NON-ABELIAN DISCRETE GAUGE THEORY

Thesis by

Kai-Ming Lee

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1994
(Submitted May 20, 1994)

Acknowledgments

There are many people who have helped me in these years. To thank them one by one will certainly lead to the omission of many. Thus, let me just mention the ones that I have worked with in my research. First of all, I wish to express my gratitude to my advisor, Prof. John Preskill. He does not only provide me with advice and guidance, but he also shares with me many of his ideas and insights. It is always a pleasure to talk to him. Without him, graduate school would be a much less enjoyable period in my life. I wish to thank Hoi-Kwong Lo. His questions and comments on my work are always valuable. I would also like to thank Martin Bucher. We worked together on the subject presented in chapter 3.

Abstract

Gauge theory with a finite gauge group (or with a gauge group that has disconnected components) is systematically studied, with emphasis on the case of a non-Abelian gauge group. An operator formalism is developed, and an order parameter is constructed that can distinguish the various phases of a gauge theory. The non-Abelian Aharonov-Bohm interactions and holonomy interactions among cosmic string loops, vortices, and charged particles are analyzed; the detection of Cheshire charge and the transfer of charge between particles and string loops (or vortex pairs) are described. Non-Abelian gauge theory on a surface with non-trivial topology is also discussed. Interactions of vortices with "handles" on the surface are discussed in detail. The electric charge of the mouth of a "wormhole" and the magnetic flux "linked" by the wormhole are shown to be non-commuting observables. This observation is used to analyze the color electric field that results when a colored object traverses a wormhole.

Table of Contents

Chapter 1: Introduction

It is well known that a classical black hole can carry only a few kinds of hair. They are the mass, angular momentum and electric charge. This is the "no hair theorem." However, a black hole may also carry quantum hair if there is a gauge symmetry that is spontaneously broken.[1] The simplest model has $U(1)$ gauge symmetry and two charged scalar fields. One scalar has charge N and the other has charge 1. If the charge N field condenses, the gauge symmetry will be broken from $U(1)$ to Z_N. All $U(1)$ charges will be screened modulo N. This means that a particle that has charge $n + N$ will behave the same as a particle with charge n after the symmetry breaking. The black hole now can carry Z_N charge. Since the unbroken gauge group is discrete, there is no low energy local excitations of the gauge field. We could conclude that classically the black hole carries no charges. We have to rely on the quantum interference to detect the Z_N charge.

Let us recall the Aharonov-Bohm effect.[2] When we do a double slit experiment using electrons instead of photons, we would observe the interference patterns similar to the case of photons. If we put a long solenoid between the slits, the interference patterns will be shifted according to the magnitude of the magnetic flux inside the solenoid. This is because the wavefunction of the electrons will acquire different quantum phases for different paths around the solenoid. The phase difference will depend on, apart from the geometric setup, the magnetic flux. If we deform the paths continuously without crossing the solenoid, this part of the phase difference does not change. It is a topological interaction. From the relation between the phase shift and the magnetic flux inside the solenoid, we can deduce the charge of the electrons. This is a quantum effect and it does not depend on whether the gauge group is discrete or not.

In the model discussed in the first paragraph, there is a stable topological line defect (in 3+1 dimensions), called a cosmic string,[3] carrying the Z_N magnetic flux. By the Aharonov-Bohm effect, we can detect the Z_N charge of the black hole by sending the black hole around the cosmic string and observing the interference pattern. As discussed above, this is a quantum effect and has no classical counterpart. The Z_N charge also affects the Hawking temperature of the black hole.[4] As expected, when

the black hole carries only a few units of Z_N charge compared to N, increasing the charge will decrease the Hawking temperature.

The unbroken gauge group of the above model is Abelian;[1,5,4] it could be non-Abelian for other models.[5,6,7,8,9] We can also consider $2+1$ dimensional spacetime or even spacetime with non-trivial topology. And the role of the black hole in the above model is to illustrate that we don't have to "touch" the tested object to find out its charge. The argument applies to any charged particles. Thus, we now consider a more general model which, at high energy, has continuous gauge symmetry. At low energy, the gauge symmetry is spontaneously broken to a symmetry with finite gauge group or a gauge group with discrete components. There will be stable topological defects. In $3+1$ dimensions, the defect is one-dimensional, and called a cosmic string. In $2+1$ dimensions, the defect is pointlike, and called a vortex. The cosmic string or vortex will have Aharonov-Bohm interactions with the charged particles.

If the unbroken gauge group is Abelian, the physics will be similar to the physics of the black hole and the cosmic string discussed above. If the unbroken gauge group is non-Abelian, there will be the non-Abelian analogue of the Aharonov-Bohm effect which means that a charged particle winding around a cosmic string loop or a vortex will be transformed by an element of the unbroken gauge group. Also for the non-Abelian theory, a cosmic string loop or a pair of vortex-antivortex may carry electric charge, the "Cheshire charge." [6] It has no localized source but can be detected by the Aharonov-Bohm effect.

Besides the interactions with the string loops or vortices, charged particles could have similar interactions with the non-trivial topology of the space.[10] This can be easily visualized in $2+1$ dimensional spacetime because the topology of surfaces is well understood. The handles of a surface can carry magnetic flux and Cheshire charge. A charged particle that travels through a handle will be transformed just as if it has wound around a vortex.

The $3+1$ dimensional analogue of a handle is a wormhole. The above analysis applied to a wormhole tells us that the electric charge of a wormhole mouth and the magnetic flux "linked" by the wormhole are not commuting observables.[11] We can specify the values of one or the other to fully specify the state of the wormhole.

In this thesis, the main topic is non-Abelian discrete gauge theory. The interactions between vortices, charged particles and the topology of the space will all be considered. In chapter 2,[8] a charge operator is constructed for quantum field theory with non-Abelian discrete gauge symmetry. A non-local order parameter is formulated to distinguish the various phases. The construction of the operators in a lattice is explicitly described. Also, we investigated the Aharonov-Bohm interactions between string loops or vortices and charged particles, holonomy interactions between string loops, transfer of Cheshire charge, decay of domain walls and the Aharonov-Bohm interactions of magnetic monopoles with electric flux tubes. The chapter is a generalization to the non-Abelian theory of the work done on the Abelian theory in Ref. 5.

In chapter 3,[9] we will concentrate on the detection and the transfer of Cheshire charge. In particular, we will describe the transfer of electric charge from a particle to a string loop, and will describe the detection of the Cheshire charge on that loop, using the language of gauge-invariant correlation functions.

We consider the effect of the topology of the space in chapter 4.[10] Since the topology of two-dimensional space is easier to understand, we investigate the interactions of vortices on surfaces. If the genus of the surface is greater than zero, the handles can carry magnetic flux and Cheshire charge. The motion of the vortices can be described by the braid group of the surface. How the motion of the vortices affects the state is analyzed in detail. We also describe the symmetry algebra, the quantum double, and the most general point like excitations, dyons, allowed by the theory.

In chapter 5,[11] we use the technique of chapter 4 to resolve some puzzles in wormhole electrodynamics and chromodynamics. We discuss how to measure the charge and flux of a wormhole by the Aharonov-Bohm effect. We find that after a colored object traverses a wormhole, its state will be entangled with the state of the wormhole. This is related to the fact that the charge and flux of a wormhole are not commuting observables.

There is also some final remarks in chapter 6.

REFERENCES

1. L.M. Krauss and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1221; L.M. Krauss, Gen. Rel. Grav. **22** (1990) 253.
2. Y. Aharonov and D. Bohm, Phys. Rev. **119** (1959) 485.
3. J. Preskill, Vortices and monopoles, *in* Architecture of the fundamental interactions at short distances, ed. P. Ramond and R. Stora (North-Holland, Amsterdam, 1987).
4. S. Coleman, J. Preskill and F. Wilczek, Nucl. Phys. **B378** (1992) 175.
5. J. Preskill and L. Krauss, Nucl. Phys. **B341** (1990) 50.
6. M. Alford, K. Benson, S. Coleman, J. March-Russell and F. Wilczek, Phys. Rev. Lett. **64** (1990) 1632 [Erratum: **65** (1990) 668]; Nucl. Phys. **B349** (1991) 414.
7. A. S. Schwarz, Nucl. Phys. **B208** (1982) 141; M. Bucher, Nucl. Phys. **B350** (1991) 163; M. Alford, S. Coleman and J. March-Russell, Nucl. Phys. **B351** (1991) 735; M. Alford and J. March-Russell, Nucl. Phys. **B369** (1992) 276; M. Bucher, H.-K. Lo and J. Preskill, Nucl. Phys. **B386** (1992) 3; H.-K. Lo and J. Preskill, Phys. Rev. **D48** (1993) 4821.
8. M.G. Alford, K.-M. Lee, J. March-Russell and J. Preskill, Nucl. Phys. **B384** (1992) 251.
9. M. Bucher, K.-M. Lee and J. Preskill, Nucl. Phys. **B386** (1992) 27.
10. K.-M. Lee, Phys. Rev. **D49** (1994) 2030.
11. H.-K. Lo, K.-M. Lee and J. Preskill, Phys. Lett. **B318** (1993) 287.

Chapter 2: Non-Abelian Strings and Vortices

In many cases, a gauge theory that undergoes the Higgs mechanism will contain topologically stable defects. Among the defects that can occur are line defects ("cosmic strings") in 3+1 dimensions, and pointlike vortices in 2+1 dimensions.[1]

If the symmetry breaking pattern and the matter content of the theory are suitable, the strings or vortices may have long range interactions with various particles. These interactions are a consequence of the Aharonov-Bohm phenomenon—the wave function of a particle acquires a non-trivial phase when the particle is covariantly transported around the string.[2,3]

This simple observation has interesting implications.[4–6] Because the Aharonov-Bohm interaction has infinite range, and no local operator can destroy an object that has an infinite-range interaction with another object, gauge theories with such interactions always respect non-trivial superselection rules. The structure of the superselection sectors can be invoked to distinguish among the various possible phases of a gauge theory. Furthermore, the existence of infinite range interactions that are fundamentally quantum mechanical exposes the limitations of the "no-hair" theorems of black hole physics. Though a black hole may have no *classical* hair, it *can* carry quantum numbers that are detectable by means of the interaction between the black hole and a cosmic string.[4,7–9]

The "Aharonov-Bohm phase" acquired by an object that circumnavigates a string is, in general, a gauge transformation contained in the unbroken gauge group. When the manifest gauge symmetry is non-Abelian, this gauge transformation might not lie in the center of the group. In that case, we say that the string is "non-Abelian." [10] The physical properties of a non-Abelian string are qualitatively different from the properties of an Abelian string.[5,6,11–13] In particular, a loop of non-Abelian string (or a *pair* of non-Abelian vortices, in 2+1 dimensions) can carry a non-trivial gauge charge, so that the loop has an Aharonov-Bohm interaction with other strings. Moreover, non-Abelian Aharonov-Bohm interactions involve transfer of charge between string loops and charged particles. Remarkably, the charge carried by a loop (which has a topological origin) cannot be localized anywhere on the string loop or in its vicinity. Following Ref. 13, we will refer to such unlocalized charge as "Cheshire charge."

If we are to appeal to the Aharonov-Bohm effect to probe the phase structure of a gauge theory, or to investigate the quantum physics of black holes, we must be able to discuss interference phenomena in a language that does not rely on weak-coupling perturbative methods; we need a framework that (at least in principle) takes full account of the effects of virtual pairs and of the fluctuations of quantum fields. Such a framework was erected, for Abelian strings (or vortices), in Ref. 5. There, operators were constructed that create a loop of string, or that introduce (as a classical source) the closed world sheet of a string. The correlation functions of these operators can be studied to investigate the properties of the strings, and their Aharonov-Bohm interactions in particular.

Our main objective in this chapter is to generalize the work of Ref. 5 to the case of non-Abelian strings. Because of the subtle and elusive physics of non-Abelian strings, this generalization is not entirely straightforward.[14] Our primary motivation comes from two considerations. First, we seek assurance that the exotic physics of the non-Abelian Aharonov-Bohm effect, previously inferred in the weak-coupling limit, actually survives in a fully quantum field-theoretic treatment. Second, we hope to construct (non-local) order parameters that can be used to classify the phases of a gauge theory.

Let us formulate the order-parameter problem more explicitly, and in so doing, review some of the principal results of Ref. 5.

A gauge theory can have an interesting phase diagram. Depending on its Higgs structure and on the parameters of the Higgs potential, the theory may be in a Coulomb (massless) phase, a Higgs phase, or a confinement phase. A Higgs phase is, roughly speaking, characterized by the existence of stable magnetic flux tubes, and a confinement phase by the existence of stable electric flux tubes.

Non-local gauge-invariant order parameters can be devised that distinguish among the various phases. The expectation value of the Wilson loop operator[15] exhibits area-law behavior if there are stable electric flux tubes, and perimeter-law behavior otherwise. The expectation value of the 't Hooft loop operator[16] exhibits area law behavior if there are stable magnetic flux tubes, and perimeter-law behavior otherwise.

These order parameters are not sufficient, however, to distinguish among all pos-

sible phases of a general gauge theory. Consider, for example, the case of an $SU(N)$ gauge theory with matter in the fundamental representation. In this theory, the Wilson loop always obeys the perimeter law, because an electric flux tube can break via nucleation of a pair. An 't Hooft loop operator can also be defined, but always obeys the area law.* Yet the theory can have a non-trivial phase diagram. Using adjoint Higgs fields, it is possible to break the gauge group down to its center Z_N. This phase, which admits free Z_N charges, is distinguishable from the confinement phase.

In Ref. 5, an order parameter was described that can distinguish the free-charge phase from the confinement phase in an $SU(N)$ gauge theory.† The idea is that the free-charge phase supports stable magnetic flux tubes (cosmic strings), and these strings have an *infinite-range* Aharonov-Bohm interaction with Z_N charges. No such interaction can exist if Z_N charges are confined, or if Z_N is spontaneously broken by a Higgs field that transforms as the fundamental representation. (Indeed, the confinement phase and the "Higgs" phase with Z_N completely broken are *not* distinguishable;[19] this is an instance of "complementarity."

To construct the order parameter, we must first devise an operator $F(\Sigma)$ that introduces a cosmic string world sheet on the closed two-dimensional surface Σ. (As we will discuss later, this operator is closely related to the 't Hooft loop operator.) Then consider the quantity

$$A(\Sigma, C) = \frac{F(\Sigma)W(C)}{\langle F(\Sigma)\rangle\langle W(C)\rangle} , \tag{2.1}$$

where $W(C)$ is the (fundamental representation) Wilson loop. In the free-charge phase, we have

$$\lim \langle A(\Sigma, C)\rangle = \exp\left(\frac{2\pi i}{N} k(\Sigma, C)\right) . \tag{2.2}$$

Here the limit is taken with Σ and C increasing to infinite size, and with the closest approach of Σ to C also approaching infinity; $k(\Sigma, C)$ denotes the linking number of

---

⋆ This is actually an oversimplification of the status of the 't Hooft loop, as we will discuss in Sections 4, 6, and 8.

† This order parameter was actually discussed earlier by Fredenhagen and Marcu in Ref. 17. See also Ref. 18.

the surface Σ and the loop C. On the other hand, if there are no free Z_N charges, then we have

$$\lim \langle A(\Sigma, C) \rangle = 1 \ . \tag{2.3}$$

The non-analytic behavior of $A(\Sigma, C)$ guarantees that a well-defined phase boundary separates the two phases. We will refer to $A(\Sigma, C)$ as the "ABOP," or "Aharonov-Bohm Order Parameter."

One of our objectives in this chapter is to generalize the above construction, and to further explore its consequences. In general, it is not sufficient to determine the realization of the center of the gauge group, in order to distinguish all possible phases of a gauge theory. For example, $SU(N)$ might break to a discrete subgroup H that is not contained in the center. Different unbroken groups (including non-Abelian ones) can be distinguished according to the varieties of cosmic stings that exist, and the nature of the Aharonov-Bohm interactions of these strings with free charges. To probe the phase structure more thoroughly, we need to construct a more general $F(\Sigma)$ operator. And we will need to consider carefully the interpretation of the behavior of the corresponding $A(\Sigma, C)$ operator. The interpretation involves subtleties associated with the non-Abelian Aharonov-Bohm effect.

In the example described above, there are other order parameters that can serve to distinguish the free-charge phase from the confinement phase.[17,18,20] But we believe that the properties of strings and of their Aharonov-Bohm interactions provide a more powerful method for classifying phases in more general cases.

The remainder of this chapter is organized as follows: Section 1 reviews the formalism for describing configurations of many non-Abelian vortices or strings. We emphasize the need to select an (arbitrary) basepoint for the purpose of defining the "magnetic flux" of a vortex, discuss the holonomy interactions between vortices, explain the origin of Cheshire charge, and note that non-Abelian strings generically become entangled when they cross each other.

In Section 2, we discuss the properties of domain walls that are bounded by loops of non-Abelian string, and observe that physically distinct strings can bound physically identical walls. Hence when a wall "decays" by nucleating a loop of string, several competing "decay channels" may be available.

Section 3 concerns Aharonov-Bohm interactions in Abelian gauge-Higgs systems. We explore the screening of such interactions due to Higgs condensation. The observations in Section 2 concerning domain wall decay are invoked, in order to interpret the results.

In Section 4, we begin our analysis of the effects of quantum fluctuations on the non-Abelian Aharonov-Bohm effect. We find that non-perturbative fluctuations cause ambiguities in the Aharonov-Bohm "phase" beyond those that occur in leading semiclassical theory. We construct operators that create configurations of many non-Abelian strings. Two types of operators are considered. One type introduces *classical* string sources on closed world sheets. The other type creates (or destroys) *dynamical* string loops.

