The topological theory of defects in ordered media*†

N. D. Mermin

Laboratory of Atomic and Solid State Physics Cornell University, Ithaca, New York 14853

Aspects of the theory of homotopy groups are described in a mathematical style closer to that of condensed matter physics than that of topology. The aim is to make more readily accessible to physicists the recent applications of homotopy theory to the study of defects in ordered media. Although many physical examples are woven into the development of the subject, the focus is on mathematical pedagogy rather than on a systematic review of applications.

CONTENTS

I. Introduction 592
II. Ordered Media and Defects 594
A. Examples of ordered media 594
1. Planar spins 594
2. Ordinary spins 594
3. Nematic liquid crystals 594
4. Biaxial nematics 595
5. Superfluid helium-3 595
B. A simple example of the topology of defects: planar spins in two dimensions 596
C. An even simpler illustration: ordinary spins in the plane 600
III. The Fundamental Group 600
A. The fundamental group at a point 601
1. Loops at x 601
2. Homotopies at x 601
3. The product of two loops 602
4. The product of homotopy classes of loops 602
5. The fundamental group at x, $\pi_1(R,x)$ 602
6. A few simple examples 603
a. The circle 603
b. The surface of a sphere 603
c. The figure-eight space 604
B. The fundamental group of a connected space 604
1. The isomorphism between fundamental groups based at different points 604
2. Uniqueness (or lack of uniqueness) of the path isomorphisms between based fundamental groups 605
C. The fundamental group and the classes of freely homotopic loops 606
D. The fundamental group and the combination of line defects 607
IV. Group-theoretic Structure of the Order-parameter Space 608
A. Continuous groups 608
B. Group-theoretic description of the order- 609
parameter space
C. Examples of order-parameter spaces as 611
coset spaces 611
1. Planar spins 612
2. Ordinary spins 612
3. Nematics 612
4. Biaxial nematics 612
5. Superfluid helium-3
D. Properties of the fundamental group of 613
a topological group
V. The Fundamental Group of the Order- 614
parameter Space
A. The fundamental theorem on the 614
fundamental group
1. The correspondence between loops in
coset space and the connected 615
components of the isotropy subgroup
2. Proof of the isomorphism between 616
$\pi_1(G/H)$ and H/H_0
B. Computing the fundamental group of the 617
order-parameter space 617
1. Planar spins 617
2. Ordinary spins 617
3. Nematics 618
4. Biaxial nematics 619
5. Superfluid helium-3 619
VI. Media with Non-Abelian Fundamental Groups
A. More on the nature of line defects in 620
biaxial nematics
B. The combination of defects in the non- 621
Abelian case 623
C. The entanglement of line defects
VII. The Second Homotopy Group and the
Classification of Point Defects in Three 625
Dimensions 625
A. The second homotopy group at x, $\pi_2(R,x)$ 626
B. The second homotopy group, $\pi_2(R)$
C. The fundamental theorem on the second 628
homotopy group of coset spaces
D. The action of π_1 on π_2: classes of 630
freely homotopic spheres in G/H
E. Examples of point defects 631
1. Planar spins 631
2. Ordinary spins 631
3. Nematics 631
4. Biaxial nematics 632
5. Superfluid helium-3 632
VIII. Ordered Media with Broken Translational 632
Symmetry in the Uniform State
A. The naive generalization 633
1. Order-parameter space 633
2. Defects and homotopy groups 633
3. Computing the homotopy groups 633
B. Critique of the naive generalization 633

*A preliminary version of this paper appeared as Technical Report No. 3045 of the Materials Science Center of Cornell University. It has been substantially revised at the University of Sussex, with the partial financial support of the Science Research Council of Great Britain and with the generous help and hospitality of A. J. Leggett and D. F. Brewer. The revision has also been supported in part by the National Science Foundation under Grant No. DMR 77-18329 and through the Materials Science Center of Cornell University, Technical Report No. 4021.

†A complementary review of homotopy, from the point of view of a mathematician and field theorist, by Louis Michel is scheduled to appear in a future issue of Reviews of Modern Physics.

C. Some applications 635
1. Crystals (dislocations only) 635
2. Crystals (dislocations and disclinations) 635
3. Directed stripes in the plane 637
4. Cholesteric liquid crystals 638
IX. Higher Homotopy Groups, Relative Homotopy Groups, and Exact Sequences 639
A. Higher homotopy groups 639
B. Relative homotopy groups 640
C. The exact homotopy sequence 641
D. Recovery of the fundamental theorems from an exact sequence 643
E. Uses of the third homotopy group; solitons 644
Appendix A: Glossary of Technical Terms 645
Appendix B: Summary of Elementary Group Theory 646
References 648

I. INTRODUCTION

The language, methods, and theorems of algebraic topology, particularly homotopy theory, have been used in the study of relativistic field theories for over a decade. Their systematic application to the study of defects in condensed matter physics is more recent, having started in earnest with the independent studies of Toulouse and Kleman (1976), Volovik and Mineyev (1976), and Rogula (1976).

Nevertheless the nonrelativistic applications of homotopy theory have already had a significant impact. In one case (the *A* phase of superfluid helium-3) it was a nontrivial application of the topological method (Anderson and Toulouse, 1977) that first revealed the possibility of some striking and quite unexpected hydrodynamical behavior. Applications of the method to liquid crystals have in many cases brought a new coherence, consistency, and simplicity to our understanding of defects in these intricate systems. The study of Poénaru and Toulouse (1977) on the crossing of line defects demonstrates that the topological point of view can suggest forms of physical behavior which, to my knowledge, had not been imagined before, and can furthermore give a systematic and precise description of such behavior which is far from obvious to an intuition unaided by the tools that homotopy theory provides.

At a minimum, homotopy theory provides *the* natural language for the description and classification of defects in a large class of ordered systems. Whether it will eventually gain as wide a currency among condensed matter physicists as, for example, the language and theorems of the theory of group representations, depends both on how large that class of systems proves to be, and on how many of the *nontrivial* topological insights turn out to have direct manifestations in the laboratory.

While its eventual importance is not yet clear, there is no question that the topological method is currently enjoying a vogue among condensed matter physicists, leading to an unfortunate barrier between those who speak its language and those who do not. This article is intended for physicists sufficiently interested in or curious about this novel approach to want to learn the language and even some of the techniques. I have tried to write a piece of introductory mathematical pedagogy which always focuses on the physical applications, but makes no attempt at a systematic review of the (rapidly growing) list of materials and problems to which the method is being applied.

The need for such a pedagogical article lies in the nature of the mathematical literature on homotopy theory, which can be approached only with great effort by those with the mathematical background of a typical physicist (as I can testify from painful personal experience). The trouble is that expositions which can be followed by one with such a background—i.e., those to be found in introductory undergraduate topology texts—fail to reach many of the theorems and concepts central to the physical applications. Expositions that do go far enough, however, place the subject in so general a mathematical setting and assume so high a level of mathematical expertise, that it is not at all easy to ferret out those occasional bits of the vastly expanded subject that are physically pertinent.

What follows emerged from my own recent efforts to feel my way through the mathematical thicket towards an understanding of the recent developments in the theory of defects. The reader is firmly warned that I am not at all expert in topology: Virtually everything I know (or think I know) is to be found in the pages that follow. This has the obvious drawback that my exposition may, on occasion, lapse into unnecessary clumsiness or even error. On the other hand, because I plunged into the subject as innocent and impatient as any mathematically illiterate physicist would be likely to be, the structure I constructed to support my own understanding stands a better chance of being congenial to those who approach mathematics with a physicist's temperament.

The main mathematical background I assume in the reader is an acquaintance with the most elementary terms and concepts of group theory (but not, except in an occasional inessential remark, with any of the theory of group representations). Except for references to the homorphism between SU(2) and SO(3), familiar to most physicists (in slightly different terminology) from the theory of spin $\frac{1}{2}$, the required background in group theory can be found in the first few mathematical pages of any text on physical applications of groups. This background is summarized in Appendix B.

I do not assume any familiarity with the properties of continuous groups or the concepts of topology. I have tried to treat homotopy theory and the topology of continuous groups in much the same way that elementary calculus is dealt with by physicists: I rely heavily on the reader's firm intuitive grasp of the notion of continuity, and invite readers possessing the appropriate blend of ingenuity and perversity to add whatever assumptions of regularity are needed to exclude whatever pathological counterexamples they may come up with. This is, admittedly, a dangerous game to play, but it has had a long and honorable history of successful practice. In my opinion the substantial gain in clarity it achieves more than compensates for the reduction in certainty. Bridges would not be safer if only people who knew the proper definition of a real number were allowed to design them.

The exposition is organized as follows:

Section II introduces ordered media and their order-parameter spaces in the conventional way. Several examples are given which are used repeatedly as illustrations in the sections that follow. The relevance of topological analysis to the theory of defects is developed in a special case simple enough to require none of the formalism of homotopy theory, but general enough to permit many essential ideas to be fully expounded.

Section III abstracts the essence of the informal topological argument of Sec. II into the central notion of the fundamental group (or first homotopy group) of the order-parameter space. The fundamental group is described for quite general a space. When the space happens to be an order-parameter space, its fundamental group is shown to be directly related to the description of line defects in three-dimensional media or point defects in two dimensions.

Section IV gives a common group-theoretic characterization to the various examples of order-parameter spaces introduced in Sec. II. The characterization is essentially that of the Landau–Lifshitz theory of broken symmetry. Although its group-theoretic statement might appear perversely abstract, it turns out to be by far the most natural formulation for the treatment of defects.

Section V computes the fundamental group for any space that can be given the group-theoretic characterization of Sec. IV. The implications of this computation for the theory of defects are illustrated with the standard examples.

Section VI describes some striking behavior that can arise in media whose order-parameter spaces have non-Abelian fundamental groups. In my opinion this is the most interesting feature yet to emerge from the topological approach.

Section VII introduces the second homotopy group, and computes it for spaces having the group-theoretic structure described in Sec. IV. Its relevance to point defects in 3-dimensional media is described, and illustrated with the standard examples.

Section VIII discusses whether the class of systems to which the topological method applies includes media with broken translational symmetry. (All of the standard examples are translationally invariant.) I warn the reader that my views on this point are not completely orthodox: Much of the basic literature treats media with broken translational symmetry on precisely the same footing as media in which only rotational symmetries are broken, and the reservations I describe are dealt with casually, if at all. I indicate what problems I believe must be resolved before the method can be used with confidence when there is broken translational symmetry. Some examples are given in which the method does indeed reproduce well known results, when used cautiously.

Section IX introduces some further topological abstractions: higher homotopy groups, relative homotopy groups, and exact homotopy sequences. The exact sequence is a powerful computational tool, but is probably of little interest to readers who are not actively engaged in topological studies. Relative homotopy groups, however, have recently been given some direct physical applications in their own right, which are briefly mentioned. The relevance of the first, second, *and* third homotopy groups to the classification of solitons is mentioned.

Appendix A consists of a glossary of technical terms introduced in the text, with references to where they are first defined. Words or phrases corresponding to such citations appear in italics in the text.

Appendix B lists the definitions and theorems of elementary group theory that are used in the text, and summarizes the relevant features of the homomorphism between SO(3) and SU(2).

A few remarks about supplementary mathematical references: I have found little written after 1960 of much help for reasons already alluded to. Texts that try to be clear do not go nearly far enough, while texts that do go far enough go so far that the task of extracting useful results from the very general setting is at least as difficult as deriving them from scratch by oneself. Of the earlier postwar books, I have found Hilton (1953) and Steenrod (1951) to be the most helpful. Steenrod has more than one wants to know (but not orders of magnitude more) and Hilton, in the relevant chapters, has not quite enough. However, both treatises, unlike most of their successors, are illuminated by somewhat austere but unquestionably human presences. My favorite books are those of Pontryagin (1966) and Weyl (1946). (The dates in both cases are those of the second editions.) Unfortunately they only bear peripherally on the mathematics of interest, many important developments having taken place after the appearance of their first editions. However, they are so beautifully and humanely written that I commend them to readers for preliminary background and would, if I could, make them required reading for anyone publishing what passes these days for mathematical prose.

Citations of the recent literature on physical applications are only given when the sources cited might amplify or clarify the points made here. Because this is a pedagogical rather than an historical review, I have not cluttered the text with citations whose only purpose is to acknowledge sources. I have, however, included among the references those papers from which I acquired my own understanding of the subject, even though not all of them are cited in the text itself.

Since few readers are likely to persevere to the end of my essay I conclude this introductory section with some personal acknowledgments. Jason Ho and Steve Shenker persuaded me that I could learn homotopy theory, Joel Mermin convinced me that I could lecture on it, Eric Siggia helped me to understand what Poénaru and Toulouse (1977) and Kleman (1977a) were saying, and Andrew Somese came through with friendly mathematical wisdom at some desperate moments.[1]

[1] It was M. E. Fisher who first suggested and repeatedly insisted that I should publish my lecture notes, but I am not sure he deserves thanks for this.

But my real teachers have been M. Kleman, V. P. Mineyev, G. Toulouse, and G. E. Volovik, who have sent me preprints and reprints from the beginning, as well as occasional stimulating personal messages. Finally, on almost every page the benefits can be discerned of eight years of ongoing arguments on the art of pedagogy with N. W. Ashcroft.

II. ORDERED MEDIA AND DEFECTS

For almost all of our purposes here an *ordered medium* can be regarded as a region of space described by a function $f(\mathbf{r})$ that assigns to every point of the region an *order parameter*. The possible values of the order parameter constitute a space known as the *ordered-parameter space* (or *manifold of internal states*).

We will generally take the region of space to be all of ordinary three-dimensional space, though two-dimensional regions can provide instructive examples and be relevant to the physics of films. The medium is said to be *uniform* if the function f is a constant—i.e., if the value of the order parameter is everywhere the same. We shall be interested in nonuniform media in which the order parameter varies continuously through the space except, perhaps, at isolated points, lines, or surfaces. These singular regions of lower dimensionality constitute the *defects* to be investigated.

A few restrictions on the form of the order-parameter space will be imposed in Sec. IV.B. All of these restrictions are satisfied by the examples given below in part A of this section, and it is through the examination of these examples that the reader is urged to acquire a sense of what ordered media and order-parameter spaces are. Some important physical systems (for example crystals, and smectic or cholesteric liquid crystals) are not among the examples of part A. Such systems (characterized by a lack of perfect translational symmetry in the uniform state) present special problems for the topological method. Their consideration is deferred to Sec. VIII to keep the exposition simple and because, in my opinion, those special problems have not yet been adequately resolved.

After the illustrative examples are introduced in part A, many essential features of the theory of defects are introduced in part B in the context of a particularly simple example. Readers thoroughly familiar with the example are nevertheless urged to read part B, because the important topological ideas introduced in that limited context turn out to be completely general. Part C contains a second simple example, to illustrate how drastically conclusions can depend on the structure of the order-parameter space.

A. Examples of ordered media

1. Planar spins

The order parameter is a vector of fixed magnitude (conventionally set equal to unity) constrained to lie in a plane. The order-parameter space can therefore be taken to be the circumference of a circle, under the usual correspondence between points on the circumference and directions in the plane. If $\hat{u}$ and $\hat{v}$ are a pair of orthonormal vectors in the plane, then the function $f(\mathbf{r})$ is of the form

$$f(\mathbf{r})=\hat{u}\cos\varphi(\mathbf{r})+\hat{v}\sin\varphi(\mathbf{r}) \tag{2.1}$$

(the circle being specified by the vector coordinates of its points).

The same order-parameter space also describes superfluid helium-4, where the order parameter is a complex scalar field of fixed magnitude ψ_0 but arbitrary phase:

$$\psi(\mathbf{r})=\psi_0 e^{i\varphi(\mathbf{r})}\,. \tag{2.2}$$

The circle can now be regarded as the unit circle in the complex plane.

2. Ordinary spins

The order parameter is a unit vector free to point in any direction of three-dimensional space. The order-parameter space can therefore be taken to be the surface of the unit sphere in 3-space. If $\hat{u}$, $\hat{v}$, and $\hat{w}$ are a fixed orthonormal triad then the function f is of the form

$$f(\mathbf{r})=\mathbf{s}(\mathbf{r})=s_u(\mathbf{r})\hat{u}+s_v(\mathbf{r})\hat{v}+s_w(\mathbf{r})\hat{w}\,,\quad s_u^2+s_v^2+s_w^2=1\,. \tag{2.3}$$

[Note that there is no underlying lattice structure. The reader should either regard the medium as a ferromagnetic liquid or as a macroscopic continuum model of a ferromagnetic crystal, in which the scale of spatial variation of $\mathbf{s}(\mathbf{r})$ (taken proportional to the local spontaneous magnetization) is very large on the scale of a lattice constant. Refinements to include the lattice structure as well are certainly of interest, but are subject to the complications discussed in Sec. VIII.]

3. Nematic liquid crystals

Nematics are like example 2, except that the vector has no arrowhead (or identical arrowheads at each end). The order parameter describes the local preferred axis in a medium of long molecules with the symmetry of ellipsoids of revolution. There are various (equivalent) ways of specifying the order parameter:

(a) As a unit vector but without an associated direction. The order-parameter space is then the surface of the unit sphere, as in example 2, except that diametrically opposite points must be identified, since rotating a molecule through 180° about an axis perpendicular to the axis of continuous symmetry results in a configuration indistinguishable from the original one. This space is known as the *projective plane* (P_2). (It cannot be realized as a closed non-self-intersecting manifold in three-dimensional space, and is better regarded as the surface of an ordinary sphere in 3-space, with the appropriate identification of pairs of points, just as a circle can be regarded as a line segment with the end points identified.)

(b) If one is made uncomfortable by headless vectors, one can define the order parameter to be the dyadic

$$f(\mathbf{r})=\mathsf{M}(\mathbf{r})=\hat{n}(\mathbf{r})\hat{n}(\mathbf{r})\,,\quad (M_{ij}=n_i n_j)\,. \tag{2.4}$$

Specifying M (which is just the projection operator on $\hat{n}$) gives all the information $\hat{n}$ gives except for the sign of the direction.

(c) One can subtract from M a constant times the unit tensor to get a real traceless symmetric matrix with a pair of degenerate eigenvalues. This is probably the most physical form in which to represent the order parameter, since it can be regarded as the deviation from isotropy of any convenient tensor property of the medium (e.g., its dielectric constant).

4. Biaxial nematics

None of these have yet been produced, but they are especially interesting to examine from the point of view of the theory of defects. Biaxial nematics are like example 3, except that the symmetry of the molecule is reduced to that of a rectangular box (proper point group D_2). One can take the order parameter to be a real symmetric matrix with three (fixed) distinct eigenvalues (representing, for example, the dielectric constant of the medium, the moment of inertia of the object, or some other convenient property). One requires almost the entire three-dimensional rotation group to specify the orientation of such an object. However, configurations differing only by 180° rotations about any of the three perpendicular symmetry axes are indistinguishable, and the order-parameter space can therefore be identified with a parameter space for the full proper three-dimensional rotation group SO(3), provided the appropriate discrete sets of four points (representing quadruples of equivalent rotations) are identified. A more precise specification of this order parameter space will be given in Sec. IV.C. The only point to note now is that the order-parameter space for a conceptually quite simple medium can be rather intricate.

There is no particular reason why the point group D_2 should be singled out for special attention, and we shall also examine, under the heading of biaxial nematics, media of objects whose proper symmetry operations comprise arbitrary discrete point groups.

Note that the "objects" need not be taken too literally. The point group of symmetries of such objects can just as well be replaced by the point group of symmetries of the uniform medium, and the local orientation of the objects by the local orientation of the appropriate local symmetry axes in the nonuniform medium.

5. Superfluid helium-3

This substance has provided us with order parameters quite unlike any hitherto considered; its discovery stimulated the resort to homotopy theory, and the topological insights thereby gained have led to significant advances in our understanding. The physics underlying these peculiar order parameters and the physical implications of the topological analysis can be found, for example, in Anderson and Toulouse (1977), Anderson and Palmer (1977), Mermin (1977 and 1978a) and references cited by these authors. Here we simply note that there are systems in nature requiring order parameters considerably more intricate than those of our first five examples, and limit ourselves to describing a few of these order parameters, without inquiring into their physical interpretation. Superfluid helium-3 has several phases, and various regimes within each phase. We consider here only two examples:

a. Dipole-locked A phase

The order parameter is a pair of (distinguished) orthonormal axes, arbitrarily oriented (except for the constraint of orthogonality):

$$\hat{\phi}^{(1)}\cdot\hat{\phi}^{(1)}=\hat{\phi}^{(2)}\cdot\hat{\phi}^{(2)}=1\,,\quad \hat{\phi}^{(1)}\cdot\hat{\phi}^{(2)}=0\,. \tag{2.5}$$

The order parameter can, alternatively, be described by a single complex vector field,

$$\mathbf{e}=\hat{\phi}^{(1)}+i\hat{\phi}^{(2)}\,, \tag{2.6}$$

constrained to satisfy

$$\mathbf{e}\cdot\mathbf{e}^{*}=1\,,\quad \mathbf{e}\cdot\mathbf{e}=0\,. \tag{2.7}$$

A system whose order parameter is such an orthonormal pair is like a generalized biaxial nematic whose molecules have no proper symmetries at all, since the only proper rotation leaving a pair of orthonormal vectors fixed is the identity. The order-parameter space can therefore be identified with a parameter space for the full proper three-dimensional rotation group, SO(3). We shall have more to say about this order-parameter space below. For the moment we only remark that the order-parameter space for SO(3) can be taken to be the surface of a unit four-dimensional sphere with diametrically opposite points identified[2] (known as the *projective space* P_3). Thus the order-parameter space for dipole-locked ^{3}He-A is the analogue of that for an ordinary nematic, one dimension higher up.

b. Dipole-free A phase

The order-parameter field is now of the form

$$\mathsf{A}(\mathbf{r})=\hat{n}(\mathbf{r})\mathbf{e}(\mathbf{r})\,, \tag{2.8}$$

where $\hat{n}$ is a real unit vector and $\mathbf{e}$ is the complex nilpotent unit vector of Eq. (2.6). The relative orientations of $\hat{n}$ and $\mathbf{e}$ can vary, as well as their absolute orientations. The order-parameter space must be taken as the product of the surface of the unit sphere in 3-space (S_2) (to represent $\hat{n}$) with projective 3-space (P_3) (to represent $\mathbf{e}$) with the identification of points necessary to take account of the fact that Eq. (2.8) is unaffected if both $\hat{n}$ and $\mathbf{e}$ change sign.

With this example one begins to understand why help was sought from the topologists. Since S_2 is locally two dimensional and P_3 is locally three dimensional, the order-parameter space is an object that is locally five dimensional, complicated by rules specifying the identification of certain discrete sets of points.

A small cautionary note should be added to this list of examples: From a formal point of view the range of values of the order parameter is, in fact, defined

[2] See the discussion at the end of Appendix B.

by the order-parameter space. However, in practice one usually starts with a definite family of objects in mind (e.g., vectors, projection operators, cubes) and only then specifies an abstract order-parameter space as a convenient means of representing the possible orientations of the members of the family. It is then essential to verify that points in the order-parameter space are indeed in one-to-one correspondence with members of the family, and that the representation is a continuous one, with infinitesimally different configurations of the object being specified by infinitesimally separated points of the order-parameter space. Conventional coordinates (for example, spherical coordinates to specify the orientation of a unit 3-vector, or Eulerian angles to specify the orientation of a rigid body) often fail one or both of these tests, and their use can seduce one into erroneous representations of the order-parameter space. For the same reasons one must also guard against, for example, representing the orientation of a rigid body by the product of the surface of a 3-sphere (to specify the direction of an axis fixed in the body) with a circle (to specify the orientation of the body in the plane perpendicular to this axis). Here one has not introduced a coordinate system, but the fact is that there is no way of using this particular representation to specify all orientations of the body without running afoul of one or both of the requirements of single-valuedness and continuity.[3]

B. A simple illustration of the topology of defects: Planar spins in two dimensions

Before embarking upon a general analysis of defects, we consider the simple example of a medium of planar spins (example 1) in a two-dimensional physical space—i.e., a field $\mathbf{s}(x, y)$ of unit vectors in the plane. Suppose we are told that $\mathbf{s}(\mathbf{r})$ is continuous everywhere in the plane except, perhaps, at the point P, and, in addition, we are given the explicit form of $\mathbf{s}(\mathbf{r})$ at all points $\mathbf{r}$ farther from P than some distance d. Are there circumstances under which we can conclude that $\mathbf{s}(\mathbf{r})$ is indeed singular at P, without any information about the forbidden region other than continuity? The answer is yes.

Consider any circle centered on P with radius larger than d. The field $\mathbf{s}(\mathbf{r})$ is then known on the circle, and we can easily measure the total angle with respect to some fixed direction through which the vector $\mathbf{s}(\mathbf{r})$ turns as $\mathbf{r}$ traverses the complete circular contour. (Let us traverse the circle in a counterclockwise sense, and count counterclockwise increments in angle as positive, and clockwise increments, as negative.) Since $\mathbf{s}(\mathbf{r})$ is continuous on the circle this angle must be an integral multiple of 2π. The integer n is known as the *winding number*. Configurations of various winding number are illustrated in Fig. 1.

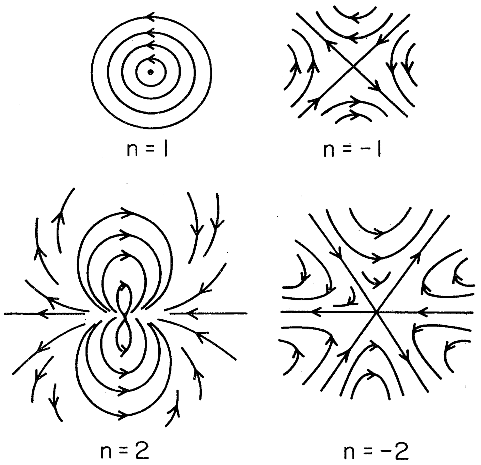

FIG. 1. Point singularities of planar spins in two dimensions with winding numbers ±1 and ±2.

Now let the circle about P around which the winding number is measured shrink continuously down to an infinitesimal circle about P, far inside the forbidden region. Because $\mathbf{s}$ is continuous except at P itself, and because the winding number is discrete (i.e., it can only change discontinuously), we can conclude that the winding number on every circle of the family must continue to have the same value n it had on the original large circle. If n is nonzero, this requires $\mathbf{s}$ to turn through an angle of at least 2π no matter how small the circle becomes; the derivative of $\mathbf{s}$ therefore diverges at the point P, and $\mathbf{s}$ is indeed singular at P.

This is an old and familiar argument, but one should pause to admire it: if there is a singularity with nonzero n at the point P, it leaves its signature on the field arbitrarily far away from P. Its presence, and the value of n itself, can be determined by measurements made as far from P as desired.

Thus n being zero on any encircling contour is a necessary condition for there to be no singularity at P. Is it also sufficient? Obviously not, for we can also manufacture a singularity far inside a circle with winding number 0 by suitably pinching the vector field within a small region, as illustrated in Fig. 2. Such a

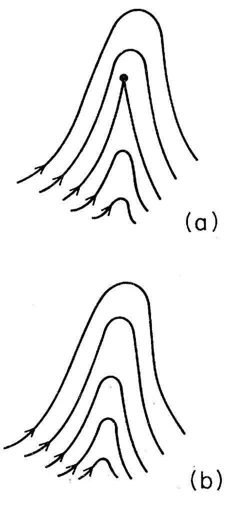

FIG. 2. (a) A planar spin singularity with zero winding number. (b) The removal of the singularity in (a) by purely local alterations in the spin configuration.

[3] Thus the entire surface of the 3-sphere together with that part of the circle specifying a 0° rotation corresponds to a single orientation of the object, violating the requirement that the correspondence be one-to-one.

singularity, however, is distinguished from those required by nonzero n in that it can be smoothed out again without affecting the continuity in the far region.

Contrast this with what must be done to remove a singularity with nonzero n: The winding number must be reduced to zero on every contour surrounding P. Since, however, the winding number is constant for any continuous variation of $\mathbf{s}(\mathbf{r})$, this requires that at some stage in the obliteration of the singularity at P, singularities pass across every contour encircling P, no matter how remote.

The removability of the $n=0$ singularity of Fig. 2 is evident from the figure itself. It can be shown, however that any $n=0$ singularity can be deformed away without doing violence to (i.e., producing singularities in) the order parameter outside of an arbitrarily small neighborhood of P. Indeed, the removal can be achieved without altering in any way the value of s outside of the small neighborhood. The argument that demonstrates this is typical of many that will follow, so we give it in some detail. Many of the concepts and definitions introduced there will be used, without further elaboration, in the considerably more intricate subsequent analysis.

We first give a slightly different way of looking at the winding number. The order-parameter space for planar spins is a circle: any possible value $\mathbf{s}$ might have can be specified by an angle, and this angle, in turn, can be represented as a point on the circular order-parameter manifold (see Fig. 3). *Specifying the order parameter along a contour in real space therefore determines a mapping of that contour into order-parameter space,* in which the point $\mathbf{r}$ on the contour is taken into that point in the order-parameter space that represents $\mathbf{s}(\mathbf{r})$. This simple fact and a few of its straightforward generalizations lie at the heart of the topological theory of defects.

When the contour in real space is closed, the values of the order parameter on the contour determine a mapping of a closed curve (the contour) into a circle (the order-parameter space). The winding number n is just the number of times that the mapping wraps the closed curve around the circle (with appropriate attention to the opposite signs of clockwise vs. counterclockwise wrappings). It is helpful, in visualizing this, to imagine the real-space contour being lifted out of its plane and carefully laid on top of the circular order-parameter manifold, each point $\mathbf{r}$ of the contour being matched against the point of the circle determined by the value of $\mathbf{s}(\mathbf{r})$ (Fig. 4). It should be clear to anyone who has ever wrapped a rubber band around a cylinder, that any mapping with winding number n can be deformed into any other mapping with winding number n, but that two mappings with distinct winding numbers cannot be deformed into one another.[4]

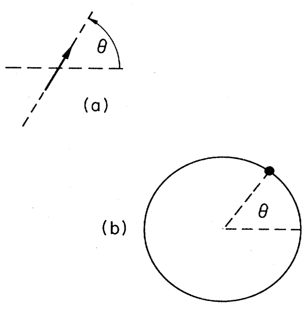

FIG. 3. (a) A planar spin in a given orientation. (b) The representation of that orientation by a point in the order-parameter space.

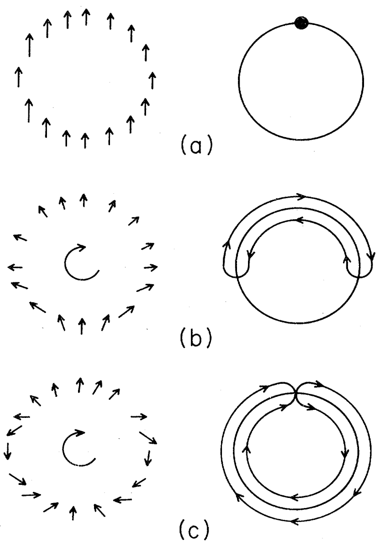

FIG. 4. Spin configurations on circular contours (left) and the maps they determine of the contours into order-parameter space (right). (a) The spin is uniform over the entire contour. The contour is therefore mapped into a single point of order-parameter space. (b) The spin is nonuniform with zero winding number. The resulting map of the contour into order-parameter space can be shrunk to a point. (c) The spin is nonuniform with winding number 2. The resulting map wraps the contour twice around the circular order-parameter space.

This question of whether one mapping is or is not continuously deformable into another, plays a central role in the general theory of defects. Some nomenclature is necessary to avoid phrases as clumsy as the first clause of the preceding sentence. Two mappings of the closed contour into the order-parameter circle (and, more generally, two mappings of a given space into another space) are said to be *homotopic* if one can be continuously deformed into the other. Any explicit construction of such a deformation is called a *homotopy*. Thus, for example, if $f_0(\mathbf{r})$ and $f_1(\mathbf{r})$ are two homotopic mappings taking the points $\mathbf{r}$ of some contour into the circle, then a homotopy would be a continuous one-parameter family of mappings $h_t(r)$ which agreed with f_0 at $t=0$, agreed with f_1 at $t=1$, and was continuous in both r and t.

It is sometimes helpful to think of t as the time, and $h_t(\mathbf{r})$ as a time-dependent continuous deformation of one image of the contour into the other. Alternatively,

[4] Mathematicians devote a lot of strenuous effort to proving this point, because if the only limitation on the order parameter is continuity, one can construct some quite bizarre mappings. Physical ordered media being, at least in this respect, like rubber bands, we shall cheerfully take the result to be obvious. Readers who disapprove of this can extract a proof as a trivial corollary of the much more general results to be derived later on.

one sometimes imagines "spreading the homotopy out in space" by regarding t as an additional spatial variable. Thus a homotopy between two maps of a circular contour into order-parameter space could be regarded as a single map of a cylindrical shell, the lowest circle being mapped by f_0, the highest by f_1, and the circle at height t by h_t. (Evidently this is related to the view of the homotopy as "time-dependent" by the usual physicist's trick of depicting events graphically in time by the introduction of another spatial dimension.)

We now return to our singularity of zero winding number, and show how it can be removed without at all altering (much less tearing) the configuration of the order parameter beyond a small distance d from the singular point P. It is convenient to introduce polar coordinates with P as origin, writing the unit vector field as $\mathbf{s}(r, \theta)$. Note that for fixed r, $\mathbf{s}(r, \theta)$ determines a map of a circle into the order-parameter space. If we regard r as a parameter (on the same footing as the parameter t in a homotopy) then the order-parameter configuration $\mathbf{s}(r, \theta)$ determines in a natural way a one-parameter family of maps of a single circle into the order-parameter space. Since the order parameter is continuous except at $r=0$, so is the one-parameter family of maps. Note that this family gives a homotopy between the maps of a circle into order-parameter space provided by the function $\mathbf{s}(r, \theta)$ at any two distinct values of r. Since homotopic maps have the same winding number, we recover on a rather higher level of abstraction, our earlier observation on the invariance of the winding number.

One can invert this rather banal observation to construct from a homotopy of two maps of a circle, a continuous order-parameter field in real space. This kind of trick permits us to remove the $n=0$ singularity in the following way. First note that a map that takes the entire circle into a single point of order-parameter space clearly has winding number $n=0$. Call such a map $\mathbf{s}_0(\theta)$. Let $\mathbf{s}_1(\theta)$ be the map provided by the order parameter $\mathbf{s}(d, \theta)$ on the circle of radius d about P. Since we are dealing with an $n=0$ singularity, $\mathbf{s}_1$ has zero winding number, and is therefore homotopic to $\mathbf{s}_0$, via a homotopy $\mathbf{s}_t(\theta)$. We can remove the singularity in the order parameter by simply spreading this homotopy out over the space from $r=0$ to $r=d$, thereby creating a singularity-free patch that joins continuously onto the original field at $r=d$. Formally, we define $\overline{\mathbf{s}}(r, \theta)$ by

$$\overline{\mathbf{s}}(r, \theta)=\mathbf{s}_t(\theta), \big|_{t=r/d} \,. \tag{2.9}$$

By construction $\overline{\mathbf{s}}$ agrees with $\mathbf{s}$ at $r=d$, it is continuous for $0 \leq r \leq d$, and it approaches the constant $\mathbf{s}_0$ as r approaches zero. The singularity has been removed.

The $n=0$ singularity was removed by performing corrective surgery in an arbitrarily small neighborhood of the singular point; in contrast, singularities with $n \neq 0$ cannot be removed without tampering with the order parameter at arbitrarily great distances from the singular point. An $n=0$ type of singularity is said to be *removable* or *topologically unstable*. Singularities with $n \neq 0$ are called *topologically stable.*

It must be stressed that topological instability does not necessarily imply physical instability. To investigate the latter it is necessary to know the free energy associated with all of the configurations connecting the singular with the nonsingular one. If it is impossible to get from one to the other without passing through configurations of higher free energy then either, then the topologically unstable singularity may in fact possess a considerable degree of physical metastability.[5] However, a topologically stable singularity cannot be obliterated by a mere fluctuation in the local configuration and is, in this sense, physically stable as well.

The result we have proved about $n=0$ singularities is a special case of the following more general result:

Let $\mathbf{s}(\mathbf{r})$ and $\mathbf{s}'(\mathbf{r})$ be two configurations of the order parameter, both singular only at P and with the same winding number. Then the structure of $\mathbf{s}$ in the neighborhood of P can be replaced by the structure of $\mathbf{s}'$ by purely *local surgery*, i.e., one singular region can be transformed into the other without any alterations outside of a neighborhood of P—there is no *topological barrier* against transforming one singular structure into the other.

