinity. Not much progress has been achieved in this direction, but at least for small breaking the calculations are feasible and some lessons have been learned.

#### 1. The instanton measure and the perturbative beta function

The simplest ( $N=1$ ) supersymmetric non-Abelian gauge theory is SU(2) SUSY gluodynamics, defined by

$$\mathcal{L} = -\frac{1}{4g^2} F_{\mu\nu}^a F_{\mu\nu}^a + \frac{i}{2g^2} \lambda^a (\mathbb{D})_{ab} \lambda^b, \quad (364)$$

where the gluino field  $\lambda^a$  is a Majorana fermion in the adjoint representation of SU(2). More complicated theories can be constructed by adding additional matter fields and scalars.  $N$  extended supersymmetry has  $N$  gluino fields as well as additional scalars. Clearly, supersymmetric gluodynamics has  $\theta$  vacua and instanton solutions. The only difference as compared to QCD is that the fermions carry adjoint color, so there are twice as many fermion zero modes. If the model contains scalar fields that acquire a vacuum expectation value, instantons are approximate solutions, and the size integration is automatically cut off by the scalar vacuum expectation value. These theories usually resemble electroweak theory more than they do QCD.

At first glance, instanton amplitudes seem to violate supersymmetry: the number of zero modes for gauge fields and fermions does not match, while scalars have no zero modes at all. However, one can rewrite the tunneling amplitude in manifestly supersymmetric form (Novikov *et al.*, 1983). We shall not do this here, but stick to the standard notation. The remarkable observation is that the determination of the tunneling amplitude in SUSY gauge theory is actually simpler than in QCD. Furthermore, with some additional input, one can determine the complete perturbative beta function from the tunneling amplitude.

The tunneling amplitude is given by

$$n(\rho) \sim \exp\left(-\frac{2\pi}{\alpha}\right) M^{n_g - n_f/2} \left(\frac{2\pi}{\alpha}\right)^{n_g/2} \times d^4x \frac{d\rho}{\rho^5} \rho^k \prod_f d^2\xi_f, \quad (365)$$

where all factors can be understood from the 't Hooft calculation discussed in Sec. II.C.4. There are  $n_g = 4N_c$  bosonic zero modes that have to be removed from the determinant and these give one power of the regulator mass  $M$  each. Similarly, each of the  $n_f$  fermionic zero modes gives a factor  $M^{1/2}$ . Introducing collective coordinates for the bosonic zero modes gives a Jacobian  $\sqrt{S_0}$  for every zero mode. Finally,  $d^2\xi$  is the integral over the fermionic collective coordinates, and  $\rho^k$  is the power of  $\rho$  needed to give the correct dimension. Supersymmetry now ensures that all non-zero-mode contributions exactly cancel. More precisely, the subset of SUSY transformations that does not rotate the instanton field itself mixes fermionic and bosonic non-zero modes but annihilates zero modes. This is why all non-zero modes cancel, but zero modes can be unmatched. Note that as a result of this cancellation, the power of  $M$  in the tunneling amplitude is an integer.

Renormalizability demands that the tunneling amplitude be independent of the regulator mass. This means that the explicit  $M$  dependence of the tunneling amplitude and the  $M$  dependence of the bare coupling have to cancel. As in QCD, this allows us to determine the one-loop coefficient of the beta function  $b = (4 - N)N_c - N_f$ . Again note that  $b$  is an integer, a result that would appear very mysterious if we did not know about instanton zero modes.

In supersymmetric theories one can even go one step further and determine the beta function to all loops (Novikov *et al.*, 1983; Vainshtein *et al.*, 1986). For that purpose let us write down the renormalized instanton measure

$$n(\rho) \sim \exp\left(-\frac{2\pi}{\alpha}\right) M^{n_g - n_f/2} \left(\frac{2\pi}{\alpha_R}\right)^{n_g/2} Z_g^{n_g/2} \times \left(\prod_f Z_f^{-1/2}\right) d^4x \frac{d\rho}{\rho^5} \rho^k \prod_f d^2\xi_f, \quad (366)$$

where we have introduced the field renormalization factors  $Z_{g,f}$  for the bosonic/fermionic fields. Again, non-renormalization theorems ensure that the tunneling amplitude is not renormalized at higher orders (the cancellation between the non-zero-mode determinants persists beyond one loop). For gluons the field renormalization (by definition) is the same as the charge renormalization  $Z_g = \alpha_R/\alpha_0$ . Furthermore, supersymmetry implies that the field renormalization is the same for gluinos and gluons. This means that the only new quantity in Eq. (366) is the anomalous dimension of the quark fields,  $\gamma_\psi = d \log Z_f/d \log M$ .

Again, renormalizability demands that the amplitude be independent of  $M$ . This condition gives the Novikov-Shifman-Vainshtein-Zakharov beta function (Novikov *et al.*, 1983) which, in the case  $N=1$ , reads

$$\beta(g) = -\frac{g^3}{16\pi^2} \frac{3N_c - N_f + N_f\gamma_\psi(g)}{1 - N_c g^2/8\pi^2}. \quad (367)$$

The anomalous dimension of the quarks has to be calculated perturbatively. To leading order, it is given by

$$\gamma_\psi(g) = -\frac{g^2}{8\pi^2} \frac{N_c^2 - 1}{N_c} + O(g^4). \quad (368)$$

The result (367) agrees with explicit calculations up to three loops (Jack, Jones, and North, 1997). Note that the beta function is scheme dependent beyond two loops, so in order to make a comparison with high-order perturbative calculations, one has to translate from the Pauli-Villars scheme to a more standard perturbative scheme, e.g.,  $\overline{MS}$ .

In theories without quarks, the Novikov-Shifman-Vainshtein-Zakharov result determines the beta function completely. For  $N$ -extended supersymmetric gluodynamics, we have

$$\beta(g) = -\frac{g^3}{16\pi^2} \frac{N_c(4 - N)}{1 - (2 - N)N_c g^2/(8\pi^2)}. \quad (369)$$

One immediately recognizes two interesting special cases. For  $N=4$ , the beta function vanishes, and the theory is conformal. In the case  $N=2$ , the denominator vanishes, and the one-loop result for the beta function is exact.

## 2. $N=1$ supersymmetric QCD

In the previous section we used instantons only as a tool to simplify a perturbative calculation. Naturally, the next question is whether instantons cause important dynamic effects in SUSY theories. Historically, there has been a good deal of interest in SUSY breaking by instantons, first discovered in the quantum-mechanical model discussed in Sec. II.B. The simplest field theory in which SUSY is broken by instantons is the  $SU(2) \times SU(3)$  model (Affleck, Dine, and Seiberg, 1984a, 1985; Vainshtein, Shifman, and Zakharov, 1985). However, nonperturbative SUSY breaking does not take place in supersymmetric QCD, so we shall not discuss it any further.

An effect that is more interesting in our context is gluino condensation in SUSY gluodynamics [Eq. (364); Novikov *et al.*, 1983]. For  $SU(2)$  color, the gluino condensate is most easily determined from the correlator  $\langle \lambda_\alpha^a \lambda^{a\alpha}(x) \lambda_\beta^b \lambda^{b\beta}(0) \rangle$ . In  $SU(N)$ , one has to consider the  $N$ -point function of  $\lambda_\alpha^a \lambda^{a\alpha}$ . In supersymmetric theories, gluinos have to be massless, so the tunneling amplitude is zero. However, in the  $N$ -point correlation function all zero modes can be absorbed by external sources, similar to the axial anomaly in QCD. Therefore there is a nonvanishing one-instanton contribution. In agreement with SUSY Ward identities, this contribution is  $x$  independent. Therefore one can use cluster decomposition to extract the gluino condensate  $\langle \lambda\lambda \rangle = \pm A \Lambda^3$  [in  $SU(2)$ ], where  $A$  is a constant that is fixed from the single-instanton calculation.

There are a number of alternative methods for calculating the gluino condensate in SUSY gluodynamics. For example, it has been suggested that  $\langle \lambda\lambda \rangle$  can be calculated directly using configurations with fractional charge (Cohen and Gomez, 1984). In addition to that, one can include matter fields, make the theory Higgs-like, and then integrate out the matter fields (Novikov *et al.*, 1985a; Shifman and Vainshtein, 1988). This method gives a different coefficient for the gluino condensate, a problem that was recently discussed by Kovner and Shifman (1997).

The next interesting theory is  $N=1$  SUSY QCD, where we add  $N_f$  matter fields (quarks  $\psi$  and squarks  $\phi$ ) in the fundamental representation. Let us first look at the Novikov-Shifman-Vainshtein-Zakharov beta function. For  $N=1$ , the beta function blows up at  $g_*^2 = 8\pi^2/N_c$ , so the renormalization-group trajectory cannot be extended beyond this point. Recently, Kogan and Shifman have suggested that at this point the standard phase meets the renormalization-group trajectory of a different (nonasymptotically free) phase of the theory (Kogan and Shifman, 1995). The beta function vanishes at  $g_*^2/(8\pi^2) = [N_c(3N_c - N_f)]/[N_f(N_c^2 - 1)]$ , where we

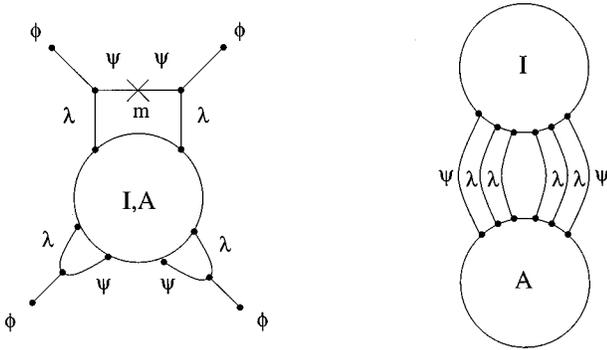


FIG. 41. Instanton contributions to the effective potential in supersymmetric QCD.

have used the one-loop anomalous dimension. This is reliable if  $g_*$  is small, which we can ensure by choosing  $N_c \rightarrow \infty$  and  $N_f$  in the conformal window  $3N_c/2 < N_f < 3N_c$ . Recently, Seiberg has shown that the conformal point exists for all  $N_f$  in the conformal window (even if  $N_c$  is not large) and has clarified the structure of the theory at the conformal point (Seiberg, 1994).

Let us now examine the vacuum structure of SUSY QCD for  $N_c=2$  and  $N_f=1$ . In this case we have one Majorana fermion (the gluino), one Dirac fermion (the quark), and one scalar squark (or Higgs) field. The quark and squark fields do not have to be massless. We shall denote their mass by  $m$ , while  $v$  is the vacuum expectation value of the scalar field. In the semiclassical regime  $v \gg \Lambda$ , the tunneling amplitude (366) is  $O(\Lambda^5)$ . The 't Hooft effective Lagrangian is of the form  $\lambda^4 \psi^2$ , containing four quark and two gluino zero modes. As usual, instantons give an expectation value to the 't Hooft operator. However, due to the presence of Yukawa couplings and a Higgs vacuum expectation value, they can also provide expectation values for operators with fewer than six fermion fields. In particular, combining the quarks with a gluino and a squark tadpole, we can construct a two-gluino operator. Instantons therefore lead to gluino condensation (Affleck, Dine, and Seiberg, 1984b). Furthermore, using two more Yukawa couplings one can couple the gluinos to external quark fields. Therefore, if the quark mass is nonzero, we get a finite density of individual instantons  $O(m\Lambda^5/v^2)$ ; see Fig. 41.

As in ordinary QCD, there are also instanton/anti-instanton molecules. The contribution of molecules to the vacuum energy can be calculated either directly or indirectly, using the single-instanton result and arguments based on supersymmetry. This provides a nice check on the direct calculation because in this case we have a rigorous definition of the contribution from molecules. Calculating the graph shown in Fig. 41(b), one finds (Yung, 1988)

$$\epsilon_{\text{vac}}^{\text{IA}} = \frac{32\Lambda^{10}}{g^8 v^6}, \quad (370)$$

which agrees with the result originally derived by Affleck *et al.* (1984b) using different methods. The result implies that the Higgs expectation value is driven to in-

finiteness, and  $N_f=1$  SUSY QCD does not have a stable ground state. The vacuum can be stabilized by adding a mass term. For nonzero quark mass  $m$  the vacuum energy is given by

$$\epsilon_{\text{vac}} = 2m^2 v^2 - \frac{16m\Lambda^5}{g^4 v^2} + \frac{32\Lambda^{10}}{g^8 v^6} = 2 \left( mv - \frac{4\Lambda^5}{g^4 v^3} \right)^2, \quad (371)$$

where the first term is the classical Higgs mass term, the second is the one-instanton contribution, and the third is due to molecules. In this case, the theory has a stable ground state at  $v^2 = \pm 2\Lambda^{5/2}/(g^2 m^{1/2})$ . Since SUSY is unbroken, the vacuum energy is exactly zero. Let us note that in the semiclassical regime  $m \ll \Lambda \ll v$  everything is under control: instantons are small  $\rho \sim v^{-1}$ , individual instantons are rare, molecules form a dilute gas,  $nR^4 \sim (v/\Lambda)^{10}$ , and the instanton and anti-instanton inside a molecule are well separated,  $R_{\text{IA}}/\rho \sim 1/\sqrt{g}$ . Supersymmetry implies that the result remains correct even if we leave the semiclassical regime. This means, in particular, that all higher-order instanton corrections [ $O(\Lambda^{15})$  etc] have to cancel exactly. Checking this explicitly might provide a very nontrivial check of the instanton calculus.