A subtle aspect of the construction is that, for both types, the many string configurations are *coherent.* This means the following: The "magnetic flux" carried by a string can be characterized by an element a of the unbroken gauge group H. If an object that transforms as the representation (ν) of H is transported about this string, the Aharonov-Bohm phase that it acquires, averaged over a basis for the representation, is

$$\frac{1}{n_\nu}\chi^{(\nu)}(a) \ , \tag{2.4}$$

where n_ν is the dimension of (ν) and $\chi^{(\nu)}$ is its character.[6,21] Now if an a string and a b string are combined *incoherently*, then the averaged phase acquired by an object that traverses the two strings in succession is

$$\frac{1}{n_\nu}\chi^{(\nu)}(a)\ \frac{1}{n_\nu}\chi^{(\nu)}(b) \ ; \tag{2.5}$$

that is, it is the product of the Aharonov-Bohm factors associated with the two individual strings. But if the two strings are patched together *coherently*, then the averaged phase becomes

$$\frac{1}{n_\nu}\chi^{(\nu)}(ab) \ . \tag{2.6}$$

This property, that the Aharonov-Bohm factor associated with a pair of coherently combined strings is *not* just the product of the Aharonov-Bohm factors of the two individual strings, is a distinguishing feature of the non-Abelian Aharonov-Bohm

effect, and one of the goals of our work has been to better understand how this coherence is maintained when quantum fluctuations of fields are fully taken into account.

Correlation functions of the operators constructed in Section 4 are used in Section 5 to investigate Aharonov-Bohm interactions in a pure gauge theory, at both strong and weak coupling. In Section 6, we use these operators to study the quantum-mechanical mixing between different string states first discussed in Ref. 21.

Holonomy interactions between pairs of vortices, or pairs of string loops, are considered in Section 7. Correlation functions are used to demonstrate holonomy scattering and string entanglement. We also analyze a correlation function that realizes the transfer of charge from a charged source to a loop of string, and the subsequent detection of the (Cheshire) charge on the loop by means of its Aharonov-Bohm interaction with yet another loop of string.

In Section 8, we consider non-Abelian gauge-Higgs systems; we analyze the consequences of the Higgs mechanism for the stability of strings, and for the screening of Aharonov-Bohm interactions. We discuss how non-local order-parameters can be used to explore the phase structure of such a system.

Section 9 concerns the properties of dynamical magnetic monopoles, in a confining gauge theory. We construct generalizations of the Wilson and 't Hooft operators for a theory with dynamical monopoles, and use these operators to demonstrate the Aharonov-Bohm interaction between monopoles and electric flux tubes.[22,23]

The Appendix provides some additional details concerning some of the lattice calculations that are mentioned in the text of the chapter.

2.1. Non-Abelian vortices and strings

We will briefly review the formalism[24] for describing configurations of many vortices (in two spatial dimensions) or many strings (in three spatial dimensions). Our purpose is to remind the reader of two peculiar features of non-Abelian strings. First, there is an ambiguity when two or more loops of string are patched together to construct a multi-string configuration. Second, a loop of string can carry charges,

and can exchange charge with other objects by means of the non-Abelian Aharonov-Bohm effect. An understanding of these features is needed in order to interpret the behavior of the correlation functions that we will construct.

Many-Vortex Configuration We will consider vortices first, and discuss strings later. A single isolated vortex can be associated with an element of the unbroken gauge group H, according to

$$a(C, x_0) = P \exp\left(i \int_{C, x_0} A \cdot dx \right) . \tag{2.7}$$

Here C is a closed oriented path, far from the vortex core, that encloses the vortex and begins and ends at the point x_0. This group element $a(C, x_0)$ is invariant under deformations of the path C that keep x_0 fixed and that avoid the vortex core. An object that transforms as the irreducible representation (ν) of H acquires the "Aharonov-Bohm" phase $D^{(\nu)}(a(C, x_0))$ when covariantly transported around the vortex. We require $a \in H$, because the Higgs fields that drive the symmetry breakdown must return to their original values when so transported. The vortices can be topologically classified, with the topological charge taking values in $\pi_0(H)$; that is, a vortex with "flux" $a \in H$ can be smoothly deformed to another vortex with flux $b \in H$ if and only if a and b lie in the same connected component of H.[⋆] If H is a discrete group (as we will usually assume in this chapter), then, the topological charge is specified by an element of H.

Similarly, we may associate n elements of H with a configuration of n vortices. To do so, we must choose n standard paths, all beginning and ending at the same point x_0, that circumnavigate the various vortices.[24]

This description of the n-vortex configuration evidently suffers from some ambiguities. First, it is not gauge invariant. Under a gauge transformation that takes the value $h \in H$ at the point x_0 (and so preserves the Higgs order parameter at that

⋆ Here, we have implicitly defined H as the unbroken subgroup of a *simply connected* underlying (spontaneously broken) gauge group. The global topology of the underlying group is irrelevant, however, for the purpose of classifying the Aharonov-Bohm interactions of the vortices.

point), the elements $a_1, a_2, \ldots, a_n$ transform according to

$$a_i \to h a_i h^{-1} \ . \tag{2.8}$$

Thus, the configurations of a single vortex are labeled by the conjugacy classes of H. But the gauge freedom for many vortices involves just one overall conjugation. This means that specifying the positions and classes of all vortices is not sufficient to characterize the multi-vortex configuration. Single vortices can be patched together in various ways to construct different multi-vortex states.

Another ambiguity concerns the choice of the paths C_i that enclose the vortices. The paths that begin and end at x_0 and avoid the cores of n vortices fall into homotopy classes; these classes are the elements of $\pi_1(M_n, x_0)$, the fundamental group of M_n, the plane with n punctures. (This group is F_n, the free group with n generators.) By assigning elements of H to each of the generators of $\pi_1(M_n, x_0)$, we have defined a homomorphism from $\pi_1(M_n, x_0)$ into H. This homomorphism (modulo the one overall conjugation) characterizes the configuration of n vortices.

But this homomorphism is still ambiguous, because the n generators can be chosen in various ways. Consider in particular the case of two vortices, as shown in Fig. 1. Standard paths α_1 and α_2 have been chosen in Fig. 1a that wind counterclockwise around the two vortices. A topologically distinct choice for the path around vortex 1 is shown in Fig. 1c, and a distinct choice for the path around vortex 2 is shown in Fig. 1d. Suppose now that vortex 1 winds around vortex 2 (in the sense defined by the path α_2), and returns to its original position, as in Fig. 1b. We may deform our paths during the winding so that no vortex ever crosses any path; then each path is mapped to the same group element after the winding as before the winding. But after the winding, the final deformed path is not homotopically equivalent to the initial path. Therefore, the homomorphism that maps $\pi_1(M_2, x_0)$ to the group has changed.

Suppose, for example, that initially α_1 is mapped to a_1 and α_2 is mapped to a_2. We wish to determine the final values, after the winding, of the group elements associated with the paths α_1 and α_2. For this purpose, it is convenient to notice that, during the winding, the path shown in Fig. 1c is "dragged" to α_1. Therefore, the group element associated with this path before the winding is the same as the group

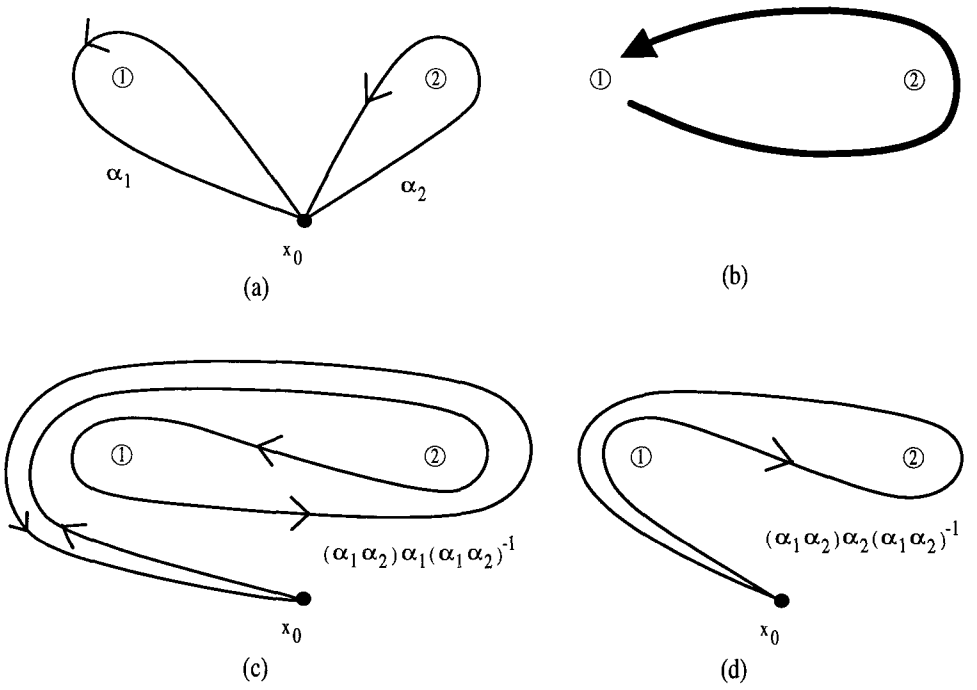

Fig. 1: (a) Standard paths α_1 and α_2, based at x_0, that enclose vortex 1 and vortex 2. (b) Vortex 1 winds around vortex 2. (c) Path that, during the winding of vortex 1 around vortex 2, gets dragged to α_1. (d) Path that gets dragged to α_2 during the winding.

element associated with α_1 after the winding. We see that this path can be expressed as $(\alpha_1\alpha_2)\alpha_1(\alpha_1\alpha_2)^{-1}$, where $\alpha_1\alpha_2$ denotes the path that is obtained by traversing α_2 first and α_1 second. The initial homomorphism maps this path to a path-ordered exponential with the value

$$a_1' = (a_1a_2)a_1(a_1a_2)^{-1} \ . \tag{2.9}$$

Similarly we note that the path shown in Fig. 1d is dragged during the winding to α_2. This path is $(\alpha_1\alpha_2)\alpha_2(\alpha_1\alpha_2)^{-1}$, and is mapped by the initial homomorphism to

$$a_2' = (a_1a_2)a_2(a_1a_2)^{-1} \ . \tag{2.10}$$

We conclude that the final homomorphism after the winding maps α_1 to a_1' and maps α_2 to a_2'. Winding vortex 1 around vortex 2 has changed the two-vortex state (if a_1 and a_2 do not commute)—both group elements have become conjugated by a_1a_2.

This change in the two-vortex state is not a mere mathematical curiosity. The physical interpretation is that there is a long-range interaction between non-commuting vortices.[25,24] We refer to this as the "holonomy interaction." It is a type

of Aharonov-Bohm interaction, but we prefer not to use that name in this context. The holonomy interaction is really a classical effect, while the term "Aharonov-Bohm" is best reserved for an intrinsically quantum-mechanical interaction. When the number of vortices is three or more, these holonomy interactions can be quite complicated.

Finally, another source of ambiguity concerns the choice of the basepoint x_0. Winding the basepoint around some of the vortices (or some of the vortices around the basepoint) also changes our description of the multi-vortex configuration. This feature is not of much physical interest. When it is convenient, we will consider the basepoint to be far from all vortices, so that winding of vortices around the basepoint need not be considered.

The gauge-invariant content of the classical n-vortex configuration can be encoded in the Wilson loop operators

$$W^{(\nu)}(C) = \frac{1}{n_\nu}\chi^{(\nu)}\left(P\exp\left(i\oint_C A\cdot dx\right)\right) , \tag{2.11}$$

where $\chi^{(\nu)}$ denotes the character, and n_ν the dimension, of irreducible representation (ν). If C is a closed path around a single vortex, then $W^{(\nu)}(C)$, for all (ν), provides sufficient information to identify the class to which the vortex belongs. Likewise, $W^{(\nu)}(C)$, for all (ν) and all C in $\pi_1(M_n)$ suffices to determine the n-vortex configuration, up to one overall conjugation. It is not sufficient though, to know $W^{(\nu)}(C_i)$ for all the generators C_i of $\pi_1(M_n)$; this determines only the class of each vortex, but not how the vortices are patched together. It is therefore more convenient and efficient to fix the gauge at the basepoint and assign group elements to standard paths around the vortices than to give a fully gauge-invariant description in terms of Wilson loops.

When the unbroken gauge group H is discrete, one sometimes says that the theory respects a "local discrete symmetry."[4-6,26-29] This terminology is used because a field that is covariantly constant outside the vortex core is typically not single valued—on a closed path that encloses a vortex, a covariantly constant field is periodic only up to the action of an element of H. (And, conversely, a field that *is* periodic cannot be covariantly constant.) But the phrase "local discrete symmetry" should not be misunderstood. Often, when we say that a symmetry is local, we mean that all

physical states are required to be invariant under the symmetry. That is not what is meant here. Isolated H charges can exist, and can, in fact, be detected at infinite range via their Aharonov-Bohm interactions with vortices.

Cheshire Charge Aside from their topological charges, vortices can also carry ordinary H quantum numbers.[5,12,13] These are easiest to discuss if we consider a vortex-antivortex pair, with trivial total topological charge. In this configuration, the vortex is described by a group element a, and the antivortex by a^{-1}. But under H transformations, a mixes with other representatives of the conjugacy class to which it belongs.

The action of H on the members of a class α defines a (reducible) representation which we denote as $D^{(\alpha)}$. In $D^{(\alpha)}$, each element of H is represented by a permutation of the class, according to

$$D^{(\alpha)}(h): \quad |a\rangle \rightarrow \left|hah^{-1}\right\rangle \ , \quad a \in \alpha \ . \tag{2.12}$$

This representation can be decomposed into irreducible representations of H. The physical interpretation is that the vortex pair can carry H charge. This charge is a property of the vortex-antivortex composite system and is not localized on either vortex or antivortex. It can be detected by means of the Aharonov-Bohm interaction of the composite with another vortex.

For each class α in H there is a unique state that can be constructed that is uncharged (transforms trivially under H); it is the superposition of group eigenstates

$$\frac{1}{\sqrt{n_\alpha}} \sum_{a \in \alpha} |a\rangle \tag{2.13}$$

(where n_α denotes the order of the class). Let us consider what happens to this state when an object that transforms as the irreducible representation (ν) of H passes between the two vortices.

The composite system consisting of the vortex pair and the charged projectile has a well-defined H-charge; the composite transforms according to a representation of H that could be inferred by studying the Aharonov-Bohm interaction of the composite with other vortices. (It is to ensure this that we consider a *pair* of vortices with trivial

total topological charge.) Interactions between the vortex pair and the projectile cannot change this total H-charge. From this observation, we can infer how the charge of the vortex pair changes when a charged object winds around one of the vortices.[30]

It is convenient to imagine that the vortex pair actually interacts with a particle-antiparticle pair, combined in the trivial representation of H, so that the total H-charge of particles plus vortices is trivial. Then, if one particle winds around one vortex as in Fig. 2, the charge transferred to the vortex pair is opposite to the charge transferred to the particle pair; the vortices are in a state transforming as the representation $(\mu)^*$, if the particle pair transforms as (μ).

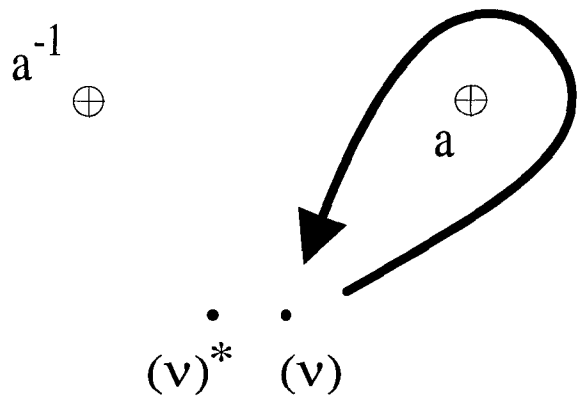

Fig. 2: A particle transforming as the representation (ν) of H interacts with a vortex pair. The particle and its antiparticle are initially prepared in an uncharged (gauge-singlet) state. After the particle winds around the vortex, the particle-antiparticle pair has acquired a non-trivial charge.

The charge-zero state of the particle-antiparticle pair has the group-theoretic structure

$$\frac{1}{\sqrt{n_\nu}} \left| e_i^{(\nu)*} \otimes e_i^{(\nu)} \right\rangle \ , \tag{2.14}$$

(summed over i) where $\{e_i^{(\nu)}\}$ denotes a basis for the vector space on which the representation (ν) acts, and n_ν is the dimension of that representation. After the particle winds around a vortex with flux $a \in H$, this state becomes

$$\frac{1}{\sqrt{n_\nu}} \left| e_i^{(\nu)*} \otimes e_j^{(\nu)} \right\rangle D_{ji}^{(\nu)}(a) \ . \tag{2.15}$$

We may regard this state as a vector in the space on which the representation $D^{(\nu)*} \otimes$

$D^{(\nu)}$ acts, expanded in terms of the natural basis $\left|e_i^{(\nu)*} \otimes e_j^{(\nu)}\right\rangle$. This vector can also be expanded in terms of the basis that block-diagonalizes the representation, as

$$\frac{1}{\sqrt{n_\nu}} D^{(\nu)}(a) = \sum_{\mu,y} c_{\mu,y} v^{(\mu),y} \ . \tag{2.16}$$

Here, $v^{(\mu),y}$ is a unit vector in the irreducible subspace acted on by the representation $(\mu), y$; the index y occurs because a given irreducible representation may occur in $D^{(\nu)} \otimes D^{(\nu)*}$ more than once.