To prove this one requires a configuration $\mathbf{s}''(\mathbf{r})$ which agrees with $\mathbf{s}$ when r exceeds a distance d_1, agrees with $\mathbf{s}'$ when r is less than $d_0 < d_1$, and is continuous for r between d_0 and d_1. Now the map of a circle into order-parameter space $\mathbf{s}_1(\theta)$ provided by $\mathbf{s}(d_1, \theta)$ is homotopic to the map $\mathbf{s}_0(\theta)$ provided by $\mathbf{s}'(d_0, \theta)$. If $\mathbf{s}_t(\theta)$ is the homotopy, then the required interpolation between $\mathbf{s}$ and $\mathbf{s}'$ is given by

$$\mathbf{s}''(r, \theta)=\mathbf{s}_t(\theta)\big|_{t=(r-d_0)/(d_1-d_0)} \,, \quad d_0 \leq r \leq d_1 \,. \tag{2.10}$$

Colloquially (and, I hope, more clearly) the core of $\mathbf{s}'$ is fitted into $\mathbf{s}$ by an interpolation provided by spreading the homotopy across the intervening region. This is illustrated in Fig. 5.

I emphasize again that singular configurations with different winding numbers cannot be transformed into one another by local surgery, since such a transformation requires altering the discrete winding number at arbitrarily great distances from the singular point.

We have therefore succeeded in grouping all singularities into classes (indexed by the winding number n) with the property that two singularities in the same class can be deformed into each other by localized alterations in the order parameter, while singularities in distinct classes cannot. Singularities in the same class are said to be *topologically equivalent.*

The establishment of this classification scheme used nothing beyond the simple fact that two mappings of a circle (the contour) into a circle (the order-parameter space) can be deformed into one another continuously

[5] Although the topological method *per se* does not focus strongly on questions of energetics, one can be led to some rather peculiar conclusions by altogether ignoring them (as indicated, for example, in Sec. VIII). Such considerations are beyond the scope of this review, but it is important to bear in mind that in many cases the topological analysis provides only a framework, into which a subsequent study of energetics must be fitted to arrive at a full understanding.

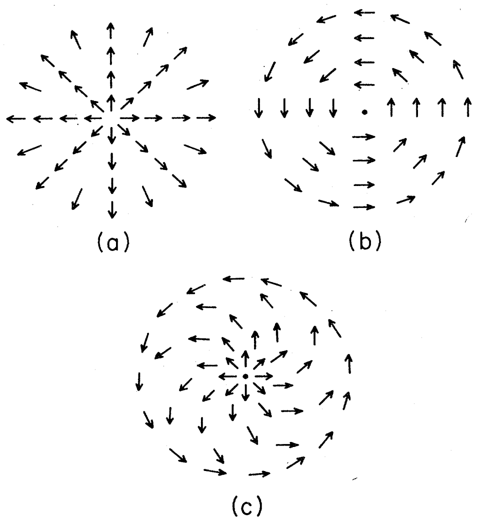

FIG. 5. (a) and (b). Two planar spin configurations with winding numbers +1. (c) Configuration (b) has been altered to coincide with configuration (a) in the core region near the singular point, but remains unaltered away from the core region.

if and only if they have the same winding number.[6] The classification scheme for the singularities was thus entirely determined by a certain topological property of the order-parameter space. We shall find that similar conclusions can be reached for a large class of ordered systems. In many cases, however, the corresponding topological features of the order-parameter space are not nearly as easy to grasp intuitively, and a certain amount of mathematical machinery can be of considerable help in extracting the analogous results.

Before leaving the example of planar spins, we note a few additional useful points of more general validity. Suppose there are two singular points P and Q, with winding numbers n and m. To find the winding number for a contour that encircles both singularities, we exploit the invariance of the winding number under deformations of the contour in the nonsingular region. Thus we can deform the contour into one that encircles P, another that encircles Q, and two pieces that cover the same ground but in opposite directions, to join the pieces around P and Q. The part around P contributes n, the part around Q contributes m, and whatever the contribution from the third piece, the fourth piece makes an equal and opposite contribution. The procedure is quite analogous to the evaluation of contour integrals by the method of residues, and is illustrated in Fig. 6.

Thus if we allow for pairs of defects, a contour with winding number n might contain a pair the sum of whose winding numbers is n, rather than just a single defect with that winding number. However, by precisely the same reasoning as we have followed above, we can argue that a pair of defects can be transformed into a single one with the total net winding number, without requiring the surgery to extend beyond the interior of any contour surrounding the pair. If we extend our notion of equivalence to pairs of defects, then any defect pair is equivalent to a single defect with winding number equal to the sum of the winding numbers of the separate members of the pair (Fig. 7). Thus we have not only classified the defects by the additive group of integers, but we have also found that defects can only combine to give ones characterized by the sum of the characterizing integers. It is the generalization of this group-theoretic description of defects and their combination laws that we shall be constructing in the sections that follow.

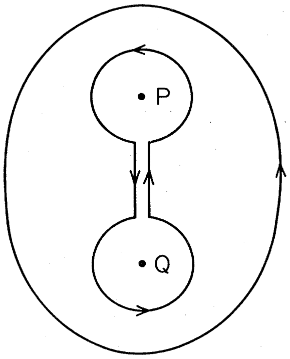

FIG. 6. Two point singularities P and Q and two surrounding contours. The winding number on the inner contour is the sum of the winding numbers determined by P and Q separately. Since the inner contour can be continuously deformed into the outer one, this is also the winding number for the outer contour.

The following point is also worth noting: a special case of the above conclusion is that a pair of defects with winding numbers n and $-n$ is equivalent to a nonsingular configuration. The physical manifestation of that equivalence is that the defects can annihilate one another within a bounded region without the need for any rearrangement of the order-parameter field at large

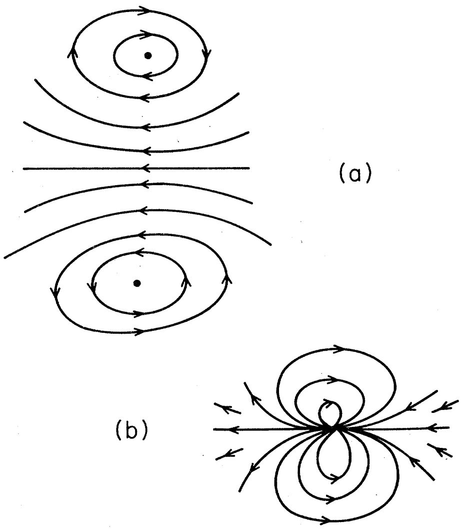

FIG. 7. (a) Two planar spin defects with winding number +1. (b) The topologically equivalent single planar spin defect with winding number +2.

[6] It is an unfortunate feature of this example that closed real-space contours (of importance in all examples) have the same topological structure as the order-parameter space itself, which in general, of course, can have any number of structures.

distances. The converse of this observation suggests a simple mechanism for eliminating a stable defect: Simply bring in from infinity another stable defect with the opposite winding number and allow the two to annihilate when they meet. Note that the process of bringing the antidefect in from infinity does indeed produce singularities at contours arbitrarily far from the first defect, as we saw that it must.

C. An even simpler illustration: Ordinary spins in the plane

To emphasize how intimately the classification scheme depends on the structure of the order-parameter space, suppose we replace the planar spins by ordinary 3-dimensional spin vectors (still keeping the physical space two-dimensional). The order-parameter space is then the surface of a three-dimensional sphere, and a simple argument shows that all defects are unstable. From the topological point of view this amounts to the assertion that any continuous map of a closed loop into the surface of a three dimensional sphere can be continuously deformed into the constant map (or, putting it more vividly, shrunk to a point). The required homotopy is constructed as follows (see also Fig. 8):

(i) Pick any point on the sphere which has a neighborhood about it through which the loop does not pass. (We assume the mapping of the loop into the sphere is not so pathological as to produce a space-filling curve.)

(ii) Punch a small hole in the sphere within that neighborhood.

(iii) Map the surface of the punctured sphere onto a circle in the plane. The image of the loop will be some loop in the interior of the circle.

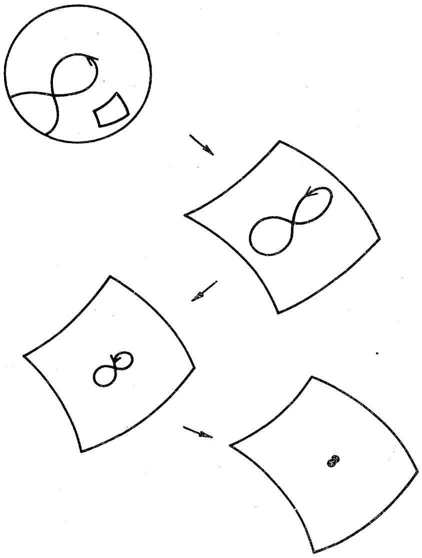

FIG. 8. Procedure for shrinking a loop on the surface of a sphere to a point. Punch a hole (square in the figure) into any part of the sphere where there is no loop, regard the sphere with a hole as a bounded simply connected portion of the plane, and shrink the loop to a point within that region.

(iv) Map the interior of the circle continuously onto the center by simply shrinking the radius to zero. The image of the loop will thereby be taken continuously into the single point at the center of the circle.

The result of these steps is an explicit specification for shrinking the original loop on the sphere continuously to a point. It amounts to simply sliding the image of the loop on the sphere away from the point picked out in (i) until it collapses down to, for example, the antipodal point.

III. THE FUNDAMENTAL GROUP

We wish to generalize the procedure just described for planar spins to arrive at a classification scheme valid for all of the media described in Sec. I. The central feature of that classification scheme emerged from examining the mappings of closed curves in physical space into the order-parameter space. Note, first, that the restriction to a two-dimensional physical space is inessential. The conclusions apply equally well in three dimensiona, provided the singular regions to be studied are not points, but lines.[7] In either case the aim is to associate a discrete invariant with closed contours surrounding the singular point (two dimensions) or line (three dimensions). The invariant is constructed by noting that the values of the order parameter on any such closed contour provide a mapping of a circle[8] into the order-parameter space. Any two contours that can be continuously deformed into each other within the nonsingular region of physical space provide homotopic maps.[9] Since any closed contours that surround the singular point (two dimensions) or line (three dimensions) exactly once can be so deformed into one another, we have the basis for a classification scheme in terms of classes of homotopic maps of closed loops into the order parameter space.

In this section we shall examine the general structure of classes of homotopic maps of closed loops into any order-parameter space R. Our aim is to show that these homotopy classes can, in fact, be given a group structure, and that the combination law for that group [known as the *fundamental group* or *first homotopy group* of R, almost invariably denoted by the symbol $\pi_1(R)$] is closely related to the combination law for physical point (two dimensions) or line (three dimensions) defects.

Since maps of real-space loops or circles into order-parameter space give closed curves (or loops) in

[7] The study of point defects in 3-dimensional space is taken up in Sec. VII.

[8] Regard the circle as the parameter space for the contour. The parametrization of the contour maps the circle onto the contour, and the values of the order parameter map the contour into order-parameter space. The combination of the two maps the circle into the order-parameter space.

[9] The deformation of the contours is given by a one-parameter family of contours starting with the first and ending with the second. Since each contour in the family provides a map of the circle into order-parameter space, we have a continuous one-parameter family of maps of the circle into order-parameter space, which is the required homotopy.

order-parameter space itself, we shall usually speak simply of loops in order-parameter space or homotopy classes of loops in order-parameter space [rather than, more clumsily, "maps of circles (loops) into order-parameter space" or "homotopy classes of maps of circles (loops) into order-parameter space"].[10]

It turns out that the group structure we need can in general be imposed only on the set of homotopy classes of loops that have a single point of order-parameter space in common. This is done in part A. In part B it is shown that the structure of the resulting group does not depend on the choice of the special point, so that the group does indeed reflect the structure of the order-parameter space as a whole. In part C we examine the relevance of the fundamental group for the classification of defects, where loops of interest need not have any points of order-parameter space in common, and in D we similarly examine the relation of the fundamental group to the combination law for defects. We conclude the section with a theorem and definition that are used in the two sections that follow on how actually to compute the fundamental group.

Throughout this section (and throughout the entire paper) we shall take the order-parameter space R to be connected, in the sense that any two points can be joined by a continuous path lying entirely in R. Putting this more formally, we shall assume that if x and y are in R, then there is a continuous map f of the interval $[0, 1]$ into R such that $f(0)=x$ and $f(1)=y$. This connectedness assumption is, in fact, no restriction, since any closed real-space path that steers clear of order-parameter singularities can only give a mapping into a single connected component of the order-parameter space. Therefore if there should be more than one connected component to R (as there often is) one simply has distinct classification schemes for each separate component. Putting the point differently, unless the order parameter is singular on a subspace large enough to cut physical space into disconnected pieces, one can produce a classification of order-parameter fields according to the connected components of the order-parameter space in which they take their values. For analytical purposes one can then regard the classes associated with distinct components as independent physical systems.[11]

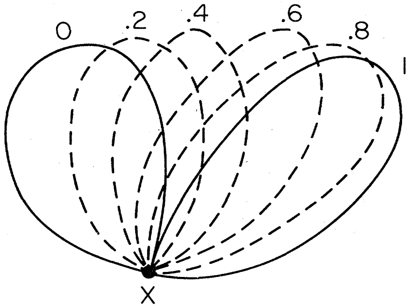

FIG. 9. Successive loops at x in a homotopy. The loops are labeled by values of the parameter t in the homotopy. The initial and final homotopic loops are solid, the intermediate ones, dashed.

A. The fundamental group at a point

Before we can describe the fundamental group of the order-parameter space as a whole, we must describe the fundamental group associated with the order-parameter space R together with any one of its points x, called the base point. The fundamental groups associated with different base points will turn out to be isomorphic to one another, and the fundamental group of R itself is then defined to be that abstract group of which the fundamental groups at the various base points are isomorphic copies.

1. Loops at x

Consider all closed continuous directed curves in R that pass through the point x. We call them loops in R at x. They can be described in terms of continuous maps f of the real interval $0 \leq z \leq 1$ into R, with $f(0)=f(1)=x$. The sense of the loop will be indicated in figures, when necessary, by an arrow indicated the direction of increasing z. Note that the words "curve" or "loop" are being used somewhat more generally than usual, since $f(z)$ identically equal to x is such a mapping. The loops can have such degeneracies, provided only that they are continuous and start and end at x.

2. Homotopies based at x

We introduce a restricted notion of homotopy, saying that two loops f and g are *homotopic at x*, if there is a continuous family of loops, all passing through x, such that f and g are members of the family (see Fig. 9). Formally, there must be a family $h_t(z)$ of mappings of $[0, 1]$ into R such that h is continuous in both t and z and

$$\begin{aligned} &\text{(i)}\ h_0=f, \\ &\text{(ii)}\ h_1=g, \qquad\qquad (3.1)\\ &\text{(iii)}\ h_t(0)=h_t(1)=x,\ \text{for all } t. \end{aligned}$$

Except for the additional restriction (iii) this is the definition of homotopy we used in Sec. II.[12] *Based*

[10] A cautionary remark: One can sometimes become confused by simply forgetting whether a particular loop is a loop in physical space or a loop in the order-parameter space. Since a real-space loop and a given order-parameter field $f(\mathbf{r})$ determine an order-parameter-space loop, it is easy carelessly to identify the two loops in one's thinking. Such carelessness should be avoided: Always know which space which loops lie in.

[11] If one is interested in defects that actually do divide physical space in pieces (line defects in two dimensions or plane defects in three) then the connectedness of the order-parameter space does play a role, but the resulting structure is trivial: Such defects are topologically stable if and only if the order parameters on either side are in disconnected pieces of the order-parameter space. This conclusion is sometimes obscured by being stated in terms of the so-called *zeroth homotopy group*, $\pi_0(R)$, which is simply the set of disconnected pieces of R, which can be given a group structure for many order parameters of interest.

[12] If one wishes to emphasize that a homotopy need not be tied to the base point, one refers to it as a *free homotopy* and to the homotopic maps as *freely homotopic*.

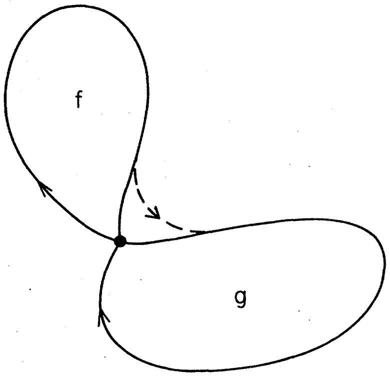

FIG. 10. The loop product $f \circ g$. First f is traversed, then g. Loops homotopic to $f \circ g$ will also start and finish at the base point, but need not return to the base point in midjourney, as indicated by the dashed interpolation.

homotopy is a more restrictive relation, since all members of the family giving the homotopy are required to be tied down at the base point x.

3. The product of two loops

We define the product, $f \circ g$, of two loops f and g as the loop obtained by first traversing f, and then g (see Fig. 10). Formally,

$$f \circ g(z) = f(2z), \quad 0 \leq z \leq \tfrac{1}{2};$$
$$= g(2z-1), \quad \tfrac{1}{2} \leq z \leq 1. \tag{3.2}$$

This composition law underlies the binary operation of the group we are constructing, and we shall refer to it repeatedly in proving many elementary theorems. However, Eq. (3.2) cannot be used to impose a group structure on the set of individual loops at x. One problem, for example, is that if we regard as distinct two loops specified by distinct maps f, then the combination law given by Eq. (3.2) is not even associative, for although $f \circ (g \circ k)$ gives the same curve in R as $(f \circ g) \circ k$, the maps are parametrized differently. [The map $f \circ (g \circ k)$ gives f for the first half of the interval, g for the third quarter, and k for the fourth; the map $(f \circ g) \circ k$ gives f for the first quarter, g for the second quarter, and k for the last half.] A more serious deficiency is that the product of two loops necessarily passes through the base point x at $z = \frac{1}{2}$, which leads to severe restrictions on how (if at all) an arbitrary loop can be represented as a product of other loops.

The way out of these (and similar) difficulties is to extend the notion of product from pairs of loops, to pairs of classes of mutually homotopic loops. (Since it is classes of homotopic loops, rather than individual loops, that characterize defects, this extension is a very natural one for our purposes.)

4. The product of homotopy classes of loops

If f is a loop, we define $[f]$ to be the set of all loops homotopic (at x) to f. Evidently we can divide all possible loops at x into distinct *classes of mutually homotopic loops*. We call any particular loop in a class a *representative* of that class. We shall denote classes of loops either by Greek letters, or in terms of a particular representative, with the bracket notation. Thus (to illustrate the nomenclature) if α is a class of loops and f and g both belong to α, then $\alpha = [f] = [g]$.

We wish to define the product of two classes by

$$[f] \circ [g] = [f \circ g]. \tag{3.3}$$

This definition makes sense, provided the homotopy class of the product $f \circ g$ does not depend on which particular representatives f and g we choose to form that product. This in turn follows from the trivial theorem that if $f \sim f'$ and $g \sim g'$, then $f \circ g \sim f' \circ g'$. [We take the symbol ~ to mean "is homotopic to at x," throughout the ensuing discussion. We shall also term a result *trivial* if (a) it follows from the underlying definitions without any trickery or ingenuity and (b) a written specification of how it follows runs the danger of suggesting that it is nontrivial.]

The definition of class multiplication [Eq. (3.3)] has many virtues. By going from the product of loops to the product of classes of homotopic loops, we have freed the definition of product from the particular parametrization used to represent the loop. (Trivial theorem: Two loops that differ only in their parametrization are indeed in the same homotopy class.) More importantly, the product loop no longer reveals telltale signs of its origins, for although the product of any particular f and g is tied to x at $z = \frac{1}{2}$, a general representative of $[f \circ g]$ certainly is not (Fig. 11). The greatest virtue, however, is that under the combination law [Eq. (3.3)] the homotopy classes of loops at x form a group.

5. The fundamental group at x, $\pi_1\ (R, x)$

To verify that the homotopy classes of loops form a group under Eq. (3.3) we must verify:

(i) Associative law: $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$. To prove this simply take any three maps f, g, and k from each of the three classes and note that (as observed above) $(f \circ g) \circ k$ differs from $f \circ (g \circ k)$ only in parametrization. Hence the two belong to the same homotopy class. Since the homotopy class of products is independent of

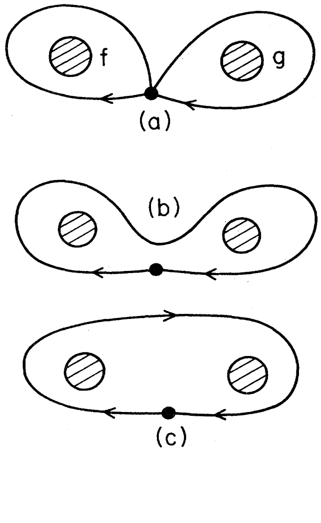

FIG. 11. (a) Two based loops f and g drawn in a plane with two holes in it. The figure also represents the loop product $f \circ g$. (b) and (c) Two other based loops homotopic at the base point to $f \circ g$.

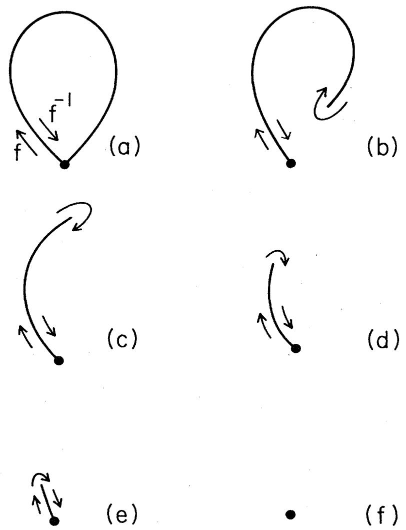

FIG. 12. (a) The loop product $f\circ f^{-1}$ is given by making a round trip from the basepoint in the clockwise sense, followed by another round trip in the counterclockwise sense. (b)–(f) Successive stages in the homotopy between $f\circ f^{-1}$ and the constant map. The loops in the homotopy (which are all degenerate loops, being mere lines) go from the base point out to the end of the line and back.

the representatives chosen to construct the classes, this establishes the associative law.

(ii) Identity. Let $e(z)$ be the constant map: $e(z)\equiv x$. (The corresponding loop is a single point.) The homotopy class $[e]$ is then the set of all loops at x which can be shrunk down to the single point x (always with at least one point tied to x throughout the shrinking). It is trivial to establish that

$$[e]\circ[f]=[f]\circ[e]=[f], \tag{3.4}$$

so that $\epsilon=[e]$ is the group-theoretic identity.

(iii) Inverse. Let f^{-1} be the loop given by traversing the loop f in the opposite direction:

$$f^{-1}(z)=f(1-z),\quad 0\le z\le 1. \tag{3.5}$$

If f and g are homotopic at x then (trivially) so are f^{-1} and g^{-1}. We can therefore associate with any homotopy class α, a class α^{-1}, defined by

$$[f]^{-1}=[f^{-1}].$$

We now prove that $\alpha^{-1}\circ\alpha=\epsilon$. Note first that since we can represent the product $\alpha^{-1}\circ\alpha$ by the product of any mapping in α with any mapping in α^{-1}, we can, in particular, choose the mapping representing α^{-1} to be the inverse of the mapping representing α. We need therefore only show that the loop $f^{-1}\circ f$ is homotopic to e (i.e., can be shrunk to a point at x) for any f. The required homotopy is pictured in Fig. 12. Analytically, the homotopy can be taken to be

$$\begin{aligned} h_t(z)&=f(2zt),\quad 0\le z\le \tfrac{1}{2};\\ &=f(2t(1-z)),\quad \tfrac{1}{2}\le z\le 1. \end{aligned} \tag{3.6}$$

This establishes that the classes of homotopic loops at x form a group. The group is called $\pi_1(R,x)$, and is known as the *fundamental group* of R at x (the "at x" usually being omitted, as a consequence of the considerations in part B below). The subscript 1 anticipates the fact that there are other groups (associated with the mappings of S_n—the surface of a sphere in Euclidean $n+1$ space—into R) called $\pi_n(R,x)$. The fundamental group is sometimes called the *first homotopy group* (and π_n is then called the nth homotopy group.)

The group π_1, need not be Abelian, as the last of the following examples reveals.

6. A few simple examples

(a) The circle. The homotopy classes of maps into the circle with base point x (like the homotopy classes of unbased maps) are specified by the winding number n. (Whether there is a base point or not is immaterial for all maps except those in the class of e, since all maps not homotopic to a constant take on all values in the circle.) The product of classes given by n and m, has winding number $n+m$ (where n and m can be positive or negative). The fundamental group is therefore Abelian and isomorphic to the additive group of the integers (known as Z). Evidently the group structure does not depend on the choice of base point. One writes

$$\pi_1(S_1)=Z. \tag{3.7}$$

(b) The surface of a sphere. Any mapping of a loop into a sphere can be shrunk to a point (as described

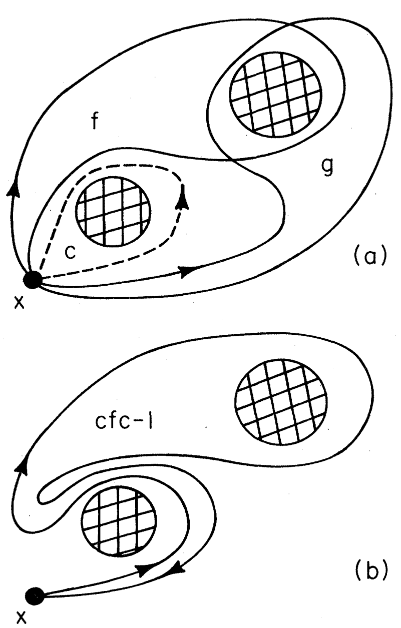

FIG. 13. The figure-eight space: a plane with two holes in it. (a) Two based loops. Although they are freely homotopic, the loops are not homotopic at the base point since one cannot be deformed into the other without becoming detached from the base point at some stage of the homotopy. (b) A loop that is homotopic at x to $c\circ f\circ c^{-1}$ [where c is the dashed loop in (a)] and that also is clearly homotopic at x to g. Since f and g are not homotopic at x, the loop products $c\circ f$ and $f\circ c$ cannot be homotopic at x.

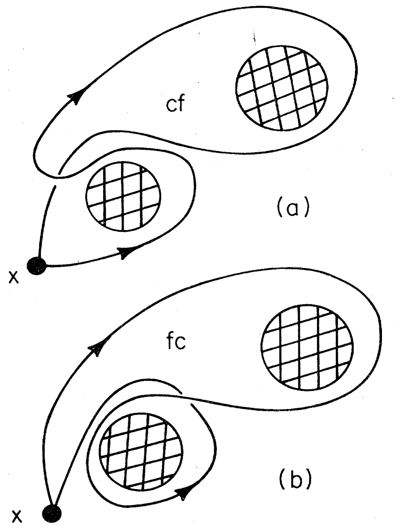

FIG. 14. The loop products (a) $c\circ f$ and (b) $f\circ c$. They are not homotopic at x, but are freely homotopic.

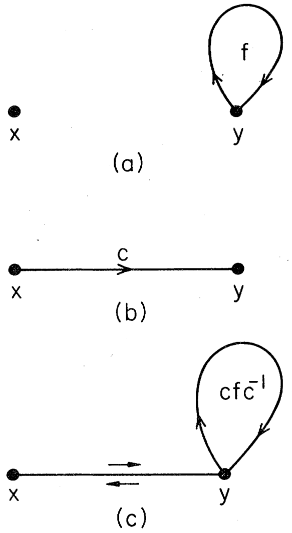

FIG. 15. (a) A loop f at y and another point x. (b) A path c joining x to y. (c) The loop $c\circ f\circ c^{-1}$ at x.

earlier). Thus the fundamental group of the surface of a sphere (at any base point) consists only of the identity. Using an additive group notation, one usually writes

$$\pi_1(S_2)=0\,. \tag{3.8}$$

(c) The figure-eight space. This is important to keep in mind as one of the simplest examples of a space with a non-Abelian fundamental group. Consider the space R consisting of a plane disc with two holes punched in it. Two loops, f and g, are shown in Fig. 13. Because they pass on opposite sides of the lower hole, there is no way to deform the loop f into the loop g while holding a point fixed at x (though, of course, the two loops are trivially freely homotopic). Note, though, that if c is the dashed loop in the figure, then we do have

$$c\circ f\circ c^{-1}\sim g\,. \tag{3.9}$$

Since f and g are not homotopic at x, it follows from this that $c\circ f$ cannot be homotopic to $f\circ c$, i.e., the fundamental group of the space is non-Abelian. Loops in the classes of $c\circ f$ and $f\circ c$ are shown in Fig. 14. Although they are trivially freely homotopic, it is impossible to construct a homotopy that stays attached to the base point at all t.

B. The fundamental group of a connected space

1. The isomorphism between fundamental groups based at different points

We first show that if x and y are any two points of R, then we can associate with any path c connecting x and y a natural group-theoretic *path isomorphism* between $\pi_1(R,x)$ and $\pi_1(R,y)$. The structure of this isomorphism may depend on the choice of path.

We first give a formal definition of a *path* in R which is the obviously generalization of our earlier definition of a loop: a path c between x and y is determined by a continuous mapping $c(z)$ of the interval $0\leq z\leq 1$ into R, such that $c(0)=x$ and $c(1)=y$. Thus a loop is a path that starts and ends at the same point.

If a path ends where another path begins then we can define the product of the two paths as the single path given by traversing first one, then the other. Should the paths happen to be loops, this reduces to our earlier definition of the loop product.

If c is a path from x to y and f is a loop at y, then cfc^{-1} is a loop at x, as shown in Fig. 15.[13] By c^{-1} we mean the path c traversed in the opposite sense. More formally

$$c^{-1}(z)=c(1-z)\,,\quad 0\leq z\leq 1\,. \tag{3.10}$$

If f is homotopic to g at y, then cfc^{-1} will be homotopic to cgc^{-1} at x, for if h_t is the homotopy between f and g at y, then ch_tc^{-1} will be the required homotopy at x. One can therefore associate with the path c a mapping between the homotopy classes of loops at x and y:

$$c([f])=[cfc^{-1}]\,, \tag{3.11}$$

where the correspondence is independent of the choice of representative f. Note that every class of loops at x is the image of a class of loops at y (for $[g]$ is the image under c of $[c^{-1}gc]$). Note also that if $c([f_1])=c([f_2])$ then $[f_1]=[f_2]$ {as a consequence of the fact that $[f_i]=c^{-1}(c([f_i]))$.} Thus the correspondence (3.11) is a one-to-one correspondence from the classes of loops at y onto the classes of loops at x.

Finally, it follows from the trivial result

$$(cfc^{-1})(cgc^{-1})\sim c(fg)c^{-1} \tag{3.12}$$

that the mapping (3.11) preserves the algebraic structure of homotopy class multiplication:

[13] From this point onward we simplify the notation for loop or path products from $f\circ g$ to simply fg. We shall reinsert the little circle should there be a danger of confusing the loop or path product with some other kind of product.

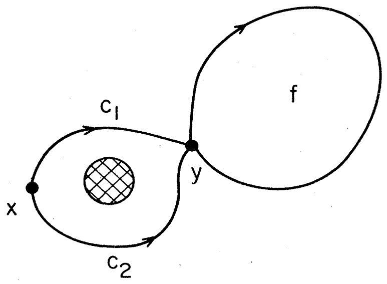

FIG. 16. A loop f at y and two paths c_1 and c_2 from x to y. The loop at x given by $c_1 \circ c_2^{-1}$ surrounds a hole, and therefore c_1 and c_2 can give distinct path isomorphisms between $\pi_1(R,y)$ and $\pi_1(R,x)$.

$$c(\alpha)c(\beta)=c(\alpha\beta)\,. \tag{3.13}$$

Taken together, these results establish that $\alpha \to c(\alpha)$ is in fact an isomorphism between $\pi_1(R,x)$ and $\pi_1(R,y)$. *There is thus a single abstract group, $\pi_1(R)$, of which the based fundamental groups are isomorphic copies.* This abstract group is known as the *fundamental group* of R.

The existence of a single abstract fundamental group characterizing the entire order-parameter space is of central importance in establishing the classification scheme for unbased mappings of loops into R, i.e., for establishing the classes of freely homotopic loops in R. We shall show that if the fundamental group of R is Abelian then there is simply a one-to-one correspondence between classes of freely homotopic loops and elements of the fundamental group; if, however, the fundamental group is non-Abelian, then the correspondence is between classes of freely homotopic loops and conjugacy classes of the fundamental group.[14] To establish these conclusions (which translate immediately into conclusions on the classification of line defects) we must examine the relation between the various isomorphisms $[f]\to c([f])$ associated with different choices of the path c linking x and y.

2. Uniqueness (or lack of uniqueness) of the path isomorphisms between based fundamental groups

The path isomorphism is independent of the choice of path, if and only if the fundamental group of the space R is Abelian. To see this, suppose first that the path isomorphism depends on choice of path. There are then two paths from x to y, c_1 and c_2, and a class of loops α in $\pi_1(R,y)$, such that

$$c_1[\alpha]\neq c_2[\alpha]\,. \tag{3.14}$$

If f is any y-based loop in α, then Eq. (3.14) asserts that the x-based loops $c_1fc_1^{-1}$ and $c_2fc_2^{-1}$ are not homotopic at x (Fig. 16). It follows from this that the y-based loop f is not homotopic to the y-based loop $(c_1^{-1}c_2)f(c_1^{-1}c_2)^{-1}$. But $c_1^{-1}c_2$ is itself a y-based loop—call it g. We then conclude that $f\neq gfg^{-1}$ at y, or, equivalently, that $fg\neq gf$ at y. Thus $\pi_1(R,y)$ and therefore $\pi_1(R)$ itself is non-Abelian.

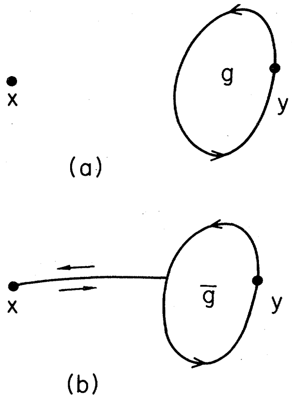

FIG. 17. Two loops at y. (a) The loop g. (b) The loop $\bar{g}$ that differs from g only by the addition of an intermediate round trip to x.

Conversely, suppose that $\pi_1(R)$ and hence $\pi_1(R,y)$ is non-Abelian. Then there are two y-based loops, f and g, with $fg\neq gf$ or, equivalently, with $f\neq gfg^{-1}$. We now replace g by the homotopic (at y) loop $\bar{g}$ which follows the same course as g, except that at a certain moment it makes a detour over to x and then retraces that detour, as shown in Fig. 17. We can decompose $\bar{g}$ into two paths c_1 and c_2 from x to y, as shown in Fig. 18:

$$\bar{g}=c_2^{-1}c_1\,. \tag{3.15}$$

The statement that $f\neq gfg^{-1}$ now becomes the statement that $c_1fc_1^{-1}\neq c_2fc_2^{-1}$. Thus the path isomorphism given by c_1 is not the same as that given by c_2.

Thus if $\pi_1(R)$ is Abelian there is a unique natural isomorphism between the $\pi_1(R,x)$ for different base points x: namely the path isomorphism c, which does not depend on the particular choice of path. If, on the other hand, $\pi_1(R)$ is non-Abelian, then the various path isomorphisms between $\pi_1(R,y)$ and $\pi_1(R,x)$ can differ, but only by an inner automorphism[15] of $\pi_1(R,x)$. For suppose c_1 and c_2 give two distinct path isomorphisms. Then $k=c_1c_2^{-1}$ is a loop at x, and the two mappings of $\pi_1(R,y)$ onto $\pi_1(R,x)$ are related by:

$$c_1fc_1^{-1}\sim k(c_2fc_2^{-1})k^{-1}\,. \tag{3.16}$$

Thus the two isomorphic images of $\pi_1(R,y)$ differ by the inner automorphism:

$$[\alpha]\to[k]\circ[\alpha]\circ[k]^{-1} \tag{3.17}$$

of $\pi_1(R,x)$.