Recently, significant progress has been made in determining the structure of the ground state of  $N=1$  supersymmetric QCD for arbitrary  $N_c$  and  $N_f$  (Intriligator and Seiberg, 1996). Seiberg showed that the situation discussed above is generic for  $N_f < N_c$ , a stable ground state can only exist for nonzero quark mass. For  $N_f = N_c - 1$ , the  $m \neq 0$  contribution to the potential is an instanton effect. In the case  $N_f = N_c$ , the theory has chiral symmetry breaking (in the chiral limit  $m \rightarrow 0$ ) and confinement.<sup>62</sup> For  $N_f = N_c + 1$ , there is confinement but no chiral symmetry breaking. Unlike QCD, the 't Hooft anomaly matching conditions can be satisfied with massless fermions. These fermions can be viewed as elementary fields in the dual (or "magnetic") formulation of the theory. For  $N_c + 1 < N_f < 3N_c/2$ , the magnetic formulation of the theory is IR free, while for  $3N_c/2 < N_f < 3N_c$  the theory has an infrared fixed point. Finally, for  $N_f > 3N_c$ , asymptotic freedom is lost and the electric theory is IR free.

### 3. $N=2$ supersymmetric gauge theory

The work of Seiberg has shown that supersymmetry sufficiently constrains the effective potential in  $N=1$  supersymmetric QCD that the possible vacuum states for all  $N_c$  and  $N_f$  can be determined. For  $N=2$  extended supersymmetric QCD, these constraints are even more powerful. Witten and Seiberg have been able to determine not just the effective potential, but the complete

<sup>62</sup>There is no instanton contribution to the superpotential, but instantons provide a constraint on the allowed vacua in the quantum theory. To our knowledge, the microscopic mechanism for chiral symmetry breaking in this theory has not been clarified.

low-energy effective action (Seiberg and Witten, 1994). Again, the techniques used in this work are outside the scope of this review. However, instantons play an important role in this theory, so we should like to discuss a few interesting results.

$N=2$  supersymmetric gauge theory contains two Majorana fermions  $\lambda_\alpha^a, \psi_\alpha^a$  and a complex scalar  $\phi^a$ , all in the adjoint representation of  $SU(N_c)$ . In the case of  $N=2$  SUSY QCD, we add  $N_f$  multiplets  $(q_\alpha, \tilde{q}_\alpha, Q, \tilde{Q})$  of quarks  $q_\alpha, \tilde{q}_\alpha$  and squarks  $Q, \tilde{Q}$  in the fundamental representation. In general, gauge invariance is broken and the Higgs field develops an expectation value  $\langle \phi \rangle = a\tau^3/2$ . The vacua of the theory can be labeled by a (gauge-invariant) complex number  $u = (1/2)\langle \text{tr } \phi^2 \rangle$ . If the Higgs vacuum expectation value  $a$  is large ( $a \gg \Lambda$ ), the semiclassical description is valid and  $u = (1/2)a^2$ . In this case, instantons are small  $\rho \sim a^{-1}$  and the instanton ensemble is dilute  $n\rho^4 \sim \Lambda^4/a^4 \ll 1$ .

In the semiclassical regime, the effective Lagrangian is given by

$$\begin{aligned} \mathcal{L}_{\text{eff}} = & \frac{1}{4\pi} \text{Im} \left[ -\mathcal{F}'(\phi) \left( \frac{1}{2} (F_{\mu\nu}^{sd})^2 + (\partial_\mu \phi)(\partial^\mu \phi^\dagger) \right. \right. \\ & \left. \left. + i\psi\partial\bar{\psi} + i\lambda\partial\bar{\lambda} \right) + \frac{1}{\sqrt{2}} \mathcal{F}'''(\phi) \lambda \sigma^{\mu\nu} \psi F_{\mu\nu}^{sd} \right. \\ & \left. + \frac{1}{4} \mathcal{F}''''(\phi) \psi^2 \lambda^2 \right] + \mathcal{L}_{\text{aux}} + \dots, \end{aligned} \quad (372)$$

where  $F_{\mu\nu}^{sd} = F_{\mu\nu} + i\tilde{F}_{\mu\nu}$  is the self dual part of the field strength tensor,  $\mathcal{L}_{\text{aux}}$  contains auxiliary fields, and “...” denotes higher derivative terms. Note that the effective low-energy Lagrangian contains only the light fields. In the semiclassical regime, this is the  $U(1)$  part of the gauge field (the “photon”) and its superpartners. Using arguments based on electric-magnetic duality, Seiberg and Witten determined the exact prepotential  $\mathcal{F}(\phi)$ . From the effective Lagrangian, we can immediately read off the effective charge at the scale  $a$ ,

$$\mathcal{F}''(a) = \frac{\tau(a)}{2} = \frac{4\pi i}{g^2(a)} + \frac{\theta}{2\pi}, \quad (373)$$

which combines the coupling constant  $g$  and the  $\theta$  angle. The Witten-Seiberg solution also determines the anomalous magnetic moment  $\mathcal{F}'''$  and the four-fermion vertex  $\mathcal{F}''''$ . In general, the structure of the prepotential is given by

$$\mathcal{F}(\phi) = \frac{i(4-N_f)}{8\pi} \phi^2 \log\left(\frac{\phi^2}{\Lambda^2}\right) - \frac{i}{\pi} \sum_{k=1}^{\infty} \mathcal{F}_k \phi^2 \left(\frac{\Lambda}{\phi}\right)^{(4-N_f)k}. \quad (374)$$

The first term is just the perturbative result with the one-loop beta functions coefficient. As noted in Sec. VIII.C.1, there are no corrections from higher loops. Instead, there is an infinite series of power corrections. The coefficient  $\mathcal{F}_k$  is proportional to  $\Lambda^{(4-N_f)k}$ , which is exactly what one would expect for a  $k$ -instanton contribution.

For  $k=1$ , this was first checked by Finnel and Pouliot (1995) in the case of  $SU(2)$  and by Ito and Sasakura (1996) in the more general case of  $SU(N_c)$ . The basic idea is to calculate the coefficient of the 't Hooft interaction  $\sim \lambda^2 \psi^2$ . The gluino  $\lambda$  and the Higgsino  $\psi$  together have eight fermion zero modes. Pairing zero modes using Yukawa couplings of the type  $(\lambda\psi)\phi$  and the nonvanishing Higgs vacuum expectation value, we can see that instantons induce a four-fermion operator. In an impressive *tour de force*, the calculation of the coefficient of this operator was recently extended to the two-instanton level<sup>63</sup> (Dorey, Khoze, and Mattis, 1996a; Aoyama *et al.*, 1996). For  $N_f=0$ , the result is

$$S_{4f} = \int dx (\psi^2 \lambda^2) \left[ \frac{15\Lambda^4}{8\pi^2 a^6} + \frac{9!\Lambda^8}{3 \times 2^{12} \pi^2 a^{10}} + \dots \right], \quad (375)$$

which agrees with the Witten-Seiberg solution. This is also true for  $N_f \neq 0$ , except in the case  $N_f=4$ , where a discrepancy appears. This is the special case in which the coefficient of the perturbative beta function vanishes. Seiberg and Witten assume that the nonperturbative  $\tau(a)$  is the same as the corresponding bare coupling in the Lagrangian. But the explicit two-instanton calculation shows that even in this theory the charge is renormalized by instantons (Dorey, Khoze, and Mattis, 1996b). In principle, these calculations can be extended order by order in the instanton density. The result provides a very nontrivial check on the instanton calculus. For example, in order to obtain the correct two-instanton contribution, one has to use the most general (ADHM) two-instanton solution, not just a linear superposition of two instantons.

Instantons also give a contribution to the expectation value of  $\phi^2$ . Pairing off the remaining zero modes, we see that the semiclassical relation  $u = a^2/2$  receives a correction (Finnel and Pouliot, 1995)

$$u = \frac{a^2}{2} + \frac{\Lambda^4}{a^2} + O\left(\frac{\Lambda^8}{a^6}\right). \quad (376)$$

More interesting are instanton corrections to the effective charge  $\tau$ . The solution of Seiberg and Witten can be written in terms of an elliptic integral of the first kind,

$$\tau(u) = i \frac{K(\sqrt{1-k^2})}{K(k)}, \quad k^2 = \frac{u - \sqrt{u^2 - 4\Lambda^4}}{u + \sqrt{u^2 - 4\Lambda^4}}. \quad (377)$$

In the semiclassical domain, this result can be written as the one-loop perturbative contribution plus an infinite series of  $k$ -instanton terms. Up to the two-instanton level, we have [for a more detailed discussion of the nonperturbative beta function, see Bonelli and Matone (1996)]

$$\frac{8\pi}{g^2} = \frac{2}{\pi} \left[ \log\left(\frac{2a^2}{\Lambda^2}\right) - \frac{3\Lambda^4}{a^4} - \frac{3 \times 5 \times 7 \Lambda^8}{8a^8} + \dots \right]. \quad (378)$$

<sup>63</sup>Supersymmetry implies that there is no instanton/anti-instanton contribution to the prepotential.

It is interesting to note that instanton corrections tend to accelerate the growth of the coupling constant  $g$ . This is consistent with what was found in QCD by considering how small-size instantons renormalize the charge of a larger instanton (Callan *et al.*, 1978a). However, the result is opposite to the trend discussed in Sec. III.C.3 (based on the instanton size distribution and lattice beta function), which suggests that in QCD the coupling runs more slowly than suggested by perturbation theory.

If the Higgs vacuum expectation value is reduced, the instanton corrections in Eq. (377) start to grow and compensate for the perturbative logarithm. At this point the expansion (378) becomes unreliable, but the exact solution of Seiberg and Witten is still applicable. In the semiclassical regime, the spectrum of the theory contains monopoles and dyons with masses proportional to  $\tau$ . As the Higgs vacuum expectation value is reduced, these particles can become massless. In this case, the expansion of the effective Lagrangian in terms of the original (electrically charged) fields breaks down, but the theory can be described in terms of their (magnetically charged) dual partners.

## IX. SUMMARY AND DISCUSSION

### A. General remarks

Finally, we should like to summarize the main points of this review, discuss some of the open problems, and provide an outlook. In general, semiclassical methods in quantum mechanics and field theory are well developed. We can reliably calculate the contribution of small-size (large-action) instantons to arbitrary Green's functions. Problems arise when we leave this regime and attempt to calculate the contribution from large instantons or close instanton/anti-instanton pairs. While these problems can be solved rigorously in some theories (as in quantum mechanics or in some SUSY field theories), in QCD-like theories we still face a number of unresolved problems and therefore have to follow a somewhat more phenomenological approach. Nevertheless, the main point of this review is that important progress has been made in this context. The phenomenological success of the instanton liquid model is impressive, and initial attempts to check the underlying assumptions explicitly on the lattice are very encouraging.

With this review, we want not only to acquaint the reader with the theory of instantons in QCD, but also to draw attention to the large number of observables, in particular hadronic correlation functions at zero and finite temperatures, that have already been calculated in the instanton liquid model. While some of these predictions have been compared with phenomenological information or lattice results, many others still await confrontation with experiment or the lattice. The instanton liquid calculations were made possible by a number of technical advances. We now have a variety of approaches at our disposal, including the single instanton, the mean-field and random-phase approximations, as

well as numerical calculations that take the 't Hooft interaction into account to all orders.

The progress made in understanding the physics of instantons in lattice calculations has been of equal importance. We now have data concerning the total density, the typical size, the size distribution, and correlations between instantons. Furthermore, there are detailed checks on the mechanism of  $U(1)_A$  violation and on the behavior of many more correlation functions under cooling. Recent investigations have begun to focus on many interesting questions, like the effects of quenching, correlations of instantons with monopoles, etc.

In the following we shall first summarize the main results concerning the structure of the QCD vacuum and its hadronic excitations, then discuss the effects of finite temperature, and finally try to place QCD in a broader context, comparing the vacuum structure of QCD with other non-Abelian field theories.

### B. Vacuum and hadronic structure

The instanton liquid model is based on the assumption that nonperturbative aspects of the gluonic vacuum, like the gluon condensate, the vacuum energy density, or the topological susceptibility, are dominated by small-size ( $\rho \approx 1/3$  fm) instantons. The density of tunneling events is  $n \approx 1$  fm<sup>-4</sup>. These numbers imply that the gauge fields are very inhomogeneous, with strong fields ( $G_{\mu\nu} \sim g^{-1} \rho^{-2}$ ) concentrated in small regions of space-time. In addition, the gluon fields are strongly polarized, the field strength locally being either self-dual or anti-self-dual.

Quark fields, on the other hand, cannot be localized inside instantons. Isolated instantons have unpaired chiral zero modes, so the instanton amplitude vanishes if quarks are massless. In order to get a nonzero probability, quarks have to be exchanged between instantons and anti-instantons. In the ground state, zero modes become completely delocalized, and chiral symmetry is broken. As a consequence, quark-antiquark pairs with the quantum numbers of the pion can travel infinitely far, and we have a Goldstone pion.