We may interpret

$$p_\mu = \sum_y |c_{\mu,y}|^2 \tag{2.17}$$

as the probability that the final state of the particle pair transforms as $D^{(\mu)}$; correspondingly, it is the probability that, after interacting with the projectile, the vortex pair transforms as $D^{(\mu)*}$ (The $|c|$'s do not depend on how the class representative a is chosen.) Of course, p_μ can be non-vanishing only if $D^{(\mu)*}$ is one of the irreducible constituents of $D^{(\alpha)}$.

Thus we see that charge is transferred to the vortex pair, via the non-Abelian Aharonov-Bohm effect, when a charged particle winds around the vortex;[5,12,13] an initially uncharged vortex pair becomes a non-trivial superposition of charge eigenstates. This is "Cheshire charge." The charge is a global property of the vortex pair, and is not localized on the vortices.

Strings The above description of vortices is easily translated into appropriate language to describe loops of string. A multi-string configuration can be characterized by assigning elements of H to each of a set of standard paths (beginning and ending at a basepoint x_0) that link the string loops. Again, the paths that begin and end at x_0 and avoid the cores of all strings fall into homotopy classes; these classes are the elements of $\pi_1(M, x_0)$, where M now denotes R^3 with all string loops removed.

Thus, when many strings are present, there is an ambiguity in the choice of the "standard path" that links a given string, and a corresponding ambiguity in the assignment of group elements to strings. The physics associated with this ambiguity is that there is a long-range holonomy interaction between string loops.

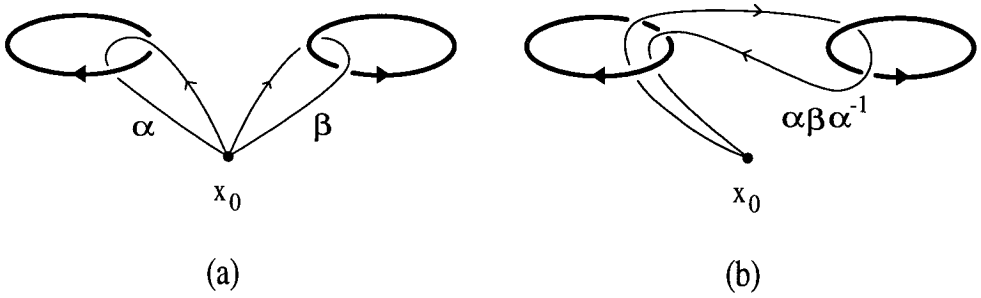

Fig. 3: (a) Standard paths α and β, based at x_0, associated with the a string and the b string. (b) Path that is dragged to the path β as the loop of b string winds through the loop of a string (in the sense defined by α).

In Fig. 3a, standard paths α and β have been introduced that wind around two string loops. Suppose that initially the homomorphism maps α to the group element a, and maps β to the group element b. Now suppose that the b string loops winds through the a loop, in the sense defined by α. During the winding, the path shown in Fig. 3b is dragged to β. Thus, the flux associated with this path before the winding is the same as the flux associated with β after the winding. This path can be expressed as $\alpha\beta\alpha^{-1}$, and so is mapped by the initial homomorphism to the group element aba^{-1}. We conclude that the final homomorphism after the winding maps β to

$$b' = aba^{-1} \ , \tag{2.18}$$

and maps α to $a' = a$. When a loop of b string winds through a loop of a string, its flux becomes conjugated—it becomes an aba^{-1} string.

For future reference, we remark that Fig. 3 has an alternative interpretation. The path shown in Fig. 3b will also be dragged to β if the a string winds around the basepoint x_0, without the b loop passing through the a loop at all. Of course, the basepoint x_0 is completely arbitrary, and selected by convention, so this process has no actual physical effect on the b loop. Yet, the effect of parallel transport around the path β will be different than before, after the a loop has lassoed the basepoint. We will return to this point in Section 4.

Furthermore, a string loop, like a pair of vortices, can carry an H charge, and can exchange charge with other charged objects by means of the Aharonov-Bohm effect.

Branching and entanglement Another property of strings will be relevant to our subsequent analysis; namely, that strings can *branch*. This feature is illustrated in

Fig. 4. There a string labeled by $c = ba$ splits into two strings with labels a and b. If the gauge group is non-Abelian, then the precise rules for composing the group elements at a branch point depend on our conventions—these (arbitrary) conventions associate each string with a path around the string that begins and ends at the basepoint, as shown in Fig. 4.

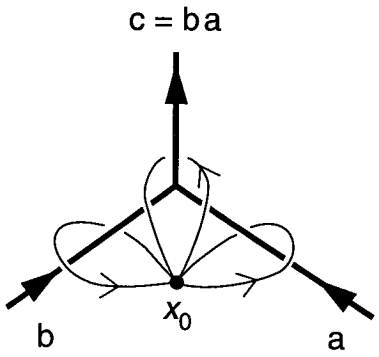

Fig. 4: The branching of a $c = ba$ string into an a string and a b string. The standard paths, based at x_0, associated with each string are indicated.

Related to the branching phenomenon is another generic feature of non-Abelian cosmic strings—strings that are labeled by non-commuting group elements become *entangled* when they cross.[31] Fig. 5 illustrates that when an a string and a b string cross, they become connected by a segment of c string. The mathematics underlying this entanglement is very similar to that underlying the holonomy interaction between two non-Abelian vortices, in two spatial dimensions.

Let us calculate the flux c carried by the string segment that connects the a string and the b string, after the crossing. First, we establish our conventions by choosing standard paths α, β, and γ that encircle the string, as in Fig. 5a. The group element c associated with γ can be determined if we demand that the bridge connecting the strings can be removed by re-crossing the strings. However, even once we have fixed all our conventions, the element c is not uniquely determined—there are two possible choices. This ambiguity arises because a crossing event involving two oriented strings has a handedness. A useful way to think about the handedness of the crossing is to imagine that the a and b strings are actually large closed loops. When the string loops cross, they become linked. But the linking number can be either $+1$ or -1.

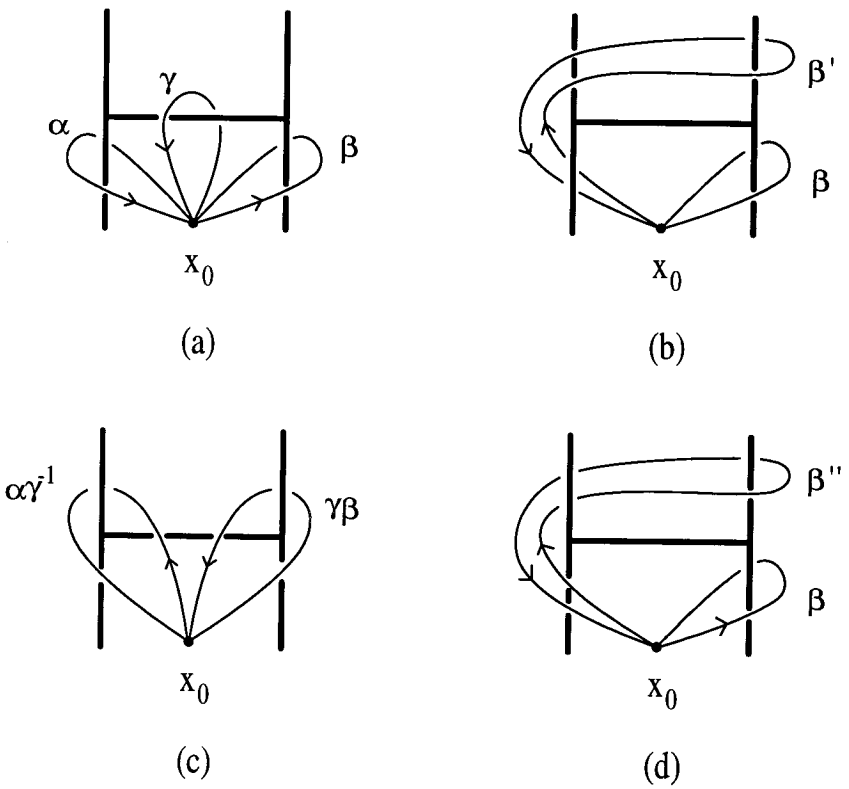

Fig. 5: (a) The standard paths α, β, and γ associated with entangled strings. These paths are mapped by the string homomorphism to group elements a, b, and c. (b) The paths β and β' are both mapped to b, if the bridge connecting the strings can be removed by crossing the strings. (c) Standard paths associated with the "upper halves" of entangled strings. (d) The paths β and β'' are both mapped to b, if the bridge can be removed by crossing the strings "the other way."

(That is, we may consider a surface bounded by one of the loops. This surface inherits an orientation from the loop, defined (say) by the right-hand rule. If the other loop pierces the surface in the same sense as the outward-pointing normal, the linking number is $+1$; otherwise, -1.) The two possible linking numbers correspond to the two possible types of crossing events.

For one of the two types of crossings (the case in which the the loops have linking number -1), the paths β and β' in Fig. 5b must be mapped to the same group element. In terms of the standard paths defined in Fig. 5a, we have $\beta' = \alpha^{-1}\gamma\beta\alpha$. (Recall our notation—in a composition of paths, the path on the right is traced first.) We therefore find $b = a^{-1}cba$, or

$$c = aba^{-1}b^{-1} \ . \tag{2.19}$$

The connection with the holonomy interaction between vortices is clarified by Fig. 5c, where standard paths $\alpha\gamma^{-1}$ and $\gamma\beta$ are shown that wind around the "upper halves" of the entangled strings. Since $ac^{-1} = (ab)a(ab)^{-1}$ and $cb = (ab)b(ab)^{-1}$, the "flux" carried by the upper half strings differs from the flux carried by the lower half strings by conjugation by ab (just as winding an a vortex counterclockwise around a b vortex causes the flux of both vortices to be conjugated by ab).

For the other type of crossing (the case in which the loops have linking number $+1$), the paths β and β'' in Fig. 5d must be mapped to the same group element. This path is $\beta'' = \alpha\beta\gamma\alpha^{-1}$, so that $b = abca^{-1}$, or

$$c = b^{-1}a^{-1}ba \ . \tag{2.20}$$

Now the upper half strings carry flux $(ab)^{-1}a(ab)$ and $(ab)^{-1}b(ab)$, which is just the change in a two-vortex state that results from winding an a vortex around a b vortex in the *clockwise* sense.

2.2. Non-Abelian walls bounded by strings

If a gauge theory respects a local discrete symmetry G, then, as we have seen, the theory admits cosmic strings that are classified by the elements of G. But if the group G is spontaneously broken to a subgroup H, then some of these strings become boundaries of domain walls.[32] Here we will consider some of the properties of domain walls bounded by strings. It is necessary to understand these properties, in order to interpret the behavior of our correlation functions, and to use them to probe the phase structure of a model.

In this section, as before, we assume that a standard choice of gauge has been made at a standard basepoint. This choice allows us to assign a definite group element to each cosmic string, and also fixes the embedding of the unbroken group H in G.

Classifying walls bounded by string If a discrete gauge group G is spontaneously broken to a subgroup H, then the cosmic strings of the theory are classified by elements of H. The elements of G that are not in H are associated not with isolated strings, but rather with strings that bound segments of domain wall.

There is not, however, a one-to-one correspondence between elements of G that are not in H and the physically distinguishable types of wall. This is because a single type of wall can be bounded by more than one type of string. As is illustrated in Fig. 6, strings labeled by a and b can bound the same wall, if $h = a^{-1}b$ is in the unbroken group H. The composite of the two strings is an object that does *not* end on a wall, which means that the wall that ends on an a string must also end on a b string (with conventions chosen as in the figure). Thus, no locally measurable property distinguishes a wall that ends on an a string from a wall that ends on a b string, if $a^{-1}b \in H$. In other words, walls are classified by the cosets $\{gH\}$ of H in G (with the trivial coset corresponding to the trivial wall, *i.e.*, no wall at all). Whether the walls are labeled by left or right cosets is a matter of convention.

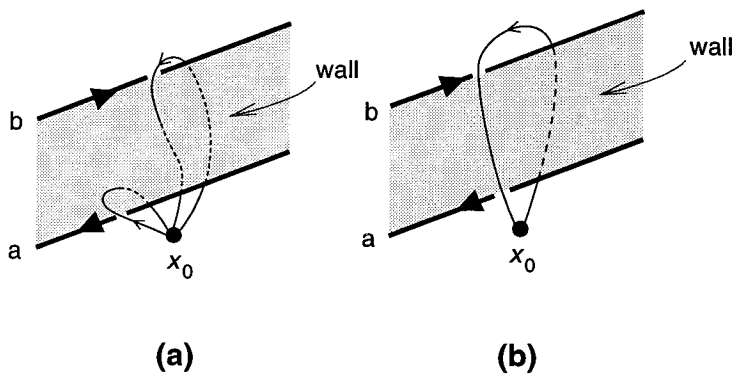

Fig. 6: (a) Standard paths that are mapped by the string homomorphism to elements $a, b \in G$ that are *not* in the unbroken group H. Each string is the boundary of a domain wall. (b) A path that is mapped to $a^{-1}b \in H$. This path does not cross the wall.

Another way to think about this classification is in terms of the discontinuity of the order parameter across the domain wall. Suppose that the Higgs field Φ that drives the breakdown from G to H transforms as the representation R of G, so that $D^{(R)}(h)\Phi = \Phi$ for $h \in H$. Across the wall labeled by a, the order parameter jumps from Φ to $D^{(R)}(a)\Phi$. This discontinuity is independent of the choice of coset representative.

To avoid confusion, we should remark that the classification that we have just described is not the usual classification of domain walls in a spontaneously broken

gauge theory. The point is that we are not classifying the domain walls that are absolutely stable, as in the standard analysis of topological defects in gauge theories. Instead, we are classifying the different walls that can end on strings.

Notice that the criterion for two strings to be boundaries of the same wall is *not* the same as the criterion for two strings to be (locally) indistinguishable objects. Strings labeled by a and b have the same structure if a and b are related by $a = hbh^{-1}$ for some $h \in H$. Two elements of G can be in the same H coset without being related in this way. Therefore, the same type of wall can have more than one type of boundary, in general.

Wall decay The walls described above, those associated with elements of G that are not in the unbroken group H, are unstable. Such a wall will decay by nucleating a loop of string by quantum tunneling; the loop then expands, consuming the wall[1].

As we have seen, it is possible for one type of wall to end on more than one type of string. Hence, the wall may be able to decay in more than one way. Furthermore, if a large segment of wall bounded by an a string decays, two different decay modes that are *locally* indistinguishable may be *globally* distinguishable. For example, it may be that b and b' are both in the same H coset as a, and also that $b' = hbh^{-1}$, for some $h \in H$. Then the wall can decay by nucleating either a b string or a b' string, and the strings look the same locally. But $a^{-1}b$ and $a^{-1}b'$ need not be the same element of H. Then the ribbon of wall that is produced by the nucleation of the loop (see Fig. 6) is different in the two cases.

If a large loop of a string is prepared, which bounds a wall, then the wall can decay in any of the available channels. The string with the lowest tension is the most likely to nucleate, but all strings that can bound the wall nucleate at some non-vanishing rate. Many holes eventually appear in the wall, and the boundaries of the holes (the strings) expand due to the tension in the wall. Eventually the holes collide.

When an a hole meets a b hole, junctions form, and a ab^{-1} string appears that bridges the hole. (See Fig. 7.) It may be that the b string has higher tension than the a string. Then the junctions will get pulled around the boundary of the hole. They annihilate, liberating a ab^{-1} string from the decaying wall system.

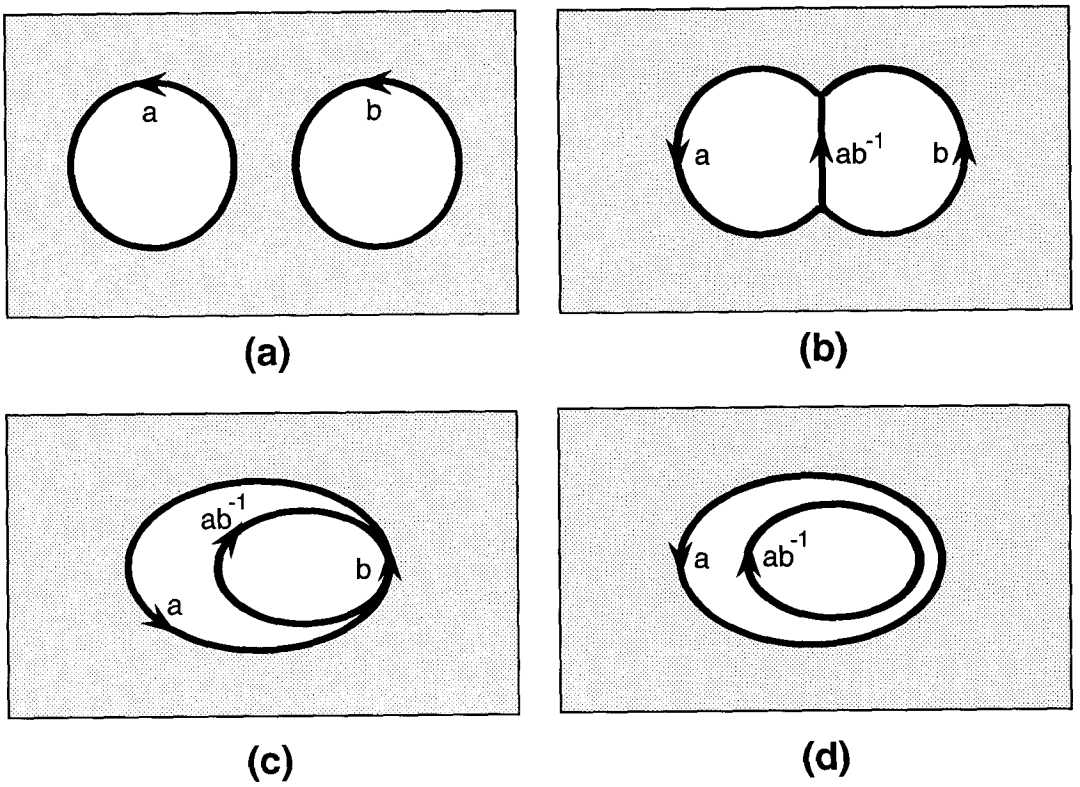

Fig. 7: (a) Two holes appear in a domain wall, due to the spontaneous nucleation of loops of a string and b string. (b) The holes meet and coalesce. A segment of ab^{-1} string appears that bridges the hole. (c) The b string has higher tension than the a string, and so pulls the string junctions toward each other. (d) A loop of ab^{-1} string breaks free from the decaying wall.