Note that given any inner automorphism of $\pi_1(R,x)$, we can take a loop from the class $[k]$ generating the inner automorphism, break it up into two paths from x to y (as in Figs. 17 and 18), and use these two paths to generate path isomorphisms between $\pi_1(R,y)$ and $\pi_1(R,x)$ that differ by the given inner automorphism.

Now the conjugacy classes of a group are invariant under inner automorphisms.[16] Thus the path isomor-

[14] Readers whose group theory is a bit rusty might consult the summary of basic group-theoretic terms in Appendix B.

[15] See the group-theoretic glossary in Appendix B.

[16] Proof: If a is an element of the group G, then the conjugacy class of a is the set of all elements of G of the form bab^{-1} for arbitrary b in G. Under the inner automorphism $g\to cgc^{-1}$ for fixed c the members of conjugacy classes can be permuted within each class, but cannot be moved from one class to another, since $bab^{-1}\to c(bab^{-1})c^{-1}=(cb)a(cb)^{-1}$.

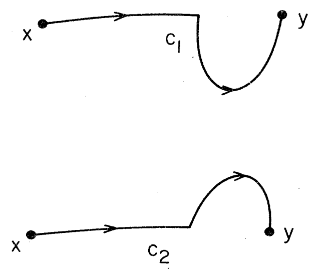

FIG. 18. The decomposition of the loop $\bar{g}$ of Fig. 17(b) into paths c_1 and c_2 from x to y, with $\bar{g}=c_2^{-1}\circ c_1$.

phisms establish a unique (i.e., path-independent correspondence between the conjugacy classes of the based fundamental groups, but not between the elements of the groups themselves, unless the fundamental group is Abelian (in which case the conjugacy classes consist of single elements).

C. The fundamental group and the classes of freely homotopic loops

As in the example of Sec. II, quite generally we can test the nature of a line defect[17] by examining the order parameter on a real-space closed contour surrounding the line. For a given configuration of the order-parameter field, such a contour yields a loop in the order-parameter space R which is just the image of the contour in order-parameter space determined by the values of the order parameter along the contour.

As the real-space contour is slid, shrunk, or otherwise deformed in the region free of order-parameter singularities, one generates a family of mutually homotopic loops in order-parameter space. In general there is no reason why all loops in this family should share a single common point in order-parameter space.[18] A line defect is therefore characterized by a set of loops equivalent under free homotopy in the order-parameter space.

When homotopies are released from the base point then loops f and g at x representing elements from the same conjugacy class of $\pi_1(R,x)$ may be freely homotopic even if the elements themselves are distinct. For if f and g represent elements from the same conjugacy class then there must be a loop b at x with

$$f\sim bgb^{-1}. \tag{3.18}$$

The required homotopy between f and g is given by retracting b and b^{-1} back into g, as in Fig. 19.

Conversely, two loops at x are freely homotopic only if they represent elements from the same conjugacy class of $\pi_1(R,x)$. This follows from a somewhat more general result:

Let f and g be two loops (which need not have a common point). The loops are freely homotopic if and only if there is a path c connecting a point x of f to a point y of g, with f homotopic at x to cgc^{-1}. (Should x and y be the same, c becomes a loop b and we are back to the more restricted case.)

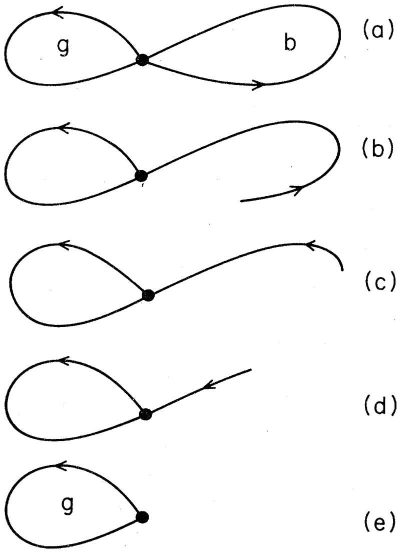

FIG. 19. (a) The loop $b\circ g\circ b^{-1}$ which first traverses b, then g, then b again in the reverse order. (b)–(d) Intermediate loops in a free homotopy between $b\circ g\circ b^{-1}$ and g itself. Each loop starts at the loose end, goes to the base point along what remains of b, traverses g, and then returns along the fragment of b to the starting point. (e) The loop g.

Now g is freely homotopic to cgc^{-1}, the free homtopy being simply a continuous retraction of the round trip from y to x along, c, back into the single point y (as illustrated in Fig. 19 in the special case where c is a loop b). Thus the existence of c with $f\sim cgc^{-1}$ at x, insures the free homotopy of f and g.

Conversely, if f and g are freely homotopic, let h be the homotopy, so that $h_t(z)$ is a continuous family of loops with $h_0=f$, $h_1=g$. As t ranges from 0 to 1, the points $h_t(0)$ trace out a path c connecting f and g [Fig. 20(a)]. But f is homotopic at the point $f(0)$ to the loop cgc^{-1}, the homotopy k_t being given by the free homotopy h_t embellished by the "umbilical cord" provided by the segment of c connecting $h_t(0)$ with $h_0(0)=f(0)$, as shown by Fig. 20(b).

This result can be stated in terms of path isomorphisms: a loop f at x is freely homotopic to a loop g at y if and only if there is a path isomorphism c taking the homotopy class $[f]$ of $\pi_1(R,x)$ into the homotopy class $[g]$ of $\pi_1(R,y)$. The set of all such path isomorphisms establishes a unique connection between the conjugacy classes of the two based fundamental groups, but can arbitrarily rearrange the elements within a given conjugacy class, depending on the path. Therefore classes of freely homotopic loops in R can be labeled by the conjugacy classes of $\pi_1(R)$. To construct such a labeling scheme single out any point x of R for reference, and assign to any loop f in R the (unique) conjugacy class of $\pi_1(R,x)$ with which f is identified by the path isomorphisms linking $\pi_1(R,y)$ with $\pi_1(R,x)$, where y is any point of f.

[17] From this point on I shall stop reminding the reader that, except for such intrinsically three-dimensional effects as entanglement, what is said about line defects in a three-dimensional space also holds for point defects in two dimensions.

[18] This complication is barely noticeable in the case we examined in Sec. II, since all but the trivial class of maps of a contour into a circular order-parameter space cover all points of the order-parameter space and therefore necessarily have common points.

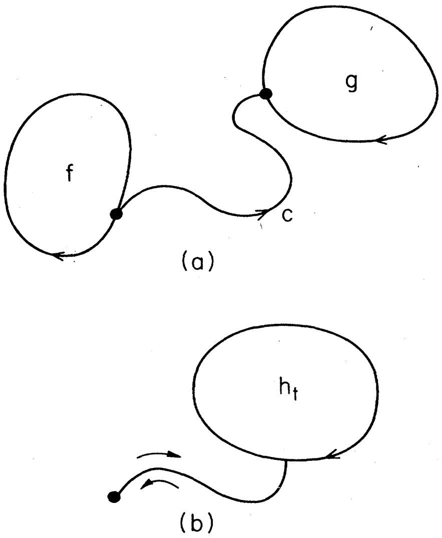

FIG. 20. (a) Two freely homotopic loops f and g. The heavy dots are $f(0)$ and $g(0)$. The path c connecting $f(0)$ to $g(0)$ is traced by the starting points $h_t(0)$ of the loops of the free homotopy between f and g. (b) A stage in the homotopy at $f(0)$ between f and $c \circ g \circ c^{-1}$.

When the fundamental group of an order-parameter space is Abelian the conjugacy classes consist of single group elements (since $aga^{-1}=g$, if a and g commute). Thus any contour encircling a line defect determines a unique element of the fundamental group of the order-parameter space (just as, in the case of planar spins, a unique winding number was determined). Line defects are thus characterized by the members of the fundamental group. Two defects characterized by distinct members of the fundamental group cannot be transformed one into the other by local surgery for precisely the same reasons as in the case of planar spins. Furthermore if two line defects *are* characterized by the same element of the fundamental group, then one can be given the core of the other by purely local surgery.[19] Thus (in the Abelian case) line defects are topologically equivalent if and only if they are characterized by the same elements of the fundamental group of the order-parameter space.

The situation in the case of a non-Abelian fundamental group is somewhat more intricate, and considerably more intriguing. We defer a discussion to Sec. VI, where all the peculiarities of media with non-Abelian fundamental groups will be examined together.

D. The fundamental group and the combination of line defects

The discussion at the end of Sec. II.B on the combination of line defects in the planar spin system applies directly to the general case. Suppose there are two (nonintersecting) line defects P and Q. A real-space contour that surrounds both lines will give a loop in order-parameter space that is freely homotopic to the product of the loops determined by contours that surround each line separately (see Fig. 21).

Since freely homotopic classes of loops in order-parameter space are associated with conjugacy classes of the fundamental group, we conclude that a single defect equivalent to the combined lines can only be characterized by conjugacy classes of the fundamental group whose members are the products of members of the classes characterizing the original pair of defects. The reader is urged not to ponder the preceding sentence at this stage; we shall return to the point in the discussion of non-Abelian fundamental groups in Sec. VI. If the fundamental group is Abelian, the combination law is simplified by the fact that conjugacy classes consist of single group elements, and the product of conjugacy classes is the unique (and order-independent) product of those elements. We therefore have a simple generalization of the law that winding numbers add when defects are combined, for media with Abelian fundamental groups.

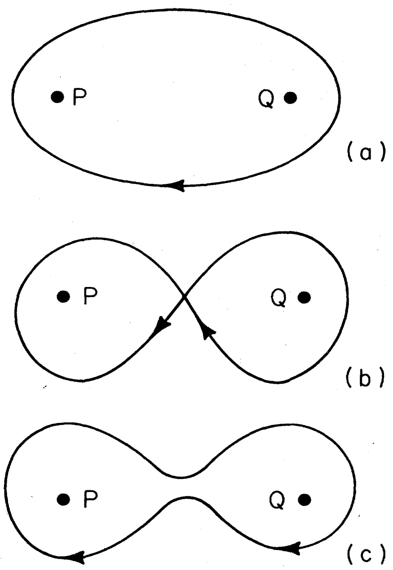

FIG. 21. (a) Two defects P and Q and an encircling contour. (b) Two contours with a common point that encircle P and Q separately. The loop in order-parameter space determined by the values of the order parameter along the two successive contours in (b) is in the homotopy class of the product of the homotopy classes determined by P and Q separately. This loop is freely homotopic to the loop in order-parameter space determined by the values of the order parameter along the contour in (a). To construct the free homotopy deform one contour into the other via intermediate contours such as that in (c). The homotopy is provided by the family of loops in order-parameter space determined by the values of the order parameter on the family of real-space intermediate contours.

[19] The reader is invited to construct the argument. It is a slightly more complicated than the corresponding argument in Sec. II, but only because of the jump from two spatial dimensions to three, and not because of the jump from winding numbers to an arbitrary (but Abelian) fundamental group. Thus the homotopies between loops surrounding the two line defects must now be appropriately stacked together along the direction of the line, to provide the desired interpolation between the outer region of one type and the core region of the other. This is more easily done than said.

If the fundamental group of the order-parameter space is Abelian, then when two defects characterized by elements α and β of the fundamental group combine, the new defect is characterized by the element $\alpha \circ \beta = \beta \circ \alpha$. Note that the process of combination envisioned here is via the same local surgery by which a trivial defect can be removed, or a nontrivial one given the core of another characterized by the same homotopy class. Given any distance r, no matter how small, we can continuously deform the medium so that the two line defects lie inside a cylinder of radius r. Circular contours on the surface of the cylinder will then give maps in the class $\alpha \circ \beta$, so that homotopies can be used to replace the configuration within the cylinder by any single line defect in the class $\alpha \circ \beta$, without in any way altering the configuration outside of the cylinder. In contrast, two defects α and β cannot be converted to one that is not homotopic to the product (in the Abelian case) without having to alter the medium at arbitrarily large distances from the defects.

Thus we conclude that in media with Abelian fundamental groups, the rules for the combination of line defects are given precisely by the multiplication table for the fundamental group. If the fundamental group is non-Abelian, the combination laws are given by the class multiplication table, a state of affairs to be elaborated upon in Sec. VI.

We have therefore reduced the problem of classifying and confining line defects to the problem of computing the fundamental group of the order-parameter space. This is taken up in the two sections that follow.

We conclude this section with a definition and a theorem.

Definition: A connected space R is said to be *simply connected* if its fundamental group contains only the identity [a property usually written in the notation of additive groups as $\pi_1(R)=0$]. In more down-to-earth terms, a space is simply connected if any loop in the space can be shrunk continuously to a point.

Theorem: Let R be the *product* of two spaces R_1 and R_2. By this we mean that R consists of pairs of points from R_1 and R_2 with a notion of continuity given by the stipulation that a sequence (x_n, y_n) in R converges to (x, y) if and only if x_n converges to x in R_1 and y_n converges to y in R_2. Thus the real plane is the product of the real line with itself, a cylinder is the product of a circle with a line, etc. More precisely, such a product space is called a *topological product* as a reminder that the property inherited by the pairs from the two spaces is that of continuity (rather than, for example, some algebraic structure). If R is the topological product of R_1 and R_2 then the fundamental group of R is simply the group-theoretic direct product of the fundamental groups of R_1 and R_2:

$$\pi_1(R_1 \times R_2) = \pi_1(R_1) \times \pi_1(R_2) . \tag{3.19}$$

The proof of this follows trivially from the equally trivial[20] result that two loops are homotopic in R if and only if their projections in R_1 and R_2 are independently homotopic in R_1 and R_2. [The projection in R_1 of the loop $(x(z), y(z))$ is the loop $x(z)$, etc.]

IV. GROUP-THEORETIC STRUCTURE OF THE ORDER-PARAMETER SPACE

In the preceding section we saw that a group could be associated with an order-parameter space—its fundamental group—whose algebraic structure was intimately related to the behavior of line defects in the ordered medium. In this section we shall make further use of the language and elementary theorems of group theory to describe ordered media, but in quite a different way from Sec. III. Indeed, except for the final subsection, this section is quite independent of the topological concepts developed in the preceding section. The reason for this independent excursion into group-theoretic terrain is this: The computation of the fundamental group of an order-parameter space can be enormously simplified—in fact reduced to a simple algorithm—provided one expresses the structure of the order-parameter space in terms of the structure of certain groups of continuous transformations which can act on the uniform medium.

In part A the notion of a continuous group is introduced and some basic properties of such groups are extracted. In part B the representation of order-parameter spaces in terms of the appropriate continuous groups is described, and it is illustrated with the standard examples in part C. In part D we show that the fundamental group of a space has some special properties when the space is itself a continuous group.

A. Continuous groups

A *continuous group* (also known unfortunately as a *topological group*) is a group (an infinite group, if the continuity is to lead to nontrivial structure) which in addition to its group multiplication table has been endowed with enough structure to enable one to apply the usual notions of continuity to sequences or sets of group elements. (This might require defining a metric giving the distance between any pair of group elements, or, more generally, specifying a topology by characterizing the open sets.) For us the notion of continuity will be obvious since the groups we shall examine are always groups of transformations and it will be clear what is meant by two transformations differing only by an infinitesimal transformation.

The group structure and the topological space structure (i.e., the structure associated with continuity) are not independent of one another, but are fused by the further requirement that the group operations themselves be continuous. This means that if a_n and b_n are two sequences of group elements converging to a and b, then the sequence of products $a_n b_n$ should converge to ab; furthermore, the sequence of inverses, a_n^{-1} should converge to a^{-1}. Both requirements can be merged into the single requirement that $a_n b_n^{-1}$ converge to ab^{-1} whenever a_n converges to a and b_n converges to b.

In most physical applications the term *Lie group* is used much more often than "continuous group" or "topological group." A Lie group is a continuous group

[20] "Trivial" is used here in the technical sense defined in Sec. III.A.4.

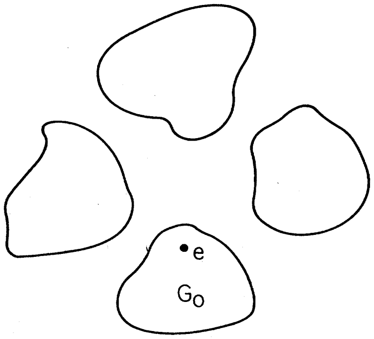

FIG. 22. Schematic picture of a four-component continuous group G. The piece of G containing the identity e is called G_0.

satisfying the somewhat stronger regularity condition that some neighborhood of the identity should admit of a smooth parametrization by variables $t_1, t_2, t_3, \ldots$. All of the continuous groups described here (and all I can imagine ever arising in condensed matter physics) are Lie groups as well. However, except in a few rather technical theorems, it is continuity rather than the existence of a suitable parametrization that plays the essential role for us. We shall therefore retain the more intuitive term of "continuous group," except when quoting theorems which explicitly require the additional structure afforded by a parametrization.

To indicate the kind of arguments one can produce when dealing simultaneously with group structure and continuity, we consider a few simple theorems, which will also be of considerable use in the development that follows.

1. Let G be a (not necessarily connected) topological group and let G_0 be that subset of G that is connected to the identity (known as the connected component of the identity—see Fig. 22. Then G_0 is a normal subgroup of G.

To prove that G_0 is a subgroup we must show that if a and b belong to G_0 then so does ab^{-1}. Since G_0 is the connected component of the identity e, there must be a continuous path a_t in G such that $a_0 = e$ and $a_1 = a$, and similarly for b. But if a_t and b_t are continuous paths, then so is $a_t b_t^{-1}$. Since this last path starts at e and ends at ab^{-1}, we have established that ab^{-1} is also in G_0.

For G_0 to be a *normal* subgroup we must have cac^{-1} in G_0 for any a in the subgroup G_0 and any c in the full group G. This follows from the fact that if a_t is a continuous path from the identity e to a, then ca_tc^{-1} will be a continuous path from $cec^{-1} = cc^{-1} = e$, to cac^{-1}. Thus cac^{-1} is also connected to the identity.

2. The disjoint connected components of G are just the cosets of the subgroup G_0.

Recall that if G is a group and G_0 is a subgroup, then the coset[21] aG_0 is defined for any a in G to be the set of all elements in G of the form ah, where h is any element of G_0. It is an elementary purely algebraic theorem that two cosets are either identical or have no elements in common at all. To show that the set of disjoint cosets of G_0 is identical to the set of connected pieces of G, we must show (a) that each coset is connected and (b) that any elements in G that can be joined by a continuous path of elements are in the same coset of G_0. Crucial to proving both (a) and (b) is the observation that if g_t is a continuous family of group elements and a is some fixed group element, then ag_t is also continuous. Using this fact, we proceed as follows:

(a) If b_0 and b_1 are both in the coset aG_0 then there are elements h_0 and h_1 in G_0 with $b_0 = ah_0$, $b_1 = ah_1$. Since G_0 is connected there is a continuous path h_t of group elements entirely in G_0 connecting h_0 to h_1. Since every h_t is in G_0, the set of elements given by ah_t is entirely in the coset aG_0. But this set is also a continuous path, and it connects ah_0 with ah_1. Thus aG_0 is indeed connected.

(b) Let b_t be a continuous path connecting two elements b_0 and b_1 of G. Then $b_0^{-1}b_t$ is a continuous path connecting the identity $(b_0^{-1}b_0)$ to the point $b_0^{-1}b_1$. Thus $b_0^{-1}b_1$ belongs to G_0, the connected component of the identity. Hence $b_1(=b_0(b_0^{-1}b_1))$ belongs to the coset b_0G_0. But so does $b_0(=b_0e, e$, the identity). Hence b_0 and b_1 are indeed in the same coset.

These two theorems have the combined consequences that one can impose a group structure on the disconnected pieces of a topological group G. For these are just the cosets of the normal subgroup G_0. By another elementary and purely algebraic theorem, the cosets of a subgroup G_0 can be themselves given the structure of a group (known as the quotient group and written G/G_0) provided the subgroup is a normal subgroup. In the quotient group the product of two cosets is simply the unique[22] coset containing the product of any two members of the two cosets.

The quotient group G/G_0 formed by the connected pieces of G is sometimes also written as $\pi_0(G)$. The notation is intended to suggest that it be regarded as a *zeroth homotopy group* of G. The basis for the analogy is that $\pi_1(G)$ characterizes the sets of equivalence classes of mappings of loops (one-dimensional) into G, whereas π_0 characterizes the sets of equivalence classes of mappings of points (zero-dimensional).

B. Group-theoretic description of the order-parameter space

All of the order parameters described in Sec. II.A and most of the order parameters commonly encountered in condensed matter physics, have associated with them a group of transformations G, with the property that if f_1 and f_2 are possible values of the

[21] Actually what is described here is a left coset. However, the left and right cosets of a normal subgroup are the same. Note also (I hope) that the proof of theorem 2 is trivial. I have run the risk here of confusing the reader with a trivial proof, because I believe it is important to illustrate the kinds of manipulations one can subject continuous groups to.

[22] To establish the uniqueness group-theoretically one needs the fact that G_0 is a normal subgroup. From the topological point of view, however, it is immediately obvious from simple continuity considerations that the products of pairs of elements from two connected pieces of G must all lie in a single connected piece.

order parameter, then there is a transformation g in G which takes f_1 into f_2: $f_2 = gf_1$. We shall illustrate this in some detail in part C of this section. A group G of transformations on a space R with this property is said to *act transitively* on R.

For given f_1 and f_2 the transformation taking f_1 into f_2 need not be unique. There may (and in general there will) be many g in G satisfying $f_2 = gf_1$. Indeed, the group G itself need not be unique. If, for example the order parameter is a three-dimensional unit vector, then the group of proper rotations SO(3) contains many different operations which take a given unit vector $\mathbf{s}_1$ into another specified unit vector $\mathbf{s}_2$. So does the group O(3) of proper *and* improper rotations. The description that follows will be valid for any group G that acts transitively on R. The most convenient choice of G is not necessarily the "smallest" such group.

If f is any given value of the order parameter we define H_f to be the set of all transformations in G that leave f unchanged. Thus a transformation g belongs to H_f if and only if

$$gf = f\,. \tag{4.1}$$

Evidently H_f is a subgroup, for if a and b leave f fixed so does ab^{-1}. It is variously known as the *isotropy subgroup* of f or the *fixer* of f or the *little group of f*. If the order parameter is a unit three-vector $\mathbf{s}$ and G is SO(3) then $H_{\mathbf{s}}$ is the subgroup of rotations about the axis $\mathbf{s}$. If G were taken to be O(3) for the same order parameter, then $H_{\mathbf{s}}$ would also include the produces of rotations about $\mathbf{s}$ with mirrorings in the plane of $\mathbf{s}$.

In general H_f is not a normal subgroup. Indeed, if $f_2 = gf_1$, then it is readily verified that

$$H_{f_2} = gH_{f_1}g^{-1}\,. \tag{4.2}$$

The example of three-dimensional spins makes it clear that this is not in general equal to H_{f_1}: The subgroups of rotations about distinct axes are distinct subgroups.

We shall characterize the order parameter space in terms of the group G, and the isotropy subgroup H_f for a particular value f of the order parameter, chosen arbitrarily but thereafter fixed. We shall call that fixed value the *reference order parameter* or *standard order parameter.* In the case of 3-spins, for example, the standard order parameter might be taken to be a unit vector along the z axis. When no confusion can result we shall drop the subscript "f" from H, which should then be understood to be the isotropy subgroup for the reference order parameter.

The structure we shall describe is independent of the arbitrary choice of reference order parameter. For changing the reference order parameter from f to f' changes H into gHg^{-1} [where $f' = gf$—see Eq. (4.2)]. Whatever structures we build out of G and H_f can therefore be converted to the corresponding structures built from G and H_f, by the inner automorphism $G \to gGg^{-1}$. This transformation preserves all group-theoretic structure; furthermore it is a continuous transformation and it therefore preserves all topological structure. If we can characterize the order-parameter space R entirely in terms of the algebraic and topological properties of G and H, that characterization will not depend on the particular choice of reference order parameter.

There is, in fact, a very simple such characterization: *The order parameter space R can be taken to be the space of cosets of H in G.* Denoting the coset space by G/H, this assertion is compactly summarized in the formula[23]

$$R = G/H\,. \tag{4.3}$$

To establish that (4.3) gives a representation of the order-parameter space we must show that there is a correspondence between cosets of H in G and values of the order parameter which (1) is one-to-one and (2) is continuous. The correspondence itself is set up as follows:

Let f be the reference order parameter, so that H is the set of elements g of G satisfying $gf = f$. Any other value f' of the order parameter is of the form $f' = gf$ for some (not necessarily unique) g in G that is not in H. The correspondence associates with this f' the coset gH.

We now prove that the correspondence has the desired properties:

1. One-to-one If af and bf are both equal to f' for two elements a and b of G, then $f = a^{-1}bf$, which shows that $a^{-1}b$ is a member of the isotropy subgroup H. As a result $b\,[=a(a^{-1}b)]$ belongs to the coset aH. So does a $(=ae)$. Thus a and b are in the same coset: The choice of coset is independent of the particular group element chosen to convert f into f', so that a given f' determines a unique coset. Conversely, given the coset we can recover a unique order parameter f' by letting any member of the coset act on the reference order parameter f; for if $af \neq bf$, then $a^{-1}b$ is not a member of the isotropy subgroup H, so that b cannot belong to the coset aH that contains a.

2. Continuity We have yet to define what is meant by continuity in a space of cosets. The definition is, in fact, constructed to ensure that the one-to-one representation of order-parameter values by cosets does indeed preserve continuity. A sequence of cosets is taken to be a convergent sequence if and only if it can be represented as a sequence g_nH, where g_n is a convergent sequence in the group G.[24] Thus two cosets are nearby if and only if they can be constructed by multiplying the subgroup H by two nearby elements of G. It follows at once that a convergent sequence of cosets corresponds to a conver-

[23] A cautionary remark: When H is a *normal* subgroup the coset space is itself a group (the quotient group) and we have already introduced the notation G/H for that group. In the present context H is not necessarily a normal subgroup, but the coset space can still be defined and is given the name G/H even when a group structure cannot be imposed upon it. The order-parameter space is a space of cosets, but not, except in very special cases, a quotient group.

[24] Not any representation of a convergent sequence of cosets will have this property. For example, the trivially convergent sequence $H, H, H, \ldots$ can be represented as h_1H, h_2H, $h_3H, \ldots$ where h_n is a completely random (and hence nonconvergent) sequence of group elements all of which belong to H itself. On the other hand the sequence can also be represented by one in which all the h_n are the same element of H, which clearly does converge. The crucial point is that there must be *some* representation in terms of a convergent sequence of elements of G.

gent sequence of order parameters $f_n = g_n f$.[25] Conversely, a convergent sequence of order parameters is represented by a convergent sequence of cosets. For suppose the sequence of order parameters f_n converges to $\bar{f}$. Then for an arbitrarily small neighborhood of the identity element e in G, there must be an integer N such that values of f_n for n greater than N can be represented in the form $f_n = e_n \bar{f}$, where the e_n are in the small neighborhood of e. If $\bar{f}$ is represented by the coset K, then the f_n will be represented by the cosets $e_n K$, which do indeed converge to $eK = K$.

Although this description of the order-parameter space as a space of cosets may seem unfamiliar, it is important to note that it is nothing but a rather abstract formalization of the commonplace practice of describing ordered media in terms of broken symmetry. Consider a uniform ordered medium and let the reference order-parameter f be the value the order parameter everywhere assumes. The group G must contain enough transformations to convert f into any other possible value of the order parameter. For this purpose the full symmetry group of empty physical space will certainly suffice, though it might be convenient to chose a smaller or a larger group, as we shall see. In any event, the full symmetry of the group G is broken by the ordering, since not all elements of G leave f invariant. The isotropy subgroup H is the set of transformations in G that do leave the system invariant, even after the ordering has set in, i.e., H is the symmetry group of the ordered phase. The fact that the ordering breaks the underlying symmetry is expressed in the fact that H is only a subgroup of the underlying group G. If the symmetry were completely broken (so that H consisted of the identity alone) then all transformations of G would yield distinct order parameters, and we could identify the order-parameter space with G itself. If, however, the ordered phase retains some residual symmetry (characterized by the group H) then unique values of the order parameter will correspond to whole sets of elements of G. These sets *are* the cosets of H in G, and the collection of all these sets *is* the coset space G/H.

To summarize the conventional picture, G is the symmetry group of the disordered phase (more correctly, "a disordered phase," since we need not choose G to be the full symmetry group of empty space) and H is the subgroup of G that describes the symmetry of the ordered phase. Both the formal and the intuitive content of the representation of the order parameter by a space of cosets should be illuminated by converting the descriptions of the standard examples of Sec. II.A. into this language.

C. Examples of order-parameter spaces as coset spaces

We now illustrate the ideas developed above with the various examples of Sec. II.A. The reader should note that important general remarks will be made in the discussions of particular examples, which should therefore be perused even if the reader feels fully in command of a given case.

1. Planar spins

The order parameter is a unit vector in the plane. A suitable group G is therefore the two-dimensional (proper) rotation group SO(2). No transformation other than the identity in SO(2) leaves a vector fixed, so H is the trivial subgroup consisting of the identity alone, regardless of the choice[26] of reference order parameter f. The cosets of the subgroup consisting of the identity alone are the group elements themselves. Thus the order-parameter space for planar spins can be taken to be SO(2) itself. From this point of view the circle we used in Sec. II to represent the order-parameter space plays the role of a convenient representation for SO(2). But the order-parameter space is SO(2) itself.

It is important to realize that any other choice of G would lead back to the same order-parameter space. We could, for example, have taken O(2) rather than SO(2) as the group G, including improper as well as proper rotations of the plane. If the reference order parameter were taken as a unit vector along the x axis, then H would consist of the identity and the operation

$$m = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

of reflection in the y axis. The cosets of H are pairs of elements, consisting of a proper rotation and that same rotation followed by the reflection m. Evidently the coset space can be parametrized by parametrizing the proper members of the pairs, and we are back to the original circle.

A more important example of this flexibility in description is the following: Suppose we took the group G to be not the proper rotation group SO(2) but the one-dimensional translation group, T(1). This would be a natural choice if we represented the spins in the form $\mathbf{s} = \hat{x}\cos\theta + \hat{y}\sin\theta$. The operations of G would consist of the transformations $\theta \to T_\phi \theta = \theta - \phi$. The isotropy subgroup would be the subgroup of T(1) consisting of translations through 2π, independent of the choice of reference order parameter. Since T(1) is Abelian this is a normal subgroup, and therefore the order-parameter space G/H would itself be a group. This group—the one-dimensional translation group with translations differing by 2π identified—would be isomorphic to SO(2), and we would again recover the original order-parameter space. The importance of this example will emerge later on.

Both of these alternative descriptions suggest a useful way of looking at coset spaces: Taking elements of the order-parameter space to be cosets of the group G is the same as taking the order-parameter space to be G itself, with the proviso that elements in g belonging to

[25] We naturally define continuity in the group G itself so that a convergent sequence of transformations acting on any particular value of the order parameter yields a convergent sequence of order parameters.

[26] Note that H will in general be independent of the choice of reference order parameter if and only if it is a normal subgroup of G. If H *is* a normal subgroup, then G/H is itself a group, and the order-parameter space itself can be given a group structure.

the same coset are to be identified. Thus closed loops in coset space can be represented by open paths in G, provided the paths start and stop at points in the same coset. Further aspects of this point of view will emerge in the examples that follow.

2. Ordinary spins

The order parameter can be any unit vector in 3-space, so the full proper three-dimensional rotation group SO(3) is required to take any value of the order parameter into any other. If the reference order parameter is taken to be a unit vector along $\hat{z}$, the isotropy subgroup is the two-dimensional proper rotation group SO(2) with respect to the z axis, and the order-parameter space is $R = G/H = \mathrm{SO}(3)/\mathrm{SO}(2)$. It may seem a bit pompous to replace the simple surface of a sphere S_2 by the space of cosets of SO(2) in SO(3), but, as we shall see, this is in many ways the simpler and more natural representation to use. Still simpler and more natural, as we shall also see, is the representation given by replacing SO(3) by SU(2); i.e., we take advantage of the homomorphism[27] between SU(2) and SO(3) to take the larger group SU(2) as the one that acts transitively on the order-parameter space. The isotropy subgroup H is now the subgroup of SU(2) that leaves the z axis fixed. This consists of all unitary matrices of the form

$$\exp(i\theta\sigma_z) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

and is evidently isomorphic to U(1), the group of one-dimensional unitary transformations. The topological properties of three-dimensional spins turn out to follow most directly from viewing the order-parameter space as $R = \mathrm{SU}(2)/\mathrm{U}(1)$.

3. Nematics

If the group G is taken to be SO(3) then the isotropy subgroup consists of the group D_∞ of rotations about the molecular axis and 180° rotations about axes perpendicular to the molecular axis. Thus the order-parameter space is $R = \mathrm{SO}(3)/D_\infty$. If we were to include inversions by taking G to be O(3), we should have to add improper operations to the isotropy subgroup H, extending D_∞ to $D_{\infty h}$. The coset space would retain the same structure. Here again (and in all the other cases as well) we shall find it most convenient to take G to be SU(2), and H the inverse image in SU(2) [known as the *lift* in SU(2)] of the subgroup D_∞ of SO(3).

Note that the unadorned sphere (representing 3-spins—arrows) and the sphere with diametrically opposite points identified (representing nematics—headless arrows) appear on a more symmetric footing in this description. One is SO(3)/SO(2) [or, in more conventional point group notation, $\mathrm{SO}(3)/C_\infty$] and the other is $\mathrm{SO}(3)/D_\infty$. By starting the description in both cases with the group $G = \mathrm{SO}(3)$ instead of directly with the surface of the sphere many points must be identified in both cases. The distinction between the two cases now lies in the isotropy subgroup: $\mathrm{SO}(3)/C_\infty = S_2$, the 2-sphere; $\mathrm{SO}(3)/D_\infty = P_2$, the projective plane.

4. Biaxial nematics

If G is taken to be SO(3) then the isotropy group of the rectangular box is only the four-element group consisting of the identity and 180° rotations about three mutually perpendicular axes (D_2). Order-parameter space is $R = \mathrm{SO}(3)/D_2$. We shall see that the lift of D_2 in SU(2) is the group Q of quaternions.[28] The natural representation for the order-parameter space R of a biaxial nematic turns out to be $R = \mathrm{SU}(2)/Q$. More generally, if the "molecules" of the biaxial nematic have a proper point group H_0, then the natural order-parameter space turns out to be $\mathrm{SU}(2)/H$, where H is the lift of H_0.

5. Superfluid helium-3

The order parameter in the dipole-locked A phase is a pair of orthonormal axes. If we take the reference order parameter to be the pair $\hat{x}, \hat{y}$, then there is a unique correspondence between pairs of orthonormal axes and proper rotations of $\hat{x}, \hat{y}$, i.e., the group G can be taken to be SO(3) and the isotropy subgroup H consists of the identity alone. Thus the order-parameter space is $R = G/H = \mathrm{SO}(3)$.

If, instead, we take G to be SU(2), then the isotropy subgroup H is independent of choice of reference order parameter and consists of the two elements of SU(2) that map onto the identity of SO(3) under the homorphism. These are the two elements represented by the 2×2 matrices 1 and -1, corresponding in the conventional picturesque but confusing language to the identity and "the 360° roation."

The dipole-free A phase affords an unusual example of a case where G *must* be bigger than SO(3). The order parameter [see Eq. (2.8)] is the product of an arbitrary unit 3-vector $\hat{n}$ and a complex 3-vector of the form $\hat{u} + i\hat{v}$, where $\hat{u}$ and $\hat{v}$ are an orthonormal pair. The orientations of $\hat{n}$ and $\hat{u} + i\hat{v}$ are uncoupled. If we take the reference order parameter to be $A_{ij} = z_i(x_j + iy_j)$ then we can generate an arbitrary order parameter by letting one rotation act on the index i and another in general distinct rotation act on the index j. Thus G can be taken to be the direct product of SO(3) with itself: $G = \mathrm{SO}(3) \times \mathrm{SO}(3)$, elements of G consisting of pairs (R, R') of distinct rotations. The isotropy subgroup H contains all elements of the form $R(\hat{z}, \theta) \times 1$ (where 1 is the identity rotation), since left rotations about the z axis leave the reference order parameter unchanged. Although any distinct rotations take $\hat{x} + i\hat{y}$ into distinct pairs, the rotation that changes the sign of $\hat{x} + i\hat{y}$ will leave the order parameter A_{ij} unaltered if it is paired with a rotation that changes the sign of $\hat{z}$. Therefore the isotropy subgroup also contains all elements of the form $R(\hat{u}, \pi)$,

[27] This homomorphism, though introduced here in an inessential way, will eventually play a central role in our analysis. It is reviewed in Appendix B.