This difference in the distribution of vacuum fields leads to significant differences in gluonic and fermionic correlation functions. Gluonic correlators are much more short range, and as a result the mass scale for glueballs,  $m_{0^{++}} \approx 1.5$  GeV, is significantly larger than the typical mass of non-Goldstone mesons,  $m_\rho = 0.77$  GeV. The polarized gluon fields lead to large spin splittings for both glueballs and ordinary mesons. In general, we can group all hadronic correlation functions into three classes: (i) Those that receive direct instanton contributions that are attractive,  $\pi, K, 0^{++}$  glueball,  $N, \dots$ , (ii) those with direct instantons effects that are repulsive,  $\eta', \delta, 0^{+-}$  glueball,  $\dots$ , and (iii) correlation functions with no direct instanton contributions  $\rho, a_1, 2^{++}$  glueball,  $\Delta, \dots$ . As we have repeatedly emphasized throughout this review, already this simple classification based on

first-order instanton effects gives a nontrivial understanding of the bulk features of hadronic correlation functions.

In addition, the instanton liquid allows us to go into much greater detail. Explicit calculations of the full correlation functions in a large number of hadronic channels have been performed. These calculations require only two parameters to be fixed. One is the scale parameter  $\Lambda$  and the other characterizes the scale at which the effective repulsion between close pairs sets in. With these parameters fixed from global properties of the vacuum, we not only find a very satisfactory description of the masses and couplings of ground-state hadrons, but we also reproduce the full correlation function whenever they are available. Of course, the instanton model reaches its limits as soon as perturbative or confinement effects become dominant. This is the case, for example, when one attempts to study bound states of heavy quarks or tries to resolve high-lying radial excitations of light hadrons.

How does this picture compare with other approaches to hadronic structure? As far as the methodology is concerned, the instanton approach is close to (and to some extent a natural outgrowth of) the QCD sum-rule method. Moreover, the instanton effects explain why the OPE works in some channels and fails in others (those with direct instanton contributions). The instanton liquid provides a complete picture of the ground state, so that no assumptions about higher-order condensates are required. It also allows the calculations to be extended to large distances, so that no matching is needed.

In the quark sector, the instanton model provides a picture similar to the Nambu and Jona-Lasinio model. There is an attractive quark-quark interaction that causes quarks to condense and binds them into light mesons and baryons. However, the instanton liquid provides a more microscopic mechanism, with a more direct connection to QCD, and relates the different coupling constants and cutoffs in the Nambu and Jona-Lasinio model.

Instead of going into comparisons with the plethora of hadronic models that have been proposed over the years, let us emphasize two points that we feel are important. Hadrons are not cavities which are empty inside (devoid of nonperturbative fields) as the bag model suggests. Indeed, matrix elements of electric and magnetic fields inside the nucleon can be determined from the trace anomaly (Shuryak, 1978b; Ji, 1995), showing that the density of instantons and the magnitude of the gluon condensate inside the nucleon is only reduced by a few percent. Hadrons are excitations of a very dense medium, and this medium should be understood first. Furthermore, spin splittings in glueballs ( $2^{++}-0^{++}, \dots$ ) and light hadrons ( $\rho-\pi, \dots$ ) are not small, so it makes no sense to treat them perturbatively. This was directly checked on the lattice: spin splittings are not removed by cooling, which quickly eliminates all perturbative contributions.

What is the prospect for future work on hadronic structure? Of course, our understanding of hadronic

structure is still very far from being complete. Clearly, the most important question concerns the mechanism of confinement and its role in hadronic structure. In the meantime, experiments continue to provide interesting new puzzles: the spin of the nucleon, the magnitude and polarization of the strange sea, the isospin asymmetry in the light  $u, d$  quark sea, etc.

### C. Finite temperature and chiral restoration

Understanding the behavior of hadrons and hadronic matter at high temperature is the ultimate goal of the experimental heavy-ion program. These studies complement our knowledge of hadronic structure at zero temperature and density and provide an opportunity to observe rearrangements in the structure of the QCD vacuum directly.

Generalizing the instanton liquid model to finite temperatures is straightforward in principle. Nevertheless, the role of instantons at finite temperatures has been reevaluated during the past few years. There is evidence that instantons are not suppressed near  $T_c$ , but disappear only at significantly higher temperatures. Only after instantons disappear does the system become a perturbative plasma.<sup>64</sup>

In addition, we have argued that the chiral transition is due to the dynamics of the instanton liquid itself. The phase transition is driven by a rearrangement of the instanton liquid, going from a (predominantly) random phase at small temperature to a correlated phase of instanton/anti-instanton molecules at high temperature. Without having to introduce any additional parameters, this picture provides the correct temperature scale for the transition and agrees with standard predictions concerning the structure of the phase diagram.

If instantons are bound into topologically neutral pairs at  $T > T_c$ , they no longer generate a quark condensate. However, they still contribute to the gluon condensate, the effective interaction between quarks, and the equation of state. Therefore instanton effects are potentially very important in understanding the plasma at moderate temperatures  $T = (1-3)T_c$ . We have begun to explore some of these consequences in greater detail, in particular the behavior of spatial and temporal correlation functions across the transition region. While spacelike screening masses essentially agree with the results of lattice calculations, interesting phenomena are seen in temporal correlation functions. We find evidence that certain hadronic modes survive in the high-temperature phase. Clearly, much work remains to be

<sup>64</sup>We should mention that even at asymptotically high temperature there are nonperturbative effects in QCD, related to the physics of magnetic (three-dimensional) QCD. However, these effects are associated with the scale  $g^2 T$ , which is small compared to the typical momenta of the order  $T$ . This means that the corrections to quantities like the equation of state are small.

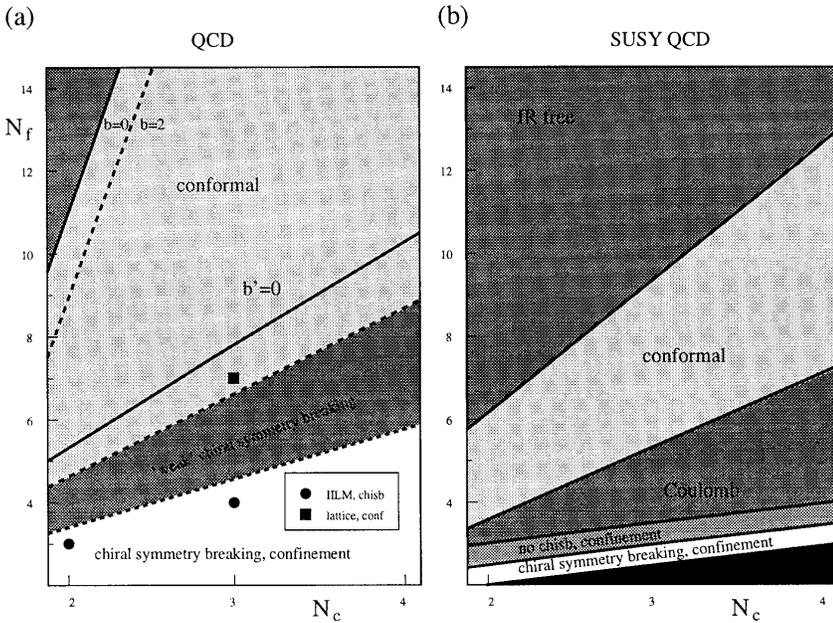


FIG. 42. Schematic phase diagram of QCD (a) and supersymmetric QCD (b) as a function of the number of colors  $N_c$  and the number of flavors  $N_f$ .

done in order to improve our understanding of the high-temperature phase.

#### D. The big picture

Finally, we should like to place QCD in a broader perspective and discuss what is known about the phase structure of non-Abelian gauge theories (both ordinary and supersymmetric) based on the gauge group  $SU(N_c)$  with  $N_f$  quark flavors. For simplicity, we shall restrict ourselves to zero temperature and massless fermions. This means that the theory has no dimensional parameters other than  $\Lambda$ . The phase diagram of ordinary and SUSY QCD in the  $N_c$ - $N_f$  plane is shown in Fig. 42. For simplicity, we have plotted  $N_c$  and  $N_f$  as if they were continuous variables.<sup>65</sup> We should emphasize that, while the location of the phase boundaries can be rigorously established in the case of SUSY QCD, the phase diagram of ordinary QCD is just a guess, guided by some of the results mentioned below.

Naturally, we are most interested in the role of instantons in these theories. As  $N_f$  is increased above the value 2 or 3 (relevant to real QCD), the two basic components of the instanton ensemble, random (individual) instantons and strongly correlated instanton/anti-instanton pairs (molecules), are affected in very different ways. Isolated instantons can only exist if the quark condensate is nonzero, and the instanton density contains the factor  $(|\langle \bar{q}q \rangle| \rho^3)^{N_f}$  which comes from the fermion determinant. As a result, small-size instantons are strongly suppressed as  $N_f$  is increased. This suppression factor does not affect instantons with a size larger than  $\rho \sim |\langle \bar{q}q \rangle|^{-1/3}$ . This means that as  $N_f$  is increased, ran-

dom instantons are pushed to larger sizes. Since in this regime the semiclassical approximation becomes unreliable, we do not know how to calculate the rate of random instantons at large  $N_f$ .

For strongly correlated pairs (molecules) the trend is exactly opposite. The density of pairs is essentially independent of the quark condensate and only determined by the interaction of the two instantons. From purely dimensional considerations one expects the density of molecules to be  $dn_m \sim d\rho \Lambda^{2b} \rho^{2b-5}$ , which means that the typical size becomes smaller as  $N_f$  is increased (Shuryak, 1987). If  $N_f > 11N_c/2$ , we have  $b < 2$ , and the density of pairs is ultraviolet divergent [see the dashed line in Fig. 42(a)]. This phenomenon is similar to the UV divergence in the  $O(3)$  nonlinear  $\sigma$  model. Both are examples of UV divergencies of a nonperturbative nature. Most likely they do not have significant effects on the physics of the theory. Since the typical instanton size is small, one would expect that the contribution of molecules at large  $N_f$  could be reliably calculated. However, since the binding inside the pair increases with  $N_f$ , the separation of perturbative and nonperturbative fluctuations becomes more and more difficult.

Rather than speculate about these effects, let us go back and consider very large  $N_f$ . The solid line labeled  $b=0$  in Fig. 42 corresponds to a vanishing first coefficient of the beta function,  $b = (11/3)N_c - (2/3)N_f = 0$  in QCD and  $b = 3N_c - N_f = 0$  in SUSY QCD. Above this line, the coupling constant decreases at large distances, and the theory is IR free. Below this line, the theory is expected to have an infrared fixed point (Belavin and Migdal, 1974; Banks and Zaks, 1982). As discussed in Sec. III.C.3, this is due to the fact that the sign of the second coefficient of the beta function  $b' = 34N_c^2/3 - 13N_cN_f/3 + N_f/N_c$  is negative while the first one is positive. As a result, the beta function has a zero at  $g_*^2/(16\pi^2) = -b/b'$ . This number determines the limiting value of the charge at large distances. Note that the

<sup>65</sup>In a sense, at least the number of flavors is a continuous variable. One can gradually remove a massless fermion by increasing its mass.

fixed point is not destroyed by higher-order perturbative effects, since we can always choose a scheme where higher-order coefficients vanish. The presence of an IR fixed point implies that the theory is conformal, which means correlation functions show a power-law decay at large distances. There is no mass gap, and the long-distance behavior is characterized by the set of critical exponents.

Where is the lower boundary of the conformal domain? A recent perturbative study based on the  $1/N_f$  expansion (Gracey, 1996) suggested that the IR fixed point might persist all the way down to  $N_f=6$  (for  $N_c=3$ ). However, the critical coupling would become larger and nonperturbative phenomena might become important. For example, Appelquist, Terning, and Wijewardhana (1996) argued that if the coupling constant reached a critical value the quark-antiquark interaction could be sufficiently strong to break chiral symmetry. In their calculation, this would happen for  $N_f^c \approx 4N_c$  [see the dashed line in Fig. 42(a)].

It was then realized that instanton effects could also be important (Appelquist and Selipsky, 1997). If the critical coupling is small, even large instantons have a large action,  $S=8\pi^2/g_*^2 \gg 1$ , and the semiclassical approximation is valid. As usual, we expect random instantons to contribute to chiral symmetry breaking. According to estimates made by Appelquist and Selipsky (1997), the role of instantons is comparable to that of perturbative effects in the vicinity of  $N_f=4N_c$ . Chiral symmetry breaking is dominated by large instantons with size  $\rho \sim |\langle \bar{q}q \rangle|^{-1/3} > \Lambda^{-1}$ , while the perturbative regime  $\rho < \Lambda^{-1}$  contributes very little. For even larger instantons,  $\rho \gg |\langle \bar{q}q \rangle|^{-1/3}$ , fermions acquire a mass due to chiral symmetry breaking and effectively decouple from gluons. This means that for large distances the charge evolves as in pure gauge theory, and the IR fixed point is only an approximate feature, useful for analyzing the theory above the decoupling scale.

Little is known about the phase structure of multiflavor QCD from lattice simulations. Lattice QCD with up to 240 flavors was studied by Iwasaki *et al.* (1996), who showed that, as expected, the theory is trivial for  $b > 0$ . The paper also confirms the existence of an infrared fixed point for  $N_f \geq 7$  ( $N_c=3$ ). In Fig. 42(a) we have marked these results by open squares. Other groups have studied QCD with  $N_f=8$  (Brown *et al.*, 1992), 12 (Kogut and Sinclair, 1988) and 16 (Damgaard *et al.*, 1997) flavors. All of these simulations find a chirally asymmetric and confining theory at strong coupling and a bulk transition to a chirally symmetric phase (at  $\beta = 2N_c/g^2 > 4.73, 4.47, \text{ and } 4.12$ , respectively).