This scenario shows that, when the wall bounding a very large loop of string decays, we may regard the decay process as dominated by the strings on which the wall can end that have the smallest possible tension. This insight will help us to interpret the behavior of the order parameter $A(\Sigma, C)$. If an operator creates a world sheet of an a string, which bounds a wall, then the Aharonov-Bohm interaction with a charge whose world line winds around the string world sheet will be the same as the Aharonov-Bohm interaction of the charge with the $ab^{-1} \in H$ string, if the b string is the lightest one that can nucleate on the wall.

For completeness, we note another structure that can arise in walls bounded by strings. We have seen that a ba string can branch into an a string and a b string. It may be that all three strings are boundaries of walls. Then a possible configuration is shown in Fig. 8. Here the ba string is a vein in the wall, where the type of wall

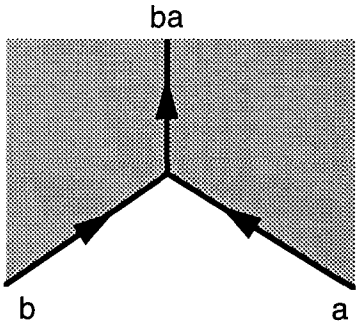

Fig. 8: A vein in a wall.

bounded by the a string joins the type of wall bounded by the b string.

2.3. The Abelian order parameter

The "Abelian" order parameter introduced in Ref. 5 can be used to probe the realization of the center of the gauge group. Here we will briefly describe how this order parameter is used to distinguish the various phases of a Z_N gauge-spin system on the lattice.

The model we consider has gauge variables U_l residing on the links (labeled by l) of a cubic lattice, and spin variables ϕ_i residing on the sites (labeled by i). Both gauge and spin variables take values in

$$Z_N \equiv \{\exp(2\pi ik/N),\ k = 0, 1, 2, \ldots, N-1\} \ . \tag{2.21}$$

The (Euclidean) action of the model is

$$S = S_{\text{gauge}} + S_{\text{spin}} \ ,$$

where

$$S_{\text{gauge}} = -\beta \sum_P (U_P + c.c.) \ , \tag{2.22}$$

and

$$S_{\text{spin}} = -\sum_{m=1}^{N-1} \gamma_m \sum_l \left([(\phi^{-1} U \phi)_l]^m + c.c. \right) \ . \tag{2.23}$$

Here $U_P = \prod_{l \in P} U_l$ associates with each elementary plaquette (labeled by P) the (ordered) product of the four U_l's associated with the (oriented) links of the plaquette,

and $(\phi^{-1}U\phi)_{ij} = \phi_i^{-1}U_{ij}\phi_j$, for each pair ij of nearest-neighbor sites. In eq. (2.23), we have introduced an independent spin coupling constant corresponding to each non-trivial irreducible representation of Z_N.

Now we must consider how the operator $F(\Sigma)$ is to be defined in this lattice model.* Recall that inserting $F(\Sigma)$ into a Green function is supposed to be equivalent to introducing a classical cosmic string source on the world sheet Σ. On the lattice (in 3+1 dimensions), we consider Σ to be a closed surface made up of plaquettes of the *dual* lattice. There is a set Σ^* of plaquettes of the original lattice that are dual to the plaquettes of Σ. (See Fig. 9.) The operator $F_n(\Sigma)$ may be expressed as

$$F_n(\Sigma) = \prod_{P\in\Sigma^*} \exp\left(\beta(e^{2\pi in/N}U_P - U_P + c.c.)\right) . \tag{2.24}$$

In other words, to evaluate the path integral for a Green function with an insertion of $F_n(\Sigma)$, we modify the plaquette action on the plaquettes that are dual to Σ, according to

$$U_P \to e^{2\pi in/N}U_P, \quad P \in \Sigma^* . \tag{2.25}$$

This transformation of S_{gauge} is equivalent to introducing n units of Z_N magnetic flux on each of the plaquettes in Σ^*.

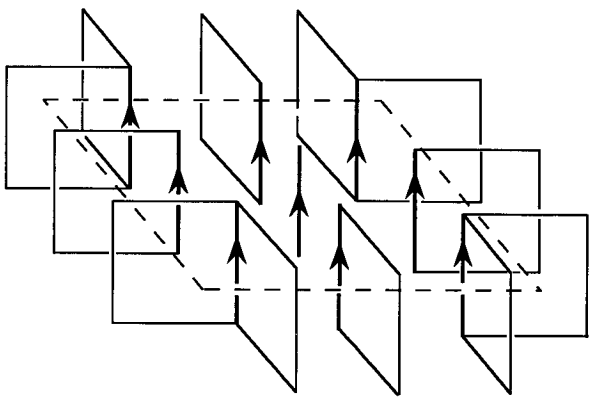

Fig. 9: The closed curve Σ in 2+1 dimensions. The dashed line is Σ, comprised of links of the dual lattice. The plaquettes shown are those in Σ^*, which are dual to the links of Σ. The links marked by arrows are those in Ω^*; they are dual to the plaquettes in a surface Ω that is bounded by Σ. (We may also interpret the dashed line as a slice through the closed surface Σ, in 3+1 dimensions.)

* The analogous construction of the 't Hooft loop operator on the lattice was first discussed in Ref. 33 and Ref. 34.

When the surface Σ is large, the vacuum expectation value of F_n behaves like

$$\langle F_n(\Sigma)\rangle \sim \exp\left(-\kappa_n^{(\mathrm{ren})}\mathrm{Area}(\Sigma)\right) \ . \tag{2.26}$$

We may interpret $\kappa_n^{(\mathrm{ren})}$ as the renormalization of the tension of the string source, due to quantum fluctuations of the charged matter fields. Eq. (2.26) also has an alternative interpretation. We may think of Σ as a surface lying in a time slice, rather than as the world sheet of a string propagating through spacetime. Then $F_n(\Sigma)$ is the operator

$$F_n(\Sigma) = \exp\left(\frac{2\pi i n}{N} Q_\Sigma\right) \ , \tag{2.27}$$

where Q_Σ is the Z_N charge contained inside the surface Σ. Virtual pairs of charged particles near Σ cause the charge Q_Σ to fluctuate. If the theory has a mass gap, then the charge fluctuations near two elements of Σ become very weakly correlated when the elements are distantly separated. Thus, charge fluctuations cause the characteristic "area-law" falloff of $\langle F_n(\Sigma)\rangle$ in eq. (2.26).

Consider the case $\beta >> 1$ and $\gamma_m << 1$ (for all m). In this case, the gauge variables are highly ordered, and so we expect that Z_N charge is not *confined.* Furthermore, the spins are disordered, so there is no Higgs condensate to *screen* Z_N charge either. Thus, free Z_N charges should exist. We can check whether this expectation is correct by using perturbative expansions to evaluate $\langle A(\Sigma, C)\rangle$.

Because the gauge coupling is weak, a frustrated plaquette (one with $U_P \neq 1$) is very costly. An insertion of $F_n(\Sigma)$ tends to frustrate the plaquettes in Σ^*. These frustrations can be avoided, though, if the gauge variables assume a suitable configuration. Choose a three-dimensional hypersurface Ω whose boundary is Σ. This hypersurface consists of a set of cubes of the dual lattice. Dual to these cubes is a set Ω^* of links of the original lattice. (See Fig. 9.) By choosing

$$\begin{aligned} U_l &= e^{2\pi i n/N}, \quad l \in \Omega^* \ , \\ U_l &= \ 1 \qquad , \quad l \notin \Omega^* \ , \end{aligned} \tag{2.28}$$

we can avoid frustrating any plaquettes. (This is the lattice analog of a "singular gauge transformation" that removes the string "singularity" on Σ.) With the links

in the configuration eq. (2.28), the fundamental representation Wilson loop operator

$$W(C) = \prod_{l \in C} U_l \tag{2.29}$$

acquires the value $\exp(2\pi i n k/N)$, where k is the linking number of C and Σ. (See Fig. 10.)

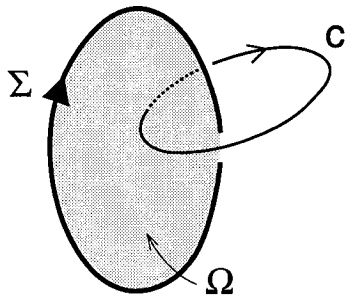

Fig. 10: A slice through a closed surface Σ that links once with the loop C. C intersects the hypersurface Ω once.

Of course, this choice of link variables changes the nearest neighbor interactions for each pair of spins that is joined by a link contained in Ω^*. But the spins are strongly coupled and highly disordered, so that they are nearly indifferent to this change; it is much more costly to frustrate a plaquette than to frustrate a link. Therefore, the expectation value of the ABOP

$$A_n^{(\nu)}(\Sigma, C) = \frac{F_n(\Sigma)\ [W(C)]^\nu}{\langle F_n(\Sigma)\rangle\ \langle [W(C)]^\nu\rangle} \tag{2.30}$$

is dominated by small fluctuations about the "background" eq. (2.28), and we conclude that $\left\langle A_1^{(1)}(\Sigma, C)\right\rangle$ satisfies eq. (2.2). Thus, there is an infinite range Aharonov-Bohm interaction, and free Z_N charges exist. (Further details of this analysis may be found in Ref. 5.)

If the gauge coupling is strong ($\beta << 1$), then there are no free Z_N charges. The Wilson loop introduces a Z_N charge as a classical source, but confinement causes a pair of Z_N charges to be produced, so that the charge is dynamically shielded. We expect that eq. (2.3) is satisfied, and this can be verified in the small-β expansion[5].

It is also interesting to consider the case in which β is large and some of the spin couplings γ_m are also large. Then the matter field is ordered and the matter "condensate" screens the charge.

For example, suppose that

$$\begin{aligned} \gamma_{m'} &>> 1 \ , \quad m' = m \ , \\ \gamma_{m'} &<< 1 \ , \quad m' \neq m \ . \end{aligned} \tag{2.31}$$

In effect, then, the operator ϕ^m condenses, and Z_N is broken down to the kernel of the representation (m); this is Z_M, where M is the greatest common factor of N and m. We anticipate, therefore, that the operator $F_n(\Sigma)$ introduces a stable cosmic string world sheet provided that $nm \equiv 0 \ (mod N)$ (so that the flux carried by the string is in the unbroken group Z_M). Otherwise, the string introduced by $F_n(\Sigma)$ is the boundary of a domain wall. This wall is unstable and decays by nucleating a loop of string.

If we assume that the flux of the string created by $F_n(\Sigma)$ combines with the flux of the nucleated string to give a trivial total flux, then we anticipate that the ABOP behaves as

$$\begin{aligned} \lim \langle A_n^{(\nu)}(\Sigma, C)\rangle &= \exp\left(\frac{2\pi i n \nu}{N} k(\Sigma, C)\right) \ , \quad mn \equiv 0 \ (\text{mod } N) \ , \\ \lim \langle A_n(\Sigma, C)\rangle &= \qquad 1 \ , \qquad\qquad\qquad mn \not\equiv 0 \ (\text{mod } N) \ . \end{aligned} \tag{2.32}$$

From this behavior, we could easily infer that the unbroken symmetry is Z_M; a string has a non-trivial Aharonov-Bohm interaction with a charge if and only if the flux carried by the string is in Z_M.

However, as we emphasized in subsection of wall decay, even if the string introduced by F_n is the boundary of a domain wall, and the wall decays by nucleating a string loop, it need *not* be the case that the nucleated string combined with the classical string source has trivial flux. Thus, eq. (2.32) does *not*, in general, correctly describe the behavior of the ABOP in the limit eq. (2.31). To see what actually happens, let us analyze the consequences of eq. (2.31) using perturbative expansions.

For $\gamma_m >> 1$, spins with non-trivial Z_M charge are highly ordered, and frustrating these spins is very costly. Now there is a competition between the reluctance of the

system to frustrate a plaquette (when $\beta >> 1$) and its reluctance to frustrate a Z_M spin. If the operator $F_n(\Sigma)$ is inserted, then, as we have seen, frustrated plaquettes can be avoided if the links of Ω^* are excited. But, if $\exp(2\pi in/N) \notin Z_M$, we can excite the U_l's on these links only at the cost of frustrating the spins there. The number of links in Ω^* increases like the volume enclosed by Σ, and the number of plaquettes in Σ^* increases only as the area of Σ. Thus, for Σ sufficiently large, frustrated plaquettes are favored over frustrated links. This is just a realization on the lattice of the decay of a domain wall by nucleation of a loop of string, where the frustrated spins comprise the wall, and the frustrated plaquettes comprise the nucleated string.

But as we discussed before, it is sometimes possible for a domain wall to decay in more than one way. To illustrate the possibilities, we will consider two different special cases.

$\underline{Z_4 \to Z_2}$ Consider, for example, a Z_4 model. For $\beta >> 1$ and $\gamma_2 >> 1$, the Z_4 symmetry will break to Z_2. One finds, indeed, that $\left\langle A_2^{(\nu)}(\Sigma, C)\right\rangle$ behaves as in eq. (2.32).

Understanding the behavior of $A_1^{(\nu)}$ and $A_3^{(\nu)}$ involves a subtlety. The $n = 3$ string is the anti-string of the $n = 1$ string; therefore, they both have the same tension. Furthermore, $\exp(2\pi i/4)$ and $\exp(-2\pi i/4)$ belong to the same coset of Z_2 in Z_4, so the $n = 1$ and $n = 3$ strings are both boundaries of the same domain wall. When the operator $F_1(\Sigma)$ is inserted, the resulting domain wall can decay in two different ways. One way, the composite string that is created has trivial Z_2 flux; the other way, it has non-trivial Z_2 flux.

Correspondingly, when $F_1(\Sigma)$ is inserted, we may choose either $U_l = 1$ or $U_l = -1$ on the links of Ω^*; the weak spin coupling in the action (the γ_2 term) depends only on U_l^2, and so is not frustrated either way. But we also need to consider the dependence on the strong couplings γ_1 and γ_3. Expanding in these small parameters, one finds that the effective tension of the composite string is renormalized by spin fluctuations. The renormalization raises the tension of the composite string with non-trivial Z_2 flux relative to that with trivial flux. Thus, the configuration such that the composite string has trivial Z_2 flux really does dominate when $F_1(\Sigma)$ (or $F_3(\Sigma)$) is inserted. Therefore, $A_1^{(\nu)}$ and $A_3^{(\nu)}$ do behave as in eq. (2.32).

$\underline{Z_6 \to Z_3}$ Now consider a Z_6 model. For $\beta >> 1$ and $\gamma_3 >> 1$, the Z_6 symmetry will break to Z_3. $A_2^{(\nu)}$ and $A_4^{(\nu)}$ behave as in eq. (2.32); so do $A_1^{(\nu)}$ and $A_5^{(\nu)}$.

Now consider $A_3^{(\nu)}$. The $n = 3$ string that is introduced by F_3 is the boundary of a domain wall. But this wall can terminate on an $n = 1$ or $n = 5$ string, as well as on an $n = 3$ string. Furthermore, the $n = 1$ (or $n = 5$) string has *lower* tension than the $n = 3$ string, so that nucleation of this string is favored.

In other words, the configurations that dominate, when $F_3(\Sigma)$ is inserted, have $U_l = \exp(2\pi i/3)$ (or $U_l = \exp(-2\pi i/3)$) for $l \in \Omega^*$. By exciting the links on Ω^*, we reduce the degree of frustration of the plaquettes of Σ. And we do so at negligible cost, because the $m = 3$ term in S_{spin} cannot distinguish $U_l = \exp(\pm 2\pi i/3)$ from $U_l = 1$.

Thus, the nucleated string, combined with the classical string source, has non-trivial flux. The combined flux can be either $n = 2$ or $n = 4$, and both occur with equal probability. And, therefore, even though F_3 inserts a string that decays, A_3 shows non-trivial behavior; we have

$$\lim \left\langle A_3^{(\nu)}(\Sigma, C) \right\rangle = \frac{1}{2}\left(e^{2\pi i\nu/3} + e^{-2\pi i\nu/3}\right) = \cos(2\pi\nu/3) \tag{2.33}$$

(if Σ and C have linking number $k = 1$).