[28] Translation: the subgroup of SU(2) that is carried into D_2 under the homomorphism between SU(2) and SO(3) is isomorphic to that eight-element non-Abelian group known as the quartenion group Q. The quartenion group will be examined explicitly in Sec. V.B.4.

$R(\hat{z}, \pi)$ where $\hat{u}$ is any axis in the x-y plane.

In all of these cases we have gone from a representation of the order-parameter space by a geometric object (a circle, a sphere, a projective space, etc.) to a representation by a coset space of a topological group. From the group-theoretic point of view, the geometric object is simply a particular way of parametrizing the coset space. Because there is a continuous one-to-one mapping of the coset space onto the geometric object and vice versa, the topological properties of one are identical to the topological properties of the other. The fundamental groups of the coset spaces of interest, however, can all be computed at once, by virtue of a single powerful theorem, which will be developed and applied in Sec. V.

Before turning to the matter of central interest, however, we must examine some features of the fundamental group $\pi_1(R)$ of a space R, when that space happens to be a continuous group G.

D. Properties of the fundamental group of a topological group

A continuous group is also a space, with continuity defined in it. Ignoring the additional algebraic structure, we can examine the general topological properties of that space, and, in particular, its fundamental group. The algebraic superstructure possessed by the space by virtue of its being a group can be of help in such topological investigations, and can lead to some quite general simplifications in the topological structure. Perhaps the most important such simplification is the fact that the fundamental group of a continuous group is always Abelian. We prove this result below because it is of some use in the development that follows, because some of the subsidiary concepts and results are also of interest, and because it furnishes an especially simple example of the study of the homotopy groups of continuous groups—a study that will become rather more elaborate in subsequent sections.

Because the fundamental group of a space is isomorphic to any of the based fundamental groups, to compute $\pi_1(G)$ it suffices to compute the group $\pi_1(G, e)$ of classes of loops based at the identity. Now a loop based at the identity is simply a map $f(z)$ of the interval $0 \leq z \leq 1$ into the continuous group G, which starts and finishes at the identity

$$f(0) = f(1) = e\,. \tag{4.4}$$

The path product of two loops $f(z)$ and $g(z)$ is defined in the usual way [Eq. (3.2)] as the loop at e given by first traversing f and then g. Because G is a group, in addition to the loop product $f \circ g$, we can also define a *group product*, $f \times g$, which is given at any z by the group-theoretic product of the values of f and g at that z:

$$f \times g|_z = f(z) g(z)\,. \tag{4.5}$$

Since $e^2 = e$, the group product of two loops at e is also a loop at e.

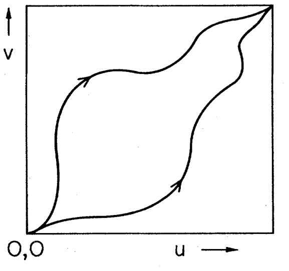

FIG. 23. Various homotopic paths in the unit square connecting 0,0 to 1,1. Such paths determine homotopic loops in G at e, given by the values of $f(u)g(v)$ as u and v traverse the path.

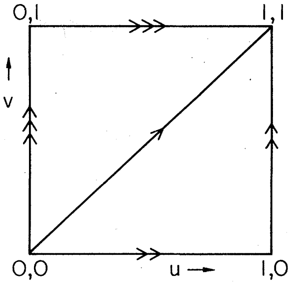

FIG. 24. Three paths connecting 0,0 to 1,1 in the unit square. The corresponding loops $f(u)g(v)$ are all homotopic at e in G. The diagonal path gives the group product [Eq. (4.5)]. The path with double arrows yields the loop product $f \circ g$. The path with triple arrows gives $g \circ f$.

The group product of two loops, or rather a slight generalization of the group product, provides the basis for a quite simple proof that $\pi_1(G)$ is Abelian. The basic result is this:

If f and g are two loops at e in G, then the loop product $f \circ g$, the loop product $g \circ f$, and the group product $f \times g$, are all mutually homotopic at e.

Since $\pi_1(G, e)$ consists of classes of homotopic loops at e, this establishes that $[f] \circ [g] = [g] \circ [f]$—i.e., that the multiplication of such classes is commutative.

To prove the basic result, consider the continuous map of the unit square $0 \leq u, v, \leq 1$ into G, given by

$$u, v \to f(u) g(v)\,. \tag{4.6}$$

The image in G of any line within the square connecting 0, 0 to 1, 1 is evidently homotopic to the image of any other line, the homotopy being provided by any convenient deformation of one line into the other within the square (Fig. 23). The image of any such line is a loop at e in G, so by tracing various paths from 0, 0 to 1, 1 in the square, we produce various loops in G that are homotopic at e. Now the loop produced by going along the diagonal of the square is just the group product (4.5). However (Fig 24), by taking a route that goes along the edges of the square, one produces either $f \circ g$ or $g \circ f$, depending on which pair of edges one chooses.[29] This completes the proof.

Although the proof that $\pi_1(G)$ is Abelian for any continuous group is quite elementary, the result is, from some points of view, quite startling. Recall, for example (Sec. III.A.6.c) that the fundamental group of a figure eight is non-Abelian. It follows that although a circle can be taken as the parameter space for a continuous group, a figure eight cannot. Elementary as our proof was, I doubt that many people, first introduced to the notion of a group, would realize that "non-figure-eight-parametrizable" was one of its very basic attributes.

[29] Remember that along the edges of the square either the f or the g in Eq. (4.6) is equal to the identity e, since $f(0) = f(1) = g(0) = g(1) = e$.

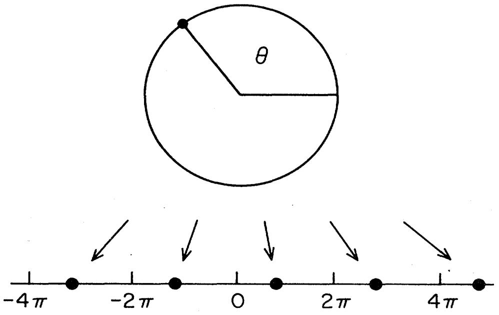

FIG. 25. The universal covering group for the two-dimensional proper rotation group SO(2) is the one-dimensional translation group T(1). A point at the angle θ in the circular order-parameter space for SO(2) is the image of all the points $x=\theta+2\pi n$ along the linear parameter space for T(1), under the covering homomorphism.

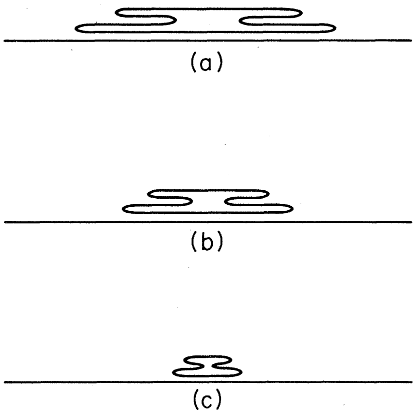

FIG. 26. The real line is simply connected. (a) A loop in the real line. Points on the loop are taken into the point on the real line directly below. The squashing of the loop into the line has not been carried to completion in the figure to make it clear that the loop is a loop. (b) and (c) Successive stages in the shrinking of the loop to a point.

V. THE FUNDAMENTAL GROUP OF THE ORDER-PARAMETER SPACE

A complete classification of the line defects of an ordered medium and their combination laws is provided by the fundamental group of the order-parameter space. We now give a simple rule for finding the fundamental group for any ordered medium whose order-parameter space has the structure described in the preceding section. The rule is developed in general terms in part A, and applied to the standard examples in part B.

For the reasons given at the beginning of Sec. III, it suffices to consider connected order-parameter spaces, and therefore to consider only connected transformation groups G. For rather more subtle reasons it also suffices to consider only simply connected[30] transformation groups. This follows from a theorem that any continuous group can be imbedded in a larger group (known as its *universal covering group*) that *is* simply connected. The precise nature of the imbedding is that the group G_1 is the homomorphic image[31] of its simply connected universal covering group G_2.

The most familiar example (to physicists) of a group and its universal covering group is the three-dimensional group of proper rotations SO(3) (which, as we shall see, is not simply connected) and its universal covering group, the special unitary group SU(2) (which is). The relation between the two is a homomorphic mapping of SU(2) onto SU(3) which takes a pair of 2×2 unimodular unitary matrices (differing by an overall minus sign) into each 3×3 real orthogonal unimodular matrix. (See Appendix B for a more detailed description.)

A simpler example is the two-dimensional group of proper rotations SO(2). A parameter space for SO(2) is the circle, so the fundamental group of SO(2) is isomorphic to the integers. The universal covering group is the one-dimensional translation group T(1). The homomorphism (see Fig. 25) associates with the rotation through θ all the translations through $\theta+2\pi n$ for any integral n (positive, negative, or zero). A parameter space for T(1) is evidently the entire real line. This is simply connected, since any continuous image of a loop in the real line can be scaled continuously down to a point (see Fig. 26).

In applications the only universal covering groups we shall make use of are those for SO(3) and SO(2). We shall therefore dispense with a proof of the general theorem that such covering groups always can be constructed,[32] the construction being explicitly given in the cases of interest.

Note that if one can describe with a group G_1 the transformations taking the standard order parameter into any particular value, then one can equally well construct a description using the universal covering group G_2. One merely replaces the set of transformations in G_1 giving any particular value of the order parameter with the (larger) set of transformations in G_2 that correspond, under the covering homomorphism, to the first set. The discussion of planar spins in Sec. IV.C.1 serves as an explicit example. The point to keep in mind is that if G is enlarged then so is the isotropy subgroup H. Such changes "factor out" of the coset space G/H, which represents the order-parameter space precisely because it eliminates any such redundancies.

We now proceed to the fundamental theorem.

A. The fundamental theorem on the fundamental group

Theorem: Let G be a connected, simply connected continuous group. Let H by any subgroup of G. Let H_0 be the set of points in H that are connected to the identity by con-

[30] Recall that a space is simply connected if its fundamental group contains only the identity (i.e., if any loop can be shrunk to a point).

[31] More explicitly, there is a map ϕ of G_2 onto G_1 associating with each element of G_1 a distinct pair (or triple, or, for an n to one homomorphism, n-tuple) of elements of G_2. The map preserves the group structure, in that $\phi(a)\ \phi(b)=\phi(ab)$.

[32] A straightforward and old fashionedly readable exposition of the theorem can be found in Pontryagin (1966), p. 351ff.

tinuous paths lying entirely in H. Then H_0 is a normal subgroup of H, and the quotient group H/H_0 is isomorphic to the fundamental group $\pi_1(G/H)$ of the coset space G/H.

That H_0 is a normal subgroup of H is the content of the theorem proved in Sec. IV.A.1. When going through the proof of the main part of the theorem, it is helpful to keep in mind three different possibilities for the structures of H, H_0, and the quotient group H/H_0:

(i) The subgroup H may be discrete, i.e., the set of members of H is a discrete set in G. (Formally, a set is discrete in G if each of its points has a neighborhood in G containing none of the other points, i.e., none of the points are too close together.) If H is discrete then H_0, the connected component of the identity in H, must consist of the identity alone. The quotient group H/H_0 is then just the subgroup H itself, and the theorem identifies the fundamental group with the isotropy subgroup.

(ii) The subgroup H may be a connected subset of G. This is the opposite extreme from (i). If H is connected then H_0 is all of H, the quotient group H/H_0 is the one-element group, and the fundamental group is 0: The order-parameter space G/H is simply connected.

(iii) The subgroup H may consist of two or more disjoint connected components. If these components are single points we are back to case (i). If they are not then H_0 is a proper subgroup of H, and H/H_0 is neither the full group H nor the trivial one-element group. The elements of H/H_0—the cosets of H_0 in H—are the connected components of H.

In all three cases the order[33] of H/H_0 (and hence the order of the fundamental group) is just the number of connected components of H. Since cases (i) and (ii) are clearly special cases of (iii), case (iii) is the one to keep primarily in mind in following the proof.

Note, finally, before embarking on the proof, the usefulness of the final result. It reduces the problem of loops in the coset space G/H to some elementary algebraic features of the isotropy subgroup H. One needs only to count up the connected pieces of H (which constitute the elements of H/H_0) and work out the multiplication table for these pieces by noting the (unique) piece that contains the product of any two representative members of each pair of pieces. That multiplication table *is* the multiplication table for the fundamental group.

The proof is given in two stages. First, the precise correspondence between loops in coset space and components of H is described in some detail. It is essential to understand that correspondence if one wishes to make full use of the result. Then the isomorphism asserted by the theorem is proved. Readers willing to take my word for it can skip the second part.

1. The correspondence between loops in coset space and the connected components of the isotropy subgroup

First note that if $g(z)$ is a continuous path in the group G, then the path in coset space given by

$$K(z)=g(z)H \tag{5.1}$$

is continuous by the very definition of continuity in coset space (point 2 of Sec. IV.B). The converse proposition is more subtle but also true: If $K(z)$ is a continuous path in coset space then it can be represented by some (clearly not unique) path $g(z)$ in the group, through Eq. (5.1). An honest proof of this is a fairly delicate matter, and all the proofs I have seen require the continuous group G to be a Lie group as well. For our purposes, however, the validity of the converse is amply demonstrated by the following observation:

Given a continuous family of cosets, $K(z)$, let $f(z)$ be the corresponding continuous family of order parameters with $f(z_0)$ given by the action on the reference order parameter f of any of the members of $K(z_0)$ (the result being independent of the particular choice of member). It is intuitively clear that one can trace out the continuous trajectory $f(z)$ in order-parameter space by applying a continuous sequence of transformations $g(z)$ to the reference order parameter f. But this means that each $g(z)$ itself belongs to the coset $K(z)$, so that $K(z)$ can indeed be represented in the form of Eq. (5.1).

We need not apply Eq. (5.1) to arbitrary loops in coset space, for the structure of the fundamental group is given by any based fundamental group. It is convenient to take as base point in coset space the subgroup H itself. To represent a loop at H in coset space via Eq. (5.1), the path $g(z)$ in the group G must start and finish in the subgroup H, though, of course, the path in the group need not be a closed loop, since $gH=H$ for any g in H. Note that for any group element h in H, $g(z)h$ gives the same coset loop via (5.1) as does $g(z)$ itself. Hence by picking h to be the inverse of $g(1)$ we can ensure that the representative path ends at the identity e.

We therefore can represent loops at H in coset space by continuous paths in G connecting points in H to the identity e. We classify the representative paths according to which of the connected pieces H_i of H they start from (Fig. 27).

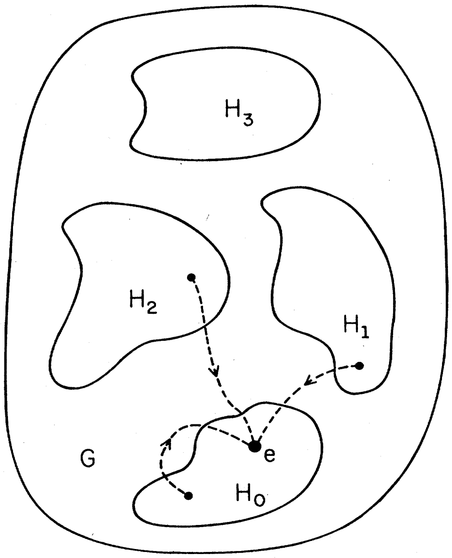

FIG. 27. Various paths in the group G that connect points in the subgroup H with the identity e.

[33] The order of a group is simply the number of elements it contains.

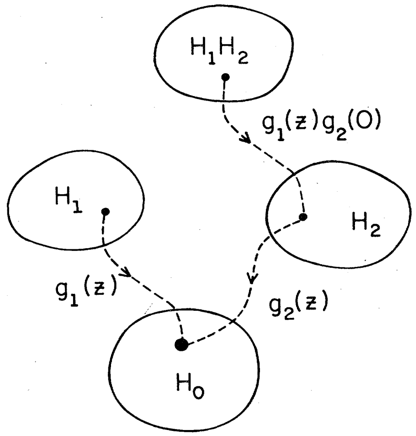

FIG. 28. Construction for determining the product of two paths g_1 and g_2 in G, connecting points in the components H_1 and H_2 of the subgroup H to the identity e. Shift the path g_1 by multiplying each of its points by $g_2(0)$. This gives a path $g_1g_2(0)$ that terminates at the starting point of g_2. The path product $g_1 \circ g_2$ is defined to be the composite path given by first traversing $g_1g_2(0)$ and then continuing on to e along g_2.

Consider now two loops at H in coset space, $K_1(z)$ and $K_2(z)$, that can be represented by paths $g_1(z)$ and $g_2(z)$, respectively connecting points in H_1 and H_2 to the identity. We construct a path in G representing the loop product $K_1 \circ K_2$ as follows:

Since $g_2(0)$ belongs to H_2 and hence to H itself, the path $g_1(z)\,g_2(0)$ also represents the coset loop $K_1(z)$. This path, however, connects the point $g_1(0)\,g_2(0)$ of G to the point $g_1(1)\,g_2(0) = eg_2(0) = g_2(0)$, which is the starting point of the path that represents $K_2(z)$. Thus we can put the two paths $g_1(z)\,g_2(0)$ and $g_2(z)$ together (first traversing the former, then the latter) to get a single continuous path that starts at $g_1(0)\,g_2(0)$ and ends at e (see Fig. 28). By its construction this path represents the loop product $K_1 \circ K_2$. Its starting point, $g_1(0)\,g_2(0)$ lies in the component of H that corresponds in the quotient group H/H_0 to the product of H_1 with H_2.[34]

That constitutes the essence of the theorem: Loops in coset space based at H correspond to paths in G connecting the elements of the factor group H/H_0 (i.e., the connected pieces of H) to the identity. The loop product of two loops at H in coset space can be represented by a path in G originating from the piece of H which is the product (when considered as an element of H/H_0) of the two pieces connected to e by the paths representing the two loops.

2. Proof of the isomorphism between $\pi_1\,(G/H)$ and H/H_0

To complete the argument given above we require a demonstration that the correspondence just described between homotopy classes of loops at H in coset space, and connected pieces of H in G, is in fact a one-to-one correspondence. We must thus show (a) that any two paths in G connecting a given component of H to the identity yield via (5.1) homotopic loops at H in G/H and (b) that two homotopic coset loops in G/H at H can only be represented by paths in G connecting the same component of H to the identity. We take up the points one by one.

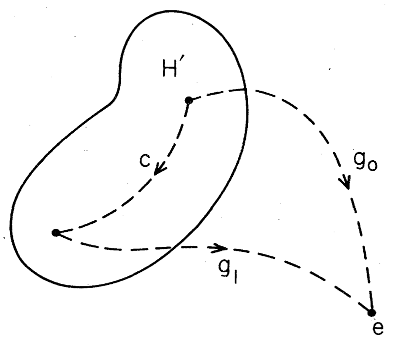

FIG. 29. Two paths g_0 and g_1 connecting points in the same component H' of the subgroup H to the identity e. A third path c can be drawn entirely in H' connecting the starting points of g_0 and g_1.

(a) Consider two paths $g_0(z)$ and $g_1(z)$, each of which joins a point in the same component H' of H to the identity e. If we can find a continuous interpolation $g_t(z)$ between the two paths which starts in H' and ends at e for each t, then we are done, for the required homotopy in coset space will be just $K_t(z) = g_t(z)H$. We construct the required interpolation as follows:

Because H' is a connected piece of H, we can find a path c in G that connects $g_0(0)$ to $g_1(0)$ and lies entirely in H' (see Fig. 29). The path given by first traversing c, then traversing g_1, and then traversing g_0 in reverse order, is a closed loop in G. But G, by assumption, is simply connected.[35] Hence that loop can be shrunk to a point in G. The shrinking homotopy can be viewed (Fig. 30) as a continuous map of a triangle into G, the edges of the triangle going into g_0, g_1, and c. By simply reparametrizing that triangle (Fig. 31) so that t describes paths connecting various points along c to e, we can construct from it the required interpolation.

(b) Let $K_t(z)$ be a homotopy at H between two coset loops at H, K_0, and K_1. For any fixed value of t let $f(z)$ and $g(z)$ be two paths in G terminating at the identity and representing $K_t(z)$, so that

$$K_t(z) = f(z)H = g(z)H\,. \tag{5.2}$$

It follows from Eq. (5.2) that $f(z)^{-1}g(z)$ belongs to the subgroup H for every choice of z. But $f^{-1}g$ is itself a continuous path in G that terminates at the identity. Since it never leaves H its starting point $f(0)^{-1}g(0)$ must lie in the connected component of the identity H_0. On purely algebraic grounds it follows that $f(0)$ and $g(0)$ belong to the same coset of H_0. Since the cosets of H_0 are the connected pieces of H, the paths f and g must start in the same component of H.

We have thus established that for any given t there is a unique component of H from which any path to e in G representing $K_t(z)$ must start. Since the family of coset loops $K_t(z)$ is continuous in t, this unique piece of H

[34] See the remarks following the proof of theorems 1 and 2 in Sec. IV.A, if this is not obvious.

[35] It is here (and only here) that we require G to be the universal covering group.

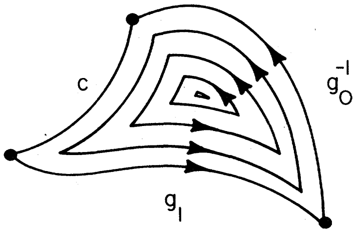

FIG. 30. The loop $cog_1og_0^{-1}$ in G formed by the paths in Fig. 29. Because G is simply connected the loop can be shrunk to a point. Successive loops in a shrinking homotopy are shown.

must vary continuously with t. But the pieces of H constitute a discrete set, so the piece associated with $K_t(z)$ cannot vary at all with t. Letting t vary all the way from 0 to 1 we conclude that paths representing K_0 via Eq. (5.1) must connect the same connected piece of H to the identity as the paths representing K_1, which is what we set out to prove.

We now illustrate the use of the theorem by computing the fundamental groups of the order-parameter spaces of our standard examples.

B. Computing the fundamental group of the order-parameter space

To apply the theorem to the examples of Sec.II.A we must represent their order-parameter spaces as coset spaces, as described in Sec.IV.C, being sure to choose for the group G one that is connected and simply connected.

1. Planar spins

The representation $\mathbf{s}=\hat{x}\cos\theta+\hat{y}\sin\theta$ permits us to take the group G to be the full one-dimensional translation group $T(1)$: $T_\varphi(\theta)=\theta-\varphi$. Here $T(1)$, being parametrized by the entire real axis, is simply connected. (Note that the more "natural" choice for G, the two-dimensional rotation group, has a circle as its parameter space and is therefore not simply connected.) The isotropy subgroup H consists of translations through integral multiples of 2π:

$$H=\{T_{2\pi n},\ n=0;\pm1,\pm2,\cdots\}\,. \tag{5.3}$$

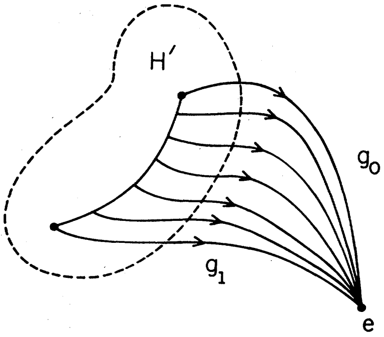

FIG. 31. A reparametrization of the shrinking homotopy of Fig. 30, showing that the homotopy also provides a continuous deformation of the path g_0 into the path g_1 via a family of intermediate paths, all of which start in the subgroup H' and end at e.

This is a discrete group, i.e., the connected component of the identity consists of the identity alone. Hence H/H_0 is equal to H itself, and the fundamental group of the planar spins is isomorphic to H, which in turn is isomorphic to the additive group of the integers, Z:

$$\pi_1(R)=Z\quad\text{(planar spins)}\,. \tag{5.4}$$

We recover our old classification by integral winding numbers.

2. Ordinary spins

Because SO(3) is not simply connected, to apply the theorem we must represent a general order parameter by the action of the simply connected group SU(2) on a reference order parameter.[36] If the reference spin is taken to be along the z axis, then a general spin is given by

$$\mathbf{s}=R(\hat{n},\theta)\hat{z}\,, \tag{5.5}$$

where R is the rotation in SO(3) through the angle θ about the axis $\hat{n}$. To apply the theorem we must use for G not SO(3), but SU(2), in terms of which Eq. (5.5) becomes

$$\mathbf{s}\cdot\sigma=u^\dagger(\hat{n},\theta)\sigma_z u(\hat{n},\theta)\,, \tag{5.6}$$

where

$$u(\hat{n},\theta)=\exp[i(\theta/2)\hat{n}\cdot\sigma]\,. \tag{5.7}$$

[The two-to-one nature of the relation between SU(2) and SO(3) can be seen from the fact that $u(\hat{n},\theta+2\pi)=-u(\hat{n},\theta)$ would serve as well as $u(\hat{n},\theta)$ in Eq. (5.6).]

The reference order parameter is left invariant [i.e., Eq. (5.6) gives $\mathbf{s}=\hat{z}$] for just those u with $\hat{n}=\hat{z}$ (as is evident, if one thinks about the corresponding rotations). Thus H is the subgroup of SU(2) of 2×2 matrices of the form:

$$u(\hat{z},\theta)=e^{i(\theta/2)\sigma_z}=\begin{pmatrix}e^{i\theta/2} & 0\\ 0 & e^{-i\theta/2}\end{pmatrix} \tag{5.8}$$

Evidently this is a connected subgroup: Any two such matrices can be joined by a continuous family of them. Thus $H_0=H$ and H/H_0 is the trivial group 0.[37] We recover our old conclusion that there are no stable line defects:

$$\pi_1(R)=0\,,\quad\text{(3-spins)}\,. \tag{5.9}$$

3. Nematics

We continue to represent the order parameter as in case 2, but the isotropy subgroup of SU(2) must now be expanded to include transformations that take $\hat{z}\to-\hat{z}$, i.e., rotations through π about arbitrary axes perpendicular to $\hat{z}$. Such a rotation can always be represented as a 180° rotation about a particular axis (say $\hat{y}$) fol-

[36] The connectivity of SO(3) and SU(2) and the relationship between them is reviewed in Appendix B.

[37] When the fundamental group is Abelian the general convention is to describe it in the language of additive (rather than multiplicative) groups. In particular the one-element group is always named 0 (rather than 1).

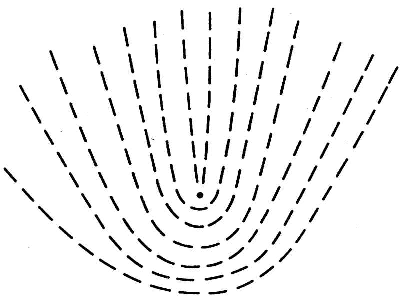

FIG. 32. The only topologically nontrivial line singularity in a nematic: the 180° disclination.

lowed by a suitable rotation about $\hat{z}$. Since rotations through π about $\hat{y}$ are represented by

$$\pm u(\hat{y},\pi) = \pm i\sigma_y = \pm\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{5.10}$$

the isotropy subgroup H includes, besides u of the form (5.8), products of these with $i\sigma_y$. [The products with $-i\sigma_y$ are then automatically included, since $\theta \to \theta + 2\pi$ sends $u \to -u$ in Eq. (5.8).] Thus in addition to the set (5.8), H will include:

$$v(\hat{z},\theta) = u(\hat{z},\theta)(i\sigma_y) = \begin{pmatrix} 0 & e^{i\theta/2} \\ -e^{i\theta/2} & 0 \end{pmatrix}. \tag{5.11}$$

Evidently the connected set (5.11) is not in the connected component of the set (5.8). Since the identity is of the form (5.8), the matrices (5.8) give H_0. The set (5.11) is of the form $H_0(i\sigma_y)$. The full isotropy subgroup H is the union of these two sets, H_0 being a normal subgroup of H and $H_0(i\sigma_y)$ being the (single) coset not equal to H_0. The group structure of H/H_0 is therefore that of the two-element group consisting of the identity and $i\sigma_y$ which (like any two element group) is isomorphic to the integers modulo 2, Z_2. We conclude that

$$\pi_1(R) = Z_2, \quad \text{(nematics)}. \tag{5.12}$$

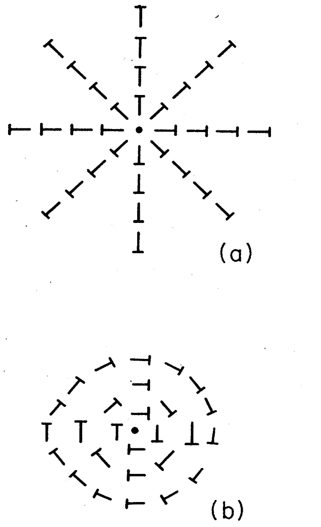

FIG. 33. Escape in the third dimension of two 360° disclinations. The lines have been given "nail heads" for ease in description, but one end is still to be considered indistinguishable from the other. The escape is achieved by rotating each nail about a perpendicular line lying in the plane of the page, until the head points straight out of the page. The resulting configuration is uniform.

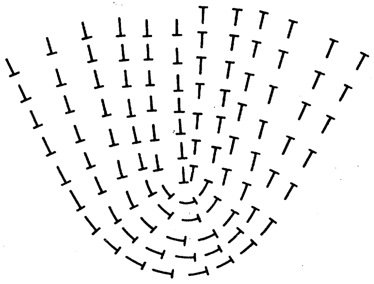

FIG. 34. The escape route of Fig. 33 fails when applied to the 180° disclination of Fig. 32, for nails on opposite sides of the line extending vertically from the singular point would have to rotate in opposite senses, and the intermediate configurations in the escape could not be everywhere continuous away from the singularity.

There is thus precisely one class of nonremovable line singularities in a nematic.

Note that the structure of the nonremovable singularity is implicit in the foregoing analysis. It corresponds to a loop in coset space represented by a path in SU(2) connecting $H_0(i\sigma_y)$ to the identity. One of the simplest such paths connects $i\sigma_y$ itself to the identity, and can be taken as the family of rotations through θ about the y axis, where θ runs from π down to zero. Thus a nonremovable singularity can be represented as one in which the headless vector $\hat{n}$ characterizing nematic order rotates in a plane through π as the singular line is encircled as shown in Fig. 32. Such a singularity is called a 180° *disclination.*

Note that when the net change in angle is 2π the singularity is topologically trivial. This follows algebraically from the fact that a 2π singularity can be viewed as the loop product of two π singularities, and is therefore described by the square of the coset $H_0(i\sigma_y)$. But the square gives back H_0 itself (the square of the nontrivial element in the two-element group must be the identity) which corresponds to the class of removable singularities. For the 2π singularities depicted in Fig. 33 the removal can be simply achieved by everywhere continuously rotating the local $\hat{n}$ to an orientation perpendicular to the page. The fate thus suffered by the removable 2π singularity is known as "escape in the third dimension." Note that the π singularity of Fig. 32 cannot be eliminated in this way; if the attempt is made to escape out of the page, a singular line is produced throughout the "collapse" extending from the singular point out to infinity (Fig. 34).

4. Biaxial nematics

The only rotations that take a rectangular box into itself besides the identity are three 180° rotations about three mutually perpendicular axes. If G is taken to be SO(3) the isotropy subgroup is therefore the four-element point group, D_2. If we take G to be SU(2), then H is expanded to the eight-element subgroup of SU(2) which is taken into D_2 under the homomorphism. We have

$$\begin{aligned} &\pm 1 \to 1, \\ &\pm u(\hat{n},\pi) = \pm i\hat{n}\cdot\sigma \to R(\hat{n},\pi), \end{aligned} \tag{5.13}$$

and therefore

$$H=\{1,\,-1,\,i\sigma_x\,,\,-i\sigma_x\,,\,i\sigma_y\,,\,-i\sigma_y\,,\,i\sigma_z\,,\,-i\sigma_z\}\,. \qquad (5.14)$$

This is isomorphic to the *quaternion group*[38] Q, and since it is a discrete subgroup of SU(2), $H/H_0=H$. Thus

$$\pi_1(R)=Q\,, \quad \text{(biaxial nematics)}\,. \qquad (5.15)$$

Note that Q is non-Abelian. The biaxial nematics are therefore an especially interesting system from the topological point of view. We defer a discussion of their defects until the general discussion of the non-Abelian case in Sec. VI.

Note the form analogous to Eq. (5.15) for the generalized biaxial nematics. A medium that can be viewed as a field of objects with a discrete symmetry group P whose subgroup of proper rotations[39] is P_0, has a fundamental group given by the lift of P_0 in SU(2)—i.e., the subgroup of SU(2) whose order is twice that of P_0 which is carried into P_0 under the homomorphism. Such subgroups of SU(2) are important in the theory of magnetism in crystals, where they are known as *double groups*.

5. Superfluid helium-3

As noted in Sec.IV.C.5, the isotropy subgroup of the dipole-locked A phase is the identity alone if G is taken to be SO(3), and is the two-element group (1, −1) if G is taken to be SU(2). Since H is a discrete subgroup of SU(2), H_0 is the identity, and $H/H_0=H$, the two-element group. Thus

$$\pi_1(R)=Z_2\,, \quad (\text{dipole-locked } {}^3\text{He-}A)\,. \qquad (5.16)$$

In the dipole-free A phase we noted that G can be taken to be the direct product of SO(3) with itself, in which case H consists of elements of the form $[R(z,\theta),1]$ and $[R(\hat{n},\pi),R(\hat{z},\pi)]$ for any axis $\hat{n}$ in the x-y plane. To construct a simply connected[40] G, we must replace each SO(3) by SU(2). To compute the fundamental group we must therefore determine the lift of H from SO(3)×SO(3) to SU(2)×SU(2).[41] The elements $[R(z,\theta),1]$ are lifted to elements

$$[u(\hat{z},\theta),1],[u(\hat{z},\theta),-1], \quad 0\leq\theta\leq 4\pi\,, \qquad (5.17)$$

of SU(2)×SU(2). [We do not need to specify an additional pair with a minus sign attached to $u(\hat{z},\theta)$ since $-u(\hat{z},\theta)=u(z,\theta+2\pi)$.] In treating the elements $[R(\hat{n},\pi),R(\hat{z},\pi)]$ it is convenient to represent rotations through π about arbitrary axes $\hat{n}$ perpendicular to $\hat{z}$ in the equivalent form of rotations through π about a fixed axis (taken here as the x axis) compounded with arbitrary rotations about $\hat{z}$. Doing this, we find that these elements of H lift to

$$\begin{aligned}&[u(\hat{x},\pi)u(\hat{z},\theta),u(\hat{z},\pi)]\,,\\&[u(\hat{x},\pi)u(\hat{z},\theta),-u(\hat{z},\pi)]\,, \quad 0\leq\theta\leq 4\pi\,.\end{aligned} \qquad (5.18)$$

If we now define an element g of SU(2)×SU(2) by

$$g=[u(\hat{x},\pi),u(\hat{z},\pi)]\,, \qquad (5.19)$$

then we can characterize the pieces of the subgroup H of SU(2)×SU(2) given by Eqs. (5.17) and (5.18) as follows:

The connected component of the identity H_0 is the first of the two sets of elements given in Eq. (5.17); the first of the two sets given in Eq. (5.18) is just the coset gH_0, the second in Eq. (5.17) is the coset g^2H_0, and the second in Eq. (5.18) is the coset g^3H_0. Thus the quotient group H/H_0 has the same structure as the cyclic group of order 4 generated by e, g, g^2, and g^3, and we conclude that

$$\pi_1(R)=Z_4\,, \quad (\text{dipole-free } {}^3\text{He-}A)\,. \qquad (5.20)$$

VI. MEDIA WITH NON-ABELIAN FUNDAMENTAL GROUPS

Two loops in order-parameter space are freely homotopic (i.e., one can be continuously slid about until it coincides with the other) if and only if they are characterized by the same conjugacy class of the fundamental group. The discussions in Secs. II.B and III.D on whether line defects can be transformed into one another by local surgery used no feature of order-parameter space topology beyond the organization of loops into equival-

[38] As originally defined by Hamilton, the *quaternion group* contains, in addition to ±1, elements $\pm i$, $\pm j$, and $\pm k$ satisfying $i^2=j^2=k^2=ijk=-1$. These relations suffice to determine the entire 8×8 multiplication table. They are satisfied by the identifications $i\longleftrightarrow i\,\sigma_x$, $k\longleftrightarrow i\,\sigma_y$, and $j\longleftrightarrow i\,\sigma_z$.