It may appear that these results are in contradiction with the results of Appelquist *et al.* mentioned above, according to which chiral symmetry should be broken for  $N_f < 12$  ( $N_c=3$ ), but this is not the case, since the condensate is expected to be exponentially small. In other words, in order to reproduce the subtle mechanism of chiral breaking by large-distance Coulomb ex-

changes or large-size instantons, the lattice has to include the relevant scales,<sup>66</sup> which is not the case in the present lattice simulations.

The present lattice results more closely resemble the results in the interacting instanton liquid discussed in Sec. VII.B.4. There we found a line of (rather robust first-order) transitions that touches  $T=0$  near  $N_f \approx 5$ . A large drop in the quark condensate in going from  $N_f=2,3$  to  $N_f=4$  observed by the Columbia group (Chen and Mawhinney, 1997) may very well be the first indication of this phenomenon.

Again, there is no inconsistency between the interacting instanton calculation and the results of Appelquist *et al.* The interacting instanton liquid model calculation takes into account the effects of small instantons only. If small instantons do not break chiral symmetry, then long-range Coulomb forces or large instantons can still be responsible for chiral symmetry breaking.<sup>67</sup> This mechanism was studied by a number of authors (Barducci *et al.*, 1988; Aoki *et al.*, 1990), and the corresponding quark condensate is about an order of magnitude smaller<sup>68</sup> than the one observed for  $N_f=2$ . We would therefore argue that, in a practical sense, QCD has two different phases with chiral symmetry breaking, one in which the quark condensate is large and generated by small-size instantons and one in which the condensate is significantly smaller and due to Coulomb forces or large instantons. The transition regime is indicated by a wavy line in Fig. 42(a). Further studies of the mechanisms of chiral symmetry breaking for different  $N_f$  are needed before final conclusions can be drawn.

Finally, although experiment tells us that confinement and chiral symmetry breaking go together for  $N_f=2,3$ , the two are independent phenomena, and it is conceivable that there are regions in the phase diagram where only one of them takes place. It is commonly believed that confinement implies chiral symmetry breaking, but not even that is entirely clear. In fact, SUSY QCD with  $N_f=N_c+1$  provides a counterexample.<sup>69</sup>

For comparison we also show the phase diagram of ( $N=1$ ) supersymmetric QCD [Fig. 42(b)]. As discussed in Sec. VIII.C, the phase structure of these theories was recently clarified by Seiberg and collaborators. In this

<sup>66</sup>This cannot be done by simply tuning the bare coupling to the critical value for chiral symmetry breaking  $\beta \approx 1$  because lattice artifacts create a chirally asymmetric and confining phase already at  $\beta \approx 4-5$ . Therefore one has to start at weak coupling and then go to sufficiently large physical volumes to reach the chiral symmetry-breaking scale.

<sup>67</sup>The situation is different at large temperatures because in that case both Coulomb forces and large instantons are Debye screened.

<sup>68</sup>This can be seen from the fact that these authors need an unrealistically large value of  $\Lambda_{\text{QCD}} \approx 500$  MeV to reproduce the experimental value of  $f_\pi$ .

<sup>69</sup>It is usually argued that anomaly matching shows that confinement implies chiral symmetry breaking for  $N_c > 2$ , but again SUSY QCD provides examples in which anomaly-matching conditions work out in subtle and unexpected ways.

case, the lower boundary of the conformal domain is at  $N_f = (3/2)N_c$ . Below this line the dual theory based on the gauge group  $SU(N_f - N_c)$  loses asymptotic freedom. In this case, the excitations are IR free “dual quarks” or composite light baryons in terms of the original theory. Remarkably, the ’t Hooft matching conditions between the original (short-distance) theory and the dual theory based on a completely different gauge group work out exactly. The theory becomes confining for  $N_f = N_c + 1$  and  $N_f = N_c$ . In the first case, chiral symmetry is preserved, and the low-energy excitations are massless baryons. In the second case, instantons modify the geometry of the space of possible vacua, the point where the excitations are massless baryons is not allowed, and chiral symmetry has to be broken. Note that, in ordinary QCD, the ’t Hooft matching conditions cannot be satisfied for a confining phase without chiral symmetry breaking (for  $N_c \neq 2$ ). For an even smaller number of flavors,  $0 < N_f \leq N_c - 1$ , massless SUSY QCD does not have a stable ground state. The reason is that instanton/anti-instanton molecules generate a positive vacuum energy density, which decreases with the Higgs expectation value, so that the ground state is pushed to infinitely large Higgs vacuum expectation value.

We have not shown the phase structure of SUSY QCD with  $N > 1$  gluinos. In some cases (e.g.,  $N = 4$ ,  $N_f = 0$  or  $N = 2$ ,  $N_f = 4$ ), the beta function vanishes and the theory is conformal, although instantons may still cause a finite charge renormalization. As already mentioned, the low-energy spectrum of the  $N = 2$  theory was recently determined by Seiberg and Witten. The theory does not have chiral symmetry breaking or confinement, but it contains monopoles/dyons which become massless as the Higgs vacuum expectation value is decreased. This can be used to trigger confinement when the theory is perturbed to  $N = 1$ .

To summarize, there are many open questions concerning the phase structure of QCD-like theories and many issues to be explored in future studies, especially on the lattice. The location of the lower boundary of the conformal domain and the structure of the chirally asymmetric phase in the domain  $N_f = 4 - 12$  should certainly be studied in more detail. Fascinating results have clarified the rich (and sometimes rather exotic) phase structure of SUSY QCD. To what extent these results will help our understanding of nonsupersymmetric theories remains to be seen. In any event, it is certainly clear that instantons and anomalies play a very important, if not dominant, role in both cases.

*Note added in proof.*

Since this manuscript was prepared, the subject of instantons in QCD has continued to see many interesting developments. We would like to briefly mention some of these, related to instanton searches on the lattice, the relation of instantons with confinement, instantons and charm quarks, instantons at finite chemical potential, and instantons in supersymmetric theories.

Significant progress was made studying topology on the lattice using improved operators, renormalization group techniques, and fermionic methods. Also, first re-

sults of studying instantons in the vicinity of the phase transition in full QCD were reported. As an example for the use of improved actions we mention results of Colorado group (DeGrand, Hasenfratz, and Kovacs, 1997) in pure gauge  $SU(2)$ . They find that chiral symmetry breaking and confinement are preserved by inverse blocking (unlike “cooling”) preserves even close instanton–antiinstanton pairs. The instanton size distribution is peaked at  $\rho \approx 0.2$  fm, and large instantons are suppressed.

Low-lying eigenstates of the Dirac operator were studied by the MIT group (Ivanenko and Negele, 1997). They find that the corresponding wave functions are spatially correlated with the locations of instantons, providing support for the picture of the quark condensate as a collective state built from instanton zero modes. In addition to that, they studied the importance of low-lying states in hadronic correlation functions. They demonstrate that the lowest  $\sim 100$  modes (out of  $\sim 10^6$ ) are sufficient to quantitatively reproduce the hadronic ground state contribution to the  $\rho$  and  $\pi$  meson correlation functions.

In addition to that, first attempts were made to study instantons in the vicinity of the finite temperature phase transition in full QCD (de Forcrand, Perez, Hetrick, and Stamatescu, 1998). Using the cooling technique to identify instantons, they verified the  $T$  dependence of the instanton density discussion in Sec. VII.A. They observe polarized instanton–antiinstanton pairs above  $T_c$ , but these objects do not seem to dominate the ensemble. This point definitely deserves further study, using improved methods and smaller quark masses.

In the main text we stated that there is no confinement in the instanton model. Recently, it was claimed that instantons generate a linear potential with a slope close to the experimental value 1 GeV/fm (Fukushima, Sukanuma, Tanaka, Toki, and Sasaki, 1997). This prompted Chen, Negele, and Shuryak (1998) to reinvestigate the issue, and perform high statistics numerical calculations of the heavy-quark potential in the instanton liquid at distances up to 3 fm. The main conclusions are (i) the potential is larger and significantly longer range than the dilute gas result Eq. (208); (ii) a random ensemble with a realistic size distribution leads to a potential that is linear even for large  $R \approx 3$  fm; (iii) the slope of the potential is still too small,  $K \approx 200$  MeV/fm. This means that the bulk of the confining forces still has to come from some (as of yet?) unidentified objects with small action. These objects may turn out to be large instantons, but that would still imply that the main contribution is not semiclassical. Nevertheless, the result that the heavy quark potential is larger than expected is good news for the instanton model. It implies that even weakly bound states and resonances can be addressed within in the model.

We also did not discuss the role of charm quarks in the QCD vacuum. However, the color field inside a small-size instantons  $G_{\mu\nu} \approx 1$  GeV<sup>2</sup> is comparable to the charm quark mass squared  $m_c^2 \approx 2$  GeV<sup>2</sup>, so one might expect observable effects due to the polarization of

charm quark inside ordinary hadrons. Recently, it was suggested that CLEO observations of an unexpectedly large branching ratio  $B \rightarrow \eta' K$  (as well as inclusive  $B \rightarrow \eta' + \dots$ ) provide a smoking gun for such effects (Halperin and Zhitnitsky, 1998). The basic idea is that these decays proceed via Cabibbo-unsuppressed  $b \rightarrow \bar{c}cs$  transition, followed by  $\bar{c}c \rightarrow \eta'$ . The charm content of the  $\eta'$  in the instanton liquid was estimated by Shuryak and Zhitnitsky (1997), and the result is consistent with what is needed to understand the CLEO data. Another interesting possibility raised in Halperin and Zhitnitsky (1997) is that polarized charm quarks give a substantial contribution to the spin  $\Sigma$  of the nucleon. A recent instanton calculation only finds a contribution in the range  $\Delta c/\Sigma = -(0.08-0.20)$  (Blotz and Shuryak, 1997), but the value of  $\Delta c$  remains an interesting question for future deep inelastic scattering experiments (e.g., Compass at CERN). In the context of (polarized) nucleon structure functions we should also mention interesting work on leading and nonleading twist operators (see Balla, Polyakov, and Weiss, 1997, and references therein).

Finally, Bjorken (1997) discussed the possibility that instantons contribute to the decay of mesons containing real  $\bar{c}c$  pairs (Bjorken, 1997). A particularly interesting case is the  $\eta_c$ , which has three unusual 3-meson decay channels ( $\eta' \pi \pi$ ,  $\eta \pi \pi$ , and  $KK\pi$ ), which contribute roughly 5% each to the total width. This fits well with the typical instanton vertex  $\bar{u}u\bar{d}d\bar{s}s$ . In general, all of these observables offer the chance to detect nonperturbative effects deep inside the semi-classical domain.

Initial efforts were made to understand the instanton liquid at finite chemical potential (Schäfer, 1997). The suggestion made in this work is that the role that molecules play in the high temperature phase is now played by more complicated “polymers” that are aligned in the time direction. More importantly, it was suggested that instanton lead to the formation of diquark condensates in high density matter (Alford, Rajagopal, and Wilczek, 1997; Rapp, Schäfer, Shuryak, and Velkovsky, 1997). In the high density phase chiral symmetry is restored, but SU(3) color is broken by a Higgs mechanism.

Instanton effects in SUSY gauge theories continue to be a very active field. For  $N = 2$  (Seiberg-Witten) theories the  $n$ -instanton contribution was calculated explicitly (Dorey, Khoze, and Mattis, 1997; Dorey, Hollowood, Khoze, and Mattis, 1997), and as a by-product these authors also determine the (classical)  $n$ -instanton measure in the cases  $N = 1$  and  $N = 0$  (non-SUSY). Yung also determined the one-instanton contribution to higher derivative operators beyond the SW effective Lagrangian (Yung, 1997). Another very interesting result is the generalization of the Seiberg-Witten solution to an arbitrary number of colors  $N_c$  (Douglas and Shenker, 1995). This result sheds some light on the puzzling problem of instantons in the large  $N_c$  limit. The large  $N_c$  limit is usually performed with  $g^2 N_c$  held fixed, and in that case instantons amplitudes are suppressed by  $\exp(-N_c)$ . However Douglas and Shenker (1995) found

that (at least in the case  $N=2$ ) this is not the correct way to take the large  $N_c$  limit (if we want to keep the physics unchanged).

A comparison of the running of the effective charge in  $N=2, 1$  SUSY QCD and QCD was performed by Randall, Rattazzi, and Shuryak (1998). In  $N=2$  SUSY QCD, the Seiberg-Witten solution shows that instantons accelerate the growth of the coupling. As a result, the coupling blows up at a scale  $\Lambda_{NP} > \Lambda_{QCD}$  where the perturbative coupling is still small. A similar phenomenon takes place in QCD if the instanton correction to the running coupling is estimated from the formula of Callan, Dashen, and Gross. This might help to explain why in QCD the nonperturbative scale  $\Lambda_{NP} \approx \Lambda_{\chi SB} \approx 1$  GeV  $> \Lambda_{QCD} \approx 200$  MeV.