Thus, for dynamical reasons, eq. (2.32) is not satisfied. Nevertheless, the behavior of $A_n^{(\nu)}$ has an unambiguous interpretation. For example, the ν-dependence of $\left\langle A_3^{(\nu)} \right\rangle$ shows that the *effective* string introduced by F_3 has flux that takes values in the Z_3 subgroup of Z_6. The only possible interpretation is that the unbroken subgroup is Z_3.

Indeed, the behavior of $\left\langle A_n^{(\nu)} \right\rangle$ always contains sufficient information to unambiguously identify the unbroken subgroup of an Abelian discrete local symmetry group.

Other order parameters have been suggested that probe the realization of the center of the gauge group.[17,20,18] We will comment on the efficacy of these in section 8.

2.4. Non-Abelian strings on the lattice

We have seen in the previous section how the Aharonov-Bohm effect can be used to probe the phase structure of an Abelian gauge theory. We now want to extend this procedure, so that it can be used in non-Abelian theories. The basic strategy will be the same as before—we will introduce strings and charges as classical sources, and investigate the dynamical response of the theory to these sources. But the implementation of this strategy is more delicate in the non-Abelian case. The main stumbling block is the problem of introducing non-Abelian cosmic strings in a lattice gauge theory, which we address in this section.

We will also construct operators that create or annihilate *dynamical* string loops; correlation functions of these can be used to study the dynamical properties of strings.

String calibration We consider a theory with (discrete) gauge group G. We are interested in how the local G symmetry is realized. Specifically, we wish to identify the subgroup H of G that admits free charges. (G quantum numbers may be confined, or may be screened by a Higgs condensate.)

We investigate the realization of G by assembling a laboratory that is equipped with cosmic string loops. As described in Ref. 21 we can calibrate the string loops with a beam of particles that transform as some faithful (not necessarily irreducible) representation (R) of G. We choose an arbitrary basepoint x_0, and a basis for the representation (R) at that point. We direct the beam from the basepoint to a beam splitter, allow the two beams to pass on either side of the string, and then recombine the beams and study the resultant interference pattern.

If the string is in a "group eigenstate" with flux (as defined in eq. (2.7)) $a \in G$, and $|u\rangle$ is the wave-function in internal-symmetry space of a particle at the basepoint, then, when the particle is transported around a closed path that begins and ends at x_0, the wave-function is modified according to

$$|u\rangle \to D^{(R)}(a)\ |u\rangle\ . \tag{2.34}$$

By observing the interference pattern, we can measure

$$\langle u|\, D^{(R)}(a)\, |u\rangle\ . \tag{2.35}$$

By varying $|u\rangle$, we can then determine all matrix elements of $D^{(R)}(a)$, and hence a

itself. Given our choice of basepoint, and a choice of a basis for $D^{(R)}$, this procedure allows us to associate a well-defined group element with each string.

In the case where free G charges may not exist, we must be careful that the separation between the two beams remains small compared to any confinement distance scale or Higgs screening length. At the same time, of course, the separation must be large compared to the thickness of the string core; we assume that the core is small compared to the characteristic distance scale of the dynamics that we wish to study.

Once we have calibrated the strings by measuring their Aharonov-Bohm interactions with nearby charges, we probe the dynamics of the theory by measuring the Aharonov-Bohm interactions of strings with distant charges. In this way, we hope to learn what quantum numbers are confined or screened, and to infer the "unbroken" subgroup H.

(As we have noted, each string loop associated with an element of G that is not in H will become the boundary of a domain wall. Thus it might seem that a good way to identify H is to observe which strings are attached to walls. Indeed, at sufficiently weak coupling, this is a sensible procedure, because the walls are very long lived. But at intermediate coupling this procedure may fail, because the walls decay rapidly by nucleating string loops. The time scale for the decay may be comparable to the time required to assemble and calibrate a string loop.)

Some problems with the calibration procedure should be pointed out. The first is that a string loop associated with a definite group element is not an eigenstate of the Hamiltonian of the theory. An a string and a b string will mix with each other if a and b are in the same H conjugacy class,[21] and the energy eigenstates will be linear combinations of group eigenstates that transform as definite representations of the unbroken group H. (We will have more to say about this mixing in Section 6.) But the time scale for this mixing increases exponentially with the size of the string loop. For our purposes, it will usually be legitimate to ignore the mixing and regard the strings as group eigenstates.

But there is another more serious problem. While the quantum fluctuations that change the identity of a string loop are very rare (when the loop is large), there are other, much less rare, quantum fluctuations that can change the Aharonov-Bohm

phase that is detected. Suppose, for example, that a virtual b loop nucleates, lassoes the basepoint x_0 and then reannihilates. Naturally, this process has no physical effect on an a loop that is far from the basepoint. Yet it changes the Aharonov-Bohm phase acquired by a particle in representation (ν) that winds around the string, beginning and ending at x_0; we saw in Section 1 that the phase becomes $D^{(\nu)}(bab^{-1})$ rather than $D^{(\nu)}(a)$. (This process is depicted, in 2+1 dimensions, in Fig. 11.) Such fluctuations are suppressed at sufficiently weak coupling, but they are present, at some level, for any finite coupling. They result in an ambiguity in the Aharonov-Bohm phase associated with a string, even if we fix the basepoint and a basis for the faithful representation (ν).

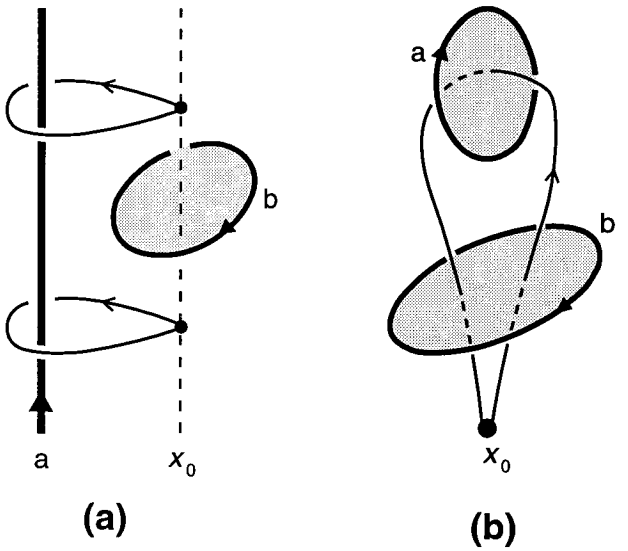

Fig. 11: (a) Spacetime diagram (in 2+1 dimensions) showing a virtual b vortex-antivortex pair that nucleates, winds around the basepoint x_0, and reannihilates. If the path based at x_0 at time t_1 is assigned the flux a, then the path at time t_2 is assigned flux bab^{-1}. (b) The same process, but with x_0 now taken to be a fixed point in Euclidean spacetime. The flux of the classical a vortex is measured to be bab^{-1}, due to the effect of the virtual b vortex pair. Shaded areas are surfaces bounded by the vortex world lines.

To avoid this ambiguity, we are forced to take a trace of the representation; the character $\chi^{(\nu)}(a)$ is unaffected by these fluctuations. Thus, the existence of free G charges (and of an unscreened Aharonov-Bohm interaction) can be probed by the

Wilson loop operator eq. (2.11). (And since $W^{(\nu)}(C)$ is gauge invariant, there is no need for the loop C to contain the basepoint x_0.) Let $F_a(\Sigma, x_0)$ denote the operator that inserts a string world sheet on Σ; the string is associated with $a \in G$, relative to the basepoint x_0. (We consider below how this operator is constructed.) Then we may define the ABOP

$$A_a^{(\nu)}(\Sigma, x_0; C) = \frac{F_a(\Sigma, x_0) W^{(\nu)}(C)}{\langle F_a(\Sigma, x_0)\rangle \langle W^{(\nu)}(C)\rangle} \ . \tag{2.36}$$

In a phase with free G charges, we expect that[14]

$$\lim \langle A_a^{(\nu)}(\Sigma, x_0; C)\rangle = \frac{1}{n_\nu} \chi^{(\nu)}(a^{k(\Sigma,C)}) \ , \tag{2.37}$$

where k denotes the linking number. (Even this statement requires a qualification, for the loop C must not be permitted to come close to retracing itself on successive passages around Σ.)

We will consider in Section 8 how $A_a^{(\nu)}(\Sigma, x_0; C)$ is expected to behave for other realizations of the local G symmetry.

Inserting string world sheets Now we will consider in more detail the problem of constructing an operator that inserts non-Abelian strings on specified world sheets. The construction will be guided by the discussion in Section 1 of the general formalism, and by the discussion above of the calibration procedure.

Given a set of disjoint closed two-dimensional surfaces $\Sigma_1, \Sigma_2 \ldots$, this operator is to introduce string world sheets on these surfaces, where the strings are associated with the group elements $a_1, a_2, \ldots$. As we have emphasized, the strings must be referred to a common basepoint x_0, and each associated group element depends on the choice of a standard path that begins and ends at x_0 and winds around the string. The group elements $a_1, a_2, \ldots$ are defined up to one overall conjugation, which corresponds to a gauge transformation at the basepoint x_0.

Furthermore, the definition of our operator must be insensitive to the possible confinement or screening of G quantum numbers; it must specify the short-distance structure of the string, and leave it up to the dynamics of the theory whether the

string can be detected at long range. It must realize the calibration procedure described above, in which the short-range Aharonov-Bohm interactions of the string are determined.

We wish to define an operator $F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)$ so that an insertion of $F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)$ in a Green function introduces strings on world sheets $\Sigma_1,\Sigma_2,\ldots,\Sigma_n$. We will first describe the operator in heuristic terms, and then give a more precise description in the context of lattice gauge theory. Loosely speaking, an insertion of $F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)$ in a Green function imposes a restriction on the gauge field configurations that are included in the path integral. Make a choice of a standard path that begins at x_0, winds around the surface Σ_1, and returns to x_0. Now consider a path P_1 homotopic to the standard path that runs from x_0 to a point on Σ_1, traverses an infinitesimal closed loop that encloses Σ_1, and then retraces itself, returning to x_0. When $F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)$ is inserted, the gauge field configuration is restricted to satisfy

$$P\ \exp\left(i\oint\limits_{P_1,x_0} A\cdot dx\right)=a_1\ , \tag{2.38}$$

for any such P_1. Similar restrictions apply for paths $P_2,P_3,\ldots,P_n$ that wind around $\Sigma_2,\Sigma_3,\ldots,\Sigma_n$, as in Fig. 12. These restrictions on the path integral define the operator $F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)$.

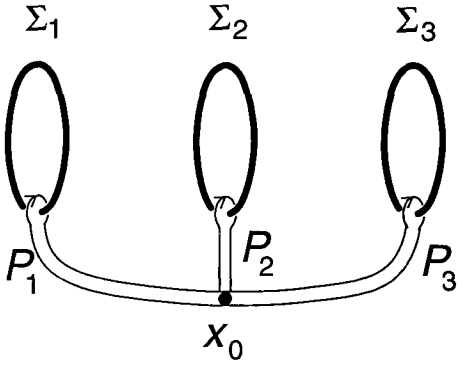

Fig. 12: Paths P_1, P_2, P_3 used in the construction of the operator $F_{a_1,a_2,a_3}(\Sigma_1,\Sigma_2,\Sigma_3;x_0)$.

The operator so constructed is not gauge invariant. Of course, if it is inserted in a Green function with gauge-invariant operators, its gauge-invariant part will be projected out. Alternatively we may obtain an explicitly gauge-invariant operator by averaging over gauge transformations at the basepoint x_0, obtaining

$$\frac{1}{n_G}\sum_{g\in G} F_{ga_1g^{-1},ga_2g^{-1},\ldots,ga_ng^{-1}}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)\ , \tag{2.39}$$

where n_G is the order of the group.

We may now consider the operator

$$A^{(\nu)}_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0;C)=\frac{F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)W^{(\nu)}(C)}{\langle F_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0)\rangle\langle W^{(\nu)}(C)\rangle}\ . \tag{2.40}$$

In a phase with free G charge, if the loop C is homotopic to $P_iP_jP_k\ldots$, we have

$$\lim\ \langle A^{(\nu)}_{a_1,a_2,\ldots,a_n}(\Sigma_1,\Sigma_2,\ldots,\Sigma_n,x_0;C)\rangle=\frac{1}{n_\nu}\chi^{(\nu)}(a_ia_ja_k\ldots)\ . \tag{2.41}$$

Thus, when a charge passes around several string loops in succession, the Aharonov-Bohm phases acquired in each successive passage are combined *coherently.* The coherence is maintained because we have defined the various loops in reference to the same basepoint x_0. If different basepoints had been chosen instead, then the group elements associated with the various string loops would have been averaged over conjugacy classes independently. Since

$$\frac{1}{n_G}\sum_{g\in G} D^{(\nu)}(gag^{-1})=\frac{1}{n_\nu}\chi^{(\nu)}(a)\ \mathbf{1} \tag{2.42}$$

(which follows from Schur's lemma), we find, for example,

$$\lim\frac{\langle F_{a_1}(\Sigma_1,x_0)F_{a_2}(\Sigma_2,y_0)W^{(\nu)}(C)\rangle}{\langle F_{a_1}(\Sigma_1,x_0)F_{a_2}(\Sigma_2,y_0)\rangle\langle W^{(\nu)}(C)\rangle}=\frac{1}{n_\nu}\chi^{(\nu)}(a_1)\ \frac{1}{n_\nu}\chi^{(\nu)}(a_2)\ , \tag{2.43}$$

if the loop C winds around world sheets Σ_1 and Σ_2 in succession.

If $F_{a_1}(\Sigma_1,x_0)$ and $F_{a_2}(\Sigma_2,y_0)$ have distinct basepoints ($x_0\neq y_0$), then each by itself introduces a gauge-singlet object. Inserting both operators combines two string loops as (trivial) charge eigenstates, rather than as group eigenstates. This is the

reason for the lack of coherence in the Aharonov-Bohm interaction characterized by eq. (2.43). To combine two strings that are not both gauge singlets, we must include in our operator some non-local construct that bridges the gap between the string loops, just as a string of electric flux must be included in a gauge-invariant operator that creates a widely separated quark-antiquark pair. In the case of two (or more) string loops, this non-local connection between the loops is provided by referring the loops to a common basepoint.

It is important, actually, that the Aharonov-Bohm interactions of an additional loop combine incoherently with the Aharonov-Bohm interactions of existing loops, when the group element associated with the additional loop is averaged over a conjugacy class. It is this incoherence property that ensures that the effects of *virtual* string loops do not spoil the "factorization up to a phase" represented by eq. (2.41).

We will now describe the construction of $F_a(\Sigma, x_0)$ in more detail, by specifying how the construction is carried out in a lattice theory with discrete gauge group G. As discussed in Section 3, Σ is to be regarded as a closed surface consisting of plaquettes of the dual lattice; these plaquettes are dual to a set Σ^* of plaquettes of the original lattice. For each plaquette $P \in \Sigma^*$, we choose a path l_P on the lattice that connects the basepoint x_0 to one of the corners of P. These paths are chosen so that each closed loop $l_P P l_P^{-1}$ is "homotopic" to the standard loop that links Σ. The various paths may be chosen arbitrarily, except that the union of all the paths should not contain any closed loops.

The effect of the operator $F_a(\Sigma, x_0)$ is to modify the plaquette action on each plaquette in Σ^*. Suppose, for example, that the plaquette action is

$$S^{(R)}_{\text{gauge},P} = -\beta\chi^{(R)}(U_P) + c.c. \tag{2.44}$$

(where R is a representation of G that must be specified to define the theory). Then an insertion of $F_a(\Sigma, x_0)$ modifies the action according to

$$S^{(R)}_{\text{gauge},P} \longrightarrow -\beta\chi^{(R)}\left(V_{l_P} a V_{l_P}^{-1} U_P\right) + c.c., \quad P \in \Sigma^*, \tag{2.45}$$

where

$$V_{l_P} = \prod_{l \in l_P} U_l \ . \tag{2.46}$$

Alternatively, we may write

$$F_a(\Sigma, x_0) = \prod_{P \in \Sigma^*} \exp\left(\beta\chi^{(R)}\left(V_{l_P} a V_{l_P}^{-1} U_P\right) - \beta\chi^{(R)}(U_P) + c.c.\right) . \qquad (2.47)$$

The operator $F_{a_1,a_2,\ldots,a_n}(\Sigma_1, \Sigma_2, \ldots, \Sigma_n, x_0)$ that inserts many string loops is constructed by a straightforward generalization of this procedure.

As constructed, $F_a(\Sigma, x_0)$ is not gauge invariant. When inserted in a Green function with gauge-invariant operators, though, it has the same effect as the explicitly gauge-invariant operator in which a is averaged over a conjugacy class, as in eq. (2.39).

It is also instructive to consider the correlator of $F_a(\Sigma, x_0)$ with the operator

$$U^{(\nu)}(C, x_0) = D^{(\nu)}\left(\prod_{l \in C} U_l\right) , \qquad (2.48)$$

where the product is taken over a closed set of links that begins and ends at x_0. The trace of $U^{(\nu)}(C, x_0)$ is (n_ν times) the Wilson loop operator $W^{(\nu)}(C)$. But $U^{(\nu)}(C, x_0)$ itself, like $F_a(\Sigma, x_0)$, is not invariant under a gauge transformation that acts non-trivially at the basepoint x_0.