[39] Only the proper point symmetries are relevant because P_0 plays the role of the isotropy subgroup H of the full group G. To apply the theorem G must be connected, and can therefore include only proper rotations. The improper symmetries of the object are therefore irrelevant.

[40] The theorem (proved at the end of Sec. III) that the fundamental group of the product of two spaces is the product of their fundamental groups assures us that SU(2)×SU(2) is simply connected. (The product of the trivial group with itself remains the trivial group.)

[41] A cautionary note may be in order at this point. When dealing with SU(2)×SU(2) one must not be seduced into error by a notational convenience that is perfectly safe when used in a single SU(2). If g is a member of SU(2) then the member $-g$ is not the real number −1 times the group element g [though it may be in a particular representation of SU(2)]. For SU(2) is not an algebra but an abstract group and scalar multiplication is not defined. The group-theoretic meaning of the notation "$-g$" is this: there is an element f in SU(2) which is not the identity, but commutes with all the other group elements and satisfies $f^2=1$. The symbol "$-g$" is simply a shorthand notation for fg. The notation is used, because it follows from the properties of f that $(fg)(fh)=gh$ and $f(fg)=g$, and these are automatically taken care of by the minus sign notation. However, one cannot identify the element $[-g,-h]$ in the direct product SU(2)×SU(2) with the element $[g,h]$. Indeed, $[-g,-h]$ is more properly written as $[fg,fh]$. The two f's in this expression appear with different members of the direct product and cannot be combined to give unity. This may appear to be an obvious point, but in practice one almost always treats SU(2)×SU(2) as the product of the two corresponding groups of unitary matrices. Since these matrix groups happen to be algebras as well, it is very tempting (but wrong) to endow the direct product with this additional structure, and collapse two −1's associated with distinct elements. The confusion is compounded by the habit of identifying SU(2) (the abstract group) with SU(2) (the faithful representation of that abstract group by 2×2 matrices). Calling the latter structure "SU(2)" the problem arises from the fact that "SU(2)"×"SU(2)" is not a faithful representation of SU(2)×SU(2).

ence classes under free homotopy. Thus the criterion for the topological equivalence of planar spin defects in terms of winding number, which generalized to any Abelian medium[42] in terms of elements of the fundamental group, applies in the most general case in terms of conjugacy classes of the fundamental group:

Two line defects are topologically equivalent (in the sense that one can be given the core of the other by purely local surgery) if and only if they are characterized by the same conjugacy class of the fundamental group.

The implications of this in non-Abelian media are rather subtle. We shall examine the non-Abelian case in this section, using the biaxial nematic as an illustrative example.[43] In part A we describe in more detail the categories of biaxial nematic line defects; in part B we describe the combination law for line defects in non-Abelian media; and in part C we examine some curious things that can happen when one tries to pull one line defect across another in a non-Abelian medium.

A. More on the nature of line defects in biaxial nematics

We have noted (at the end of Sec. V.B.4) that a medium of objects with a discrete point group symmetry has for its fundamental group the lift in SU(2) of the proper subgroup of the point group. The smallest non-Abelian fundamental group one can construct in this way is the eight-element quaternion group Q, so the biaxial nematics are the simplest such non-Abelian medium.[44]

The elements of the quaternion group are given in terms of their representation by Pauli matrices in Eq. (5.14). They can be grouped into five conjugacy classes[45]:

$$C_0=\{1\},\ \overline{C}_0=\{-1\},$$
$$C_x=\{\pm i\sigma_x\},\ C_y=\{\pm i\sigma_y\},\ C_z=\{\pm i\sigma_z\}\ . \tag{6.1}$$

The class C_0 contains removable defects; $\overline{C}_0$ contains defects in which the object rotates about 360° as the line is encircled; the classes C_x, C_y, and C_z contain defects in which the rotation is through 180° about each of the three distinct symmetry axes.

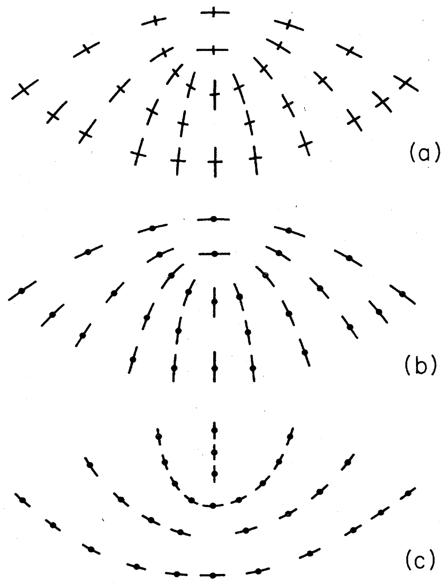

FIG. 35. Representatives of three distinct classes of line defects in the biaxial nematic. The object with the symmetry of a rectangular box is here represented as two mutually bisecting perpendicular sticks of unequal length. (a) C_z: a 180° disclination in both sticks. (b) C_x: a 180° disclination in the long stick at uniform short stick. (c) C_y: a 180° disclination in the short stick at uniform long stick.

Examples of these defects are shown in Fig. 35. For simplicity the figure shows not a rectangular box, but a pair of sticks of unequal length that perpendicularly bisect one another. The symmetry group is the same. The four nontrivial classes can then be described as a 360° disclination ($\overline{C}_0$), a 180° disclination in the long stick with no variation in the short (C_x), a 180° disclination in the short stick with no variation in the long (C_y), and a 180° disclination in both sticks (C_z). One might wonder why the 360° disclinations do not come in three varieties. The reason is that the trick that produced escape in the third dimension in the ordinary nematic (Sec. V.B.3) thereby rendering its 360° disclination trivial, has the effect in the biaxial nematic of converting the various candidates for distinct 360° disclinations into one another (Fig. 36). Note, also, that two defects described by distinct elements of the fundamental group

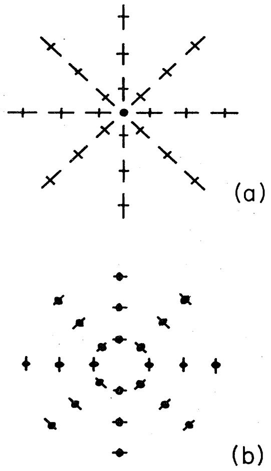

FIG. 36. The fourth class of line defects in the biaxial nematic, $\overline{C}_0$. It is shown in (a) as the 360° analogue of Fig. 35(a), and in (b) as the 360° analogue of Fig. 35(c). An attempt to make (a) escape in the third dimension by a 90° rotation about the short stick simply results in (b). An alternate escape route via a 90° rotation about the long axis results in the 360° analogue of Fig. 35(b).

[42] I shall sacrifice accuracy for brevity, referring to a medium whose order-parameter space has an Abelian (non-Abelian) fundamental group, as an "Abelian (non-Abelian) medium."

[43] Biaxial nematics have not yet been made in the laboratory. The claim has been made, however, that line defects in cholesteric liquid crystals have the same fundamental group. The validity of this, and similar assertions about media with broken translational symmetry, is considered in Sec. VIII.

[44] The non-Abelian abstract group of lowest order is the group D_3 of order 6. However, the only proper point group of order 3 is C_3, which lifts to the Abelian group C_6 in SU(2). Similar observations reveal that Q is the only non-Abelian group that can be reached by lifting a four-element proper point group to SU(2).

[45] This follows directly from the fact that the Pauli matrices anticommute and give unity when squared. Thus, for example, $(i\sigma_y)^{-1}=-i\sigma_y$, and hence $(i\sigma_y)^{-1}(i\sigma_x)(i\sigma_y)=-i\sigma_x$.

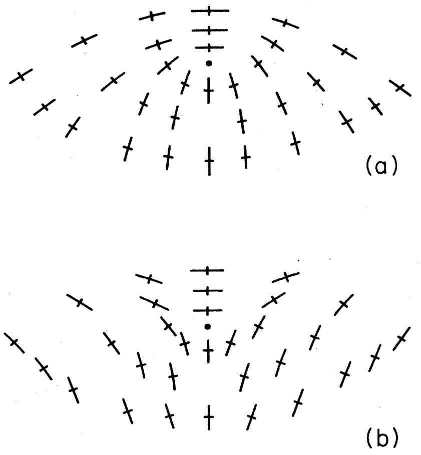

FIG. 37. Positive (a) and negative (b) 180° disclinations of the C_x type. One can be continuously transformed into the other by applying to every molecule the same 180° rotation about a horizontal axis in the plane of the page. Note that (a) and (b) are associated with distinct elements of the fundamental group, but the same conjugacy class.

in the same conjugacy class, can indeed be continuously converted into one another, as illustrated in Fig. 37.

With defects characterized by conjugacy classes of the fundamental group instead of distinct group elements, one might expect the combination law to be related to a multiplication table for conjugacy classes rather than to the group multiplication table itself. This expectation is confirmed in part B. Before turning to this we note some purely algebraic features of class multiplication.

The product of two conjugacy classes of a group is defined to be the set containing all the products of all the pairs of elements from the two classes. If a given group element occurs in more than one way as such a product it is taken to appear that many times in the product set (so that the product set contains as many elements as the product of the number of elements in the two classes). It is a simple exercise in elementary group theory to show that the product of any two classes consists itself of one or more whole classes, and that class multiplication is commutative, even though the group itself may be non-Abelian. The product of any two classes can therefore be specified by indicating the number of times each class of the entire group is present in the product set. This is usually written in an additive notation, so that, for example, the equation $C_1C_2 = C_3+4C_6+3C_8$ means that in the set of all products of pairs from the classes C_1 and C_2 each element of class C_3 is to be found once, each element of C_6 is to be found four times, each element of C_8 is to be found three times, and no elements from any other classes are to be found.

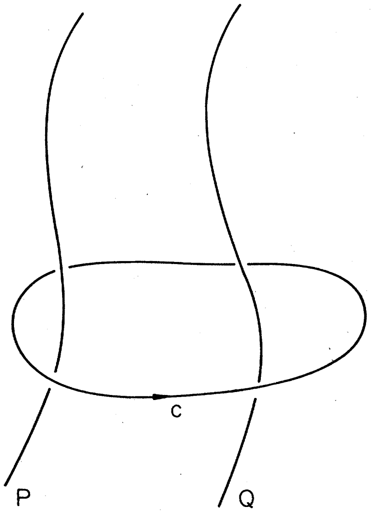

FIG. 38. Two line defects P and Q and a surrounding real-space contour c.

The class multiplication table for the quaternion group is particularly simple:

	C_0	$\overline{C}_0$	C_x	C_y	C_z
C_0	C_0	$\overline{C}_0$	C_x	C_y	C_z
$\overline{C}_0$	$\overline{C}_0$	C_0	C_x	C_y	C_z
C_x	C_x	C_x	$2C_0+2\overline{C}_0$	$2C_z$	$2C_y$
C_y	C_y	C_y	$2C_z$	$2C_0+2\overline{C}_0$	$2C_x$
C_z	C_z	C_z	$2C_y$	$2C_x$	$2C_0+2\overline{C}_0$

(6.2)

Note that the product of two classes determines a unique class in all cases except for the product of any of the three distinct 180° disclinations with itself, where the result is a combination of the trivial class and the 360° class. This ambiguity is to be expected. On the one hand two identical 180° defects will clearly combine to give a 360° defect. On the other hand a 180° defect and the −180° defect that annihilates it are in the same class. These observations account for the ambiguity, but the question remains of how one is to tell what results when such a pair of defects merge.

B. The combination of defects in the non-Abelian case

That this is a matter of some delicacy in the non-Abelian case can already be seen when one draws a loop around the pair. Who is to say whether the loop is to be viewed as first encircling P, then Q, or the other way around? The characterization of defects by conjugacy classes resolves this dilema, for ba is in the same conjugacy class as ab $(=a(ba)a^{-1})$.

We may therefore draw a real space contour surrounding both defects (Fig. 38) without concern for "the order in which they are surrounded." The image of that contour in order-parameter space provided by the values of the order parameter along the contour is characterized by a unique conjugacy class of the fundamental group. Without altering the configuration outside the loop we can replace the configuration of the order parameter inside the loop by that of any single defect in that conjugacy class.[46] To arrive at a combination law we need to know how the conjugacy class characterizing the image of the encircling contour is related to the conjugacy classes characterizing the pair of defects the contour surrounds.

Ambiguities are absent when one deals with based fundamental groups. If all loops in order-parameter space are required to share a common point then the combination law is simply given by ordinary group multiplication of individual elements in that particular based fundamental group. We therefore single out a point x_0 on the encircling real-space contour, and take the value of the order parameter $f(x_0)=f_0$ as our base point in order-parameter space. The contour is then mapped into a loop in $\pi_1(R, f_0)$ that is in the same homotopy class as the product of the two loops at f_0 determined by the two separate singularities. When the base point is abandoned each identification of a loop with a distinct element of the fundamental group must be replaced by an identification with the entire conjugacy class containing that element. Since the conjugacy class containing a product of elements must be contained in the product of the conjugacy classes containing those elements, we can conclude that when two line defects are combined the resulting defect must be characterized by a conjugacy class contained in the product of the two conjugacy classes characterizing the original defects. When these two conjugacy classes have a unique class in their product, the combination law is unambiguous.

It can happen, however (as in the case of two 180° disclinations of the same kind in the biaxial nematic), that the conjugacy classes characterizing a pair of defects combine to give more than a single conjugacy class. Since defects in different conjugacy classes cannot be transformed into each other by local surgery, for a given pair of defects each surrounding contour must be characterized by a unique conjugacy class. An ambiguity in class multiplication can therefore only mean that the class of the combined defect can depend on the choice of the surrounding contour. When the class multiplication table does not specify a unique product, the class of the composite defect depends on where one chooses to perform the local surgery.

In considering this fact it helps to think of the local surgery as being performed simply by bringing the two line defects closer and closer together until they coalesce into a single one. Specifying the surrounding contour sets limits on the choice of paths along which the defects can be brought together; they must be confined to the interior of the contour. Instead of specifying a contour within which local surgery is to be performed, we can just as well specify a path along which the two defects are to be brought together. The contour, if we wished to have it back, could then be taken to be any one that encircled both the defects and the line joining them.

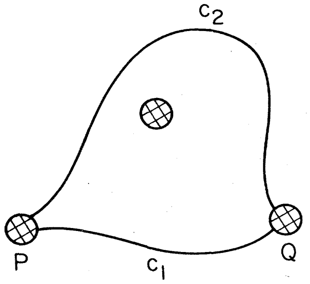

FIG. 39. Two paths c_1 and c_2 connecting defects P and Q on opposite sides of a third defect.

As long as two paths joining the defects can be deformed into one another, the class of the combined defect cannot depend on the choice of path. Such a deformation will be possible unless (Fig. 39) the two paths are on opposite sides of some third defect. We conclude that when class multiplication fails to provide a unique product class, the various possible forms for the product defect are associated with the various ways in which the path along which the defects are brought together winds its way among whatever other defects are present.

Rather than spelling this out further in confusing generality, we illustrate the point in the case of the biaxial nematic. Two identical 180° x disclinations will, of course, combine to give the 360° disclination in the absence of any other line defects. If, however, a 180° y or z disclination is also present, the x disclination can be converted into its antidefect by transport about a closed path surrounding the y or z disclination. [This is the topological content of the algebraic identity: $-(i\sigma_x)=(i\sigma_y)(i\sigma_x)(i\sigma_y)^{-1}$.] Thus looping one x disclination around the y or z disclination before combining it with the other alters the result from the 360° defect to the trivial one. The general point is much the same: If many defects are present, the element of a based homotopy group representing a given defect will depend on the particular way in which the based loop that surrounds it weaves its way through the forest of other defects; this may lead, in the non-Abelian case, to a corresponding path dependence in some of the combination laws.

Note, in passing, that this observation can be turned upside down, leading to the conclusion, in the biaxial nematic, that any 180° disclination can catalyze the topological decay of a 360° one. For we need only to dissociate the 360° disclination into two identical 180° ones of a type different from the catalyst, and then bring these together around opposite sides of the catalyst, thereby bringing about their mutual annihilation.

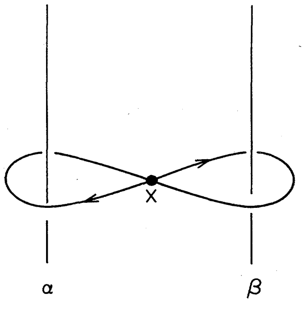

FIG. 40. Two line defects surrounded by two contours with a common point. If the value of the order parameter at x is fixed at f_0, then the values of the order parameter on each contour determine homotopy classes α and β in $\pi_1(R, f_0)$ which are used to label the defects.

[46] To avoid clumsy complications irrelevant to the point at hand, we revert to a nomenclature appropriate to point defects in two dimensions for the rest of this subsection. In three dimensions one must deal with a cylindrical locus of loops surrounding a line defect. The point to be made about non-Abelian order-parameter spaces remains the same.

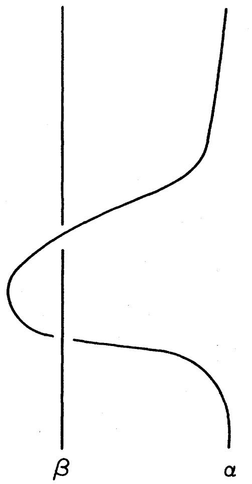

FIG. 41. The line α is moved across the line β.

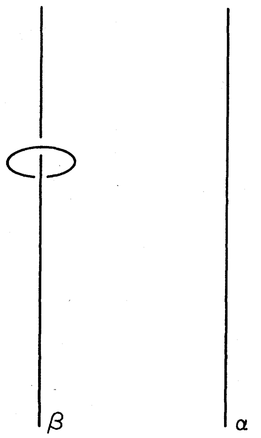

FIG. 43. Configuration equivalent to that of Fig. 42 if the contour in Fig. 42 has a nulhomotopic image in order-parameter space.

C. The entanglement of line defects

Poênaru and Toulouse (1977) have pointed out some intriguing behavior that can arise when two line defects are made to cross one another in a medium with a non-Abelian fundamental group. Two such defects, initially rectilinear and far apart, are shown in Fig. 40. We chose a point **x** of physical space at which the order parameter has the value f. We assume (and this can easily be arranged without diminishing the generality of our conclusions) that **x** and the value f of the order parameter at **x** are fixed throughout the manipulations that follow. The values of the order parameter on real-space contours passing through **x** determine maps of those contours into order-parameter space based at f. Those loops in order-parameter space represent classes α and β of the based fundamental group $\pi_1(R,f)$ and the line defects themselves are characterized as being of type α and β, with respect to the base point f.

Suppose one now attempts physically to deform the order parameter in such a way that the top part of α moves over β while the bottom part of α moves under β, leading to the configuration in Fig. 41. If the two lines in Fig. 40 are not strictly coplanar (for example, if they are at 90° to one another) precisely this type of entanglement will result when they try to pass one another.

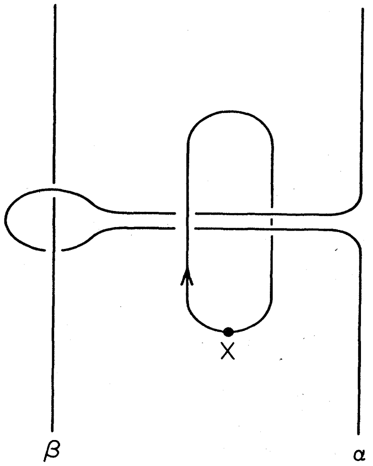

FIG. 42. An attempt, after the crossing shown in Fig. 41, to reconstitute two separate line defects. Whether or not this can be done depends on whether or not the image in order-parameter space of the contour at x is homotopic to a constant.

Let us now attempt to make the resulting configuration look as much as possible like a simple interchange of the original pair of rectilinear lines. We deform Fig. 41 to Fig. 42 which can be regarded as two straight lines β and α, a pair of horizontal lines joining them, and two somewhat complicated but completely localized regions in which the pair of horizontal lines join up with β and α. If the two horizontal lines can annihilate one another, then with purely local surgery we can arrive at the configuration in Fig. 43. (The little loop aroung β can be removed by additional local surgery.) The lines will then have passed through one another without leaving any topologically stable traces, and can continue on their way as if they had passed without getting entangled.

If, however, the pair of horizontal lines in Fig. 42 do not jointly constitute a topologically trivial linear defect, then one cannot avoid an additional singular line connecting α and β, as in Fig. 44. In this case the crossing leaves a spectacular scar in the medium, in the form of a third topologically stable line singularity

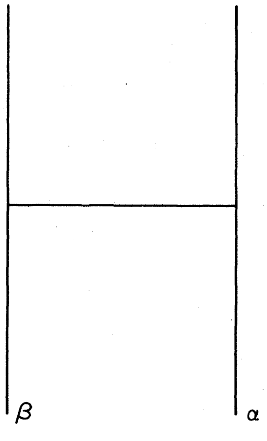

FIG. 44. Result of the attempt at moving two lines across one another if the contour in Fig. 42 has an image in order-parameter space that is not homotopic to a constant.

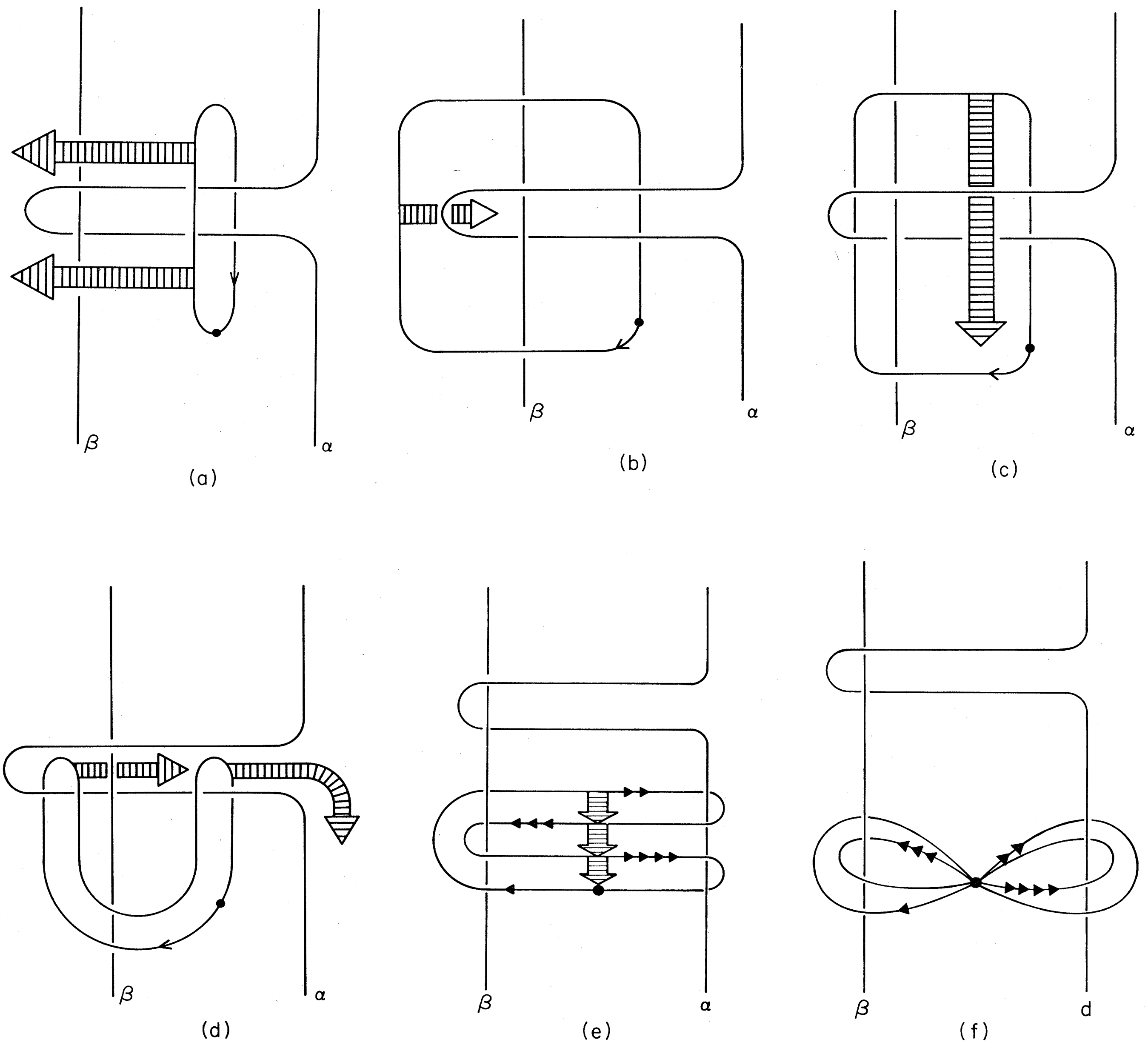

FIG. 45. A series of continuous deformations of the contour in Fig. 42 demonstrating that its image in order-parameter space is homotopic to the loop $\beta \circ \alpha \circ \beta^{-1} \circ \alpha^{-1}$. The straight line β and the bent line α should be regarded as rigid and immobile. The contour should be thought of as flexible and elastic. The striped arrows indicate successive deformations of the contour with the usual convention that at intersections the solid line lies over the broken line. A point of the contour has been left tied to the point x at all stages, so the homotopies are based at x. The final step from (e) and (f) pinches four points of the contour together at x. Comparison with Fig. 40 reveals that the final contour in (f) is just $\beta \circ \alpha \circ \beta^{-1} \circ \alpha^{-1}$.

connecting the other two. The energy associated with this third singular line will grow linearly with any additional separation between α and β, and will constitute a considerable physical barrier against their further separation. Since the mobility of line defects can play an essential role in determining the macroscopic properties of a medium (a spectacular example being the role played by dislocations in the deformation of crystals) it is important to be able to determine whether a given pair of line defects can (Fig. 43) or cannot (Fig. 44) be freely passed through one another.

Whether Fig. 43 or Fig. 44 results depends entirely on whether the loop in order-parameter space provided by the contour encircling the double line in Fig. 42, is or is not homotopic to a constant. The homotopy class of this loop in $\pi_1(R,f)$ can be related to the homotopy classes α and β characterizing the original pair of lines through the series of x-based deformations of the contour shown in Fig. 45.[47] These deformations demonstrate that the contour in Fig. 42 is homotopic at **x** to the contour in Fig. 45(f). Comparing the latter contour with those in Fig. 40 used to define the homotopy classes α and β of the original lines, we conclude that the homotopy class in $\pi_1(R,f)$ determined by the contour in Fig. 45(f) (and hence by the contour in Fig. 42) is just the product $\beta\alpha\beta^{-1}\alpha^{-1}$.

We conclude that α and β can cross without the production of an additional line defect connecting them, if and only if $\beta\alpha\beta^{-1}\alpha^{-1}$ is in the homotopy class of the identity

$$\beta\alpha\beta^{-1}\alpha^{-1}=1\,. \tag{6.3}$$

This, in turn, will hold if and only if α and β are commuting elements of $\pi_1(R,f)$. Thus two line defects can be made to cross one another without leaving traces (in the form of a connecting umbilical cord) if and only if they are characterized by commuting elements of the fundamental group.[48]

In media with Abelian fundamental groups all line defects can cross one another without producing the umbilical cord of Fig. 44. In the non-Abelian case of the biaxial nematic, an examination of the multiplication table for the quaternion group reveals that the only values assumed by $\beta\alpha\beta^{-1}\alpha^{-1}$ are 1 and -1, for any pair of elements. The homotopy class -1 characterizes the 360° disclination, and noncommuting pairs correspond to 180° disclinations of distinct types. We conclude that line singularities in a biaxial nematic can cross without the production of a connecting line except for the case of two 180° disclinations of distinct types, which are necessarily joined after crossing by a 360° disclination.

Arriving at these conclusions without the aid of homotopy groups requires a higher order of geometrical imagination than I, at least, possess; I commend them to the attention of those who suspect that the use of homotopy groups simply obscures with intricate and arid formalism what would otherwise be intuitively clear.

VII. THE SECOND HOMOTOPY GROUP AND THE CLASSIFICATION OF POINT DEFECTS IN THREE DIMENSIONS

We have focused exclusively on the question of line defects in three dimensions (or point defects in two). However similar considerations can be brought to bear on classifying and determining the combination laws for point defects in three-dimensional media. If a medium in three-dimensional space is everywhere continuous except, perhaps, at a single point P, then on any spherical surface surrounding P the order parameter will be continuous, thereby providing a continuous mapping of a sphere into the order-parameter space. By straightforward repetitions of the arguments used for line defects, one sees that point defects will be characterized by freely homotopic classes of maps of spheres into order-parameter space. Trivial (or "removable" or "topologically unstable") point defects are associated with mappings that can be deformed to the constant map—i.e., the image of the sphere in the order-parameter space R provided by the surrounding field can be continuously shrunk to a point. More generally, one point defect can be given the core of another using purely local surgery, if and only if they correspond to the same homotopy class.

It turns out that a group structure can also be imposed on the homotopy classes of maps of spheres into order-parameter space. This is the so-called *second homotopy group*, $\pi_2(R)$. As with the fundamental group $\pi_1(R)$, the second homotopy group is best introduced through the intermediary of the second homotopy groups $\pi_2(R,x)$ associated with a base point x in order-parameter space. These are described in part A. The second homotopy group $\pi_2(R)$ itself is introduced in part B, along with certain "path automorphism classes" into which it can be sectioned, which play a role analogous to conjugacy classes in $\pi_1(R)$. In part C we give an algorithm for the computation of $\pi_2(R)$ analogous to that given in Sec. V.A for the computation of $\pi_1(R)$. Because the classes of elements of $\pi_2(R)$ associated with a given type of point defect are not simply conjugacy classes[49] a further algorithm, given in part D, is required for their computation. In part E these results are illustrated through applications to the standard examples.

A. The second homotopy group at x, $\pi_2(R,x)$

Though point defects in three dimensions are characterized by freely homotopic maps of spheres into the order-parameter space R, it is again useful to introduce the intermediate notion of based mappings and homotopy with respect to a base point. We therefore consider continuous maps of a sphere into R with the restriction that the image of the sphere should contain the point x of order-parameter space; i.e., we consider spheres[50] tied down to x, just as we considered loops at x in constructing the based fundamental group.

In the case of loops, it was convenient to view a loop based at x as a map $f(z)$ of the unit interval, $0\le z\le 1$ into R, subject to the restriction $f(0)=f(1)=x$; i.e., the loop was formed by joining together the two ends of a line segment. In a similar way, we shall regard the images of spheres in R as being given by mappings $f(u,v)$ of the unit square, $0\le u,v\le 1$ into R, subject to the restriction that f take the entire circumference of the square into the single point x: $f(0,v)=f(1,v)=f(u,0)=f(u,1)=x$. The sphere is thus represented by closing

[47] Readers who have difficulty following the figure are urged to construct the lines α and β out of stiff wire, introduce the contour in the form of a loop of string about the horizontal wires, and execute by hand the motions of the contour leading to Fig. 45(e).

[48] Since fundamental groups at different base points are related by path isomorphisms, the validity of the condition $\alpha\beta=\beta\alpha$ is independent of basepoint and can be determined from the structure of the fundamental group itself.

[49] Indeed, we shall see (part A) that second homotopy groups are always Abelian, so that conjugacy classes are just individual group elements.

[50] Lacking a word which is to "sphere" as "loop" is to "circle" I shall simply continue to use the word "sphere" itself, after warning the reader that spheres in order-parameter space can be very floppy spheres, just as loops are very floppy circles.

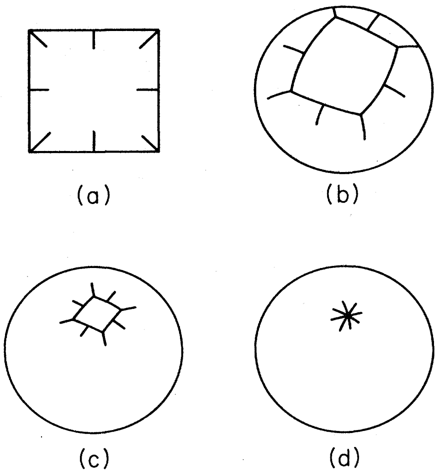

FIG. 46. A sphere can be represented as a square with all boundary points identified. The square is shown in (a) with some stripes to identify its interior. In (b) the square has been puffed up (in three dimensions) into the surface of a sphere with a square hole, given by the original boundary. In (c) the square hole has been shrunk and in (d) it is reduced to a single point. Because all boundary points of the square are identified, the point in (d) is nonsingular.

up the border of the square, as if by pulling on purse strings (Fig. 46).

The convenience of this representation is that it provides a natural definition of the *product of two maps of a sphere* into R at x: One simply joins together the two squares representing the domains of the two maps f and g, and compresses the resulting oblong back into a square, to get the map $f\circ g$. Analytically, the definition is almost exactly the same as for the product of two loops [cf. Eq. (3.2)]:

$$\begin{aligned} f\circ g(u,v) &= f(2u,v), \quad 0\leq u\leq \tfrac{1}{2}, \\ &= g(2u-1,v), \quad \tfrac{1}{2}\leq u\leq 1. \end{aligned} \tag{7.1}$$

This rule for forming the product can be represented pictorially as in Fig. 47, whose conventions we shall revert to on occasion for the provision of intuitive pictorial "proofs" of results whose verbal or analytical demonstration could be quite cumbersome.

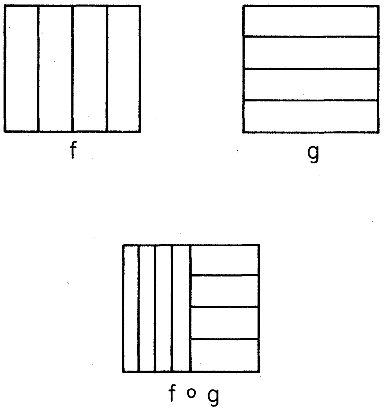

FIG. 47. Rule for forming the product of two maps of squares into R at x.

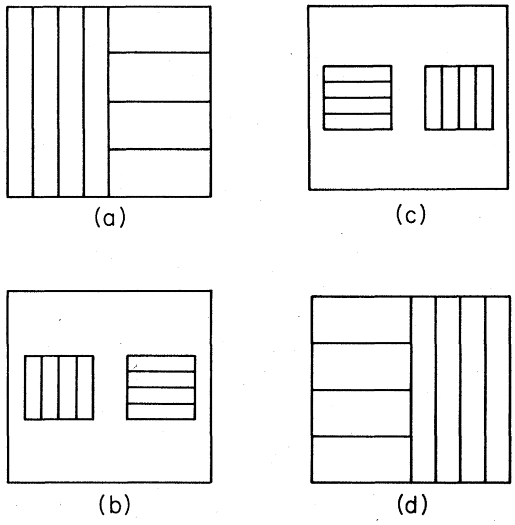

FIG. 48. Proof that the product rule in Fig. 47 is commutative. The product $f\circ g$ is shown in (a). The boundaries of both the vertically and horizontally striped rectangles are taken into x. In (b) both rectangles have been continuously deformed into smaller squares, imbedded in a sea of white, all of which is taken into the single point x. Next (b) may be further deformed by moving the little squares about within the big one. In this way their positions can be interchanged and the resulting figure (c) can then be reexpanded to give the product $g\circ f$ (d). Note that the argument can also be used to show that the homotopy class of the product defined in Fig. 47 does not depend on whether the squares are joined side to side or top to bottom, etc.

Routine extensions of the arguments given in Sec. III.A establish that multiplication can be defined for homotopy classes of maps of spheres into R at x, the class containing the product of any two representative maps being independent of the choice of representatives. The identity in the group is represented by the map that takes the sphere into a single point, and the inverse of the map $f(u,v)$ is $f(-u,v)$. This group of homotopy classes of maps of spheres into R at x is called the second homotopy group, $\pi_2(R,x)$.

In contrast to the fundamental group, the second homotopy group is always commutative. This is demonstrated in Fig. 48, which illustrates how to construct an explicit homotopy between $f\circ g$ and $g\circ f$. Unfortunately, as we shall see, this does not imply that the group multiplication table by itself gives the combination law for point defects. Classes of group elements again play the central role, but they are no longer conjugacy classes. Their nature emerges in the course of examining the relation between the based second homotopy groups $\pi_2(R,x)$, and the second homotopy group $\pi_2(R)$.