## ACKNOWLEDGMENTS

It is a pleasure to thank our friends and collaborators who have contributed to this review or to the work presented here. In particular, we should like to thank our collaborators J. Verbaarschot, M. Velkovsky, and A. Blotz. We should like to thank M. Shifman for many valuable comments on the original manuscript. We have also benefitted from discussions with G. E. Brown, D. I. Diakonov, J. Negele, M. Novak, A. V. Smilga, A. Vainshtein, and I. Zahed.

## APPENDIX: BASIC INSTANTON FORMULAS

### 1. Instanton gauge potential

We use the following conventions for Euclidean gauge fields: The gauge potential is  $A_\mu = A_\mu^a(\lambda^a/2)$ , where the SU( $N$ ) generators satisfy  $[\lambda^a, \lambda^b] = 2if^{abc}\lambda^c$  and are normalized according to  $\text{tr}[\lambda^a\lambda^b] = 2\delta^{ab}$ . The covariant derivative is given by  $D_\mu = \partial_\mu - iA_\mu$  and the field-strength tensor is

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]. \quad (\text{A1})$$

In our conventions, the coupling constant is absorbed into the gauge fields. Standard perturbative notation corresponds to the replacement  $A_\mu \rightarrow gA_\mu$ . The single-instanton solution in regular gauge is given by

$$A_\mu^a = \frac{2\eta_{a\mu\nu}x_\nu}{x^2 + \rho^2}, \quad (\text{A2})$$

and the corresponding field strength is

$$G_{\mu\nu}^a = -\frac{4\eta_{a\mu\nu}\rho^2}{(x^2 + \rho^2)^2}, \quad (\text{A3})$$

$$(G_{\mu\nu}^a)^2 = \frac{192\rho^4}{(x^2 + \rho^2)^4}. \quad (\text{A4})$$

The gauge potential and field strength in singular gauge are

$$A_\mu^a = \frac{2\bar{\eta}_{a\mu\nu}x_\nu\rho^2}{x^2(x^2+\rho^2)}, \quad (\text{A5})$$

$$G_{\mu\nu}^a = -\frac{4\rho^2}{(x^2+\rho^2)^2} \times \left( \bar{\eta}_{a\mu\nu} - 2\bar{\eta}_{a\mu\alpha} \frac{x_\alpha x_\nu}{x^2} - 2\bar{\eta}_{a\alpha\nu} \frac{x_\mu x_\alpha}{x^2} \right). \quad (\text{A6})$$

Finally, an  $n$ -instanton solution in singular gauge is given by

$$A_\mu^a = \bar{\eta}_{a\mu\nu} \partial_\nu \ln \Pi(x), \quad (\text{A7})$$

$$\Pi(x) = 1 + \sum_{i=1}^n \frac{\rho_i^2}{(x-z_i)^2}. \quad (\text{A8})$$

Note that all instantons have the same color orientation. For a construction that gives the most general  $n$ -instanton solution, see Atiyah *et al.* (1977).

## 2. Fermion zero modes and overlap integrals

In singular gauge, the zero-mode wave function  $i\mathcal{D}\phi_0=0$  is given by

$$\phi_{a\nu} = \frac{1}{2\sqrt{2}\pi\rho} \sqrt{\Pi} \left[ \theta \left( \frac{\Phi}{\Pi} \right) \right]_{\nu\mu} U_{ab} \epsilon_{\mu b}, \quad (\text{A9})$$

where  $\Phi = \Pi - 1$ . For the single-instanton solution, we get

$$\phi_{a\nu}(x) = \frac{\rho}{\pi} \frac{1}{(x^2+\rho^2)^{3/2}} \left[ \left( \frac{1-\gamma_5}{2} \right) \frac{\not{x}}{\sqrt{x^2}} \right]_{\nu\mu} U_{ab} \epsilon_{\mu b}. \quad (\text{A10})$$

The instanton-instanton zero-mode density matrices are

$$\begin{aligned} \phi_I(x)_{i\alpha} \phi_J^\dagger(y)_{j\beta} &= \frac{1}{8} \varphi_I(x) \varphi_J(y) \left( \not{x} \gamma_\mu \gamma_\nu \not{y} \frac{1-\gamma_5}{2} \right)_{ij} \\ &\quad \otimes (U_I \tau_\mu^- \tau_\nu^+ U_J^\dagger)_{\alpha\beta}, \end{aligned} \quad (\text{A11})$$

$$\begin{aligned} \phi_I(x)_{i\alpha} \phi_A^\dagger(y)_{j\beta} &= -\frac{i}{2} \varphi_I(x) \varphi_A(y) \left( \not{x} \gamma_\mu \not{y} \frac{1+\gamma_5}{2} \right)_{ij} \\ &\quad \otimes (U_I \tau_\mu^- U_A^\dagger)_{\alpha\beta}, \end{aligned} \quad (\text{A12})$$

$$\begin{aligned} \phi_A(x)_{i\alpha} \phi_I^\dagger(y)_{j\beta} &= \frac{i}{2} \varphi_A(x) \varphi_I(y) \left( \not{x} \gamma_\mu \not{y} \frac{1-\gamma_5}{2} \right)_{ij} \\ &\quad \otimes (U_A \tau_\mu^+ U_I^\dagger)_{\alpha\beta}, \end{aligned} \quad (\text{A13})$$

with

$$\varphi(x) = \frac{\rho}{\pi} \frac{1}{\sqrt{x^2(x^2+\rho^2)^{3/2}}}. \quad (\text{A14})$$

The overlap matrix element is given by

$$\begin{aligned} T_{AI} &= \int d^4x \phi_A^\dagger(x-z_A) i\mathcal{D} \phi_I(x-z_I) \\ &= r_\mu \text{Tr}(U_I \tau_\mu^- U_A^\dagger) \frac{1}{2\pi^2 r} \frac{d}{dr} M(r), \end{aligned} \quad (\text{A15})$$

with

$$M(r) = \frac{1}{r} \int_0^\infty dp p^2 |\varphi(p)|^2 J_1(pr). \quad (\text{A16})$$

The Fourier transform of the zero-mode profile is given by

$$\varphi(p) = \pi\rho^2 \frac{d}{dx} (I_0(x)K_0(x) - I_1(x)K_1(x)) \Big|_{x=p\rho/2}. \quad (\text{A17})$$

## 3. Properties of $\eta$ symbols

We define four-vector matrices

$$\tau_\mu^\pm = (\vec{\tau}, \mp i), \quad (\text{A18})$$

where  $\tau^a \tau^b = \delta^{ab} + i\epsilon^{abc} \tau^c$  and

$$\tau_\mu^+ \tau_\nu^- = \delta_{\mu\nu} + i\eta_{a\mu\nu} \tau^a, \quad (\text{A19})$$

$$\tau_\mu^- \tau_\nu^+ = \delta_{\mu\nu} + i\bar{\eta}_{a\mu\nu} \tau^a, \quad (\text{A20})$$

with the  $\eta$  symbols given by

$$\eta_{a\mu\nu} = \epsilon_{a\mu\nu} + \delta_{a\mu} \delta_{\nu 4} - \delta_{a\nu} \delta_{\mu 4}, \quad (\text{A21})$$

$$\bar{\eta}_{a\mu\nu} = \epsilon_{a\mu\nu} - \delta_{a\mu} \delta_{\nu 4} + \delta_{a\nu} \delta_{\mu 4}. \quad (\text{A22})$$

The  $\eta$  symbols are (anti) self-dual in the vector indices

$$\begin{aligned} \eta_{a\mu\nu} &= \frac{1}{2} \epsilon_{\mu\nu\alpha\beta} \eta_{a\alpha\beta}, & \bar{\eta}_{a\mu\nu} &= -\frac{1}{2} \epsilon_{\mu\nu\alpha\beta} \bar{\eta}_{a\alpha\beta} \\ \eta_{a\mu\nu} &= -\eta_{a\nu\mu}. \end{aligned} \quad (\text{A23})$$

We have the following useful relations for contractions involving  $\eta$  symbols

$$\eta_{a\mu\nu} \eta_{b\mu\nu} = 4\delta_{ab}, \quad (\text{A24})$$

$$\eta_{a\mu\nu} \eta_{a\mu\rho} = 3\delta_{\nu\rho}, \quad (\text{A25})$$

$$\eta_{a\mu\nu} \eta_{a\mu\nu} = 12, \quad (\text{A26})$$

$$\eta_{a\mu\nu} \eta_{a\rho\lambda} = \delta_{\mu\rho} \delta_{\nu\lambda} - \delta_{\mu\lambda} \delta_{\nu\rho} + \epsilon_{\mu\nu\rho\lambda}, \quad (\text{A27})$$

$$\eta_{a\mu\nu} \eta_{b\mu\rho} = \delta_{ab} \delta_{\nu\rho} + \epsilon_{abc} \eta_{c\nu\rho}, \quad (\text{A28})$$

$$\eta_{a\mu\nu} \bar{\eta}_{b\mu\nu} = 0. \quad (\text{A29})$$

The same relations hold for  $\bar{\eta}_{a\mu\nu}$ , except for

$$\bar{\eta}_{a\mu\nu} \bar{\eta}_{a\rho\lambda} = \delta_{\mu\rho} \delta_{\nu\lambda} - \delta_{\mu\lambda} \delta_{\nu\rho} - \epsilon_{\mu\nu\rho\lambda}. \quad (\text{A30})$$

Some additional relations are

$$\begin{aligned} \epsilon_{abc} \eta_{b\mu\nu} \eta_{c\rho\lambda} &= \delta_{\mu\rho} \eta_{a\nu\lambda} - \delta_{\mu\lambda} \eta_{a\nu\rho} \\ &\quad + \delta_{\nu\lambda} \eta_{a\mu\rho} - \delta_{\nu\rho} \eta_{a\mu\lambda}, \end{aligned} \quad (\text{A31})$$

$$\epsilon_{\lambda\mu\nu\sigma} \eta_{a\rho\sigma} = \delta_{\rho\lambda} \eta_{a\mu\nu} + \delta_{\rho\nu} \eta_{a\lambda\mu} + \delta_{\rho\mu} \eta_{a\nu\lambda}. \quad (\text{A32})$$

#### 4. Group integration

In order to perform averages over the color group, we need the following integrands over the invariant  $SU(N_c)$  measure:

$$\int dU U_{ij} U_{kl}^\dagger = \frac{1}{N_c} \delta_{jk} \delta_{li}, \quad (\text{A33})$$

$$\begin{aligned} \int dU U_{ij} U_{kl}^\dagger U_{mn} U_{op}^\dagger \\ = \frac{1}{N_c^2} \delta_{jk} \delta_{li} \delta_{no} \delta_{mp} + \frac{1}{4(N_c^2 - 1)} (\lambda^a)_{jk} (\lambda^a)_{li} \\ \times (\lambda^b)_{no} (\lambda^b)_{mp}. \end{aligned} \quad (\text{A34})$$

Additional results can be found in Creutz (1983). These results can be rearranged using the  $SU(N)$  Fierz transformation,

$$(\lambda^a)_{ij} (\lambda^a)_{kl} = -\frac{2}{N_c} \delta_{ij} \delta_{kl} + 2 \delta_{jk} \delta_{il}. \quad (\text{A35})$$