In a phase with free G charges, and in the leading order of weak coupling perturbation theory, one finds that*

$$\lim \frac{\langle F_a(\Sigma, x_0) U^{(\nu)}(C, x_0)\rangle}{\langle F_a(\Sigma, x_0)\rangle \langle \mathrm{tr}\ U^{(\nu)}(C, x_0)\rangle} = \frac{1}{n_\nu} D^{(\nu)}\left(a^{k(\Sigma,C)}\right) . \qquad (2.49)$$

This equation merely states that, once a loop of string has been calibrated, the same Aharonov-Bohm phase can be recovered again if another interference experiment is subsequently performed. But we have already emphasized that quantum fluctuations (such as the virtual string depicted in Fig. 11) can spoil this result. Indeed, when higher orders in the weak coupling expansion are included, it is seen that the correlator of $F_a(\Sigma, x_0)$ and $U^{(\nu)}(C, x_0)$ fails to "factorize" as in eq. (2.49), even when Σ and C are far apart. As stated before, we must consider the correlator of $F_a(\Sigma, x_0)$ with the gauge-invariant operator $W^{(\nu)}(C)$, in order to extract an Aharonov-Bohm "phase" that depends only on the topological linking of Σ and C (in the limit of infinite separation).

---

* This calculation is described in more detail in Section 5 and in the Appendix.

To conclude this section, we must ask how the operator $F_a(\Sigma, x_0)$ depends on the basepoint x_0 and on the choice of the paths $\{l_P\}$. (We consider correlation functions of F_a with gauge-invariant operators, so we regard F_a as gauge-invariant, with a averaged over a conjugacy class.) We note that, since the union $\cup_P\, l_P$ of all the paths contains no closed loops, we can choose a gauge with $U_l = 1$ for all $l \in l_P$. In this gauge, the group element a is inserted directly on the plaquettes of Σ^*. Thus, it is clear that the correlator of $F_a(\Sigma, x_0)$ with any local gauge-invariant operator is independent of the choice of basepoint and paths.

The correlator of $F_{a_1,a_2,\ldots,a_n}(\Sigma_1, \Sigma_2, \ldots, \Sigma_n, x_0)$ with non-local gauge-invariant operators (like Wilson loop operators) depends on the choice of path only to the extent that we have already noted. That is, it depends on the choice of the "standard paths" that enter the calibration of the string loops. We may change the paths from the basepoint x_0 to the plaquettes of Σ_2^*, by winding these paths around Σ_1, as in Fig. 13. In effect, this change alters the group element assigned to the string world sheet on Σ_2; a_2 becomes replaced by $a_1 a_2 a_1^{-1}$. As we noted in Section 1, this ambiguity in $F_{a_1,a_2,\ldots,a_n}(\Sigma_1, \Sigma_2, \ldots, \Sigma_n, x_0)$ is the origin of the holonomy interaction between string loops (or vortices).[25,24]

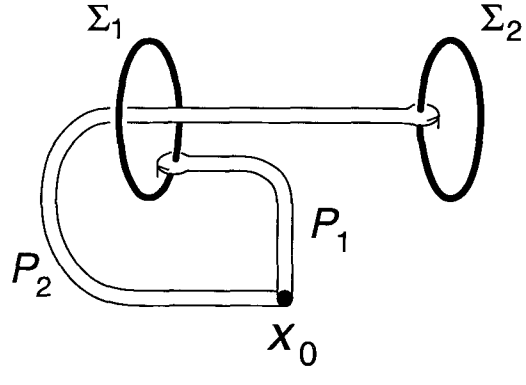

Fig. 13: A different choice for the path P_2 from the basepoint x_0 to the world sheet Σ_2.

Inserting string loops An insertion of the operator $F_a(\Sigma, x_0)$ introduces a string on the closed world sheet Σ. The string may be regarded as an infinitely heavy classical source. Thus $F_a(\Sigma, x_0)$ is closely analogous to the Wilson loop operator $W^{(\nu)}(C)$; an

insertion of $W^{(\nu)}(C)$ introduces an infinitely heavy classical source (in representation (ν)) on the closed world line C.

But we will also find use at times for an operator that creates (or annihilates) a *dynamical* cosmic string. Such an operator can be obtained by a simple modification of the construction described above—namely, the surface Σ is chosen to be, rather than a closed surface, a surface with non-trivial boundary C. (See Fig. 14.) Carrying out the same procedure as before for such a surface, we arrive at an operator $B_a(C, \Sigma, x_0)$ that creates a string (or annihilates an anti-string) on the loop C. More generally, an operator

$$B_{a_1,a_2,\ldots,a_n}(C_1, \Sigma_1, C_2, \Sigma_2, \ldots, C_n, \Sigma_n, x_0) \tag{2.50}$$

creates strings on the loops $C_1, C_2, \ldots, C_n$, with all strings referred to a common basepoint x_0.

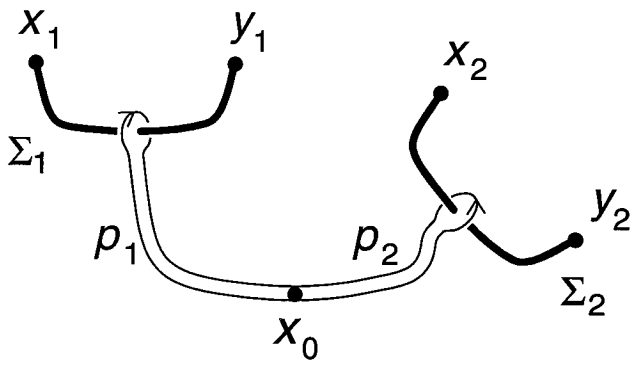

Fig. 14: The paths P_1 and P_2, based at x_0, and the open curves Σ_1 and Σ_2 used in the definition of the 't Hooft operator $B_{a,b}(x_1, y_1, \Sigma_1, x_2, y_2, \Sigma_2, x_0)$, in 2+1 dimensions.

This construction generalizes a construction devised by 't Hooft,[16] and we will refer to B as the "'t Hooft loop" operator. However, 't Hooft considered the situation in which the strings have no Aharonov-Bohm interactions with other fields. In that case, the surface Σ is an invisible gauge artifact, and $B(C)$ depends on C alone. We are interested in strings that have Aharonov-Bohm interactions, and in that case Σ is not invisible.[5]

To better understand why the 't Hooft operator $B_a(C, \Sigma, x_0)$ must depend on the surface Σ as well as on the loop C, it is helpful to think about a theory defined in 2+1 spacetime dimensions. In that case, the operator $B_a(x, y, \Sigma, x_0)$ creates a vortex at x and an anti-vortex at y; Σ is a path connecting x and y. But no gauge-invariant local operator exists that creates an isolated vortex at x. Because a vortex can be detected at infinite range via the Aharonov-Bohm effect, there is a vortex superselection rule. Hence, the operator that creates a vortex cannot be local; it has a semi-infinite string that can be seen by the fields of the theory. (Similarly, there is an electric charge superselection rule in quantum electrodynamics. No gauge-invariant local operator can create an isolated electron; the operator that creates an electron must also create a string of electric flux that ends on the electron.)

Correlation functions such as $\langle B_{a_1,a_2}(C_1, \Sigma_1, C_2, \Sigma_2, x_0)\rangle$ can be used to determine the tension of a dynamical string, or the amplitude for mixing between group eigenstates, as we will describe in Section 6.

(As an aside, we remark that an 't Hooft loop, like a Wilson loop, admits an alternative interpretation. If a Wilson loop operator acts on a timelike slice, it is natural to interpret it as an insertion of a classical charged source, as noted above. But if the Wilson loop acts on a spacelike slice, we may interpret it as an operator that creates a closed electric flux tube. We have noted that we may think of an 't Hooft operator acting on a spacelike slice as an object that creates a cosmic string. Alternatively, we may interpret the 't Hooft loop acting on a timelike slice as an insertion of a classical *magnetic monopole* source. In the situation originally considered by 't Hooft, the monopole satisfied the Dirac quantization condition, and so its "Dirac string" was invisible. We are considering a situation in which the Dirac string is visible; the surface Σ bounded by C is the world sheet of this Dirac string.)

2.5. Classical strings in the pure gauge theory

We will now analyze the behavior of the operator $A_a^{(\nu)}(\Sigma, x_0; C)$, using perturbative expansions. Here we will consider the case of a pure gauge theory with (discrete) gauge group G. In Section 8, we will consider the effects of introducing matter.

Strong coupling Although our main interest is in the physics at weak gauge coupling, we will make a few comments about the strong-coupling behavior of the pure gauge theory.

The plaquette action of the theory is taken to be

$$S^{(R)}_{\text{gauge},P} = -\ \beta\ \chi^{(R)}(U_P) + c..c\ , \tag{2.51}$$

where R is a representation of G. For $\beta << 1$, this theory confines sources that transform as certain irreducible representations of G. The criterion for a source to be confined is easiest to state in the case where R is irreducible. In that case, a source in the irreducible representation (ν) is *not* confined if and only if there are non-negative integers k_1 and k_2 such that

$$\begin{aligned} (R)^{k_1} \otimes (R^*)^{k_2} &\supset (\nu)\ , \\ (R)^{k_1} \otimes (R^*)^{k_2} &\supset (1)\ , \end{aligned} \tag{2.52}$$

where R^* denotes the complex conjugate of R, and (1) denotes the trivial representation. The point is that if eq. (2.52) is satisfied, then it is possible for a source in representation (ν) to be "screened by gluons." In other words, the electric flux tube that terminates on a (ν) source can break due to glue fluctuations. (The criterion eq. (2.52) generalizes the familiar notion that an adjoint representation source is unconfined in a strongly coupled $SU(N)$ gauge theory.) If (ν) *is* confined, then the expectation value of the corresponding Wilson loop operator decays for a large loop C like

$$\langle W^{(\nu)}(C)\rangle \sim \exp\left(-\kappa^{(\nu)}\text{Area}(C)\right)\ , \tag{2.53}$$

where Area(C) is the minimal area of a surface bounded by C, and $\kappa^{(\nu)}$ is the tension of a flux tube that carries electric flux in the representation (ν). (Of course, virtual glue may partially screen the source; $\kappa^{(\nu)}$ is the tension of the lightest flux tube that can terminate on a (ν) source.)

Since the gauge variables can see a classical string with "magnetic flux" $a \in G$, the quantum fluctuations of the gauge variables renormalize the tension of the string by an amount $\kappa_a^{(\text{ren})}$. Thus, the expectation value of the operator F decays for a large

surface Σ like

$$\langle F_a(\Sigma, x_0)\rangle \sim \exp\left(-\kappa_a^{(\mathrm{ren})} A(\Sigma)\right) \ , \tag{2.54}$$

where $A(\Sigma)$ is the area of Σ. The calculation of the leading contribution to $\kappa_a^{(\mathrm{ren})}$, for $\beta << 1$, is described in the Appendix.

Now consider the operator $A_a^{(\nu)}(\Sigma, x_0; C)$. Since there are no infinite-range Aharonov-Bohm interactions in the confining phase of the theory, we might expect that

$$\lim \langle A_a^{(\nu)}(\Sigma, x_0; C)\rangle = 1 \ . \tag{2.55}$$

This is indeed found for a representation (ν) that is not confined (such that the electric flux tube can break). Different behavior is found, however, for a representation (ν) that is confined. The operator $A_a^{(\nu)}(\Sigma, x_0; C)$ can have a non-trivial expectation value because the electric flux tube stretched across C crosses the surface Σ at certain points. In the leading order of strong-coupling perturbation theory (and assuming that no "partial screening" of the source occurs), each crossing contributes to A the factor

$$\frac{1}{n_\nu}\chi^{(\nu)}(a) \tag{2.56}$$

(or its complex conjugate, depending on the relative orientation of the flux tube and Σ at the point of crossing). Thus, even when Σ and C are far apart, $\langle A_a^{(\nu)}(\Sigma, x_0; C)\rangle$ is not a purely topological quantity that depends only on the linking number of Σ and C. Of course, in higher orders in the strong coupling expansion, the behavior of $A_a^{(\nu)}$ becomes still more complicated.

Similarly, if a loop C links with two different surfaces Σ_1 and Σ_2, the operator $A_{a_1,a_2}^{(\nu)}(\Sigma_1, \Sigma_2, x_0; C)$ acquires a factor $(1/n_\nu)\chi^{(\nu)}(a_1)$ each time the electric flux tube crosses Σ_1 and a factor $(1/n_\nu)\chi^{(\nu)}(a_2)$ each time it crosses Σ_2 (in the leading order of the strong-coupling expansion). Because of confinement, the string world sheets combine as trivial charge eigenstates rather than group eigenstates, even though both are defined with respect to a common basepoint.

Weak coupling For $\beta >> 1$ there is no confinement, and the Wilson loop operator exhibits perimeter law decay for any representation (ν).

The operator $F_a(\Sigma, x_0)$ decays as in eq. (2.54). The calculation of the leading behavior of $\kappa_a^{(\mathrm{ren})}$, for $\beta >> 1$, is described in the Appendix.

When the operator $F_a(\Sigma, x_0)$ is inserted, there is a configuration of the gauge variables such that no plaquettes are excited. This configuration can be constructed by choosing a set Ω of cubes of the dual lattice such that the boundary of Ω is Σ. Each cube in Ω is dual to a link of the original lattice. The configuration with no excited plaquettes is

$$\begin{aligned} U_l &= a \ , \quad l \in \Omega^* \ , \\ U_l &= e \ , \quad l \notin \Omega^* \ . \end{aligned} \tag{2.57}$$

This configuration is unique up to a gauge transformation. (The gauge transformations are deformations of Ω.)

Weak-coupling perturbation theory is carried out by expanding in the number of excited plaquettes, and in the degree of excitation. In the limit $\beta \to \infty$, the configurations with the minimal number of excited plaquettes dominate. By calculating the Wilson loop operator for the configuration with no excited plaquettes, we verify eq. (2.37) in the weak-coupling limit. Thus we find, as expected, that G charges are neither confined nor screened. Similarly, we may verify eq. (2.41) and eq. (2.49) in this limit.

When higher-order corrections in weak-coupling perturbation theory are computed, we find as anticipated that eq. (2.37) and eq. (2.41) continue to hold. But eq. (2.49) does not survive. These corrections are further discussed in the Appendix.

We wish to make one other remark here about the weak-coupling expansion, which might help to avoid confusion. To calculate the weak-coupling behavior of correlation functions that involve the operator $F_{a_1,a_2,\ldots,a_n}(\Sigma_1, \Sigma_2, \ldots, \Sigma_n, x_0)$, we first construct the configuration that has no excited plaquettes when $F_{a_1,a_2,\ldots,a_n}(\Sigma_1, \Sigma_2, \ldots, \Sigma_n, x_0)$ is inserted. This construction is a straightforward generalization of eq. (2.57). However, there are topologically inequivalent ways of choosing non-intersecting surfaces $\Omega_1, \ldots, \Omega_n$, that are bounded by $\Sigma_1, \ldots, \Sigma_n$ (as in Fig 15). Thus, one might get the impression that there can be two (or more) *gauge-inequivalent* configurations that both have no excited plaquettes. But this is not the case. To see why not, it is important to keep track of the basepoint, and of the paths from the basepoint to

the loops. If Ω_1 is distorted past Ω_2, as in Fig. 15, then Ω_1 crosses the paths from the basepoint to Σ_2. To avoid exciting any plaquettes, then, the links contained in Ω_2^* must now take the value $a_1 a_2 a_1^{-1}$. (This, again, is a reflection of the holonomy interaction between string loops.) Therefore, a Wilson loop that crosses Ω_1 first and Ω_2 second, in Fig. 15a, behaves exactly the same way as a Wilson loop that crosses Ω_2 first and Ω_1 second, in Fig. 15b. There is a unique gauge equivalence class of configurations with no excited plaquettes, just as there should be.

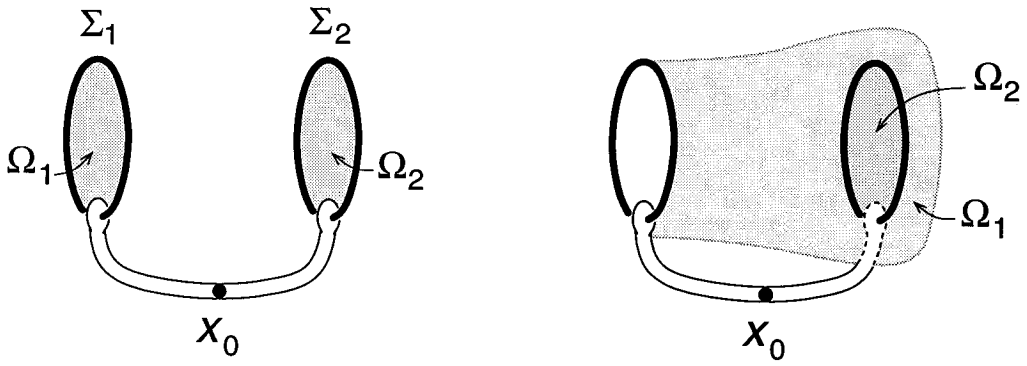

Fig. 15: Two topologically inequivalent ways of choosing the hypersurface Ω_1 that is bounded by Σ_1.

2.6. Dynamical strings and vortices

We will now consider how the 't Hooft operator can be used to investigate the properties of *dynamical* strings (in 3+1 dimensions) and vortices (in 2+1 dimensions).