B. The second homotopy group, $\pi_2(R)$

We next establish that the second homotopy groups at different base points are isomorphic. This permits the introduction of the abstract second homotopy group, $\pi_2(R)$, of which the based homotopy groups are isomorphic copies. In the case of the fundamental group $\pi_1(R)$, these isomorphisms are unique unless $\pi_1(R)$ is noncommutative. In the case of $\pi_2(R)$, even though second homotopy groups are always commutative and even if $\pi_1(R)$ should be commutative, there still need not necessarily be a unique isomorphic mapping between second homotopy groups at different base points.

The extent to which the isomorphisms between based second homotopy groups are not unique bears directly on the question of how the physically pertinent classes of freely homotopic maps of spheres into R are related to the second homotopy group. It also enters into determining the laws governing the combination of point defects. The issues are quite similar to those we encountered in Secs. III.B, C, and D, and VI.B. I shall therefore present them rather more sketchily, except insofar as the peculiar features of $\pi_2(R)$ play a special role.

Given a path $c(z)$ connecting two points x and y in the order-parameter space R, we can construct a correspondence between $\pi_2(R,x)$ and $\pi_2(R,y)$ by simply joining any sphere at x to the point y by means of the line $c(z)$, and regarding the resulting "balloon plus string" as a mapping of a sphere into R at y, which is rather degenerate along the string (Fig. 49). If f is a mapping of a sphere into R representing a homotopy class in $\pi_2(R,x)$, then we denote the mapping at y constructed in this way as $c(f)$. The notation is intended to suggest that the path c acts on the map f based at x to produce the map $c(f)$ based at y.

We can give a more formal construction of $c(f)$. Take the unit square in the $u-v$ plane and inscribe in it a square half as big at its center. On the inner square define $c(f)$ to act precisely as f acts on its entire square, so that $c(f)$ takes the inner square into exactly the same balloon at x as f takes its entire square into. Divide the remaining part of the square between the inner and outer circumferences into a family of square circumferences growing continuously from the inner to the outer one, as z goes from 0 to 1. Let $c(f)$ take the circumference parametrized by z into the single point $c(z)$, as indicated in Fig. 50.

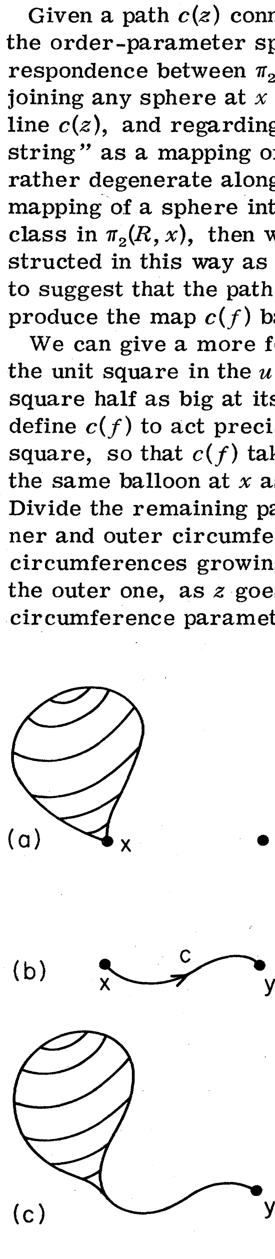

FIG. 49. (a) A sphere at x and a second point y. (b) A path c joining x to y. (c) The sphere at y given by the action of c on the sphere at x. Note that part of the sphere at y has degenerated to a line, but it can be "reinflated" by an appropriate homotopy.

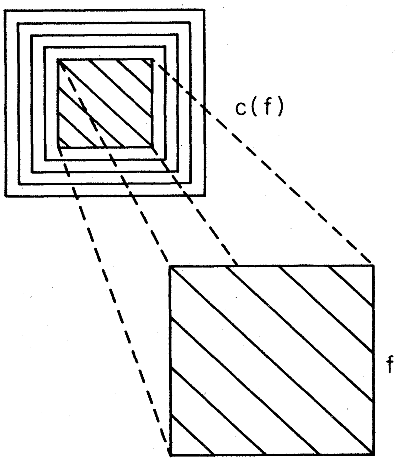

FIG. 50. The diagonally striped square in the lower right represents a sphere in order-parameter space, the entire circumference being taken into the single point x. The square in the upper left represents the sphere $c(f)$ at y. The inner part of that square is just the striped square at x, scaled down in size. It is surrounded by a family of square circumferences which expand outward to fill the rest of the square. Each circumference is taken into a single point along the curve c from x to y, the innermost going into x and the outermost to y.

By drawing pictures of the appropriate squares one immediately sees that if f and g are in the same homotopy class of $\pi_2(R,x)$ then $c(f)$ and $c(g)$ will be in the same homotopy class of $\pi_2(R,y)$, and conversely.[51] This permits one to define the operation of c on an entire homotopy class α, $c(\alpha)$ being the class containing $c(f)$, where f is any representative map of α. An equally simple result is that the product of the maps $c(f)$ and $c(g)$ at y is homotopic to the map $c(fg)$. Consequently the map of $\pi_2(R,x)$ onto $\pi_2(R,y)$ given by $\alpha\rightarrow c(\alpha)$ is an isomorphism.

This isomorphism will be unique if and only if the automorphism of $\pi_2(R,x)$ onto itself produced by any closed path starting and ending at x, is the identity automorphism; i.e., if and only if any balloon at x is homotopic at x to the balloon produced by forming a loop at x out of the balloon string as in Fig. 51. As we shall see, such a homotopy need not, in general, exist. If there are balloons and paths for which the homotopy does not exist, then $\pi_1(R,x)$ will act as a nontrivial group of automorphisms on $\pi_2(R,x)$.[52] If the group of automorphisms is trivial, consisting only of the identity—i.e., if the homotopy between balloon and balloon + looped string always exists—then R is said to be *2-simple*.

The single abstract group $\pi_2(R)$, of which the based

[51] To establish the converse regard f as $c^{-1}[c(f)]$ and similarly for g, and apply the original theorem. [The homotopy between f and $c^{-1}(c(f))$ consists simply of retracting the double string back into the balloon.]

[52] The automorphism group need not be isomorphic to π_1. However, the automorphism given by the product of two loops will coincide with the product of the corresponding automorphisms, so the automorphism group will be a homomorphic image of π_1.

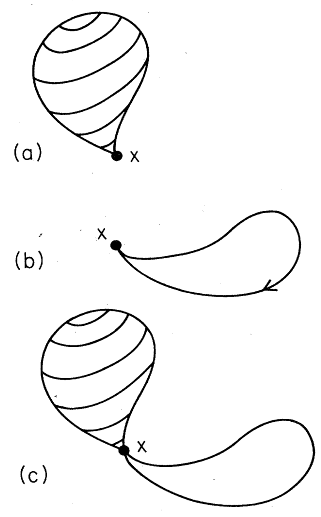

FIG. 51. (a) A sphere at x. (b) A loop at x. (c) The sphere at x formed by the action of the loop in (b) on the sphere in (a). Part of the sphere is degenerate, being only a line.

groups $\pi_2(R, x)$ are isomorphic copies, is known as the second homotopy group of R. By arguments that are simple generalizations of those in Sec. III.C, one establishes that point defects are in one-to-one correspondence with the elements of $\pi_2(R)$ (i.e., freely homotopic maps of spheres into R are in one-to-one correspondence with based maps) if and only if the space R is 2-simple [i.e., if and only if the only loop automorphism of $\pi_2(R)$ is the identity automorphism].[53] More generally, if R is not 2-simple, i.e., if $\pi_1(R)$ acts as a nontrivial group of automorphisms on $\pi_2(R)$, then the point defects are characterized by automorphism classes of elements of $\pi_2(R)$ under the group of loop automorphisms provided by π_1. By the term "automorphism classes" we mean the following: Given a group A of automorphisms of a group G, one easily verifies that under the automorphisms the elements of G split up into disjoint classes with the properties (i) that if two elements of G are in the same class then there is an automorphism in A taking one into the other and (ii) no automorphism in A takes an element in one class of G into another.

Note the similarity to conjugacy classes, which are just the automorphism classes of G under the group of inner automorphisms. From this point of view the association of distinct line defects with the conjugacy classes of π_1 is strictly analogous to the association of distinct point defects with the loop automorphism classes of π_2: The conjugacy classes of π_1 *are* its loop automorphism classes.

The bearing of these conclusions on the combination of point defects closely follows the analogous considerations on the combination of line defects given in Secs. III.D and VI.B. If the order-parameter space R is 2-simple then point defects correspond to elements of π_2, and defects corresponding to elements a and b of π_2 can only combine to give a defect corresponding to the product element ab. If, however, R is not 2-simple, then defects correspond to automorphism classes of π_2. One easily establishes that the set of all products of elements from two given automorphism classes is a union of whole automorphism classes, but in general it will contain more than one such class. When this happens the result of combining two point defects will depend upon the path along which they are brought together, it being possible to alter the result by changing the way in which the path weaves among whatever line defects are present. This behavior is quite analogous to that described in Sec. VI.B, though I emphasize once again, that it can arise even though $\pi_2(R)$ is necessarily Abelian and even if $\pi_1(R)$ is also Abelian; what matters for the path dependence is that π_1 should give a nontrivial group of automorphisms on π_2. This is particularly well illustrated in the case of the ordinary nematic, described in part E, below.

C. The fundamental theorem on the second homotopy group of coset spaces

We now examine a theorem, analogous to that described for the fundamental group in Sec. V.A, which immediately yields the second homotopy groups for all of our standard examples. We continue to represent the order-parameter space R as the coset space G/H of a simply connected group G. The theorem then asserts that $\pi_2(G/H)$ is isomorphic to $\pi_1(H_0)$, where H_0 is the connected component of the identity in the isotropy subgroup H. By using the theorem we therefore can reduce the computation of π_2 for a coset space to the computation of π_1 for a group. The latter computation, however, is easily achieved by the techniques we have already developed (though in all the cases we examine, the connectivity of H_0 is so elementary that no further computation is required).

For the theorem to hold it is necessary, as in the theorem of Sec. V, that G be simply connected. As discussed there this can always be arranged by taking G to be a universal covering group. It is also necessary (as was not the case for the theorem of Sec. V) that the second homotopy group of G should be trivial: $\pi_2(G)=0$. This additional restriction presents no practical difficulties, because of a theorem of Cartan (1936) that π_2 vanishes for any compact Lie group.

The theorem of Cartan is not an easy one; indeed, even succinct statements of it are not easily found in the mathematics literature that I have perused. However, in all the cases we shall be interested in we do not require Cartan's theorem in its full generality. The group G will be either SU(2) [the simply connected group of which SO(3) is the homomorphic image], T(1) [the simply connected group of which SO(2) is the homomorphic image], or products of these. One easily establishes that the product of two spaces with vanishing π_2 also has vanishing π_2. Any mapping of a sphere into the real line [the parameter space for T(1)] can be continuously deformed to a single point by a scale transformation. Any mapping of a sphere (S_2) into the surface of a 4-sphere [S_3—the parameter space for SU(2)] can be shrunk to a point by the construction described in Sec. II.C for maps of a circle (S_1) into S_2, escalated up by one dimension. We can therefore proceed to the

[53] We have established for each x that $\pi_1(R,x)$ is homomorphic to a group of automorphisms on $\pi_2(R,x)$. Further contemplation of path isomorphisms establishes that the structure of this automorphism group does not depend on the base point, so one can speak generally of *the action of π_1 on π_2*.

proof of the theorem, assuming that both $\pi_1(G)$ and $\pi_2(G)$ are 0.

As in the theorem on the computation of the fundamental group, it suffices here to consider a based second homotopy group in coset space G/H, taking as basepoint the isotropy subgroup H itself. Also in analogy to our earlier theorem, the basis for the proof, and for an intuitive understanding of the result, is provided by a suitable representation of the appropriate maps of squares into coset space, by corresponding maps of squares into the group G itself.

Consider a map $g(u, v)$ taking the square $0 \leq u, v \leq 1$ into the group G. This determines a map $K(u, v) = g(u, v)H$ in coset space, which $g(u, v)$ is said to represent. If K is to be a sphere in G/H at H, then g must take the entire circumference of the square into H. Since that circumference is connected, it will be taken into a single connected piece H_1 of H. If H_1 is not the connected component H_0 of the identity in H, then we can replace $g(u, v)$ by $g(u, v)h$, where h is the inverse of any element in H_1. The replacement leaves $K(u, v)$ unchanged (since $hH = H$) and gives us a new g that takes the circumference of the square into H_0.

It can be shown that any sphere at H in coset space can be represented in the form

$$K(u, v) = g(u, v)H, \tag{7.2}$$

where g takes the circumference of the square into H_0. This is almost evident when one regards coset space as order-parameter space, and notes that the problem is simply that of specifying a two-parameter continuous family of transformations in G that will spread out the reference value of the order parameter into the given sphere in order-parameter space. However, a proof, even on the primitive level of rigor I am willing to settle for, would be neither graceful nor informative. Readers disposed to pursue the point further would be well advised to start with the covering homotopy theorem, as given in the Lemma on p. 372 of the second edition of Pontryagin (1966), which has our result as a direct corollary.

We therefore take as our starting point the representation (7.2) of spheres at H in coset space by maps of squares into G which take the circumference into H_0, or, to introduce a more compact terminology, by maps of squares into G, H_0.[54] The theorem is based on the following four observations:

(1) Two homotopic maps of a sphere into G/H at H, are represented by homotopic maps of squares into G, H_0. This is "obvious" to the same degree as is the general validity of the representation (7.2); readers who require more convincing must, again, learn the covering homotopy theorem. Given this, it follows that a homotopy class of maps of spheres into G/H at H can be associated with a single homotopy class of maps of loops into H_0. A representative of that class is given by restricting the corresponding maps of squares into G, H_0 to the circumference. (The homotopy between any two such representative loops in H_0 is provided by a similar restriction of the homotopies between the corresponding squares.)

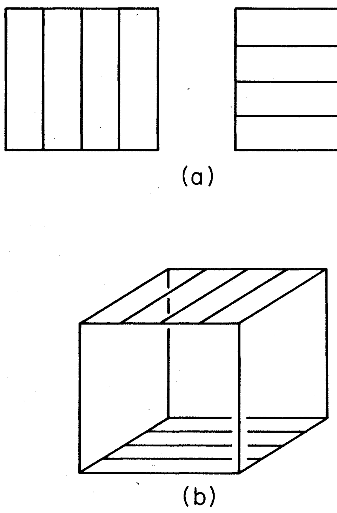

FIG. 52. (a) Two maps of squares into G, the circumference of each square being taken into H_0. If the circumferences in (a) are homotopic in H_0, then the homotopy can be used to construct a map (b) of the surface of a cube into G which agrees with the maps in (a) on the top and bottom faces and takes the four vertical faces into H_0.

(2) The preceding observation associates with each element of $\pi_2(G/H)$, a unique element of $\pi_1(H_0)$. We next establish that this association is one-to-one: homotopic loops in H_0 must come from homotopic spheres in G/H at H. This result is specific to the case $\pi_2(G) = 0$. Consider two squares that map into G, H_0, representing spheres in G/H at H. If the images of the circumferences of the squares in H_0 are homotopic, then the homotopy can be regarded as an extension of the maps of the two squares to the four walls of a rectangular cylinder of which the squares form the top and bottom (Fig. 52). Since $\pi_2(G) = 0$, the image of this surface (which is topologically equivalent to a sphere) can be deformed to a point in G. This deformation, in turn, can be regarded as an extension of the mapping from the entire surface of the rectangular cylinder to its interior. Finally, this mapping of the solid rectangular cylinder into G, which takes all four walls into H_0, can be regarded as a homotopy between the image in G, H_0 of the squares at the top and the bottom. But such a homotopy is precisely what is required to establish the homotopy of the original spheres in G/H.

(3) The preceding two observations establish a one-to-one mapping of $\pi_2(G/H)$ into $\pi_1(H_0)$. We next establish that the mapping is onto, i.e., that any loop in H_0 can be so associated with a sphere in G/H at H. This result is specific to the case $\pi_1(G) = 0$. Since G is simply connected a loop in H_0 (which is, of course, also a loop in the larger set G) can be shrunk to a point in G. The homotopy that specifies the shrinking, however, can be viewed (Fig. 53) as a map of a square into G that takes

[54] We shall also take the term "homotopy" to mean "homotopy via a family of maps that takes the circumference into H_0 at every stage" whenever the term is applied to maps of squares into G, H_0.

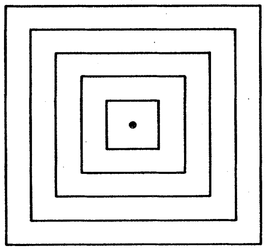

FIG. 53. The outermost square circumference gives a loop in H_0. Because H_0 lies within the simply connected group G, that loop can be shrunk to a point in G. The shrinking is specified by the images of the series of nested square circumferences starting with the outer one and ending with a single point. Viewed as a whole this shrinking homotopy provides a map of the square into G which gives the loop in H_0 when restricted to the circumference of the square.

the circumference into H_0. This is precisely what we require to associate with the loop in H_0 a sphere in coset space, via Eq. (7.2).

(4) The preceding three observations establish a one-to-one mapping of $\pi_2(G/H)$ onto $\pi_1(H_0)$. To establish that this correspondence is an isomorphism it only remains to establish that the product of homotopy classes of spheres in G/H at H is taken into the products of the corresponding homotopy classes of loops in H_0.[55]

To establish that the mapping carries the algebraic structure of $\pi_2(G/H)$ into that of $\pi_1(H_0)$ it is useful to represent each sphere in G/H at H by a map of a square into G, H_0 of a special canonical form. Given any map of the square into G, H_0 we can first continuously deform it into a map that takes at least one point on the top edge of the square into the identity. We can then continuously extend that part of the circumference which is taken into e until it includes the entire top edge and both side edges of the square, compressing the image of all of the original circumference into the bottom edge (Fig. 54).

Note that the corresponding loop in H_0 is still apparent from this canonical representation, for the bottom edge of the square provides a map of the interval $[0,1]$ into H_0 with both end points being taken into e, i.e., precisely a loop at e. Now it is evident from the defintion (7.1) of the product of homotopy classes of spheres at a point, that the product of two elements in $\pi_2(G/H, H)$ is represented by a map of a square into G, H_0 of the canonical form, given by the combination of representative maps shown in Fig. 55. The bottom edge of the combined square, however, is precisely a representation of the loop product at e of the bottom edges of the original two squares; thus the correspondence does take products in $\pi_2(G/H)$ into products in $\pi_1(H_0)$.

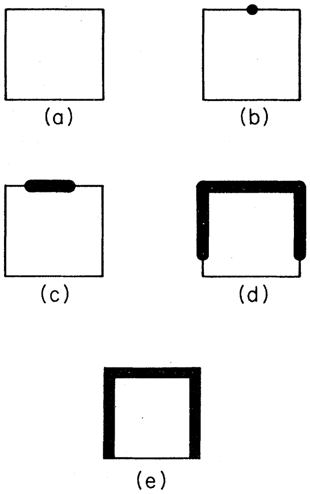

FIG. 54. (a) Map of a square into G, the circumference being taken into the connected component H_0 of H that contains the identity e. By continuously deforming the image of the square in G one can arrange for at least one point of the circumference to pass through e itself. That point is represented by the heavy dot in (b). One can then continuously extend the part of the circumference taken into e, as indicated by the thickened portions of the circumferences in (c) and (d). In this way one arrives at the map in (e), which is homotopic to the original map in (a) and which takes the interior of the square into G, the base into H_0, and the other three sides into the identity e.

D. The action of π_1 on π_2: Classes of freely homotopic spheres in G/H

The preceding discussion has established that if G is simply connected, then $\pi_2(G/H)$ is isomorphic to $\pi_1(H_0)$. The basis for the correspondence is quite intuitive: spheres in G/H can be represented by open "purses" in G, whose "mouths" wind about contours in H_0 repre-

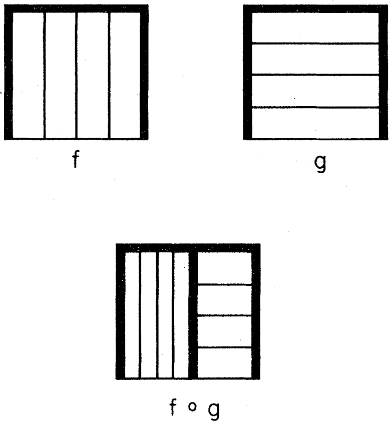

FIG. 55. Two homotopy classes in $\pi_2(G/H)$ can be represented by maps f and g of the form shown in Fig. 54 (e). Their product $f \circ g$ is then again of that form and the bottom edges of the squares, which represent loops in H_0, combine as in the ordinary loop product.

[55] We can speak unambiguously about the product of unbased loops in H_0 because H_0 is itself a continuous group, and its fundamental group is therefore Abelian. This was proved directly in Sec. IV.D. (It is about to emerge again as a consequence of the theorem we are proving and our earlier proof that second homotopy groups are always Abelian.)

senting the various homotopy classes of $\pi_1(H_0)$.

In classifying point defects in G/H it is also necessary to know the automorphism classes of $\pi_2(G/H)$ under the path automorphisms provided by the action of $\pi_1(G/H)$ on $\pi_2(G/H)$. These loop automorphisms also have a simple interpretation in terms of the algebraic structure of the isotropy subgroup H.

Two elements of $\pi_2(G/H)$ belong to the same automorphism class if they can be represented by balloons in G/H at H which differ only by a closed loop at H (see Fig. 51). Now a closed loop in G/H at H is represented in G by a path joining one of the connected components of H to the identity. Let $f(u,v)$ be the mapping of a square into G, H_0 representing the balloon without a loop in G/H at H. To represent the same balloon with a loop we proceed as follows:

Let the loop be represented by a path $g(z)$ in G, with $g(0)$ belonging to the connected component H_i of H, and $g(1)=e$. Let f_0 be the image of the circumference of the square $f(u,v)$ in H_0. Consider the family of loops in G given by $l(z)=g(z)f_0$. For each z, since f_0 lies entirely in the subgroup H_0 of $H, l(z)$ lies in a single left coset of H and therefore represents a single point in G/H. As z varies from 0 to 1, the path in G/H traced out by the cosets $l(z)H$ is just the loop at H we wish to attach to the balloon. Since the terminal point of the path is now represented by the loop f_0, which *is* the mouth of the purse, we have succeeded in attaching the loop to the balloon.

The only problem is that the set we have constructed in G representing the balloon and loop in G/H is based on the loop $g(0)f_0$, which lies in the component H_i of H. This is easily remedied, for by shifting any set of G by right multiplication by an element of H, we do not alter the corresponding set of cosets in any way. We can bring the base of the loop back into H_0 by right multiplying by $g(0)^{-1}$, for H_0 is a normal subgroup of H, and therefore $g(0)f_0g(0)^{-1}$ belongs to H_0.

We have therefore demonstrated that if f_0 is a loop in H_0 representing an element of $\pi_2(G/H)$ [via the isomorphism with $\pi_1(H_0)$] and H_i is a connected component of H representing an element of $\pi_1(G/H)$ (via the isomorphism with H/H_0), then the automorphism of $\pi_2(G/H)$ given by the element of $\pi_1(G/H)$ takes f_0 into hf_0h^{-1}, where h is any[56] element in H_i.

This conclusion can be stated compactly in the assertion that the action of $\pi_1(G/H)$ on $\pi_2(G/H)$ is given by the action on $\pi_1(H_0)$ of the inner automorphisms of H. For $\pi_1(H_0)$ contains classes of loops, g_t, in H_0 that correspond isomorphically to the classes of spheres in $\pi_2(G/H)$. Elements of $\pi_1(G/H)$ correspond isomorphically to the connected components H_i of H. The action of any element of $\pi_1(G/H)$ on an element of $\pi_2(G/H)$ is to transform that element from one corresponding to the loop g_t in H_0, to another corresponding to the loop hg_th^{-1} in H_0, where h is any element of the component H_i of H that corresponds to the element of π_1. Such transformations are the inner automorphisms of H.

Note that if the isotropy subgroup H is Abelian, then its only inner automorphism is the identity. In that case $\pi_1(G/H)$ acts trivially on $\pi_2(G/H)$ (in technical terms, G/H is "2-simple") and point defects are in one to one correspondence with the elements of $\pi_2(G/H)$, obeying combination laws given by the group multiplication table in π_2.

Note also that if H is discrete, then H_0 is a single point, which is trivially simply connected. In this case $\pi_2(G/H)=\pi_1(H_0)=0$: Ordered systems with discrete isotropy subgroups have no stable point defects.

Thus the only case in which π_1 might act nontrivially on π_2 arises when the isotropy subgroup is non-Abelian, and the connected component of the identity in the isotropy subgroup contains more than just the identity itself.

E. Examples of point defects

We illustrate these results by applying them to our standard examples.

1. Planar spins

The planar spins must be distributed in three-dimensional space for the discussion of point defects to be relevant, and the system is probably better thought of as being superfluid helium-4. The isotropy subgroup is a discrete translation group and there are therefore no stable point defects.

2. Ordinary spins

The isotropy subgroup is the subgroup $u(\hat{z},\theta)$ of SU(2) representing rotations around a single axis. This is isomorphic to the two-dimensional rotation group, and therefore its fundamental group is the group of winding numbers, Z: $\pi_2(G/H)=Z$. Since $\pi_1(G/H)=0$, G/H is 2-simple, and point defects are classified by positive or negative integers.

3. Nematics

The isotropy subgroup has two components: the subgroup $u(\hat{z},\theta)$ of SU(2), and the coset of this subgroup with $i\sigma_y$. Thus H_0 is again $u(\hat{z},\theta)$, and $\pi_2(G/H)$ is again the integers, Z. However it is a simple exercise in algebra to verify that the inner automorphism $u(\hat{z},\theta) \rightarrow (i\sigma_y)u(\hat{z},\theta)(i\sigma_y)^{-1}$ simply sends $u(\hat{z},\theta)$ into $u(\hat{z},-\theta)$. Thus loops in $\pi_1(H_0)$ with winding numbers of equal magnitude but opposite sign are taken into each other under the inner automorphisms of H. The corresponding point defects are therefore topologically equivalent to one another, and the nontrivial point defects in nematics correspond to pairs of elements in $\pi_2(G/H)$ associated with n and $-n$. With respect to a given basepoint, a defect characterized by n can be converted into one characterized by $-n$ by bringing it around a closed loop that surrounds a 180° disclination line.

The nontrivial point defects are therefore characterized by positive integers, without regard to sign. The combination law allows n and m to coalesce to either $n+m$ or $|n-m|$, depending on how the two are brought together.

Note that a nontrivial line defect (of which the 180° disclination is the only type) can catalyze the removal

[56] Since H_i is connected the loops produced by different elements h of H_i are all homotopic in H_0.

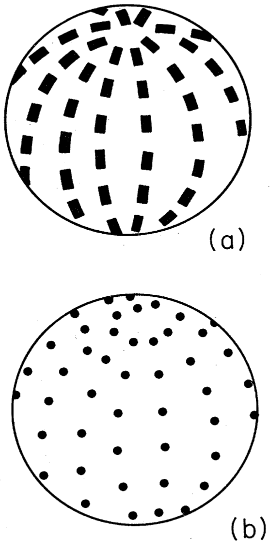

FIG. 56. (a) An attempt at a hedgehog point singularity in a biaxial nematic. Because the quills of the hedgehog have rectangular cross sections there are additional singularities on every surrounding sphere, i.e., a line singularity extends outward from the point. If (b) the quills had circular cross sections, as in an ordinary nematic, then the singularity on the spheres would vanish and an isolated point singularity could indeed be produced.

of a point defect of even index: one merely decomposes the $2n$ defect into two identical defects of type n, brings one of them around the line defect to convert it to type $-n$ (with respect to an arbitrary fixed base point), and then allows the n and the $-n$ to annihilate through recombination. More generally, a nontrivial disclination can catalyze the elimation of all point defects if the sum of their indices is even, or the transformation of all point defects into a single $n=1$ defect if the sum of their indices is odd.

4. Biaxial nematics

H is discrete so there are no stable point defects. It is worth noting why a configuration that does give a stable point defect in an ordinary nematic, fails to do so in the biaxial nematic. Consider, for example, the simple point defect given in an ordinary nematic by letting the director lie everywhere along the radial direction from the singular point P (a "hedgehog"). Suppose we try to construct such a point singularity in a biaxial nematic, by letting the z axes of the rectangular boxes point radially outward from P. This hedgehog in a biaxial nematic differs from the hedgehog in an ordinary nematic in one important respect: the nematic quills have complete rotational symmetry about the radial axis at any point; the biaxial quills do not, having a cross section of rectangular symmetry. Consequently, if one inscribes a sphere about P, the quills of the biaxial hedgehog specify a field of rectangles on that sphere. It is, however, impossible to construct a field of rectangles on a sphere without somewhere introducing a singular point on the sphere. Thus our effort to construct a *point* singularity has been unsuccessful, since the field is not regular except at the point P itself. We have instead, produced a *line* singularity (Fig. 56).

5. Superfluid helium-3

The isotropy subgroup of the dipole-locked A phase is discrete so there are no stable point singularities. As in the case of the biaxial nematic, it is worth convincing yourself that simple attempts to construct point singularities inevitably lead to line singularities emanating from the point. In the dipole-free A phase the subgroup H_0 is the subgroup of operations of the form $[u(\hat{z},\theta),1]$ in $SU(2)\times SU(2)$. Once again this has the topology of a circle, so that π_2 is isomorphic to the group of winding numbers Z. The full subgroup H consists of H_0 and the three cosets gH_0, g^2H_0, and g^3H_0, where g [see Eq. (5.19)] is the element $[u(\hat{x},\pi),u(\hat{z},\pi)]$. We have

$$\begin{aligned} &g[u(\hat{z},\theta)\times 1]g^{-1}=u(\hat{z},-\theta)\times 1\,,\\ &g^2[u(\hat{z},\theta)\times 1]g^{-2}=u(\hat{z},\theta)\times 1\,,\\ &g^3[u(\hat{z},\theta)\times 1]g^{-3}=u(\hat{z},-\theta)\times 1\,, \end{aligned} \tag{7.3}$$

and therefore the classes of equivalent point defects are as in the nematic case: Winding numbers n and $-n$ correspond to the same class of defects. Note, though, that to transform an n defect into a $(-n)$ defect it must be transported around a line singularity of the type g or g^3; transporting it around a g^2 line singularity will not change its sign.

VIII. ORDERED MEDIA WITH BROKEN TRANSLATIONAL SYMMETRY IN THE UNIFORM STATE

Our standard examples have all had complete translational invariance in the uniform state. The assumption of such translational invariance has also been implicit in much of the general discussion, though in some rather subtle ways, which I hope this section will serve to clarify.

The complications attendant upon applying the topological method to media with broken translational invariance in the uniform state (which I shall refer to generically as *crystalline media*) have not received the attention I believe they require. This may be because there is an exceedingly natural way of generalizing all of our conclusions to crystalline media.[57] This generalization (which I shall call the *naive generalization*) makes reference only to the structure of the symmetry group of the uniform medium, and it can therefore be formulated and applied without considering any of the loops, cylinders, surgery, and the like, that underlay our original formulation. Such considerations must be raised, of course, in justifying the naive generalization. Such a justification has yet to be provided, and I very much doubt that it can be, at least on the level of generality for which the method is valid for translationally invariant media.

One reason justifying the naive generalization may have received so little attention, is that in many cases it obviously works. It can give a very neat expression to many familiar, important, and intricate results. It can provide some rather novel conclusions whose validity can often be easily verified. I believe, however, that it can also produce conclusions that are at best obscure, possibly nonsensical, and, in either case, of a kind requiring a return to earlier more picturesque and *ad hoc* methods for interpretation and possible confir-

[57] The broadest statement of this generalization is that given by Kleman and Michel (1978).

mation. In short, the naive generalization of the method to the case of crystalline systems demonstrably has something of interest to tell us, but why it should have and how much it can be relied upon are very much open questions.[58]

In part A of this section the naive generalization is stated without any attempt at justification and with a minimum of critical comment. It is shown to reduce to our earlier procedure when applied to translationally invariant media. In part B I try to formulate in general terms my own reservations about the validity of the naive generalization, indicating where I think gaps in the argument remain to be filled in, or where limitations and restrictions are likely to prove necessary. In part C the naive generalization is applied to some representative cases, and characteristic triumphs, curiosities, and disasters are noted. It is beyond the scope of this review (and beyond my present capabilities) to relate these systematically to the reservations raised in part B; my aim is only to convey an impression of what I believe is the current state of the subject.

A. The naive generalization

1. Order-parameter space

Take as reference system a completely uniform specimen of the medium that fills all of space. Describe the configuration of the nonuniform medium at a point **r** by specifying a rigid body operation on the reference system that brings its local structure into coincidence with the local structure of the nonuniform system at the point **r**. The set of all these rigid body operations forms a group G containing translations as well as proper[59] rotations. The group G is the proper part of the full Euclidean group. Let H be the subgroup of G containing those rigid body operations that leave the reference system invariant (i.e., bring it everywhere into coincidence with its original configuration). The subgroup H is the proper part of the conventional space group of the uniform medium. The naive generalization takes the order-parameter space R to be the coset space G/H. Note that points of G/H are in continuous one-to-one correspondence with the physically distinguishable configurations of the uniform medium. The non-trivial (and questionable) assertion is that configurations of the nonuniform medium can still be described by maps of regions of physical space into G/H, as they can in the case of noncrystalline media.

2. Defects and homotopy groups

All our earlier conclusions relating classification schemes and combination laws for line and point defects to the first and second homotopy groups of G/H and the action of π_1 on π_2, are assumed to remain valid for crystalline media.

3. Computing the homotopy groups

The homotopy groups (and the action of π_1 on π_2) are computed by the same algorithm: One lifts the proper subgroup of the full Euclidean group to a simply connected covering group [by replacing the rotations by the corresponding operations in SU(2) (three dimensions) or T(1) (two dimensions)]. Taking that to be G and the corresponding lift of the isotropy subgroup as H, one takes, as before, $\pi_1(G/H)$ to be the quotient group H/H_0, where H_0 is the connected component of the identity in H, and one takes $\pi_2(G/H)$ to be $\pi_1(H_0)$. (The action of π_1 on π_2 is also given by our earlier prescription.)

Step 3 presents no problems; the arguments of Secs. V.A and VII.C and D were purely group theoretic in nature and applied to coset spaces of any connected, simply connected group. (The vanishing of π_2 is easily established for the covering groups of the proper two- and three-dimensional Euclidean groups.) Steps 1 and 2, however, are more doubtful. Indeed, if taken literally they are demonstrably false in almost all cases of interest, as we shall see in part B.

Note, however, that our earlier results on translationally invariant systems are correctly contained in the naive generalization, which simply expresses them from a slightly different point of view. We had regarded ordered systems as fields $f(\mathbf{r})$ of objects (vectors, headless vectors rectangular boxes, projection operators, etc.) characterized by a certain point group symmetry. The local configuration was specified by the point group operation taking a standard orientation of the object into the local one. No translational operations were mentioned or required.

We could, however, have taken as reference system not a single representative object but a complete space-filling specimen of the uniform system, as specified in step 1 of the naive generalization. Because the uniform system has full translational symmetry the translational part of the rigid body operation specified in step 1 would be completely arbitrary. Consequently the isotropy subgroup H would be the old isotropy subgroup of the point group augmented by the addition of all possible translations. (In more technical and precise terms, the new H would be the semidirect product of the old point group H with the full translation group.) Similarly, the group G would now be the old proper rotation group augmented by all possible translations (the full Euclidean group). These identical augmentations of G and H will simply cancel out when the coset space G/H is formed (just as G/H is unaltered by augmenting both G and H to include improper operations, or raising both to the universal covering group).

In contrast to the translationally invariant case, in a crystalline medium the isotropy subgroup H only contains a subgroup of the full translation group, while G continues to contain all translations. The coset space G/H can therefore have a very different structure from any coset space of a point subgroup in the proper rotation group.

B. Critique of the naive generalization

The weakness of the naive generalization stems from the fact that when translational symmetry is broken in

[58] The reader is again warned that my views on this point are rather more conservative than those I have encountered in private communications, and very much more conservative than those expressed in the published literature.