#### REFERENCES

- Abbott, L. F., and E. Farhi, 1981, *Phys. Lett. B* **101**, 69.  
 Abrikosov, A. A., 1983, *Sov. J. Nucl. Phys.* **37**, 459.  
 Actor, A., 1979, *Rev. Mod. Phys.* **51**, 461.  
 Adami, C., T. Hatsuda, and I. Zahed, 1991, *Phys. Rev. D* **43**, 921.  
 Adler, S. L., 1969, *Phys. Rev.* **177**, 2426.  
 Affleck, I., 1981, *Nucl. Phys. B* **191**, 429.  
 Affleck, I., M. Dine, and N. Seiberg, 1984a, *Phys. Lett. B* **140**, 59.  
 Affleck, I., M. Dine, and N. Seiberg, 1984b, *Nucl. Phys. B* **241**, 493.  
 Affleck, I., M. Dine, and N. Seiberg, 1985, *Nucl. Phys. B* **256**, 557.  
 Aleinikov, A., and E. Shuryak, 1987, *Yad. Fiz.* **46**, 122.  
 Alfaro, V. D., S. Fubini, and G. Furlan, 1976, *Phys. Lett. B* **65**, 163.  
 Alford, M., K. Rajagopal, and F. Wilczek, 1997, *hep-ph/9711395*.  
 Alles, B., A. DiGiacomo, and M. Gianetti, 1990, *Phys. Lett. B* **249**, 490.  
 Alles, B., M. Camprostrini, and A. DiGiacomo, 1993, *Phys. Rev. D* **48**, 2284.  
 Alles, B., M. D'Elia, and A. DiGiacomo, 1996, *Nucl. Phys. B* **494**, 281.  
 Amati, D., K. Konishi, Y. Meurice, G. C. Rossi, and G. Veneziano, 1988, *Phys. Rep.* **162**, 169.  
 Ambjorn, J., and P. Olesen, 1977, *Nucl. Phys. B* **170**, 60.  
 Andrei, N., and D. J. Gross, 1978, *Phys. Rev. D* **18**, 468.  
 Aoki, K. I., M. Bando, T. Kugo, M. G. Mitchard, and H. Nakatani, 1990, *Prog. Theor. Phys.* **84**, 683.  
 Aoyama, H., T. Harano, H. Kikuchi, I. Okouchi, M. Sato, and S. Wada, 1997, *Prog. Theor. Phys. Suppl.* **127**, 1.  
 Aoyama, H., T. Harano, M. Sato, and S. Wada, 1996, *Phys. Lett. B* **388**, 331.  
 Appelquist, T., and R. D. Pisarski, 1981, *Phys. Rev. D* **23**, 2305.  
 Appelquist, T., and S. B. Selipsky, 1997, *Phys. Lett. B* **400**, 364.  
 Appelquist, T., J. Terning, and L. C. R. Wijewardhana, 1996, *Phys. Rev. Lett.* **77**, 1214.  
 Arnold, P. B., and M. P. Mattis, 1991, *Mod. Phys. Lett. A* **6**, 2059.  
 Atiyah, M. F., N. J. Hitchin, V. G. Drinfeld, and Y. I. Manin, 1977, *Phys. Lett. A* **65**, 185.  
 Atiyah, M. F., and N. S. Manton, 1989, *Phys. Lett. B* **222**, 438.  
 Bagan, E., and S. Steele, 1990, *Phys. Lett. B* **243**, 413.  
 Bali, G. S., K. Schilling, A. Hulsebos, A. C. Irving, C. Michael, and P. W. Stephenson, 1993, *Phys. Lett. B* **309**, 378.  
 Balitskii, I. I., and V. M. Braun, 1993, *Phys. Lett. B* **314**, 237.  
 Balitskii, I., and V. Braun, 1995, *Phys. Lett. B* **346**, 143.  
 Balitsky, I. I., and A. Schäfer, 1993, *Nucl. Phys. B* **404**, 639.  
 Balitsky, I. I., and A. V. Yung, 1986, *Phys. Lett. B* **168**, 113.  
 Balla, J., M. V. Polyakov, and C. Weiss, 1997, *hep-ph/9707515*.  
 Baluni, V., 1979, *Phys. Rev. D* **19**, 2227.  
 Banks, T., and A. Casher, 1980, *Nucl. Phys. B* **169**, 103.  
 Banks, T., and A. Zaks, 1982, *Nucl. Phys. B* **82**, 196.  
 Barducci, A., R. Casalbuoni, S. D. Curtis, D. Dominici, and R. Gatto, 1988, *Phys. Rev. D* **38**, 238.  
 Belavin, A. A., and A. A. Migdal, 1974, preprint (Landau Inst.) 74-0894.  
 Belavin, A. A., A. M. Polyakov, A. A. Schwartz, and Y. S. Tyupkin, 1975, *Phys. Lett. B* **59**, 85.  
 Bell, J. S., and R. Jackiw, 1969, *Nuovo Cimento A* **60**, 47.  
 Belyaev, V. M., and B. L. Ioffe, 1982, *Sov. Phys. JETP* **83**, 976.  
 Beresinsky, V., 1971, *Zh. Eksp. Teor. Fiz.* **61**, 1144.  
 Bernard, C., T. Blum, C. DeTar, S. Gottlieb, U. M. Heller, J. E. Hetrick, K. Rummukainen, R. Sugar, D. Toussaint, and M. Wingate, 1997, *Phys. Rev. Lett.* **78**, 598.  
 Bernard, C. W., N. H. Christ, A. H. Guth, and E. J. Weinberg, 1977, *Phys. Rev. D* **16**, 2967.  
 Bernard, C., M. C. Ogilvie, T. A. DeGrand, C. DeTar, S. V. Gottlieb, A. Krasnitz, R. L. Sugar, and D. Toussaint, 1992, *Phys. Rev. Lett.* **68**, 2125.  
 Bernard, V., and U. G. Meissner, 1988, *Phys. Rev. D* **38**, 1551.  
 Bjorken, J. D., 1997, *hep-ph/9706524*.  
 Blask, W., U. Bohn, M. Huber, B. Metsch, and H. Petry, 1990, *Z. Phys. A* **337**, 327.  
 Blotz, A., and E. Shuryak, 1997a, *Phys. Rev. D* **55**, 4055.  
 Blotz, A., and E. Shuryak, 1997b, *hep-ph/9710544*.  
 Blum, T., L. Kärkkäinen, D. Toussaint, and S. Gottlieb, 1995, *Phys. Rev. D* **51**, 5133.  
 Bochkarev, A. I., and M. E. Shaposhnikov, 1986, *Nucl. Phys. B* **268**, 220.  
 Bogomolny, E. B., 1980, *Phys. Lett. B* **91**, 431.  
 Bonelli, G., and M. Matone, 1996, *Phys. Rev. Lett.* **76**, 4107.  
 Borgs, C., 1985, *Nucl. Phys. B* **261**, 455.  
 Boyd, G., 1996, preprint, *hep-lat/9607046*.  
 Boyd, G., S. Gupta, F. Karsch, and E. Laerman, 1994, *Z. Phys. C* **64**, 331.  
 Brezin, E., G. Parisi, and J. Zinn-Justin, 1977, *Phys. Rev. D* **16**, 408.  
 Brower, R. C., K. N. Orginos, and C.-I. Tan, 1996, *Phys. Rev. D* **55**, 6313.  
 Brown, F. R., F. P. Butler, H. Chen, N. H. Christ, Z.-H. Dong, W. Schaffer, L. I. Unger, and A. Vaccarino, 1990, *Phys. Rev. Lett.* **65**, 2491.  
 Brown, F. R., H. Chen, N. H. Christ, Z. Dong, R. D. Mawhinney, W. Schaffer, and A. Vaccarino, 1992, *Phys. Rev. D* **46**, 5655.

- Brown, G. E., and M. Rho, 1978, *Comments Nucl. Part. Phys.* **18**, 1.
- Brown, L. S., R. D. Carlitz, D. B. Creamer, and C. Lee, 1978, *Phys. Rev. D* **17**, 1583.
- Brown, L. S., and D. B. Creamer, 1978, *Phys. Rev. D* **18**, 3695.
- Caldi, D., 1977, *Phys. Rev. Lett.* **39**, 121.
- Callan, C. G., R. Dashen, and D. J. Gross, 1976, *Phys. Lett. B* **63**, 334.
- Callan, C. G., R. Dashen, and D. J. Gross, 1978a, *Phys. Rev. D* **17**, 2717.
- Callan, C. G., R. F. Dashen, and D. J. Gross, 1978b, *Phys. Rev. D* **18**, 4684.
- Callan, C. G., R. F. Dashen, and D. J. Gross, 1979, *Phys. Rev. D* **19**, 1826.
- Campostrini, M., A. DiGiacomo, and Y. Gunduc, 1989, *Phys. Lett. B* **225**, 393.
- Campostrini, M., A. DiGiacomo, and H. Panagopoulos, 1988, *Phys. Lett. B* **212**, 206.
- Caracciolo, S. R., A. P. Edwards, and A. D. Sokal, 1995, *Phys. Rev. Lett.* **75**, 1891.
- Carlitz, R. D., and D. B. Creamer, 1979a, *Ann. Phys. (N.Y.)* **118**, 429.
- Carlitz, R. D., and D. B. Creamer, 1979b, *Phys. Lett. B* **84**, 215.
- Carvalho, C. A. D., 1980, *Nucl. Phys. B* **183**, 182.
- Chandrasekharan, S., 1995, *Nucl. Phys. B (Proc. Suppl.)* **42**, 475.
- Chemtob, M., 1981, *Nucl. Phys. B* **184**, 497.
- Chen, D., and R. D. Mawhinney, 1997, *Nucl. Phys. B (Proc. Suppl.)* **53**, 216.
- Chen, D., J. Negele, and E. Shuryak, 1998, "Instanton-induced static potential in QCD, revisited," MIT preprint.
- Chernodub, M. N., and F. V. Gubarev, 1995, *JETP Lett.* **62**, 100.
- Chernyshev, S., M. A. Nowak, and I. Zahed, 1996, *Phys. Rev. D* **53**, 5176.
- Christ, N. 1996, talk given at RHIC Summer Study 96, Brookhaven National Laboratory.
- Christos, G. A., 1984, *Phys. Rep.* **116**, 251.
- Christov, C. V., A. Blotz, H. Kim, P. Pobylytsa, T. Watabe, T. Meissner, E. Ruiz-Ariola, and K. Goeke, 1996, *Prog. Part. Nucl. Phys.* **37**, 91.
- Chu, M. C., J. M. Grandy, S. Huang, and J. W. Negele, 1993a, *Phys. Rev. Lett.* **70**, 225.
- Chu, M. C., J. M. Grandy, S. Huang, and J. W. Negele, 1993b, *Phys. Rev. D* **48**, 3340.
- Chu, M. C., J. M. Grandy, S. Huang, and J. W. Negele, 1994, *Phys. Rev. D* **49**, 6039.
- Chu, M.-C., M. Lissia, and J. W. Negele, 1991, *Nucl. Phys. B* **360**, 31.
- Chu, M. C., and S. Schramm, 1995, *Phys. Rev. D* **51**, 4580.
- Claudson, M., E. Farhi, and R. L. Jaffe, 1986, *Phys. Rev. D* **34**, 873.
- Cohen, E., and C. Gomez, 1984, *Phys. Rev. Lett.* **52**, 237.
- Coleman, S., 1977, "The uses of instantons," Proceedings of the 1977 School of Subnuclear Physics, Erice (Italy), reproduced in *Aspects of Symmetry* (Cambridge University Press, Cambridge, England, 1985), p. 265.
- Cooper, F., A. Khare, and U. Sukhatme, 1995, *Phys. Rep.* **251**, 267.
- Corrigan, E., P. Goddard, and S. Templeton, 1979, *Nucl. Phys. B* **151**, 93.
- Creutz, M., 1983, *Quarks, gluons and lattices* (Cambridge University Press, Cambridge).
- Crewther, R. J., P. D. Vecchia, G. Veneziano, and E. Witten, 1979, *Phys. Lett. B* **88**, 123.
- Damgaard, P. H., U. M. Heller, A. Krasnitz, and P. Olesen, 1997, *Phys. Lett. B* **400**, 169.
- de Forcrand, F., and K.-F. Liu, 1992, *Phys. Rev. Lett.* **69**, 245.
- de Forcrand, P., M. G. Perez, J. E. Hetrick, and I.-O. Stamatescu, 1997, hep-lat/9802017.
- de Forcrand, P., M. G. Perez, and I.-O. Stamatescu, 1995, *Nucl. Phys. Proc. Suppl.* **47**, 777.
- de Forcrand, P., M. G. Perez, and I.-O. Stamatescu, 1997, *Nucl. Phys. B* **499**, 409.
- DeGrand, T., A. Hasenfratz, and T. G. Kovacs, 1997, hep-lat/9710078.
- DeGrand, T., R. Jaffe, K. Johnson, and J. Kiskis, 1975, *Phys. Rev. D* **12**, 2060.
- DelDebbio, L., M. Faber, J. Greensite, and S. Olejnik, 1997, *Nucl. Phys. Proc. Suppl.* **53**, 141.
- Deng, Y., 1989, *Nucl. Phys. B (Proc. Suppl.)* **9**, 334.
- Dey, M., V. L. Eletsky, and B. L. Ioffe, 1990, *Phys. Lett. B* **252**, 620.
- Diakonov, D. I., 1995, in *International School of Physics, Enrico Fermi, Course 80, Varenna, Italy*.
- Diakonov, D., 1996, *Prog. Part. Nucl. Phys.* **36**, 1.
- Diakonov, D. I., and A. D. Mirlin, 1988, *Phys. Lett. B* **203**, 299.
- Diakonov, D. I., and V. Y. Petrov, 1984, *Nucl. Phys. B* **245**, 259.
- Diakonov, D. I., and V. Y. Petrov, 1985, *Sov. Phys. JETP* **62**, 204.
- Diakonov, D. I., and V. Y. Petrov, 1986, *Nucl. Phys. B* **272**, 457.
- Diakonov, D. I., and V. Y. Petrov, 1994, *Phys. Rev. D* **50**, 266.
- Diakonov, D. I., and V. Y. Petrov, 1996, in *Continuous Advances in QCD 1996* (World Scientific, Singapore).
- Diakonov, D. I., V. Y. Petrov, and P. V. Pobylytsa, 1988, *Nucl. Phys. B* **306**, 809.
- Diakonov, D., V. Petrov, and P. Pobylytsa, 1989, *Phys. Lett. B* **226**, 372.
- Diakonov, D. I., and M. Polyakov, 1993, *Nucl. Phys. B* **389**, 109.
- Dine, M., and W. Fischler, 1983, *Phys. Lett. B* **120**, 137.
- Dorey, N., T. J. Hollowood, V. V. Khoze, and M. P. Mattis, 1997, hep-th/9709072.
- Dorey, N., V. Khoze, and M. Mattis, 1996a, *Phys. Rev. D* **54**, 2921.
- Dorey, N., V. Khoze, and M. Mattis, 1996b, *Phys. Rev. D* **54**, 7832.
- Dorey, N., V. V. Khoze, and M. P. Mattis, 1997, hep-th/9708036.
- Dorokhov, A. E., and N. I. Kochelev, 1990, *Z. Phys. C* **46**, 281.
- Dorokhov, A. E., and N. I. Kochelev, 1993, *Phys. Lett. B* **304**, 167.
- Dorokhov, A. E., N. I. Kochelev, and Y. A. Zubov, 1993, *Int. J. Mod. Phys. A* **8**, 603.
- Dorokhov, A. E., Y. A. Zubov, and N. I. Kochelev, 1992, *Sov. J. Part. Nucl.* **23**, 522.
- Douglas, M. R., and S. H. Shenker, 1995, *Nucl. Phys. B* **447**, 271.
- Dowrick, N., and N. McDougall, 1993, *Nucl. Phys. B* **399**, 426.
- Dubovikov, M. S., and A. V. Smilga, 1981, *Nucl. Phys. B* **185**, 109.
- Eguchi, T., P. B. Gilkey, and A. J. Hanson, 1980, *Phys. Rep.* **66**, 213.
- Eletskii, V. L., and B. L. Ioffe, 1993, *Phys. Rev. D* **47**, 3083.
- Eletsky, V. L., 1993, *Phys. Lett. B* **299**, 111.
- Eletsky, V. L., and B. L. Ioffe, 1988, *Sov. J. Nucl. Phys.* **48**, 384.