String tension and vortex mass The operator $B_{a,a^{-1}}(C_1, \Sigma_1, C_2, \Sigma_2, x_0)$ can be used to compute the tension of a cosmic string that carries magnetic flux a (with a caveat described below). This operator creates an a string on C_1 and annihilates it on C_2. Thus, when the loops are large and far apart, we have

$$\begin{aligned} &\langle B_{a,a^{-1}}(C_1, \Sigma_1, C_2, \Sigma_2, x_0)\rangle \\ &\sim \exp\left(-\kappa_a^{(\mathrm{ren})}(A(\Sigma_1) + A(\Sigma_2))\right) \exp\left(-\kappa_a^{(\mathrm{dyn})} A(C_1, C_2)\right) . \end{aligned} \tag{2.58}$$

Here $\kappa_a^{(\mathrm{ren})}$ is the renormalization of the tension of a "classical" string source, and $\kappa_a^{(\mathrm{dyn})}$ is the tension of a dynamical string; $A(C_1, C_2)$ is the area of the minimal surface

with boundary $C_1 \cup C_2$. (See Fig. 16.) If the loops C_1 and C_2 are chosen to be far apart compared to the correlation length of the theory, but close together compared to the size of the loops, then the dependence of $\langle B_{a,a^{-1}}(C_1, \Sigma_1, C_2, \Sigma_2, x_0)\rangle$ on the separation between the loops determines the tension $\kappa_a^{(\mathrm{dyn})}$.

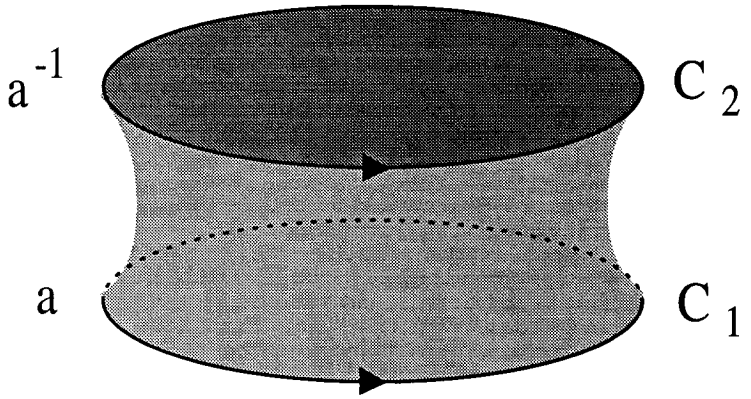

Fig. 16: Surface of minimal area bounded by the loops C_1 and C_2 (when the loops are close together).

Actually, the same information can be extracted from the behavior of the simpler operator $B_a(C, \Sigma, x_0)$. For a large loop C, we have

$$\langle B_a(C, \Sigma, x_0)\rangle \sim \exp\left(-\kappa_a^{(\mathrm{ren})} A(\Sigma)\right) \exp\left(-\kappa_a^{(\mathrm{dyn})} A(C)\right) , \qquad (2.59)$$

where $A(C)$ is the area of the minimal surface bounded by C. (See Fig. 17.) Since $A(\Sigma)$ and $A(C)$ can be varied independently, $\kappa_a^{(\mathrm{dyn})}$ can be determined. (Or, $F_a(\Sigma, x_0)$ can be used to measure $\kappa_a^{(\mathrm{ren})}$.) The calculation of $\kappa_a^{(\mathrm{dyn})}$ in weak-coupling perturbation theory is described in the Appendix. (In the strong-coupling limit, we have $\kappa^{(\mathrm{dyn})} = 0$—there are no stable magnetic flux tubes.)

Obviously, the same procedure can be used to calculate the mass of a vortex, in 2+1 dimensions.

The existence of a stable string (or vortex) can itself be used to probe the phase structure of the theory. In a confining phase, stable magnetic flux tubes do not exist; they "melt" due to magnetic disorder. If the G gauge symmetry is spontaneously broken to a subgroup H, stable a strings exist only if $a \in H$. Otherwise, an a string is the boundary of a domain wall, which decays as described in Section 2.

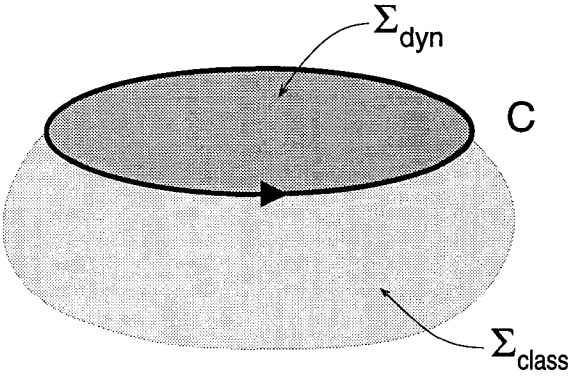

Fig. 17: The classical world sheet $\Sigma_{\rm class}$ and the dynamical world sheet $\Sigma_{\rm dyn}$ associated with the 't Hooft loop operator $B_a(C, \Sigma_{\rm class}, x_0)$.

If no stable a string exists, then $B_a(C, \Sigma, x_0)$ might not create a stable string. If it does not, its expectation value will behave, for a large loop C, like

$$\langle B_a(C, \Sigma, x_0)\rangle \sim \exp\left(-\kappa_a^{(\rm ren)} A(\Sigma)\right) \exp\left(-m_a^{(\rm ren)} P(C)\right) \ , \qquad (2.60)$$

where $P(C)$ is the perimeter of C. It may seem, then, that by measuring $\langle B_a(C, \Sigma, x_0)\rangle$, and determining whether it decays as in eq. (2.59) or as in eq. (2.60), we can find out whether a is contained in the unbroken subgroup H or not. However, there are subtleties. One problem is that there may be tradeoff between the dependence of $\langle B_a(C, \Sigma, x_0)\rangle$ on $A(\Sigma)$ and its dependence on $A(C)$. Eq. (2.60) will apply if the domain wall bounded by Σ decays by nucleating an a^{-1} string that completely cancels the flux of the classical string source on Σ. But it may be that $\kappa_a^{(\rm ren)}$ can be reduced if the nucleated string only partially screens the flux of the source. (We saw an instance of this phenomenon in the Z_6 example that was discussed in Section 3.) The advantage gained from reducing $\kappa^{(\rm ren)}$ may more than compensate for the cost of a non-vanishing $\kappa^{(\rm dyn)}$; then B_a will decay as in eq. (2.59), even though $a \notin H$.

Another complication can arise if the unbroken subgroup H is not a normal subgroup of G. For then a typical G conjugacy class contains both elements that are in H and elements that are not in H. Recall that $B_a(C, \Sigma, x_0)$ is actually averaged over the G conjugacy class that contains a. One particular H-class contained in this G-class will dominate the asymptotic behavior of $\langle B_a(C, \Sigma, x_0)\rangle$, and whether eq. (2.60) or eq. (2.59) applies depends on which class dominates.

We will return to the problem of finding a suitable order parameter, that can be used to identify H, in Section 8.

Mixing If the discrete gauge group G is unbroken, then elements of G in the same conjugacy class are associated with strings that transform into each other under the action of G. In the classical limit, these "group eigenstate" strings are degenerate energy eigenstates. (There is also a further degeneracy associated with "parity," which changes the orientation, and so transforms the a string into the a^{-1} string.) Quantum mechanically, these states mix with one another, and the degeneracy is lifted. The true energy eigenstates are "charge eigenstates" that transform according to irreducible representations of G (and parity).[21]

This mixing can be computed using the 't Hooft operator. We consider a correlation function in which an a loop is created on C_1 and a b loop is annihilated on C_2, where $b = gag^{-1}$ for some $g \in G$. It is crucial that the two strings be defined with respect to the same basepoint x_0. Otherwise, we would average a and b over the conjugacy class independently, and the correlation function would be dominated by the propagation of an a string from C_1 to C_2, rather than the mixing of an a string with a b string.

Let C_1 and C_2 be two congruent loops, one directly above the other as in Fig. 18. The separation between the loops is large compared to the correlation length of the theory, but small compared to the size of the loops. If the ab^{-1} string is stable, then, in the weak coupling limit, the correlation function will be dominated by the configuration in Fig 18a. In this configuration, the world sheets of the a and b strings join, and the loop at which they join is the boundary of the world sheet of an ab^{-1} string. If this configuration dominates, then

$$B_{a,b}(C_1, \Sigma_1, C_2, \Sigma_2, x_0) \sim \exp\left(-\kappa_a^{(\mathrm{ren})} A(\Sigma_1) - \kappa_b^{(\mathrm{ren})} A(\Sigma_2)\right) \exp\left(-\kappa_{ab^{-1}}^{(\mathrm{dyn})} A(C_1)\right) . \tag{2.61}$$

But if the ab^{-1} string is unstable, and decays to a widely separated a string and b^{-1} string, then the configuration in Fig. 18b will dominate. Here, the world sheets of the dynamical strings are stretched tightly across C_1 and C_2. If this configuration

dominates, then

$$B_{a,b}(C_1,\Sigma_1,C_2,\Sigma_2,x_0) \sim \exp\left(-\kappa_a^{(\mathrm{ren})}A(\Sigma_1) - \kappa_b^{(\mathrm{ren})}A(\Sigma_2)\right) \exp\left(-(\kappa_a^{(\mathrm{dyn})} + \kappa_b^{(\mathrm{dyn})})A(C_1)\right) . \tag{2.62}$$

Thus, the mixing amplitude in the weak-coupling limit is either

$$e^{-S_{a,b}^{(\mathrm{mix})}} \sim \exp\left(-(\kappa_a^{(\mathrm{dyn})} + \kappa_b^{(\mathrm{dyn})})A(C_1)\right) \tag{2.63}$$

or

$$e^{-S_{a,b}^{(\mathrm{mix})}} \sim \exp\left(-\kappa_{ab^{-1}}^{(\mathrm{dyn})}A(C_1)\right) , \tag{2.64}$$

whichever is larger.

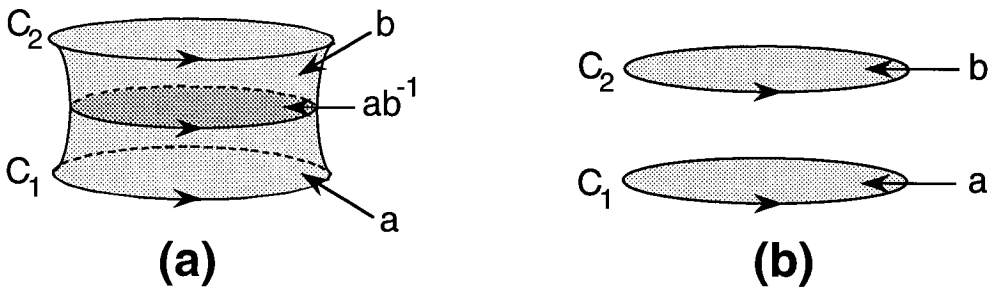

Fig. 18: Two contributions to the mixing of an a string and a b string. In (a), an ab^{-1} string spontaneously nucleates, expands, meets the a string, and converts it into a b string. In (b), the a string shrinks and annihilates, then the b string nucleates and grows.

Actually, these are not the most general possibilities, for the ab^{-1} string may be unstable, and may prefer to decay in some other channel. If the ab^{-1} string decays to n widely separated stable strings that carry flux $c_1, c_2, \ldots, c_n$, then we find

$$e^{-S_{a,b}^{(\mathrm{mix})}} \sim \exp\left(-\left(\sum_i \kappa_{c_i}^{(\mathrm{dyn})}\right)A(C_1)\right) . \tag{2.65}$$

These results were previously derived in Ref. 21.

2.7. INTERACTIONS

Entanglement and holonomy interactions The Aharonov-Bohm interaction between a classical string source and a classical charged source was studied in Section 5. Correlation functions can also be used to study the holonomy interaction between two vortices (in 2+1 dimensions), or two strings (in 3+1 dimensions), and also the entanglement of strings.

In 2+1 dimensions, two *loops* Σ_1 and Σ_2 can link, as shown in Fig. 19. Then the quantity $\langle F_{a,b}(\Sigma_1, \Sigma_2, x_0)\rangle$ probes what happens when one vortex source winds around another. As we noted in Section 1, there is a holonomy interaction between the vortices if a and b do not commute; both are conjugated by ab[25,24]. This means that the vortex world lines cannot close (the vortex pairs cannot re-annihilate) unless there is an exchange of topological quantum numbers between the two vortices.

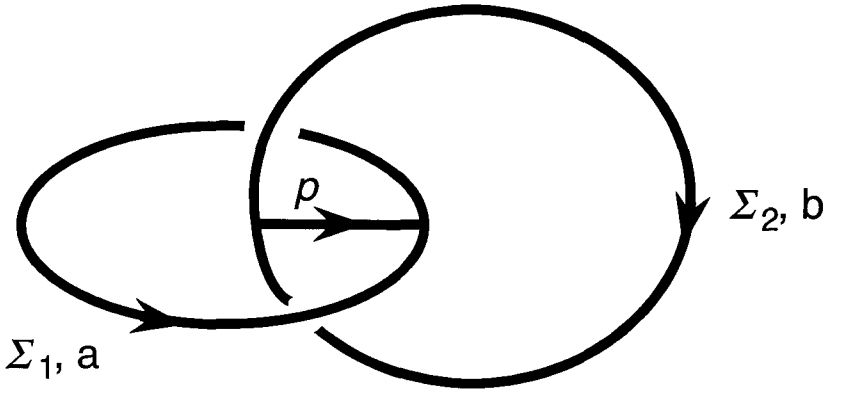

Fig. 19: Linked world lines Σ_1 and Σ_2 of a classical a vortex and a classical b vortex. If a and b do not commute, then the classical vortices must exchange a dynamical $aba^{-1}b^{-1}$ vortex. The exchange occurs along the path P, the shortest path that connects the two world lines. (Compare Fig. 5.)

An $(ab)a(ab)^{-1}$ vortex becomes an a vortex, and an $(ab)b(ab)^{-1}$ vortex becomes a b vortex, if the flux $aba^{-1}b^{-1}$ flows from the b vortex to the a vortex.* Therefore,

* Here we have used different conventions to assign group elements to the entangled vortex world lines than the conventions used in Section 1 to assign group elements to entangled strings in three spatial dimensions.

the quantum numbers that must be exchanged are those of an $aba^{-1}b^{-1}$ vortex. We thus find, in the weak coupling limit,

$$\begin{aligned}&\langle F_{a,b}(\Sigma_1,\Sigma_2,x_0)\rangle \\ &\sim \exp\left(-m_a^{(\mathrm{ren})}P(\Sigma_1)-m_b^{(\mathrm{ren})}P(\Sigma_2)\right)\exp\left(-m_{aba^{-1}b^{-1}}^{(\mathrm{dyn})}L(\Sigma_1,\Sigma_2)\right)\ .\end{aligned} \tag{2.66}$$

Here $m^{(\mathrm{ren})}$ is the renormalization of the mass of a classical vortex source, $m^{(\mathrm{dyn})}$ is the mass of a dynamical vortex, $P(\Sigma)$ is the perimeter of Σ, and $L(\Sigma_1,\Sigma_2)$ is the length of the shortest path that connects Σ_1 and Σ_2. (If the $aba^{-1}b^{-1}$ vortex is unstable, then $m_{aba^{-1}b^{-1}}^{(\mathrm{dyn})}$ is replaced by the sum of the masses of the vortices to which it decays.)

The leading behavior of weak-coupling perturbation theory on the lattice is found by identifying the configurations with the minimal number of excited plaquettes, as we described in Section 5. To find the leading contribution to $\langle F_{a,b}(\Sigma_1,\Sigma_2,x_0)\rangle$, we choose surfaces Ω_1 and Ω_2 that are bounded by Σ_1 and Σ_2 respectively, and construct a configuration with $U_l = a$ on links in Ω_1^* and $U_l = b$ on links in Ω_2^*. But where Ω_1 and Ω_2 intersect, there is a string of excited plaquettes with

$$S_{\mathrm{gauge},P}^{(R)} = -\ \beta\ \chi^{(R)}(aba^{-1}b^{-1}) + c.c. \tag{2.67}$$

(See Fig. 20.) By choosing Ω_1 and Ω_2 so that this string has minimal length, we obtain eq. (2.66).

(If the $aba^{-1}b^{-1}$ vortex is unstable, then the dominant configuration in the weak-coupling limit is found by a slightly modified procedure. Either Σ_1 or Σ_2, or both, becomes the boundary of several surfaces Ω_1^i, or Ω_2^j. The link configuration is chosen so that $U_l = c_i$ on Ω_1^{*i} and $U_l = d_j$ on Ω_2^{*j}, where $\prod c_i = a$ and $\prod d_j = b$. These surfaces intersect along several strings that connect the two world lines; thus, the classical vortices exchange several separated dynamical vortices.)