[59] As earlier, we ignore the comparatively trivial defects that can separate the medium into disconnected pieces (surface defects in three dimensions and line defects in two).

the uniform state, then the local configuration of the nonuniform system cannot be fully determined from its properties at a single mathematical point. Thus specifying the microscopic mass density at a single point $\mathbf{r}_0$ limits the possible configurations of a coinciding uniform crystal in the neighborhood of $\mathbf{r}_0$, but does not completely pin it down. To characterize a point in the nonuniform crystalline medium by a unique configuration of the reference medium one might, for example, specify the density at all points in some neighborhood of $\mathbf{r}_0$. The neighborhood could be very small. Indeed, it would have to be small, even on the scale of a single primitive cell, if the information the region contained about the local configuration were not to be rendered ambiguous by distortions resulting from the larger-scale nonuniformity itself. The best one could do would be to specify the density and suitable derivatives of the density at $\mathbf{r}_0$, reducing, as it were, the small neighborhood of $\mathbf{r}_0$ to infinitesimal dimensions.

However it is done, a certain degree of microscopic nonlocality enters into the description of the nonuniform crystalline medium: to specify the configuration at a point one must provide some partial information on how the system is changing as one departs from that point. This microscopic nonlocality of description creates problems for both steps 1 and 2 of the naive generalization.

The difficulty for step 1 stems from the fact that in a general nonuniform configuration of a crystalline medium each microscopic cell will undergo slight distortions. Thus (Fig. 57) a slightly bent simple cubic crystal cannot be built from perfect cubical blocks; if the crystal bends the blocks must also suffer some distortion. As a result of these distortions in the local microscopic structure there will be no configuration of the uniform medium that agrees precisely with the nonuniform medium in an appropriate microscopic region. Even if the region is infinitesimal, there will, in general, be a small disparity in derivatives. A slight ambiguity is therefore introduced into the specification of the medium at every point by a point in order-parameter space.

Concern over this ambiguity might seem pedantic fussiness. The distortions in individual cells will be minute, if, as is essential in a macroscopic theory, one restricts distortions to ones that vary slowly on the scale of the cellular dimensions. Nevertheless, some of the singularities of interest—dislocations, for example—can be viewed as arising precisely from the effect of many such minute distortions adding up to something of the order of a cellular dimension, when a path passing through very many cells is traversed.

I believe this difficulty can probably be dealt with by extending the group G to include not only rigid body operations on the reference system, but also an appropriate set of the tiny compressions and shears needed to bring the reference medium into unambiguously precise agreement with the local structure of the nonuniform medium. It would have to be shown that the augmented set of operations still had the topological structure of G/H. If this could be done step 1 of the naive generalization would become valid. We would have an order-parameter space with the topology of G/H which could be used to specify any configuration of the nonuniform medium in terms of a map of physical space into order-parameter space.

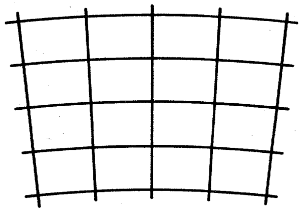

FIG. 57. Unit cells of a slightly bent crystal are themselves distorted from the form they assume in the uniform crystal.

Much more serious difficulties would still remain with step 2, which is entirely based on the converse of this last proposition: any map of physical space into order-parameter space must specify a configuration of the nonuniform system. In the case of a translationally invariant medium which can be characterized by a strictly local field, this is trivially the case. One need only give the order parameter at $\mathbf{r}_0$ the value determined by the point of order-parameter space associated with $\mathbf{r}_0$ by the mapping. If this mapping is continuous from real space into order-parameter space, then the corresponding physical field will also be continuous. In the case of crystalline systems, however, the order parameter contains a certain amount of nonlocal information. Its value at a point gives some limited information on how it is to be extrapolated to nearby points. If the order parameter is known at a point its values in the neighborhood of that point are restricted by more than just the requirements of continuity. Further *compatibility conditions* must be imposed to ensure that the values in the neighborhood are consistent with all that is implied by the value at the point. Consequently *not every continuous map of physical space into order-parameter space need correspond to a physical state of the nonuniform crystalline medium.*

This fact opens up a considerable gap between one's knowledge of freely homotopic loops and spheres in order-parameter space and the classes of defects in the physical medium. It is not, for example, even clear that a given homotopy class can be realized at all in the physical medium. And even when there do exist maps in a given homotopy class that satisfy the compatibility conditions, the question remains of whether there are homotopies between two such maps that satisfy the compatibility conditions at every stage. If there are not, then the corresponding defects will not be topologically equivalent.

Without a detailed study of compatibility conditions all one can conclude with confidence is that any defect that can be produced in the crystalline medium will (granting the validity of a suitably modified step 1) be associated with a homotopy class. The topological equivalence of two defects in the same class is an open question to be decided on a case by case basis; indeed the mere existence of any defects whatever in a given class cannot be taken for granted.

Having made all these gloomy remarks, I hasten to reemphasize that the conclusions produced by blind applications of the naive generalization are often quite interesting and instructive. I conclude this section with a survey of typical applications to illustrate both the nature of my reservations, and the elegance and power of the method, when it works.

C. Some applications

1. Crystals (dislocations only)

For simplicity we begin with an example in which rotational symmetries are ignored by imposing the additional requirement that the nonuniform crystal should everywhere have the same orientation as the uniform reference crystal.[60] The nonuniformity is then characterized entirely by displacements without rotation of the local primitive cells from the sites they would normally occupy. The full proper Euclidean group can then be replaced by the subgroup T(3) of translations, while the isotropy subgroup H becomes the subgroup of T(3) consisting of translations through Bravais lattice vectors. Since T(3) is parametrized by all of Euclidean three-space it is connected and simply connected; since H is discrete, our fundamental theorem identifies $\pi_1(G/H)$ with H itself. Thus the line defects are characterized by Bravais lattice vectors.

This is precisely the conventional description of dislocations. The Bravais lattice vector characterizing the dislocation (line defect) is known as its *Burgers vector*. Homotopy theory has landed us on familiar ground. Indeed, it has made some points that are not always emphasized in elementary descriptions of dislocations. There is, for example, a conventional distinction between screw and edge dislocations, depending on whether the Burgers vector is parallel or perpendicular to the dislocation line. The topological theory makes no reference to the orientation of the line defect: it is characterized by the Burgers vector alone. The distinction between screw and edge dislocations is thus nontopological. There should be line defects whose character alters from one type to the other as the line is traversed. How to construct such a line is shown in Fig. 58. It is encouraging that the topological method automatically brings such possibilities to our attention.

Because H is discrete $\pi_2(G/H)=0$: the theory predicts no topologically stable point defects. There are, of course, physically stable point defects of great importance in crystals, the simplest example being a *vacancy* or *void*—the absence of an ion from an isolated site. Vacancies do indeed leave no tell-tale discrete signature in the far region (Fig. 59). The fact that the removal of a void by local surgery requires the local creation of matter is, from the topological point of view, a mere quibble. Note, nevertheless, that the physical means of removing a void—its cell by cell transport to the surface or its annihilation by an interstitial point defect—are strikingly reminiscent of the nonlocal means available for the elimination of topologically stable defects. In contrast the analogous topologically unstable defect in a nematic, a bubble, can be removed by a continuous inflow of matter whose flux becomes arbitrarily small in the far region.

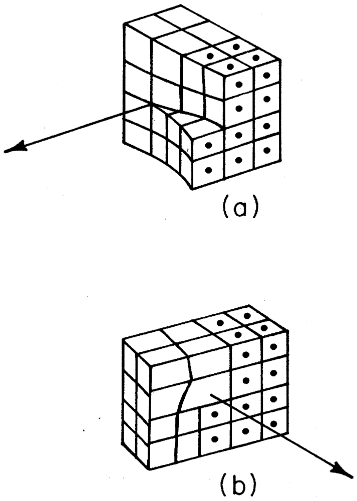

FIG. 58. (a) Portion of a cubic crystal cut away to reveal a screw dislocation (after Ashcroft *et al.*, 1976). The dislocation line and its Burgers vector are along the arrow. (b) Portion of a cubic crystal cut away to reveal an edge dislocation. The dislocation line is along the arrow; its Burgers vector is perpendicular to the line and the same as the Burgers vector of the screw dislocation. Note that the two parts of the figure can be superimposed (the groups of cells with dots on their surfaces being brought into coincidence by the superimposition). In the crystal that results there is a single dislocation line with a right-angle bend, which is a screw dislocation on one side of the bend and an edge dislocation on the other. No singularities are present on the extrapolation of either segment past the bend.

2. Crystals (dislocations and disclinations)

When the prohibition against local rotations is dropped, the fundamental group is identified with the double group that arises from the proper space group of the crystal by lifting the rotational parts of the group

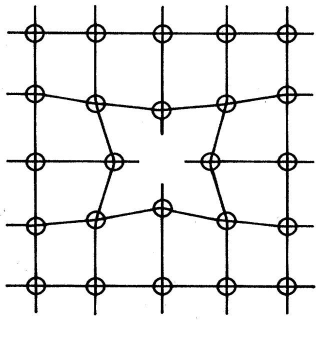

FIG. 59. A vacancy in a square lattice. Note the absence of any information about the vacancy in the far configuration.

[60] Note, already, an example of one of the difficulties mentioned in part B. If the sides, for example, of every cubic cell of the nonuniform crystal are parallel to the sides of a single cube then no nonuniformity can in fact be present. Some slight distortion of the cubic cells is essential to produce a nonuniform state. A more accurate approach would build in the possibility of such floppiness and (one hopes) show that it leaves the conclusions unchanged. I shall treat the examples in the same uncritical spirit in which the naive generalization was first put forth, focusing on problems such as this one only to illustrate some of the objections voiced in part B.

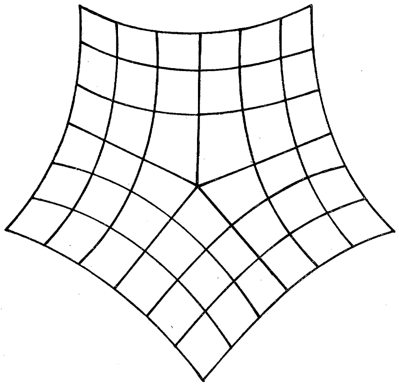

FIG. 60. A −90° disclination in a square lattice. As the singular point is circumnavigated in a counterclockwise sense, the orientation of nearby square cells rotates 90° clockwise. Considerable deviations from a locally uniform structure accompany such a defect. If the plane were an elastic sheet it would relieve such strains by buckling into the third dimension to assume a saddle shape.

operations from SO(3) to SU(2) in three dimensions, or from SO(2) to T(1) in two. For most interesting crystal structures this group is non-Abelian with an extremely intricate structure of conjugacy classes.

Discrete operations in the space group with no translational part correspond to line defects in which the local crystal structure rotates through an angle associated with a point group operation as the line is encircled. Such defects are known as disclinations. A simple disclination in a two-dimensional crystal is shown in Fig. 60. Note that the cells are severely distorted from their shape in the uniform medium at large distances from the singular point. Disclinations fail to meet the basic premise of the entire scheme that away from the singular point the medium should be locally indistinguishable from a suitably oriented uniform medium. Dislocations also violate this tenet, but not as dramatically as disclinations do.

This may well be rectifiable by expanding the group G to include the appropriate deformations of the uniform crystal, but to my knowledge nobody has yet taken up the challenge. Since isolated disclinations do involve gross distortions in large regions, they are energetically so costly as to be of little physical interest, so an incentive for repairing the basic theory might seem to be lacking. However compensating disclination pairs need not be so disastrously difficult to create, and can even be of physical interest. The naive generalization has some instructive things to say about such pairs, which a simple two-dimensional example illustrates.

Consider a square lattice in two dimensions. Elements of the lift of the isotropy subgroup (and therefore elements of the fundamental group) are associated with a pair of translations: A translation in the plane through a Bravais lattice vector **a**, and a one-dimensional translation t^n associated with a rotation through $n\pi/2$. The composition law for a pair is the usual one for space groups:

$$(\mathbf{a}, t^n)\cdot(\mathbf{b}, t^m) = (\mathbf{a} + R^n\mathbf{b}, t^{n+m}) , \tag{8.1}$$

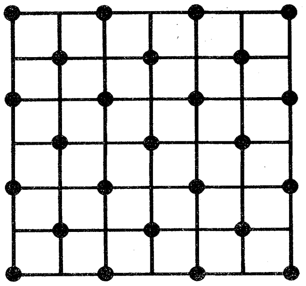

FIG. 61. The sublattice of Burgers vectors of dislocations which, in conjunction with a 90° disclination, form a single conjugacy class.

where R^n is a rotation through $n\pi/2$ in the plane of the lattice. It follows from Eq. (8.1) that inverses are given by

$$(\mathbf{a}, t^n)^{-1} = (-R^{-n}\mathbf{a}, t^{-n}) . \tag{8.2}$$

This in turn implies that

$$(\mathbf{a}, t^n)(0, t)(\mathbf{a}, t^n)^{-1} = (\mathbf{a} - R\mathbf{a}, t) . \tag{8.3}$$

Thus the conjugacy class of the pure 90° disclination $(0, t)$ contains elements superposing such a disclination with all dislocations whose Burgers vectors lie in the sublattice shown in Fig. 61. The same is true of the −90° disclination. We conclude that when two such disclinations meet they can turn into any dislocation characterized by a Burgers vector in the sublattice of Fig. 61. Conversely, any such dislocation is equivalent to a disclination pair.

This is interesting, important, and not immediately obvious to those with limited artistic talents. It is illustrated in Fig. 62. Note, though, that as the disclinations move apart the number of dislocations into which they are resolved grows. This can be regarded as a measure of the strain attendant upon the production of isolated disclinations. It also suggests that in a less naive generalization one may well have to deal with densities of defects that are very high indeed, requiring a much more careful examination of just what the appropriate length scales are.

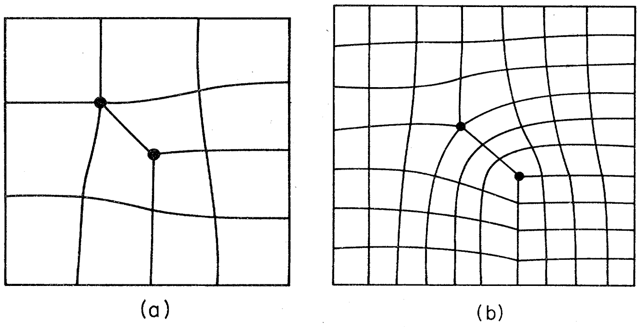

FIG. 62. (a) Nearby disclinations of +90° and −90° in a square lattice. The far configuration is that of a single dislocation with Burgers vector (1,1). (b) The +90° and −90° disclinations of (a) are moved further apart. The equivalent dislocation now has Burgers vector (3,3). Note that the strains in the figure are so evident that it seems to pop out of the page. Indeed, such a configuration is precisely that seen at the corner of a swimming pool whose bottom and sides and the surrounding ground are paved with square tiles.

3. Directed stripes in the plane

A simple illustration of the difficulties with the naive generalization is provided by a two-dimensional medium consisting in the uniform state of equidistant parallel directed lines (Fig. 63). We take the lines to be directed to the right and given by $y=0, \pm d, \pm 2d, \cdots$. The proper isotropy subgroup of the two-dimensional Euclidean group contains a discrete group Z of translations through d along the y direction and a continuous group T of arbitrary translations along the x direction. When this is lifted to the covering group another discrete translation group appears, associating the identity rotation with translations through $0, \pm 2\pi, \pm 4\pi, \cdots$. Thus the entire isotropy subgroup lifts to the Abelian group $H=Z\times Z\times T$. (The lift is Abelian because the trivial rotation commutes with all translations.) The connected part of the identity in H is $H_0=T(1)$, the subgroup of translations along x. The fundamental theorem then identifies $\pi_1(G/H)$ with $H/H_0=Z\times Z$. This is also Abelian, so that classes of defects are characterized by a pair of integers, a disclination number and a dislocation number.

In contrast to the square lattice this medium can have nontrivial nonuniform states even when we impose a condition of strict local agreement with the uniform state, by requiring that when curved the lines should still remain a distance d apart. It is then easy to associate with any point of the nonuniform medium (on or between lines) a unique (to within an operation of the isotropy subgroup) configuration of the uniform medium, bringing it into coincidence in position and slope with the nearest parts of the stripes nearest the point.

If, however, the curved stripes are everywhere equidistant, the possible singularities are quite unrelated to the topological classification scheme. For a single stripe now determines all other stripes in the family as the envelopes of all circles of radii $d, 2d, 3d\ldots$ with centers on the original stripe. The singularities of the medium are then the lines generated by the points of intersection of all the normals emanating from any given stripe (since at those and only those points the local structure will receive competing messages on how it is to be oriented).

This state of affairs is not limited to artificial models. A *smectic liquid crystal* is characterized in the uniform state by a family of equidistant parallel planes. Local agreement of the nonuniform smectic with the uniform one is ensured by the requirement that the distorted plane surfaces should remain everywhere at the uniform separation distance. The singularities of such a system turn out to be a family of surfaces. Energetic considerations rule out all such surfaces except those which degenerate into lines. Purely geometric reasoning then reveals that the only lines that can be arrived at in this way are the so-called focal conics. Such line singularities are commonly observed in smectics. Except in a few isolated cases they bear no resemblance to the line singularities indicated by the naive generalization of the topological scheme, and that scheme gives no accounting whatever for the origin of the focal conics.[61]

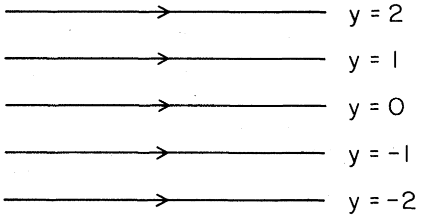

FIG. 63. The plane medium of directed stripes.

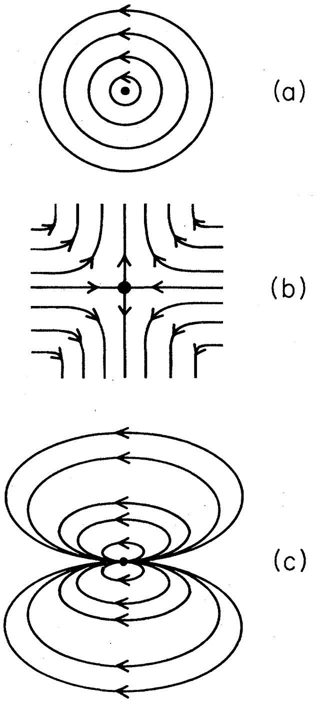

FIG. 64. (a) A pure disclination with winding number +1 in a medium of directed stripes. (b) A pure disclination of winding number −1. (c) A pure disclination of winding number +2.

The topological scheme, if it applies at all, can only hold in a regime in which the constraint of strict local agreement with the uniform medium is relaxed. Assuming that the naive generalization can be made to embrace this complication, we examine some characteristic singularities of the two-dimensional directed stripes.

Fig. 64(a) shows a pure disclination with winding number +1, which does satisfy the condition that the lines should be everywhere equidistant. A pure disclination with winding number −1 is shown in 64(b). In this configuration the constraint on uniform spacing is violated in the 45° directions even far from the singular point. The violation is, however, bounded in amplitude, the ratio of maximum to minimum interlinear distance being a factor of about $\sqrt{2}$. In Fig. 64(c) is drawn an attempt at a pure disclination with winding number +2. The constraint on interlinear spacing is much more grossly violated: the distances between some lines must, in fact, become arbitrarily large at sufficiently large distances from the singular point.

[61] Kleman and Michel (1978) give a detailed exposition of the application of the naive generalization to smectics.

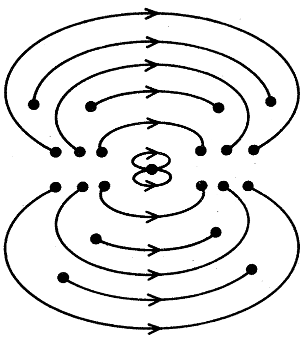

FIG. 65. Easing the gross strains in the +2 disclination of Fig. 64 (b) by the introduction of a large number of dislocations. The dislocations occur in pairs of opposite polarity.

These examples are typical. Pure disclinations with negative winding number can be incorporated into the scheme provided one introduces an appropriate range of variation for the interlinear distance. Pure disclinations with winding numbers greater than one, however, cannot be accomodated by so well controlled an extension. No matter how wide the range of interplanar distances is made, far enough away from the singular point still wider separations will be required. These come close to being examples of homotopy classes with no physically acceptable defects.

The gross distortions far from these singular points can be repaired with the aid of dislocations. By introducing enough dislocations one can keep the interplanar spacing between respectable limits, as illustrated in Fig. 65. Many dislocations are needed to do the job but, at least in the example shown, the indices of all the dislocations add to zero. This easing of disclination strain by dislocation densities will probably play an important role in a correct generalization of the topological method to crystalline media.

The striped plane also provides some simple examples of the difficulty in translating maps of physical space into order-parameter space, into acceptable configurations of the medium itself. The very well behaved $n=1$ disclination is characterized by a map of physical space into order-parameter space specifying that at an angle θ from the physical x axis (and at any distance from the origin) the reference medium is to be rotated through $\theta+90°$ (without translation) to bring it everywhere into coincidence with the local structure (Fig. 66). If we simply change this prescription by dropping the 90° from the rotation angle, we get a new map of physical space into order-parameter space homotopic to the first. (The homotopy proceeds through a succession of maps in which the 90° part of the angle drops continuously to zero.) The new map suggests a pattern of lines radiating outward from the origin. This configuration can only be realized by gross relaxations on the constraint on interlinear distance [Fig. 67(a)] or by the introduction of many dislocations whose indices do not add to zero [Fig. 67(b)].

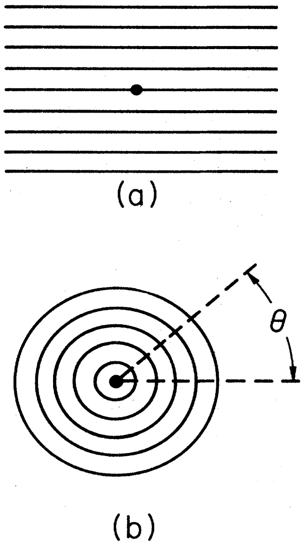

FIG. 66. (a) A uniform medium of equidistant parallel lines. The dot is the origin. (b) The configuration produced by specifying that the uniform medium is brought into coincidence with the local structure at the point with polar coordinates r and θ by counterclockwise rotation through $\theta+90°$, independent of the value of r.

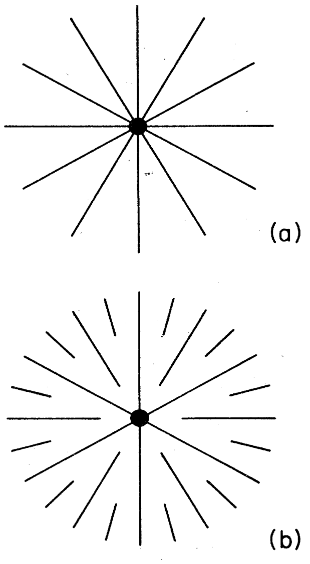

FIG. 67. If the instructions are to rotate the uniform medium of Fig. 66 (a) counterclockwise through θ to construct the configuration at r and θ then either (a) the constraint on constant interlinear spacing will be grossly violated or (b) infinitely many dislocations of the same polarity will be required.

If all nonuniform media were as simple as the plane of directed stripes, the topological method would not be worth salvaging. If, however, one merely drops the direction from the stripes, then 180° rotations appear in the space group and a non-Abelian fundamental group of considerable intricacy results. This fingerprint medium or two-dimensional smectic was introduced by Poénaru and Toulouse (1977) as a particularly simple example of a non-Abelian medium. The naive generalization of the topological method reveals the same interplay between dislocations and disclination pairs as we described for the square lattice. It also leads to all the anomalies encountered in the simpler case of directed stripes. At a minimum one should be able to characterize the sources of the difficulty in the anomalous assertions, so that the assertions that are correct and instructive can be used with confidence and without the need for case by case confirmation.

4. Cholesteric liquid crystals

This is a particularly interesting case since cholesterics actually exist, the naive generalization yields an unusually simple set of defect classes, and examples of all such defects were identified well before the topological approach was formulated.

A cholesteric is characterized in its uniform configuration by a headless vector field[62] of the form

$$\mathbf{d}=\hat{x}\cos qz+\hat{y}\sin qz\,. \tag{8.4}$$

If the wave-vector q were zero, Eq. (8.4) would describe a uniform nematic. Cholesterics are twisted nematics. In planes perpendicular to the z axis the director field is uniform, but the direction turns through 360° as the plane moves through a distance $p=2\pi/q$ along the z axis.

The naive generalization associates with such a structure a fundamental group that can be constructed as follows:

The connected component of the identity in the isotropy subgroup of the full Euclidean group is a subgroup H_0^E of operations of the form

$$(l\hat{z},R(\hat{z},ql))\,, \tag{8.5}$$

coupling translations along the z axis through l to rotations about the z axis through ql to restore the original orientation. In addition the structure (8.4) is invariant under point operations consisting of 180° rotations about the x, y, and z axes. (The latter two symmetries use the fact that $\mathbf{d}$ and $-\mathbf{d}$ are identified.) Thus the full proper isotropy subgroup H^E consists of H_0^E and its cosets with the operations

$$(0,R(\hat{x},\pi)),\ (0,R(\hat{y},\pi)),\ (0,R(\hat{z},\pi))\,. \tag{8.6}$$

When this is lifted to the covering group there is a doubling of the rotational parts. The connected component of the identity H_0 consists of the operations

$$(l\hat{z},u(\hat{z},ql)) \tag{8.7}$$

and the other seven pieces are the cosets of H_0 with the operations:[63]

$$(0,-1),\ (0,\pm i\sigma_x)\,,\ (0,\pm i\sigma_y)\,,\ (0,\pm i\sigma_z)\,. \tag{8.8}$$

Thus the quotient group H/H_0 is the quaternion group.

The naive generalization therefore endows cholesterics with the same defect structure as biaxial nematics. Four nontrivial classes of line defects are specified and no point defects. Such conclusions were, in fact, drawn some time before the development of the topological approach, simple pictures of the defects in each class being given, for example, in the text of de Gennes (1974).

On the other hand it is easy to invent simple maps of real space into order-parameter space which, as in the case of the striped plane, correspond to no configuration of the physical cholesteric. It is not at all clear that the kind of configurational flexibility will hold for cholesteric defects that permits one, in the biaxial nematic, to contemplate such processes as the crossing of defects. Maps of physical space into order-parameter space and the homotopies of such maps are almost certainly subject to further constraints.

I have neither the intent nor the ability to solve the cholesteric problem here and now, but I cite it as an important challenge which the topological method has yet to come fully to grips with.[64]

IX. HIGHER HOMOTOPY GROUPS, RELATIVE HOMOTOPY GROUPS, AND EXACT SEQUENCES

We conclude by setting some of our earlier results into a broader context. The homotopy groups π_1 and π_2 are special cases of the general nth homotopy groups. These, in turn, are special cases of the so-called relative homotopy groups. When these generalizations are made, the two fundamental theorems we have given for the construction of π_1 and π_2 can be displayed as a special application of a very general series of homomorphisms between absolute (i.e., ordinary) homotopy groups and relative homotopy groups, known as the *exact homotopy sequence.*

The higher homotopy groups are described in part A; the relative homotopy groups, in part B. The exact homotopy sequence is described in some detail in part C, and its relation to our earlier results is given in part D. We conclude with some remarks in part E on the relevance of the third homotopy group to problems somewhat broader than those posed by the theory of defects, mentioning, in particular, the relation between the topology of defects and the topology of solitons.

A. Higher homotopy groups

The definition of the *higher homotopy groups* and the demonstration that they have the appropriate properties are quite analogous to points made in our earlier treatment of the second homotopy group. The group $\pi_n(R,x)$ is the set of equivalence classes of maps of the n-dimensional unit cube, $0\leq z_j\leq 1$, $j=1\ldots n$, into the space R, such that all surface points on the cube (i.e., all points with at least one z_j equal to 0 or 1) are taken into the base point x. The composition of two maps is as in the case of π_2) given by joining the cubical domains of the maps along the faces nromal to one of the axes and rescaling the resulting oblong domain back to a cube by compression along that axis. The homotopy class of the product is independent of the choice of axis and, as in the case of the second homotopy group, the composition law for homotopy classes is commutative.

One again introduces path isomorphisms to show that the nth homotopy groups based at different points are isomorphic. The unbased group $\pi_n(R)$ is the abstract group of which the based groups are isomorphic copies. Considering just those path isomorphisms given by loops, one again realizes $\pi_1(R)$ as a group of automorphisms on $\pi_n(R)$. If the only such automorphism is the identity, then R is said to be *n-simple.* If R is n-simple then the classes of freely homotopic maps of the surface of the unit sphere in $n+1$ dimensional space (known as S_n) are in one-to-one correspondence with the elements of $\pi_n(R)$. More generally, they are in one-to-

[62] We use the term "headless vector" in the same sense as for nematics; $\mathbf{d}$ is a local anisotropy axis which has no direction associated with it.

[63] The composition law for the lift of a subgroup of the Euclidean space is: $(\mathbf{a},\,u(\hat{n}_1,\theta_1))\times(\mathbf{b},\,u(\hat{n}_2,\theta_2))=(\mathbf{a}+R(\hat{n}_1,\theta_1)\mathbf{b},\,u(\hat{n}_1,\theta_1)\,u(\hat{n}_2,\theta_2))$.

[64] Bouligand *et al.* (1978) raise some additional delicate points bearing on defects in cholesterics, but do not question the general assumptions of the naive generalization *per se.*

one correspondence with the automorphism classes of $\pi_n(R)$ under the action of the group of automorphisms provided by $\pi_1(R)$.

The proofs of these properties are straightforward restatements of the proofs for π_2, stepped up by the appropriate number of dimensions.

B. Relative homotopy groups

For concreteness, I describe the relative homotopy groups for $n=2$. The generalization to higher n should be evident. There are a few complications for $n=1$ and a few simplifications for $n>2$, which I shall state at the appropriate places.

The *relative homotopy groups* of the space R are defined with respect to a base point x *and* a subset A of R, which contains the base point. In the special case in which A contains only the base point the relative groups reduce back to the absolute ones, and textbooks which value efficiency over clarity sometimes deal with relative groups from the very start, treating the absolute groups as special cases.

The relative homotopy group $\pi_2(R,A,x)$ is constructed out of maps of the unit square into R that take three of the edges into the point x. The image of the remaining edge, however, can lie anywhere in the set A. The structure of the map is indicated schematically in Fig. 68.

The product of two such maps is defined in the manner depected in Fig. 69. The elements of $\pi_2(R,A,x)$ are the equivalence classes of such maps under homotopies which have the structure shown in Fig. 68 for each value of t. The group structure is provided by the product operation of Fig. 69, which can be shown to determine a homotopy class independent of the choice of representative maps used to form the representative product map.

The nth relative homotopy group $\pi_n(R,A,x)$ is constructed in an analogous way out of continuous maps of the n cube into R which take all of the surface into the single point x except for the cube face $z_n=0$ (which we shall refer to as the *base* of the n cube) which can be taken anywhere in the subset A of R.

One can "relativize" many of the concepts and results developed for the absolute homotopy groups. In particular the proof that $\pi_2(R,x)$ is Abelian can be carried over to $\pi_n(R,A,x)$ for $n\geq 3$. However the second relative homotopy group need not be Abelian.

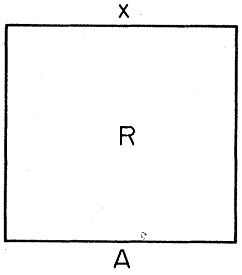

FIG. 68. Schematic representation of a map of a square into the space R, which takes the base into the subset A of R, and the other three edges into the single point x. Such a map is described as taking the square into R,A,x. Homotopy classes of such maps constitute the relative homotopy group $\pi_2(R,A,x)$.

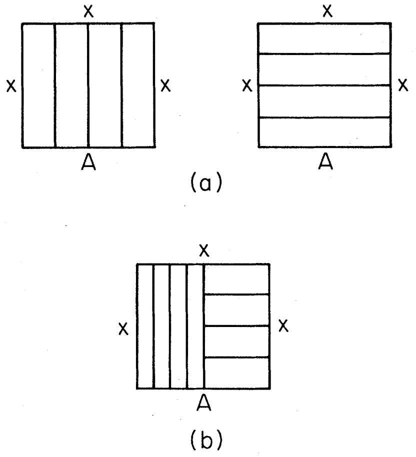

FIG. 69. (a) Two maps representing homotopy classes in $\pi_2(R,A,x)$. (b) The product of the two maps shown in (a). The interior of all squares can be taken into the full set R. The edges go into the point x or the subset A, as indicated.

The first relative homotopy group need not even be a group. It consists of maps of the interval $0\leq z\leq 1$ into R which take 1 into the base point x and 0 into any point in the set A. If R has no further structure, then the product of two such maps cannot be defined since (in contrast to the case $n=2$ of Fig. 68) there is no obvious way to put the maps side by side. If, however, R is a group G, A is a subgroup H, and x is the identity e (which is the case of major interest for us) then the product can be defined. If two classes f_i are represented by paths connecting h_i with the identity e, then to represent the product $f_1 o f_2$ we shift f_2 by uniformly multiplying all of its points by h_1 on the left, thereby constructing a path from h_1h_2 to h_1. This path can now be joined to the path f_1 from h_1 to e, giving a single path from h_1h_2 to e, as illustrated in Fig. 70. Since H is a subgroup, h_1h_2 is also in H, so the product path is again of the required form.

Note an analogous situation that arises for the absolute homotopy groups. One can define a set $\pi_0(R)$ of freely homotopic maps of single points into R. The elements of π_0 evidently correspond to the set of connected pieces of R. In general there is no natural group structure for such a set, but if R is a group G, then π_0 can be given the group structure of G/G_0, where G_0 is the connected component of the identity. (Compare the discussion of IV.A.) The zeroth homotopy group and the first relative homotopy group often fit quite naturally into the hierarchy of higher groups, when they can be defined.

Many properties of the absolute homotopy groups are better formulated in the broader context of relative homotopy groups. Perhaps the most important example of this is the discussion that follows of the exact homotopy sequence. However, the relative homotopy groups can also have direct physical applications in their own right. They will be pertinent, for example, if there is

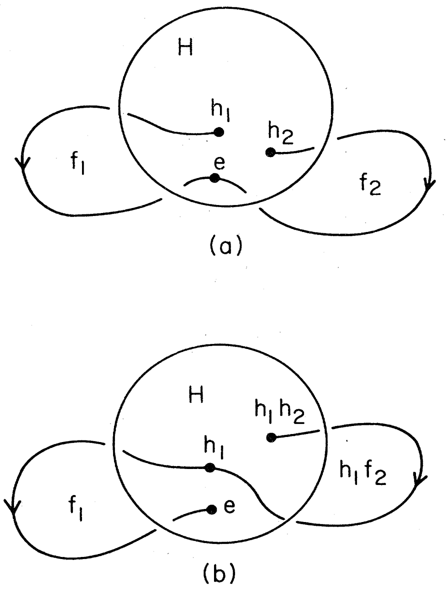

FIG. 70. (a) Two paths in a group G that connect elements of a subgroup H to the identity e. The paths represent homotopy classes in $\pi_1(G,H,e)$. (b) Rule for forming the product of such paths: shift the second by multiplying it point by point by the starting point of the first and form the combined path.

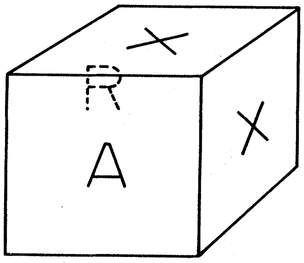

FIG. 71. Map of a 3-cube into R,A,x. Note that it carries on its base (here, a vertical face) a map of a 2-cube into A,x. This is the basis for the natural homomorphism γ_3 from $\pi_3(R,A,x)$ into $\pi_2(A,x)$.

only a restricted region S of physical space in which the order parameter can assume values in the full order-parameter space R, its value being constrained to lie in some subset A of R at points of physical space outside of S. Such a state of affairs can be produced by boundary conditions: S might be the interior of a cylinder at whose surface the order parameter was restricted to values in the subset A of R.[65] Cross sections of the cylinder would then determine homotopy classes of maps of a square into R in which the circumference was taken into the set A—i.e., elements of $\pi_2(R,A)$. Alternatively A might be the full order-parameter space, and R a still larger space containing A, from which the order parameter was ordinarily excluded on energetic grounds. Near the cores of singularities in which the order parameter ranged only through A in the far field, it might be advantageous to expand the order-parameter space to the larger space R, if divergent gradient energies could thereby be reduced. The relative homotopy groups of R and A again give a concise classification of the cases one can encounter in this context (see Mermin, Mineyev, and Volovik, 1978).