- Espinosa, O., 1990, Nucl. Phys. B **343**, 310.
- Evans, N., S. D. H. Hsu, and M. Schwetz, 1996, Phys. Lett. B **375**, 262.
- Faber, M., H. Markum, S. Olejnik, and W. Sakuler, 1995, Nucl. Phys. B (Proc. Suppl.) **42**, 487.
- Faleev, S. V., and P. G. Silvestrov, 1995, Phys. Lett. A **197**, 372.
- Feurstein, M., H. Markum, and S. Thurner, 1997, Phys. Lett. B **396**, 203.
- Feynman, R. P., and A. R. Hibbs, 1965, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York).
- Finnel, D., and P. Pouliot, 1995, Nucl. Phys. B **453**, 225.
- Forkel, H., and M. K. Banerjee, 1993, Phys. Rev. Lett. **71**, 484.
- Forkel, H., and M. Nielsen, 1995, Phys. Lett. B **345**, 55.
- Förster, D., 1977, Phys. Lett. B **66**, 279.
- Forte, S., and E. V. Shuryak, 1991, Nucl. Phys. B **357**, 153.
- Fox, I. A., J. P. Gilchrist, M. L. Laursen, and G. Schierholz, 1985, Phys. Rev. Lett. **54**, 749.
- Fried, D., and K. Uhlenbeck, 1984, *Instantons and Four-Manifolds* (Springer, New York).
- Fukushima, M., H. Suganuma, A. Tanaka, H. Toki, and S. Sasaki, 1997, hep-lat/9709133.
- Georgi, H., and A. Manohar, 1984, Nucl. Phys. B **234**, 189.
- Gerber, P., and H. Leutwyler, 1989, Nucl. Phys. B **321**, 387.
- Geshkenbein, B., and B. Ioffe, 1980, Nucl. Phys. B **166**, 340.
- Gocksch, A., 1991, Phys. Rev. Lett. **67**, 1701.
- Gonzalez-Arroyo, A., and A. Montero, 1996, Phys. Lett. B **387**, 823.
- Gracey, J. A., 1996, Phys. Lett. B **373**, 178.
- Grandy, J., and R. Gupta, 1994, Nucl. Phys. B (Proc. Suppl.) **34**, 164.
- Gribov, V. N., 1981, preprint, KFKI-1981-66, Budapest.
- Gross, D. J., R. D. Pisarski, and L. G. Yaffe, 1981, Rev. Mod. Phys. **53**, 43.
- Grossman, B., 1977, Phys. Lett. A **61**, 86.
- Gupta, R., 1992, "Scaling, the renormalization group and improved lattice actions," preprint LA-UR-92-917, Los Alamos.
- Gupta, R., D. Daniel, and J. Grandy, 1993, Phys. Rev. D **48**, 3330.
- Halperin, I., and A. Zhitnitsky, 1997, hep-ph/9706251.
- Halperin, I., and A. Zhitnitsky, 1998, Phys. Rev. Lett. **80**, 438.
- Hansson, T. H., and I. Zahed, 1992, Nucl. Phys. B **374**, 277.
- Hansson, T. H., M. Sporre, and I. Zahed, 1994, Nucl. Phys. B **427**, 545.
- Harrington, B. J., and H. K. Shepard, 1978, Phys. Rev. D **17**, 2122.
- Hart, A., and M. Teper, 1995, Phys. Lett. B **371**, 261.
- Hasenfratz, A., T. DeGrand, and D. Zhu, 1996, Nucl. Phys. B **478**, 349.
- Hatsuda, T., Y. Koike, and S. H. Lee, 1993, Nucl. Phys. B **394**, 221.
- Hatsuda, T., and T. Kunihiro, 1985, Prog. Theor. Phys. **74**, 765.
- Hatsuda, T., and T. Kunihiro, 1994, Phys. Rep. **247**, 221.
- Hecht, M., and T. DeGrand, 1992, Phys. Rev. D **46**, 2115.
- Hoek, J., 1986, Phys. Lett. B **166**, 199.
- Hoek, J., M. Teper, and J. Waterhouse, 1987, Nucl. Phys. B **288**, 589.
- Huang, Z., and X.-N. Wang, 1996, Phys. Rev. D **53**, 5034.
- Hutter, M., 1995, Z. Phys. C **74**, 131.
- Ilgenfritz, E., and M. Müller-Preussker, 1981, Nucl. Phys. B **184**, 443.
- Ilgenfritz, E.-M., M. Müller-Preußker, and E. Meggiolaro, 1995, Nucl. Phys. B (Proc. Suppl.) **42**, 496.
- Ilgenfritz, E.-M., and E. V. Shuryak, 1989, Nucl. Phys. B **319**, 511.
- Ilgenfritz, E.-M., and E. V. Shuryak, 1994, Phys. Lett. B **325**, 263.
- Intriligator, K., and N. Seiberg, 1996, Nucl. Phys. B (Proc. Suppl.) **45**, 1.
- Ioffe, B. L., 1981, Nucl. Phys. B **188**, 317.
- Irback, A., P. LaCock, D. Miller, B. Peterson, and T. Reisz, 1991, Nucl. Phys. B **363**, 34.
- Ishikawa, K., G. Schierholz, H. Schneider, and M. Teper, 1983, Phys. Lett. B **128**, 309.
- Ito, K., and N. Sasakura, 1996, Phys. Lett. B **382**, 95.
- Iwasaki, Y., K. Kanaya, S. Kaya, S. Sakai, and T. Yoshie, 1996, Z. Phys. C **71**, 343.
- Iwasaki, Y., K. Kanaya, S. Sakai, and T. Yoshie, 1991, Phys. Rev. Lett. **67**, 1491.
- Ivanenko, T. L., and J. W. Negele, 1997, hep-lat/9709130.
- Jack, I., D. R. T. Jones, and C. G. North, 1997, Nucl. Phys. B **486**, 479.
- Jackiw, R., C. Nohl, and C. Rebbi, 1977, Phys. Rev. D **15**, 1642.
- Jackiw, R., and C. Rebbi, 1976a, Phys. Rev. Lett. **37**, 172.
- Jackiw, R., and C. Rebbi, 1976b, Phys. Rev. D **14**, 517.
- Jackiw, R., and C. Rebbi, 1977, Phys. Rev. D **16**, 1052.
- Jackson, A. D., and J. J. M. Verbaarschot, 1996, Phys. Rev. D **53**, 7223.
- Ji, X., 1995, Phys. Rev. Lett. **74**, 1071.
- Kacir, M., M. Prakash, and I. Zahed, 1996, preprint, hep-ph/9602314.
- Kanki, T., 1988, in *Proceedings of the International Workshop on Variational Calculations in Quantum Field Theory*, edited by L. Polley and D. Pottinger (World Scientific, Singapore), p. 215.
- Kaplan, D. B., 1992, Phys. Lett. B **288**, 342.
- Kapusta, J., 1989, *Finite Temperature Field Theory* (Cambridge University Press, Cambridge, England).
- Kapusta, J., D. Kharzeev, and L. McLerran, 1996, Phys. Rev. D **53**, 5028.
- Karsch, F., and E. Laermann, 1994, Phys. Rev. D **50**, 6954.
- Khoze, V., and A. Yung, 1991, Z. Phys. C **50**, 155.
- Kikuchi, H., and J. Wudka, 1992, Phys. Lett. B **284**, 111.
- Kim, J. E., 1979, Phys. Rev. Lett. **43**, 103.
- Kirkwood, J. G., 1931, J. Chem. Phys. **22**, 1420.
- Klevansky, S. P., 1992, Rev. Mod. Phys. **64**, 649.
- Klimt, S., M. Lutz, and W. Weise, 1990, Phys. Lett. B **249**, 386.
- Klinkhammer, F. R., and N. S. Manton, 1984, Phys. Rev. D **30**, 2212.
- Koch, V., and G. E. Brown, 1993, Nucl. Phys. A **560**, 345.
- Koch, V., E. V. Shuryak, G. E. Brown, and A. D. Jackson, 1992, Phys. Rev. D **46**, 3169.
- Koesterlitz, J. M., and D. J. Thouless, 1973, J. Phys. C **6**, 118.
- Kogan, I. I., and M. Shifman, 1995, Phys. Rev. Lett. **75**, 2085.
- Kogut, J. B., 1979, Rev. Mod. Phys. **51**, 659.
- Kogut, J. B., and D. K. Sinclair, 1988, Nucl. Phys. B **295**, 465.
- Kovner, A., and M. Shifman, 1997, Phys. Rev. D **56**, 2396.
- Kronfeld, A. S., 1988, Nucl. Phys. B (Proc. Suppl.) **4**, 329.
- Kronfeld, A., G. Schierholz, and U. J. Wiese, 1987, Nucl. Phys. B **293**, 461.
- Kuzmin, V. A., V. A. Rubakov, and M. E. Shaposhnikov, 1985, Phys. Lett. B **155**, 36.
- Landau, L. D., and E. M. Lifshitz, 1959, *Quantum Mechanics* (Pergamon, Oxford).
- Laursen, M., J. Smit, and J. C. Vink, 1990, Nucl. Phys. B **343**, 522.

- Lee, C., and W. A. Bardeen, 1979, Nucl. Phys. B **153**, 210.
- Lee, S. H., and T. Hatsuda, 1996, Phys. Rev. D **54**, 1871.
- Leinweber, D. B., 1995a, Phys. Rev. D **51**, 6369.
- Leinweber, D. B., 1995b, Phys. Rev. D **51**, 6383.
- Levine, H., and L. G. Yaffe, 1979, Phys. Rev. D **19**, 1225.
- Lipatov, L. N., 1977, JETP Lett. **22**, 104.
- Luescher, M., 1982, Commun. Math. Phys. **85**, 39.
- Maggiore, M., and M. Shifman, 1992a, Phys. Rev. D **46**, 3550.
- Maggiore, M., and M. Shifman, 1992b, Nucl. Phys. B **371**, 177.
- Mandelstam, S., 1976, Phys. Rep. **23**, 245.
- Manousakis, E., and J. Polonyi, 1987, Phys. Rev. Lett. **58**, 847.
- McLerran, L., 1986, Rev. Mod. Phys. **58**, 1021.
- Mermin, N., and H. Wagner, 1966, Phys. Rev. Lett. **16**, 1133.
- Michael, C., and P. S. Spencer, 1994, Phys. Rev. D **50**, 7570.
- Michael, C., and P. S. Spencer, 1995, Phys. Rev. D **52**, 4691.
- Montvay, I., and G. Münster, 1995, *Quantum Fields on a Lattice* (Oxford University Press, Oxford).
- Münster, G., 1982, Z. Phys. C **12**, 43.
- Nambu, Y., and G. Jona-Lasinio, 1961, Phys. Rev. **122**, 345.
- Narayanan, R., and H. Neuberger, 1995, Nucl. Phys. B **443**, 305.
- Narayanan, R., and P. Vranas, 1997, Nucl. Phys. B **506**, 373.
- Narison, S., 1984, Z. Phys. C **26**, 209.
- Narison, S., 1989, *QCD Spectral Sum Rules* (World Scientific, Singapore).
- Narison, S., 1996, Phys. Lett. B **387**, 162.
- Novak, M., G. Papp, and I. Zahed, 1996, Phys. Lett. B **389**, 341.
- Novikov, V. A., 1987, Sov. J. Nucl. Phys. **46**, 366.
- Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1979, Nucl. Phys. B **165**, 67.
- Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1980, Nucl. Phys. B **165**, 55.
- Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1981, Nucl. Phys. B **191**, 301.
- Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1983, Nucl. Phys. B **229**, 381.
- Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1985a, Nucl. Phys. B **260**, 157.
- Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1985b, Fortschr. Phys. **32**, 585.
- Novikov, V. A., M. A. Shifman, V. B. Voloshin, and V. I. Zakharov, 1983, Nucl. Phys. B **229**, 394.
- Nowak, M., 1991, Acta Phys. Pol. B **22**, 697.
- Nowak, M. A., J. J. M. Verbaarschot, and I. Zahed, 1989a, Nucl. Phys. B **324**, 1.
- Nowak, M. A., J. J. M. Verbaarschot, and I. Zahed, 1989b, Phys. Lett. B **228**, 115.
- Nowak, M. A., J. J. M. Verbaarschot, and I. Zahed, 1989c, Nucl. Phys. B **325**, 581.
- Olejnik, S., 1989, Phys. Lett. B **221**, 372.
- Peccei, R., and H. R. Quinn, 1977, Phys. Rev. D **16**, 1791.
- Phillips, A., and D. Stone, 1986, Commun. Math. Phys. **103**, 599.
- Pisarski, R. D., and M. Tytgat, 1997, preprint, hep-ph/9702340.
- Pisarski, R. D., and F. Wilczek, 1984, Phys. Rev. D **29**, 338.
- Pisarski, R. D., and L. G. Yaffe, 1980, Phys. Lett. B **97**, 110.
- Pobylitsa, P., 1989, Phys. Lett. B **226**, 387.
- Polikarpov, M. I., and A. I. Veselov, 1988, Nucl. Phys. B **297**, 34.
- Polyakov, A. M., 1975, Phys. Lett. B **59**, 79.
- Polyakov, A., 1977, Nucl. Phys. B **120**, 429.
- Polyakov, A. M., 1978, Phys. Lett. B **72**, 477.
- Polyakov, A., 1987, *Gauge Fields and Strings* (Harwood, Chur, Switzerland).
- Preskill, J., M. B. Wise, and F. Wilczek, 1983, Phys. Lett. B **120**, 133.
- Pugh, D. J. R., and M. Teper, 1989, Phys. Lett. B **224**, 159.
- Rajagopal, K., and F. Wilczek, 1993, Nucl. Phys. B **399**, 395.
- Rajaraman, R., 1982, *Solitons and Instantons* (North-Holland, Amsterdam).
- Randall, L., R. Rattazzi, and E. Shuryak, 1998, hep-ph/9803258.
- Rapp, R., T. Schäfer, E. V. Shuryak, and M. Velkovsky, 1997, hep-ph/9711396.
- Reinders, L. J., H. Rubinstein, and S. Yazaki, 1985, Phys. Rep. **127**, 1.
- Ringwald, A., 1990, Nucl. Phys. B **330**, 1.
- Ringwald, A., and F. Schrempp, 1994, in *Proceedings of the International Seminar Quarks 94*, Vladimir, Russia, preprint, hep-ph/9411217.
- Rothe, H. J., 1992, *Lattice Gauge Theories, An Introduction* (World Scientific, Singapore).
- Salomonson, P., and J. W. V. Holton, 1981, Nucl. Phys. B **196**, 509.
- Savvidy, G. K., 1977, Phys. Lett. B **71**, 133.
- Schäfer, T., 1996, Phys. Lett. B **389**, 445.
- Schäfer, T., 1997, hep-ph/9708256.
- Schäfer, T., and E. V. Shuryak, 1995a, Phys. Rev. Lett. **75**, 1707.
- Schäfer, T., and E. V. Shuryak, 1995b, Phys. Lett. B **356**, 147.
- Schäfer, T., and E. V. Shuryak, 1996a, Phys. Rev. D **53**, 6522.
- Schäfer, T., and E. V. Shuryak, 1996b, Phys. Rev. D **54**, 1099.
- Schäfer, T., E. V. Shuryak, and J. J. M. Verbaarschot, 1994, Nucl. Phys. B **412**, 143.
- Schäfer, T., E. V. Shuryak, and J. J. M. Verbaarschot, 1995, Phys. Rev. D **51**, 1267.
- Schierholz, G., 1994, Nucl. Phys. Proc. Suppl. **37A**, 203.
- Schramm, S., and M.-C. Chu, 1993, Phys. Rev. D **48**, 2279.
- Schwarz, A. S., 1977, Phys. Lett. B **67**, 172.
- Seiberg, N., 1994, Phys. Rev. D **49**, 6857.
- Seiberg, N., and E. Witten, 1994, Nucl. Phys. B **426**, 19.
- Shifman, M. A., 1989, Sov. Phys. Usp. **32**, 289.
- Shifman, M. A., 1992, *Vacuum Structure and QCD Sum Rules* (North-Holland, Amsterdam).
- Shifman, M., 1994, *Instantons in Gauge Theories* (World Scientific, Singapore).
- Shifman, M. A., and A. I. Vainshtein, 1988, Nucl. Phys. B **296**, 445.
- Shifman, M. A., A. I. Vainshtein, and V. I. Zakharov, 1978, Phys. Lett. B **76**, 971.
- Shifman, M. A., A. I. Vainshtein, and V. I. Zakharov, 1979, Nucl. Phys. B **147**, 385, 448.
- Shifman, M. A., A. I. Vainshtein, and V. I. Zakharov, 1980a, Nucl. Phys. B **166**, 493.
- Shifman, M., A. I. Vainshtein, and V. I. Zakharov, 1980b, Nucl. Phys. B **165**, 45.
- Shifman, M., A. I. Vainshtein, and V. I. Zakharov, 1980c, Nucl. Phys. B **163**, 46.
- Shuryak, E., 1978a, Zh. Éksp. Teor. Fiz. [Sov. Phys. JETP **74**, 408].
- Shuryak, E. V., 1978b, Phys. Lett. B **79**, 135.
- Shuryak, E., 1980, Phys. Rep. **61**, 71.
- Shuryak, E. V., 1981, Phys. Lett. B **107**, 103.
- Shuryak, E. V., 1982a, Nucl. Phys. B **198**, 83.
- Shuryak, E. V., 1982b, Nucl. Phys. B **203**, 93.

- Shuryak, E. V., 1982c, Nucl. Phys. B **203**, 116.  
 Shuryak, E. V., 1982d, Nucl. Phys. B **203**, 140.  
 Shuryak, E. V., 1983, Nucl. Phys. B **214**, 237.  
 Shuryak, E. V., 1985, Phys. Lett. B **153**, 162.  
 Shuryak, E. V., 1987, Phys. Lett. B **196**, 373.  
 Shuryak, E. V., 1988a, Nucl. Phys. B **302**, 559, 574, 599.  
 Shuryak, E. V., 1988b, Nucl. Phys. B **302**, 621.  
 Shuryak, E. V., 1988c, *The QCD Vacuum, Hadrons and the Superdense Matter* (World Scientific, Singapore).  
 Shuryak, E. V., 1989, Nucl. Phys. B **319**, 521, 541.  
 Shuryak, E. V., 1993, Rev. Mod. Phys. **65**, 1.  
 Shuryak, E. V., 1994, Comments Nucl. Part. Phys. **21**, 235.  
 Shuryak, E. V., 1995, Phys. Rev. D **52**, 5370.  
 Shuryak, E. V., and J. L. Rosner, 1989, Phys. Lett. B **218**, 72.  
 Shuryak, E. V., and M. Velkovsky, 1994, Phys. Rev. D **50**, 3323.  
 Shuryak, E. V., and M. Velkovsky, 1997, preprint hep-ph/9703345.  
 Shuryak, E. V., and J. J. M. Verbaarschot, 1990, Nucl. Phys. B **341**, 1.  
 Shuryak, E. V., and J. J. M. Verbaarschot, 1991, Nucl. Phys. B **364**, 255.  
 Shuryak, E. V., and J. J. M. Verbaarschot, 1992, Phys. Rev. Lett. **68**, 2576.  
 Shuryak, E. V., and J. J. M. Verbaarschot, 1993a, Nucl. Phys. B **410**, 55.  
 Shuryak, E. V., and J. J. M. Verbaarschot, 1993b, Nucl. Phys. B **410**, 37.  
 Shuryak, E. V., and J. J. M. Verbaarschot, 1995, Phys. Rev. D **52**, 295.  
 Shuryak, E. V., and A. R. Zhitnitsky, 1997, hep-ph/9706316.  
 Skyrme, T., 1961, Proc. R. Soc. London, Ser. A **262**, 207.  
 Smilga, A., 1994a, Phys. Rev. D **49**, 5480.  
 Smilga, A., 1994b, Phys. Rev. D **49**, 6836.  
 Smilga, A. V., 1996, Phys. Rep. **291**, 1.  
 Smilga, A., and J. Stern, 1993, Phys. Lett. B **318**, 531.  
 Smit, J., and A. J. van der Sijs, 1989, Nucl. Phys. B **355**, 603.  
 Smit, J., and J. C. Vink, 1987, Nucl. Phys. B **284**, 234.  
 Smit, J., and J. C. Vink, 1988, Nucl. Phys. B **298**, 557.  
 Snyderman, N., and S. Gupta, 1981, Phys. Rev. D **24**, 542.  
 Steele, J., A. Subramanian, and I. Zahed, 1995, Nucl. Phys. B **452**, 545.  
 Suganuma, H., K. Itakura, H. Toki, and O. Miyamura, 1995, Nucl. Phys. Proc. Suppl. **47**, 302.  
 Suzuki, T., 1993, Nucl. Phys. B (Proc. Suppl.) **30**, 176.  
 Svetitsky, B., and L. G. Yaffe, 1982, Nucl. Phys. B **210**, 423.  
 't Hooft, G., 1976a, Phys. Rev. Lett. **37**, 8.  
 't Hooft, G., 1976b, Phys. Rev. D **14**, 3432.  
 't Hooft, G., 1981a, Commun. Math. Phys. **81**, 267.  
 't Hooft, G., 1981b, Nucl. Phys. B **190**, 455.  
 't Hooft, G., 1986, Phys. Rep. **142**, 357.  
 Takeuchi, S., and M. Oka, 1991, Phys. Rev. Lett. **66**, 1271.  
 Tar, C. D., and J. Kogut, 1987, Phys. Rev. D **36**, 2828.  
 Teper, M., 1986, Phys. Lett. B **171**, 86.  
 Thomas, A. W., S. Theberge, and G. A. Miller, 1981, Phys. Rev. D **24**, 216.  
 Thurner, S., M. Feurstein, and H. Markum, 1997, Phys. Rev. D **56**, 4039.  
 Ukawa, A., 1997, Nucl. Phys. B (Proc. Suppl.) **53**, 106.  
 Vafa, C., and E. Witten, 1984, Nucl. Phys. B **234**, 173.  
 Vainshtein, A. I., M. A. Shifman, and V. I. Zakharov, 1985, Sov. Phys. Usp. **28**, 709.  
 Vainshtein, A. I., V. I. Zakharov, V. A. Novikov, and M. A. Shifman, 1982, Sov. Phys. Usp. **25**, 195.  
 Vainshtein, A. I., V. I. Zakharov, V. A. Novikov, and M. A. Shifman, 1986, Sov. J. Nucl. Phys. **43**, 294.  
 Velikson, B., and D. Weingarten, 1985, Nucl. Phys. B **249**, 433.  
 Velkovsky, M., and E. V. Shuryak, 1993, Phys. Lett. B **304**, 281.  
 Velkovsky, M., and E. Shuryak, 1996, Phys. Rev. D **56**, 2766.  
 Veneziano, G., 1979, Nucl. Phys. B **159**, 213.  
 Veneziano, G., 1992, Mod. Phys. Lett. A **7**, 1666.  
 Verbaarschot, J. J. M., 1991, Nucl. Phys. B **362**, 33.  
 Verbaarschot, J. J. M., 1994a, Act. Phys. Pol. **25**, 133.  
 Verbaarschot, J. J. M., 1994b, Nucl. Phys. B **427**, 534.  
 Vink, J. C., 1988, Phys. Lett. B **212**, 483.  
 Vogl, U., and W. Weise, 1991, Prog. Part. Nucl. Phys. **27**, 195.  
 Weingarten, D., 1983, Phys. Rev. Lett. **51**, 1830.  
 Weingarten, D., 1994, Nucl. Phys. B (Proc. Suppl.) **34**, 29.  
 Wettig, T., A. Schäfer, and H. Weidenmüller, 1996, Phys. Lett. B **367**, 28.  
 Wiese, U. J., 1990, Nucl. Phys. B (Proc. Suppl.) **17**, 639.  
 Wilczek, F., 1992, Int. J. Mod. Phys. A **7**, 3911.  
 Wilson, K. G., 1969, Phys. Rev. **179**, 1499.  
 Witten, E., 1977, Phys. Rev. Lett. **38**, 121.  
 Witten, E., 1979a, Nucl. Phys. B **149**, 285.  
 Witten, E., 1979b, Nucl. Phys. B **156**, 269.  
 Witten, E., 1981, Nucl. Phys. B **188**, 513.  
 Witten, E., 1983, Nucl. Phys. B **223**, 433.  
 Wohler, C. F., and E. V. Shuryak, 1994, Phys. Lett. B **333**, 467.  
 Voit, P., 1983, Phys. Rev. Lett. **51**, 638.  
 Wolff, U., 1989, Nucl. Phys. B **322**, 759.  
 Yung, A., 1997, hep-th/9705181.  
 Yung, A. V., 1988, Nucl. Phys. B **297**, 47.  
 Zakharov, V. I., 1992, Nucl. Phys. B **371**, 637.  
 Zhitnitsky, A. R., 1980, Sov. J. Nucl. Phys. **31**, 260.  
 Zinn-Justin, J., 1981, J. Math. Phys. **22**, 511.  
 Zinn-Justin, J., 1982, "The principles of instanton calculus: A few applications," in *Recent Advances in Field Theory*, Les Houches, Session XXXIX, edited by J.-B. Zuber and R. Stora (North Holland, Amsterdam), p. 39.  
 Zinn-Justin, J., 1983, Nucl. Phys. B **218**, 333.