C. The exact homotopy sequence

Consider the relative homotopy group $\pi_n(R,A,x)$. This consists of classes of maps of the n cube into R, which take all the sides except the base into the point x. The base of the n cube [which is an $(n-1)$ cube] is taken into A with the restriction that *its* circumference [i.e., the surface of the $(n-1)$ cube] must be taken into the point x, to join continuously onto the rest of the surface of the n cube. This is pictured (in the case $n=3$) in Fig. 71.

Thus each map of the n cube into R,A,x carries with it a map of the $(n-1)$ cube into A,x.[66] The rule giving the product of maps of n cubes into R,A,x (illustrated for $n=2$ in Fig. 69) is designed so that the associated maps into A,x of the $(n-1)$ cubes forming the bases compose under the ordinary product rule for representative maps in $\pi_{n-1}(A,x)$. There is therefore a natural homomorphism γ_n from $\pi_n(R,A,x)$ into $\pi_{n-1}(A,x)$, given by simply ignoring all of the n cube except its base. The correspondence γ_n is a homomorphism but not necessarily an isomorphism for two reasons: (a) The correspondence need not be onto—there is no guarantee that every homotopy class in $\pi_{n-1}(A,x)$ can be represented by the base of a mapping from a class in $\pi_n(R,A,x)$; (b) The correspondence need not be one-to-one: It is possible that nonhomotopic maps of the n cube into R,A,x might yield homotpic maps of the $(n-1)$ cube into A,x.

The form the homomorphism γ_n can assume is limited by the topological structure of the sets R and A. These limitations are embodied in the so-called *exact sequence of homomorphisms*. To characterize the exact sequence we must first describe two other kinds of homomorphisms between homotopy groups.

1. Since A is a subset of R, any map of an n cube into A at x is also a map of the n cube into R at x. This establishes a homomorphic correspondence α_n between $\pi_n(A,x)$ and $\pi_n(R,x)$. Once again, the correspondence is a homomorphism rather than an isomorphism because of the following: (a) It need not be onto: there is no guarantee that every map of a cube into R at x is homotopic in R to a map of a cube into A at x; and (b) it need not be one-to-one: maps of the cube into A at x that are not homotopic in A might prove to be homotopic when the homotopy is allowed to range through the wider space R.

2. Since the base point x belongs to the set A, any map of an n cube into R at x is also a map of the n cube into R,A,x, in which the side going into the set A happens to be taken into the single point x of A. Once again this correspondence β_n is clearly a homomorphism between $\pi_n(R,x)$ and $\pi_n(R,A,x)$. For the same pair of reasons as in the other two cases, the homomorphism need

[65] See Bailin and Love (1978) and Volovik (1978). This case can be complicated by the fact that different subsets A of R can be required at different points of the surface.

[66] We adopt an obvious shorthand notation. A map of a cube into R,A,x is one that takes the cube into R, its base into A, and the rest of its surface into x. A map of a cube into A,x takes the cube into A and all of its surface into x.

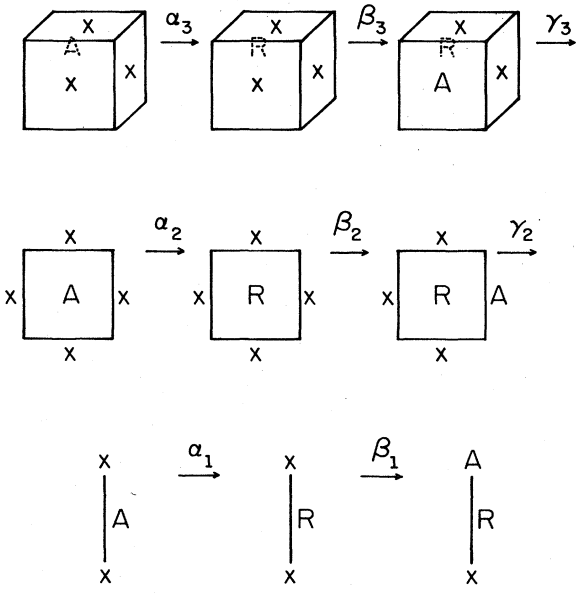

FIG. 72. A portion of the exact homotopy sequence [Eq. (9.1)].

not be an isomorphism.

We can now put together a chain of homomorphisms. Starting with the nth homotopy group of the subset A of R, we are carried by the homomorphism α_n that allows the interior of cubes in A,x to expand from A to R, into the nth homotopy group of R itself. From the nth homotopy group of R at x we are carried by the homomorphism β_n that allows the base of the cube to expand from the point x to the subset A, into the relative homotopy group $\pi_n(R,A,x)$. From this relative homotopy group we are carried by the homomorphism γ_n that ignores all but the base of the cube into the $(n-1)$th homotopy group of A. We are then in a position to repeat the cycle at the next level down in n. The sequence is summarized in Eq. (9.1)[67] and is pictured schematically in Fig. 72.

$$\begin{aligned}
&\pi_n(A,x)\xrightarrow{\alpha_n}\pi_n(R,x)\xrightarrow{\beta_n}\pi_n(R,A,x)\xrightarrow{\gamma_n}\\
&\pi_{n-1}(A,x)\xrightarrow{\alpha_{n-1}}\pi_{n-1}(R,x)\xrightarrow{\beta_{n-1}}\pi_{n-1}(R,A,x)\xrightarrow{\gamma_{n-1}}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\pi_1(A,x)\xrightarrow{\alpha_1}\pi_1(R,x)\xrightarrow{\beta_1}\pi_1(R,A,x)\xrightarrow{\gamma_1}\\
&\pi_0(A)\xrightarrow{\alpha_0}\pi_0(R)\,. \qquad (9.1)
\end{aligned}$$

In addition to the fact that it is a chain of group homomorphisms, the sequence of maps in Eq. (9.1) has one further general property: it is exact. The sequence of homomorphisms

$$G_1\xrightarrow{\varphi_1}G_2\xrightarrow{\varphi_2}G_3\xrightarrow{\varphi_3}G_4\cdots \qquad (9.2)$$

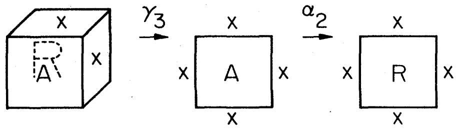

FIG. 73. The homomorphisms γ_3 and α_2.

is said to be an *exact sequence* if the subset of G_{i+1} into which G_i is taken by the homomorphism φ_i, is the kernel of the next homomorphism φ_{i+1} taking G_{i+1} into G_{i+2}.[68] Put succinctly, a sequence of homomorphisms is said to be exact if the image of each is the kernel of the next. To establish that Eq. (9.1) is an exact sequence we must show that the image of each of the three types of homomorphism is the kernel of the homomorphism that follows.[69]

(1) The image of γ_n is the kernel of α_{n-1}. Figure 73 illustrates the two homomorphisms in question. The crucial point here is that any image of an n cube in R,A,x can be regarded as a homotopy in R at x between the map of the $(n-1)$ cube into A at x provided by the base, and the trivial map of the $(n-1)$ cube into x alone provided by the face opposite the base. The parameter t of the homotopy is simply the coordinate z_n of the cube normal to the base (Figure 74). Conversely, any such homotopy provides in the same way a map of the n cube into R,A,x. It follows immediately from this that the image of γ_n is just the set of homotopy classes in $\pi_{n-1}(A,x)$ containing maps of the $(n-1)$ cube into A,x which are homotopic to the trivial map when the homotopy is allowed to range through all of R. But this last set of homotopy classes *is* the kernel of α_{n-1}, since that homomorphism acts on the homotopy classes of $\pi_{n-1}(A,x)$ by precisely such extensions of their representative maps.

(2) The image of β_n is the kernel of γ_n. Figure 75 illustrates the two homomorphisms. The central point here is that γ_n acts on homotopy classes of cubes in R,A,x by separating off their action on the base of the cube alone, so that classes taken into the identity of $\pi_{n-1}(A,x)$ by γ_n will be represented by cubes in R,A,x whose bases are homotopic to a constant in A,x. But it is precisely cubes of this type that represent classes in the image of β_n, since β_n acts on representative maps of cubes into R,x by allowing the bases to expand from the single point x into all of A. Furthermore any class of cubes of this type in R,A,x can be realized by such an expansion of a cube in R,x, as shown in Fig. 76. The image of β_n is therefore precisely the kernel of γ_n.

(3) The image of α_n is the kernel of β_n. Figure 77 illustrates the two homomorphisms. The image and the kernel in question are both homotopy classes of maps of the n cube into R,x. The assertion is proved by noting that such a map is representative of either class if

[67] The last three homomorphisms on the list make sense if R is a group G, A is a subgroup H, and x is the identity e, for in this case we can give $\pi_1(R,A,x)$, $\pi_0(A)$, and $\pi_0(R)$ a group structure. To establish that the mappings remain homomorphisms in these cases requires slightly different arguments, which are essentially those used in establishing the fundamental theorem for computing the fundamental group.

[68] The kernel of a homomorphism is the set taken into the identity by that homomorphism.

[69] The main difficulty in following the three arguments is simply keeping straight in one's mind which homomorphism is which. Readers are urged to focus their attention on the appropriate figures and their captions, using the text itself only as a guide to the figures.

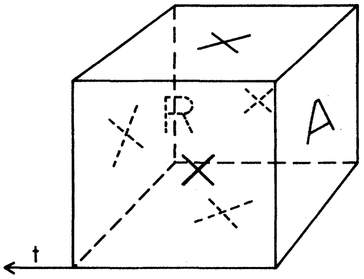

FIG. 74. The 3-cube in R,A,x on which γ_3 acts, reinterpreted as a homotopy showing that the associated map of the 2-cube into A,x is homotopic to a constant in R. This is the crucial point in establishing that the image of γ_3 is the kernel of α_2.

and only if there is a map of an $(n+1)$ cube into R which agrees with the map of the n cube into R on one of its faces [called the "side" of the $(n+1)$-cube] and takes an adjacent face (called the base) into A and the remaining faces into the point x, as shown in Fig. 78.

That the side of Fig. 78 furnishes maps of the n cube into R,x representing homotopy classes in the kernel of β_n follows from the fact that the image of the $(n+1)$ cube provides (or can be constructed out of) a homotopy between a map of the n cube into R,x and the trivial map of the n cube into x, via maps taking the n cube into R,A,x. [The parameter t in the homotopy can be taken as the coordinate along the direction normal to the side of the $(n+1)$ cube. Successive maps in the homotopy are given by vertical slices of Fig. 78.]

However, by using a different parametrization we can associate with the $(n+1)$ cube of Fig. 78 an extension of a map of an n cube into A,x [the base of the $(n+1)$ cube] to a map of an n cube into R,x [the side of the $(n+1)$ cube]. Successive stages of the extension are provided (Fig. 79) by swinging the base through 90° into the side about their common edge. Conversely, one can construct such an $(n+1)$ cube from the extension to R,x of any map of the n cube into A,x. Since the homomorphism α_n acts by just such extensions, the side of Fig. 78 furnishes all maps of the n cube into R,x that represent homotopy classes in the image of α_n.

D. Recovery of the fundamental theorems from an exact sequence

If the sets R, A, and x are a group G, a subgroup H, and the identity e, then we can formulate our earlier theorems as corollaries of the appropriate exact sequences. In doing this it is necessary to establish that $\pi_n(G/H)$—the fundamental group of the coset space G/H—can equally well be represented as the relative homotopy group $\pi_n(G,H)$. Substantial parts of the arguments given in establishing the fundamental theorems were devoted to this point. The underlying geometric idea is simply that requiring that the image of the entire surface of the cube go into the single coset H in coset space is tantamount to requiring that the entire surface of the cube go into the subgroup H, in a representative map of the cube into G. The subtlety in the argument lies in establishing that any continuous map of a cube into coset space can, in fact, be represented at all by a map of the cube into G.

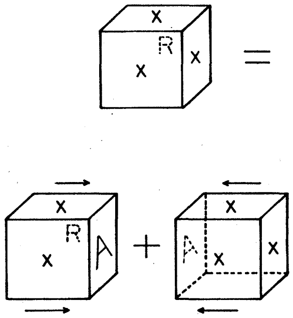

FIG. 76. How to construct a cube in R,x that β_3 takes into any cube in R,A,x that represents a map in the kernel of γ_3. The cube in R,A,x is on the lower left. Since it represents a map in the kernel of γ_3, its base is null-homotopic in A,x. The cube on the lower right represents such a homotopy. When the two cubes are combined, A face to A face, they yield the cube in R,x shown at the top. The base expanding homotopy leading from this cube back to the original cube in R,A,x consists of shaving off successive layers coming from the cube in A,x starting at the x face and working down to the A face. (The shaving is accompanied by the length rescaling necessary to keep the object a cube at each stage.)

If we identify $\pi_n(G/H)$ with $\pi_n(G,H,e)$, we arrive at the exact sequence:

$$\cdots\pi_n(G)\to\pi_n(G/H)\to\pi_{n-1}(H)\to\pi_{n-1}(G)\to\cdots. \tag{9.3}$$

Our fundamental theorems now follow from the following very general observation:

If, in the exact sequence $G_1\to G_2\xrightarrow{\varphi}G_3\to G_4$, the groups G_1 and G_4 consist of the identity alone, then the homomorphism φ between G_2 and G_3 is in fact an isomorphism. To establish that φ is an isomorphism we must show (i) that it is onto and (ii) that it is one-to-one. The first point requires that the image of G_2 in G_3 be all of G_3. Exactness tells us that the image of G_2 in G_3 is the kernel of the next homomorphism. But the next homomorphism is into the group G_4 consisting of the identity alone, so its kernel is indeed all of G_3. Point (ii) requires that the only point in G_2 taken into the identity of G_3 by φ be the identity of G_2, i.e., the kernel

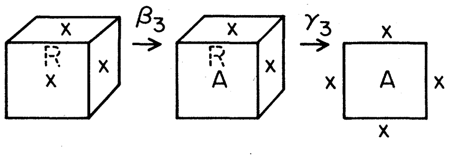

FIG. 75. The homomorphisms β_3 and γ_3.

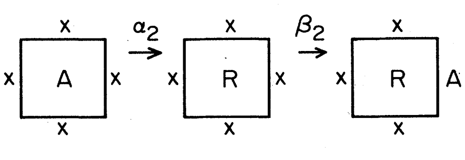

FIG. 77. The homomorphisms α_2 and β_2.

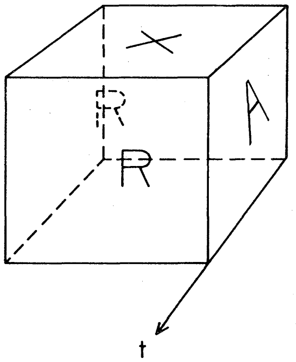

FIG. 78. A cube in R, A, x. The three unlabeled faces and the edge between the R face and the A face are all taken into the point x. The interior goes into R. The cube can be interpreted as a homotopy in R, A, x between a square in R, x (front vertical face) and the constant map (rear vertical face). The front vertical face therefore represents an arbitrary class in the kernel of β_2.

of φ is to consist of the identity alone. But exactness tells us that the kernel of φ is the image of the preceding homomorphism. The preceding homomorphism, however, acted on the group G_1 containing only the identity, and therefore its image in G_2 is indeed the identity alone.

Thus if the nth and $(n+1)$th homotopy groups of G are zero, then $\pi_n(G/H)$ is indeed isomorphic to $\pi_{n-1}(H)$, as our fundamental theorems asserted.

E. Uses of the third homotopy group; solitons

The third homotopy group provides the answer to the following classification problem:

Consider two singularity-free configurations of the order parameter, each subject to the constraint that the order parameter be uniform sufficiently far from the origin. Such a configuration provides a mapping of a solid cube into the order-parameter space, in which the entire surface of the cube is taken into the single point representing the constant asymptotic value. Thus any such configuration can be associated with a class of mappings that comprise an element of the third homotopy group. Two such configurations can be deformed into one another without altering the uniformity in the far region or introducing singularities, if and only if they are associated with the same element of π_3. Thus as π_1 classifies line singularities and π_2 classifies point singularities, π_3 classifies nonsingular configurations. (For some applications, see Shankar, 1977).

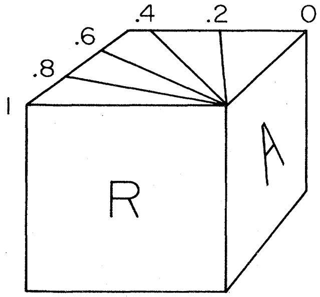

FIG. 79. The same cube as in Fig. 78. All unlabeled faces again go into x, as does the edge between the R and A faces. Another set of vertical plane sections is indicated that provide a homotopy in R between a map of a square into A, x and a map of a square into R, x. The front vertical face therefore represents an arbitrary class in the image of α_2.

Non-singular configurations can also be associated with the elements of the first and second homotopy groups, and it is into this scheme that the configurations given by π_3 fit most naturally:

Consider singularity free configurations of the order parameter that are subject to the constraint that the order parameter approaches a single value far from a given *plane*. A line drawn between the two regions of uniformity then determines a closed loop in the order parameter space, and two such configurations can be deformed into one another without the introduction of singularities if and only if the loops they determine are homotopic (at the base point determined by the fixed value far from the plane). Such configurations are therefore characterized by the elements of $\pi_1(R)$.

Similarly, singularity free configurations subject to the constraint that they are uniform far from a given *line*, are classified by the elements of $\pi_2(R)$. The π_3 configurations result when uniformity is imposed far from a single *point*.

The relation between the configurations classified by π_1 or π_2 [called planar or linear *solitons* by Mineyev and Volovik (1978)] and the corresponding line or point singularities is quite intimate. If a line (or point) singularity moves across a uniform medium it will leave a planar (or linear) soliton in its wake. Conversely, to remove a planar (or linear) soliton without relaxing the constraint far from the plane (or line) it suffices to move a line (or point) defect of the inverse type across a plane (or along a line) in the region of non-uniformity.[70]

The point solitons classified by π_3 have a somewhat tenuous physical stability, for the following reason: Associated with any deviation from uniformity there is a "bending energy" density which to leading order is generally quadratic in the gradients of the order parameter. Now any nonsingular configuration of the order parameter $f(\mathbf{r})$ that is uniform for $r > R$ can be deformed into a nonsingular configuration $f_t(\mathbf{r}) = f(\mathbf{r}/t)$ that is uniform for $r > tR$. As $t \to 0$ a singularity builds up in the neighborhood of the origin. The divergence in the bending energy density is of order $1/t^2$. However, the bending energy density is only nonzero for $r < tR$—i.e., in a region whose volume is of order t^3. The energy associated with this collapse of the configuration is therefore monotonically decreasing with t and actually vanishes at the moment of singularity. The topological singularity fails to provide an energy barrier for "phase space" reasons, and the topological classification scheme is spurious without the presence of additional stabilizing features.

[70] The planar solitons have precisely the spatial structure of the solitons beloved by students of nonlinear equations. Everything we have said about the classification and combination of line defects can be translated into the corresponding statements about such solitons. Note that planar solitons are unstable against expansion for the same energetic reasons that point (π_3) solitons are unstable against collapse. This is why they are always considered in the presence of uniform symmetry breaking fields that compress the nonuniformity into the neighborhood of a single plane.

There are ways of stabilizing the configurations. One could, for example, take into account additional higher-order gradients, which would restore the singularity at the moment of total collapse. These, however, would lead to a configuration of minimum free energy in which the equilibrium configuration would have a microscopic size determined by the ratio of the coefficients of the second and fourth order gradients. The topological scheme would again be relevant, but the kinds of objects it was describing would be rather different.

It is appropriate to conclude on this cautionary note. Not only can there be energetic barriers as well as topological ones, but—at least in the present example—topological barriers need not, in general, imply insurmountable energetic barriers. The energetics of a problem must always be examined before conclusions from the topology can be used with complete confidence.

APPENDIX A: GLOSSARY OF TECHNICAL TERMS

This short glossary lists the most important frequently recurring terms. Citations are to the definitions. If the citation is to a section number alone (e.g., III) then the definition appears at the beginning of the section before the appearance of subsection letters; if the citation is to a section number and subsection letter (e.g., III.A) then the definition appears at the beginning of the subsection before the appearance of subsubsection numbers. When a term is defined in the text it appears in italics (unless it appears in a heading) to make the references easier to locate. Technical terms of a purely group-theoretic nature are not listed in the glossary; the reader should consult Appendix B for a summary of basic group-theoretic concepts and results.

Action of π_1 on π_2	VII.B,D
Action of π_1 on π_n	IX.A
Based fundamental group $\pi_1(R,x)$	III.A.5
Based homotopy	III.A.2
Biaxial nematic	II.A.4
Burgers vector	VIII.C.1
Cholesterics	VIII.C.4
Classes of homotopic loops	III.A.4
Compatibility conditions	VIII.B
Continuous group	IV.A
Crystalline media	VIII
Defect	II
Directed stripes	VIII.C.3
Disclination in crystals	VIII.C.2
Disclination in liquid crystals	V.B.3
Dislocation	VIII.C.1, 2
Double group	V.B.4
Exact homotopy sequence	IX.C
Exact sequence of homomorphisms	IX.C
First homotopy group = fundamental group	
Fixer	IV.B
Freely homotopic	III.C
Fundamental group $\pi_1(R)$	III.B.1
Based $\pi_1(R,x)$	III.A.5
Group—see homotopy group, Lie group, topological group, etc.	
Group product of loops	IV.D
Higher homotopy groups $\pi_n(R)$	IX.A
Homotopic	II.B
at a point	III.A.2
freely	III.C
Homotopy	II.B
based	III.A.2
classes	III.A.4
Homotopy group, first, $\pi_1(R)$	III
nth, $\pi_n(R)$	IX.A
relative, $\pi_n(R,A)$	IX.B
second, $\pi_2(R)$	VII.A
zeroth, $\pi_0(R)$	III.A, IV.A, IX.B
Homotopy sequence, exact	IX.C
Isomorphism, path	III.B.1
Isotropy subgroup	IV.B
Lie group	IV.A
Lift	IV.C.3
Liquid crystal, biaxial nematic	II.A.4
cholesteric	VIII.C.4
nematic	II.A.3
smectic	VIII.C.3
Little group	IV.B
Local surgery	II.B
Loop product	III.A.3
Manifold of internal states	II
Naive generalization	VIII, VIII.A
Nematic	II.A.3
biaxial	II.A.4
n-simple	IX.A
nth homotopy group $\pi_n(R)$	IX.A
Ordered medium	II
Order parameter	II
Order parameter, reference	IV.B
Order-parameter space	II
Ordinary spins	II.A.2

Path	III.B
Path isomorphism	III.B.1
Planar spins	II.A.1
Product, classes of loops	III.A.4
Product, group, of loops	IV.D
Product, loop	III.A.3
Product space	III.D
Product, spheres	VII.A
Product, topological	III.D
Projective plane, P_2	II.A.3
Projective space, P_3	II.A.5
Quaternion group	V.B.4
Reference order parameter	IV.B
Relative homotopy group, $\pi_n(R,A)$, $\pi_n(R,A,x)$	IX.B
Removable singularity	II.B
Representative	III.A.4
Second homotopy group, $\pi_2(R)$	VII.A
based, $\pi_2(R,x)$	VII.A
Simply connected	III.D
Smectic	VIII.C.3
Solitons	IX.E
Spheres, product of	VII.A
Sphere, S_n	IX.A
Spins, ordinary	II.A.2
planar	II.A.1
Standard order parameter	IV.B
Subgroup, isotropy	IV.B
Superfluid helium-3	II.A.5
Surgery, local	II.B
Topological barrier	II.B
Topologically equivalent defects	II.B
Topological group	IV.A
Topological product	III.D
Topologically stable, unstable	II.B
Trivial	III.A.4
Two-simple	VII.B
Uniform medium	II
Universal covering group	V
Vacancy	VIII.C.1
Winding number	II.B
Zeroth homotopy group $\pi_0(R)$	III.A, IV.A, IX.B

APPENDIX B: A SUMMARY OF THE RELEVANT ELEMENTARY GROUP THEORY

A *group* G is a set of elements and a rule (or *combination law*) associating with any two elements a and b, a third, c, known as their product (or sum) and written $c=ab$ (or $c=a+b$). The combination law need not be commutative. If it is the group is said to be *Abelian* and the additive notation is often (but not always) used. If the group is non-Abelian only the multiplicative notation is used.

For a set G and a combination law to constitute a group the combination law must be *associative* [i.e., $a(bc)=(ab)c$ for all a, b, and c in G], there must be a unique *identity* element e in G satisfying $ea=ae=a$ for all a in G, and every element a in G must have *an inverse* a^{-1} satisfying $aa^{-1}=a^{-1}a=e$.

Groups can have a finite number of elements or infinitely many. In the finite case the number of elements is called the *order* of the group. Finite or denumerably infinite groups are said to be *discrete*.

Two quite different kinds of group play major roles in the topological theory of defects:

(1). Continuous groups of transformations that act on a space of vectors, tensors, etc. Such groups are commonly encountered in many branches of physics and particularly in the quantum theory. The combination law is simply given by the successive application of two transformations.

(2) Discrete groups (homotopy groups) characterizing the topological structure of spaces. Such groups are the mainstay of algebraic topology, but have played a relatively limited role in physics. The elements of such groups and their combination laws are quite different from the kinds tranditionally encountered in physical applications of group theory.

In our treatment of transformation groups the most important secondary concepts are those of subgroup, coset, normal subgroup, and quotient group. In our treatment of homotopy groups the most important secondary concepts are those of conjugacy classes, isomorphism, automorphism, and homomorphism.

A *subgroup* H of a group G is a subset of G which is itself a group. It is easily shown that H is a subgroup of G if and only if ab^{-1} lies in H for every a and b in H.

If H is a subgroup of G whose elements are h_i and g is any given element of G (which may or may not lie in H itself) then the set of all elements gh_i is called a *coset* of H. Such a coset is denoted by the symbol gH. (More precisely we have defined here a *left coset*; *right cosets* are similarly defined, but since we deal almost exclusively with left cosets, we drop the qualification except where the distinction between left and right cosets is essential.) A very important elementary theorem establishes that if g_1 and g_2 are two elements of G then the cosets g_1H and g_2H are either identical sets, or have no common elements whatever. Thus a given subgroup provides a partitioning of the group into disjoint cosets. This *space of cosets* of the subgroup H in the group G is denoted by the symbol G/H.

H is said to be a *normal subgroup* of G if the left coset gH contains the same elements as the right coset

Hg, for each g in G. If the subgroup H is a normal subgroup then one can impose a group structure on the coset space, defining the product of two cosets to be given by $(g_1H)(g_2H)=(g_1g_2)H$. (One needs H to be a normal subgroup to show that the result of this composition law is independent of the choice of elements g_1 and g_2 chosen to represent the cosets.) When H is a normal subgroup the coset space G/H is called the *quotient group* or *factor group.*

The *direct product* of two groups G and K is a set $G\times K$ of pairs (g, k) which can be shown to be itself a group under the combination law $(g_1, k_1)(g_2, k_2) = (g_1g_2, k_1k_2)$.

Two groups are said to be *isomorphic* if their group-theoretic structures are identical. More precisely, there must be a one-to-one map φ associating with each element g of the group G an element $k=\varphi(g)$ of the group K, such that every member of K is associated in this way with some element of G(i.e., φ is *onto* K) and such that the element of K corresponding to a product of elements of G is the product of the corresponding elements:

$$\varphi(g,g_2)=\varphi(g_1)\varphi(g_2)\,. \tag{B1}$$

If the groups K and G are the same group (so that φ simply permutes the elements of G) then the isomorphism is called an *automorphism.* A *homomorphism* is a correspondence between elements of two groups which is not necessarily one-to-one (i.e., it can take many elements of G into the same element of K) and which is not necessarily onto (i.e., it can leave some elements of K unassociated with any element of G) but which does satisfy the structural condition (B1). The set of elements of G taken by φ into the identity of K is called the *kernel* of the homomorphism. The set of elements of K that are associated with elements of G by φ is called the *image* of the homomorphism. For a homomorphism from G to K to be an isomorphism its kernel must consist of the identity of G alone and its image must be all of K.

Two elements a and b of G are said to be *conjugate* to one another if there is an element g of G such that $b=gag^{-1}$. The set of all elements conjugate to a given element a is called the *conjugacy class* of a. (In the theory of group representations conjugacy classes are usually refered to simply as *classes.*) It is an elementary theorem that G can be partitioned into disjoint conjugacy classes.

The operation of conjugation by a fixed element g ($a\to gag^{-1}$ for each a in G) defines an automorphism of G known as an *inner automorphism.* Only non-Abelian groups can have nontrivial inner automorphisms.

The most important continuous groups we make use of are the groups $\mathrm{T}(n)$ of all translations in n-dimensional Euclidean space, the groups $\mathrm{SO}(n)$ of all proper rotations in n-dimensional Euclidean space, and the group SU(2) of 2×2 unitary matrices with unit determinant. [We make little use of the full group $\mathrm{O}(n)$ of proper and improper rotations.]

The most important discrete groups we make use of are the groups Z_n isomorphic to the additive group of integers modulo n (also known as the *cyclic groups* of order n), the group Z, isomorphic to the additive group of the integers, and the quaternion group Q, defined in Sec. V.B.4.

We make use of the following features of SU(2) and its homomorphic correspondence to the three-dimensional proper rotation group SO(3):

Let σ_x, σ_y, and σ_z be the Pauli matrices:

$$\sigma_x=\begin{pmatrix}0&1\\1&0\end{pmatrix},\ \sigma_y=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\ \sigma_z=\begin{pmatrix}1&0\\0&-1\end{pmatrix}. \tag{B2}$$

Any unitary 2×2 matrix with unit determinant can be written in the form

$$u=a_0+i\mathbf{a}\cdot\sigma\,, \tag{B3}$$

where (a_0, a_x, a_y, a_z) is a real unit 4-vector:

$$a_0^2+a_x^2+a_y^2+a_z^2=1\,. \tag{B4}$$

The parameter space for SU(2) is thus the surface S_3 of the unit sphere in Euclidean 4-space, which is shown to be simply connected by the same argument as is given for the surface S_2 of the unit 3-sphere, in Sec. II.C.

Instead of parametrizing SU(2) with a unit 4-vector as in (B.3), one can use an equivalent parametrization in terms of a unit 3-vector $\hat{n}$ and an angle θ in the interval $0\leqslant\theta<4\pi$:

$$u(\hat{n},\theta)=\cos\tfrac{1}{2}\theta+i\sin\tfrac{1}{2}\theta(\hat{n}\cdot\sigma)=\exp(i(\theta/2)\hat{n}\cdot\sigma)\,. \tag{B5}$$

Note that $u(\hat{n},\theta+2\pi)=-u(\hat{n},\theta)$.

The clumsier parametrization (B.5) permits the homomorphic correspondence between SO(3) and SU(2) to be stated quite simply. Let $R(\hat{n},\theta)$ be a proper rotation about an axis $\hat{n}$ in 3-space through an angle θ. Then the mapping

$$\varphi\colon\ u(\hat{n},\theta)\to R(\hat{n},\theta) \tag{B6}$$

is two-to-one from SU(2) onto SO(3), since $R(\hat{n},\theta) = R(\hat{n},\theta+2\pi)$. Furthermore it can be shown to be a homomorphism in the sense of (B1).[71]

The topological significance of this relation is that SU(2) is the universal covering group for SO(3)—i.e., it is a simply connected group of which SO(3) is the homomorphic image. It is for this reason that SU(2) and its subgroups play so central a role in the topological theory of defects in three-dimensional media.

Because u and $-u$ correspond to the same rotation under the covering homomorphism, if SU(2) is parametrized by the surface of a 4-sphere [as in Eqs. (B3) and (B4)], then the parameter space for SO(3) can be taken to be the surface of the 4-sphere with the identification of diametrically opposite points. Alternatively, if SO(3) is parametrized directly through the rotation axis $\hat{n}$ and rotation angle θ, then the parameter space can be taken to be a solid 3-sphere of radius π (containing points $\theta\hat{n}$) in which diametrically opposite surface points are identified (since rotations of π and $-\pi$ about the same axis are identical).

[71] This is readily verified for infinitesimal rotation angles. If $\theta=\delta\phi$ in Eq. (B5), then $\delta(u\boldsymbol{\sigma}u^\dagger)=(i/2)\,\delta\phi[\hat{n}\cdot\sigma,\boldsymbol{\sigma}]=\delta\phi\hat{n}\,x\boldsymbol{\sigma}$, the last form following directly from the commutation relations $[\sigma_i,\sigma_j]=2i\epsilon_{ijk}\sigma_k$ obeyed by the Pauli matrices (B2). But this last form is precisely the change induced by an infinitesimal rotation about the axis $\hat{n}$ through the angle $\delta\phi$.

REFERENCES

Anderson, P. W., and R. G. Palmer, 1977, in *Quantum Fluids and Solids*, edited by S. B. Trickey, E. D. Adams, and J. W. Dufty (Plenum, New York).

Anderson, P. W., and G. Toulouse, 1977, Phys. Rev. Lett. **38**, 408.

Ashcroft, N. W., and N. D. Mermin, 1976, *Solid State Physics* (Holt, Rinehart and Winston, New York).

Bailin, D., and A. Love, 1978, J. Phys. A **11** L219.

Bouligand, Y., B. Derrida, V. Poénaru, Y. Pomeau, and G. Toulouse, 1978, J. Phys. (Paris) **39**, 863.

Cartan, E., 1936, *La Topologie des Groupes de Lie* (Hermann, Paris).

Cross, M. C., and W. F. Brinkman, 1977, J. Low Temp. Phys. **27**, 683.

de Gennes, P. G., 1974, *The Physics of Liquid Crystals* (Oxford University, New York).

Hilton, P. J., 1953, *An Introduction to Homotopy Theory* (Cambridge University, Cambridge).

Kleman, M., 1977a, J. Phys. Lett. (Paris) **38**, L-199.

Kleman, M., 1977b, "Points, lignes, parois dans les fluides anisotropes et les solids cristallins" [J. Phys. (Paris)].

Kleman, M., and L. Michel, 1978, Phys. Rev. Lett. **40**, 1387.

Kleman, M., L. Michel, and G. Toulouse, 1977, J. Phys. Lett. (Paris) **38**, L-195.

Mermin, N. D., 1977, in *Quantum Fluids and Solids*, edited by S. B. Trickey, E. D. Adams, and J. W. Dufty (Plenum, New York).

Mermin, N. D., 1978a, in *Quantum Liquids*, edited by J. Ruvalds and T. Regge (North-Holland, Amsterdam).

Mermin, N. D., 1978b, J. Math. Phys. **19**, 1457.

Mermin, N. D., V. P. Mineyev, and G. E. Volovik, 1978, J. Low Temp. Phys. **33**, 117.

Mineyev, V. P., and G. E. Volovik, 1978, Phys. Rev. B **18**, 3197.

Poénaru, V., and G. Toulouse, 1977, J. Phys. (Paris) **8**, 887.

Pontryagin, L. S., 1966, *Topological Groups*, 2nd edition (Gordon and Breach, New York).

Rogula, D., 1976, in *Trends in Applications of Pure Mathematics to Mechanics*, edited by G. Fichera (Pitman, London).

Shankar, R., 1977, J. Phys. (Paris) **38**, 1405.

Steenrod, N., 1951, *The Topology of Fibre Bundles* (Princeton University, Princeton, N.J.).

Toulouse, G., and M. Kleman, 1976, J. Phys. Lett. (Paris) **37**, L-149.

Toulouse, G., 1977, J. Phys. Lett. (Paris) **38**, L-67.

Volovik, G. E., and V. P. Mineyev, 1976, Zh. Eksp. Teor. Fiz. Pis'ma Red. **24**, 605 [JETP Lett. **24**, 561 (1976)].

Volovik, G. E., and V. P. Mineyev, 1977, Zh. Eksp. Teor. Fiz. **72**, 2256 [Sov. Phys.-JETP **45**, 1186 (1977)].

Volovik, G. E., 1978 (submitted to Zh. Eksp. Teor. Fiz. Pis'ma Red.).

Weyl, H., 1946, *The Classical Groups*, 2nd edition (Princeton University, Princeton, N.J.).