In 3+1 dimensions, strings labeled by non-commuting group elements become entangled when they cross, as we described in Section 1. String world sheets in four dimensions generically intersect at isolated points. (A point of intersection is a type of "instanton.") Consider surfaces Σ_1 and Σ_2 that cross at two points, as shown in Fig. 21. Because the strings entangle, an a world sheet on Σ_1 and a b world

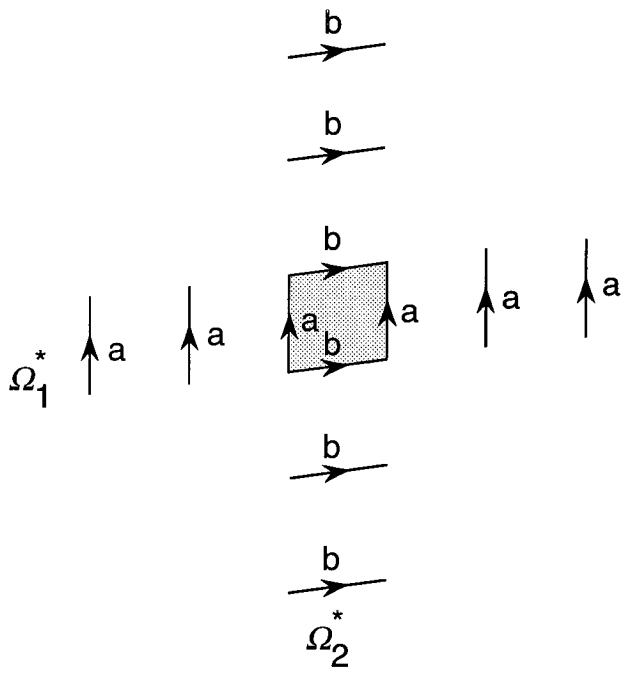

Fig. 20: A plaquette (shaded) contained in the intersection of Ω_1^* and Ω_2^*.

sheet on Σ_2 become joined by the world sheet of an $aba^{-1}b^{-1}$ string. Thus, in the weak-coupling limit, we find

$$\begin{aligned}&\langle F_{a,b}(\Sigma_1,\Sigma_2,x_0)\\&\sim \exp\left(-\kappa_a^{(\mathrm{ren})}A(\Sigma_1)-\kappa_b^{(\mathrm{ren})}A(\Sigma_2)\right)\exp\left(-\kappa^{(\mathrm{dyn})}_{aba^{-1}b^{-1}}A(\Sigma_1,\Sigma_2)\right)\ ,\end{aligned}\tag{2.68}$$

where $A(\Sigma_1,\Sigma_2)$ is the area of the minimal surface that joins Σ_1 and Σ_2. Again, this result is easily verified using weak-coupling perturbation theory on the lattice.

It is also instructive to consider $\langle F_{a,b}(\Sigma_1,\Sigma_2,x_0)\rangle$ where the surfaces Σ_1 and Σ_2 have the topology shown in Fig. 22. Here Σ_1 is a torus that links once with the sphere Σ_2. As before, Ω_1 and Ω_2 unavoidably intersect along a surface of frustrated plaquettes; we find the leading behavior of $\langle F_{a,b}(\Sigma_1,\Sigma_2,x_0)\rangle$ by choosing Ω_1 and Ω_2 so that this intersection has minimal area. The result is again eq. (2.68). But now, if the width of the torus is small compared to the size of the sphere, $A(\Sigma_1,\Sigma_2)$ is the area of the minimal slice through the torus.

To interpret this result, we recall the observation in Section 1, that when a loop of a string winds around a loop of b string, it becomes a loop of bab^{-1} string. We may

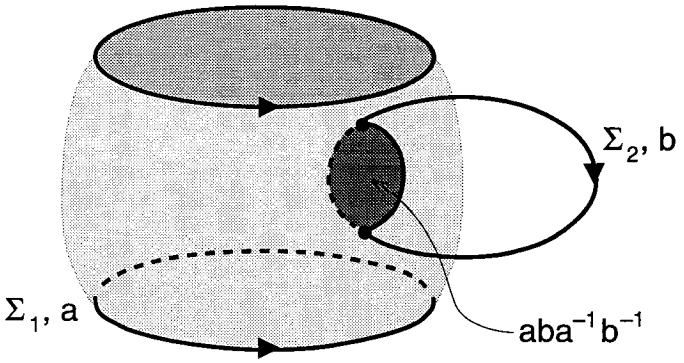

Fig. 21: World sheets Σ_1 and Σ_2 of a classical a string and classical b string that intersect at two isolated points. (Only a slice through Σ_2 is shown.) The indicated shaded region is the world sheet of a dynamical $aba^{-1}b^{-1}$ string that connects the two classical strings.

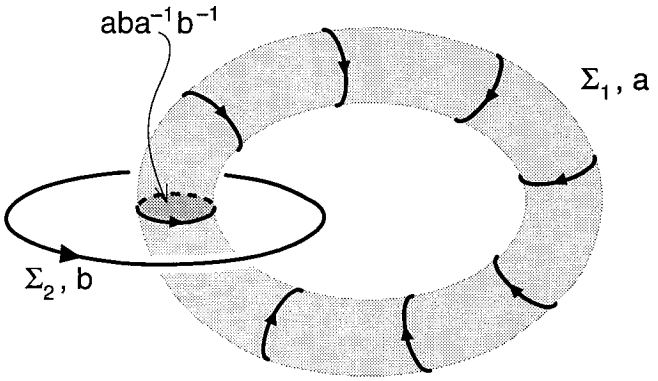

Fig. 22: World sheets Σ_1 and Σ_2 of a classical a string and a classical b string. (Only a slice through Σ_2 is shown.) Σ_1 is a torus that links with Σ_2. The shaded region is the world sheet of a dynamical $aba^{-1}b^{-1}$ string.

regard Fig. 22 as a depiction of a process in spacetime, in which loops of a and a^{-1} string are produced, the a loop winds around the b string, and the a and a^{-1} strings then annihilate. But because of the Aharonov-Bohm interaction between the strings, this process is disallowed; an a^{-1} loop cannot annihilate a bab^{-1} loop. Therefore, $\langle F_{a,b}(\Sigma_1, \Sigma_2, x_0)\rangle$ is suppressed by the amplitude for the bab^{-1} string to become, via quantum tunneling, an a string. Indeed, comparing with eq. (2.64), we see that

$$e^{-S^{(\mathrm{mix})}_{a,bab^{-1}}} \sim \exp\left(-\kappa^{(\mathrm{dyn})}_{aba^{-1}b^{-1}} A^{(\mathrm{min})}(\Sigma_1)\right) \tag{2.69}$$

is the suppression factor in eq. (2.68).

Cheshire charge and the Borromean rings As discussed in Section 1, it is an inevitable consequence of the non-Abelian Aharonov-Bohm phenomenon that a loop of string (or a pair of vortices) can carry charge, and can exchange charge with other charged objects.[5,12,13] Let us see how this property is reflected in the behavior of the correlation functions of our string operators.[30]

Consider the following process (in 2+1 dimensions), depicted in Fig. 23. First, a pair consisting of an a vortex and its anti-vortex spontaneously appears; the total charge of the pair is trivial. Then a charged particle in the representation (ν) winds counterclockwise around the a vortex; thus, charge is transferred to the vortex pair, as described in Section 1. Next, a b vortex winds around the (charged) vortex pair, acquiring an Aharonov-Bohm phase that is sensitive to the charge of the pair. Then the (ν) particle winds clockwise around the a vortex, discharging the pair. Finally, the a vortex and anti-vortex reunite and annihilate.

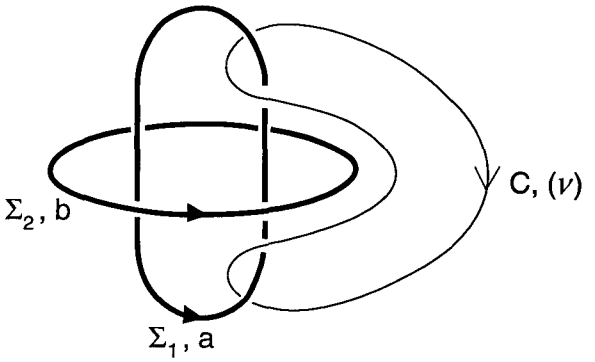

Fig. 23: The Borromean rings. Σ_1 is the world line of an a vortex, Σ_2 is the world line of a b vortex, and C is the world line of a charged particle that transforms as the representation (ν). The charged particle transfers charge to the a vortex-antivortex pair, and the charge is subsequently detected via the Aharonov-Bohm interaction of the pair with the b vortex.

If the vortices and charged particle are treated as classical sources, this process is captured by the correlation function $\left\langle F_{a,b}(\Sigma_1, \Sigma_2, x_0)\ W^{(\nu)}(C) \right\rangle$, where the three loops Σ_1, Σ_2 and C are configured as in Fig. 23. This is a topologically non-trivial joining of three loops known as the "Borromean rings"; no two loops are linked, yet the loops cannot be separated without crossing. Because this correlation function

has the spacetime interpretation described above, we may anticipate that, when the three loops are joined but far apart, $\left\langle F_{a,b}(\Sigma_1, \Sigma_2, x_0)\ W^{(\nu)}(C)\right\rangle$ differs from its value for large unjoined loops by a topological multiplicative factor—the Aharonov-Bohm phase acquired by the b vortex that winds around the charged vortex pair.

To calculate $A^{(\nu)}_{a,b}(\Sigma_1, \Sigma_2, x_0; C)$, we proceed, as usual, by finding the configuration of the link variables such that, when $F_{a,b}$ is inserted, there are no frustrated plaquettes. We pick surfaces Ω_1, bounded by Σ_1, and Ω_2, bounded by Σ_2, that do not intersect; such surfaces are shown in Fig. 24. We then choose (up to a gauge transformation)

$$\begin{aligned} U_l &= a\ , \quad l \in \Omega_1^*\ , \\ U_l &= b\ , \quad l \in \Omega_2^*\ , \\ U_l &= e\ , \quad l \notin \Omega_1^* \cup \Omega_2^*\ . \end{aligned} \tag{2.70}$$

Now we compute $W^{(\nu)}(C)$ in this configuration. As is clear in Fig. 24, the loop C crosses first Ω_1 in a positive sense, then Ω_2 in a negative sense, then Ω_1 in a negative sense, and finally Ω_2 in a positive sense, before closing. The corresponding path-ordered exponential is $ba^{-1}b^{-1}a$, and taking a trace yields[30]

$$\lim\ \left\langle A^{(\nu)}_{a,b}(\Sigma_1, \Sigma_2, x_0; C)\right\rangle \sim \frac{1}{n_\nu}\ \chi^{(\nu)}(aba^{-1}b^{-1})\ . \tag{2.71}$$

We should verify that the factor eq. (2.71) can be interpreted as the Aharonov-Bohm phase acquired by a b vortex that winds around a vortex pair with Cheshire charge. The interpretation is easiest if we explicitly average a over the representatives of the class to which it belongs. If this averaging is not performed, then there is a correlation between the choice of the class representative that is used to measure the charge, and the choice of the class representative for the string that is being measured; this correlation makes the interpretation of the measurement more complicated. Of course, averaging a and b over class representatives independently is equivalent to defining a and b with reference to distinct basepoints; we have

$$\lim\ \frac{\left\langle F_a(\Sigma_1, x_0)\ F_b(\Sigma_2, y_0)\ W^{(\nu)}(C)\right\rangle}{\langle F_a(\Sigma_1, x_0)\rangle\ \langle F_b(\Sigma_2, y_0)\rangle\ \langle W^{(\nu)}(C)\rangle} = \frac{1}{n_G}\ \sum_{g\in G} \frac{1}{n_\nu}\ \chi^{(\nu)}(gag^{-1}bga^{-1}g^{-1}b^{-1})\ . \tag{2.72}$$

We recall from the discussion in Section 1 that, when a particle in representation (ν) winds around one of the vortices of a pair that is initially uncharged, the final

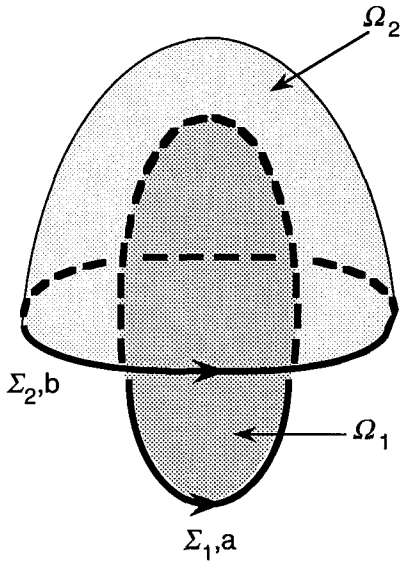

Fig. 24: Non-intersecting surfaces Ω_1 and Ω_2 that are bounded by the vortex world lines Σ_1 and Σ_2.

state of the vortex pair is a superposition of states with various values of the charge. Eq. (2.72) gives the *expectation value* of the Aharonov-Bohm phase acquired by a b vortex that winds around the charged pair. If p_μ is the probability that the vortex pair transforms as the irreducible representation μ, then we have

$$\sum_\mu p_\mu \frac{1}{n_\mu}\chi^{(\mu)}(b) = \frac{1}{n_G}\sum_{g\in G}\frac{1}{n_\nu}\chi^{(\nu)}(gag^{-1}bga^{-1}g^{-1}b^{-1}) \ . \tag{2.73}$$

Indeed, eq.(2.73) can be verified directly, with p_μ given by eq. (2.17) and (2.16). Details will be presented elsewhere.[30]

The above discussion can be applied, with hardly any modification, to the case of charged string loops, in 3+1 dimensions. There is an analog of the Borromean ring configuration, in which two disjoint closed surfaces are joined by a closed loop, although the loop is not linked with either surface. Eq. (2.72) holds, if $\Sigma_{1,2}$ are surfaces, and C is a loop, in this configuration.

2.8. GAUGE-HIGGS SYSTEMS

Weak coupling In this section, we will consider models with Higgs fields coupled to gauge fields. We wish to investigate the screening of gauge charges—and of Aharonov-Bohm interactions—due to the Higgs mechanism.

For a gauge system with (discrete) gauge group G, we introduce a Higgs field ϕ_i, defined on sites, that takes values in G. We then include in the action of the model the term

$$S_{\text{Higgs}} = -\sum_{\mu} \gamma_\mu \sum_{l} \left(\chi^{(\mu)} \left((\phi^{-1} U \phi)_l \right) + c.c \right) \ , \tag{2.74}$$

where the sum runs over all irreducible representations (μ) of G. (Compare eq. (2.23).) Suppose that some of the γ_μ's are large ($>> 1$), while all others are small (or zero). Then, we can analyze the behavior of this model using perturbative methods.

In the weak-coupling limit $\gamma_\mu \to \infty$, $\phi^{-1} U \phi$ becomes restricted to the kernel of the representation (μ) at each link, and so U_P is also "frozen" to the kernel at each plaquette. Thus, G becomes "spontaneously broken" to the subgroup $H = \text{Ker}(D^{(\nu)})$. By choosing several of the γ_μ's to be large, we can break G to the kernel of a reducible representation. Indeed, breakdown to any *normal* subgroup of G can be achieved in this way, since every normal subgroup is the kernel of some representation.

Of course, we know from the usual continuum weak-coupling analysis that more general patterns of symmetry breakdown are possible. In that analysis, the Higgs field resides in the vector space on which a representation of G acts, and, in principle, the unbroken subgroup could be the stability group of any vector in this space. In the model with S_{Higgs} given by eq. (2.74), these more general patterns are not obtained when all of the Higgs couplings either vanish or are very large. They might, of course, be obtained at intermediate coupling. After suitable block-spin transformations are performed, the effective theory that describes the infrared behavior of the model would be a "continuum" theory for which the usual analysis could apply.

Anyway, with the breaking of G to a normal subgroup H implemented as described above, we can proceed to calculate the ABOP $A_a^{(\nu)}(\Sigma, x_0; C)$ in weak-coupling perturbation theory, and so probe the fate of the Aharonov-Bohm interaction in the Higgs model. The calculation gives the expected results. But there is one difficulty.

The problem is that the irreducible representation (ν) of G is typically a reducible representation of the subgroup H. In general, just one of the irreducible representations of H that is contained in (ν) will dominate the asymptotic behavior of $W^{(\nu)}(C)$ when the loop C is large. So it is the Aharonov-Bohm interaction of this H representation with an a string that is probed by $A_a^{(\nu)}$. In fact, if (ν) contains the trivial H representation, then this will dominate at weak coupling, and $\left\langle A_a^{(\nu)} \right\rangle$ will behave trivially even though (ν) may contain other H representations that *do* have Aharonov-Bohm interactions with the string.

We have encountered a general problem with the interpretation of $\left\langle A_a^{(\nu)} \right\rangle$ that arises whenever G is non-Abelian and is partially broken. If $\left\langle A_a^{(\nu)} \right\rangle$ behaves non-trivially for some choice of a, then we know that an unscreened charge must dominate $W^{(\nu)}$; by varying a, we can obtain information about the representation according to which this charge transforms. But if $\left\langle A_a^{(\nu)} \right\rangle$ behaves trivially for all a, we know only that $W^{(\nu)}$ is dominated by a screened charge. We have no *a priori* knowledge of how this charge transforms, or of how other H representations contained in (ν) interact with strings. This problem complicates the task of inferring H from the behavior of our correlation functions.

String stability We noted in Section 6 that, if G is a discrete group, then the existence (or not) of a stable string with flux $a \in G$ provides a criterion for determining whether a is contained in the unbroken group H. It may seem, then, that the operator $B_a(C, \Sigma, x_0)$, which creates an a string, can be used to probe the realization of the G gauge symmetry. But there are two problems.

First, when inserted in a Green function with gauge-invariant operators, B_a is actually averaged over the G conjugacy class to which a belongs. This class may contain some elements that are in the unbroken group H and some that are not, which complicates the interpretation of $\langle B_a \rangle$.

Second, even if no element of the class that contains a is in H, $B_a(C, \Sigma, x_0)$ may nevertheless create a stable string. This can happen, as we saw in Section 3, due to the competition between the renormalization of the tension of the classical string that propagates on Σ and the tension of the dynamical string whose world sheet is bounded by C.

Let us ignore the second problem for the moment, and address the first. We first note that the situation is relatively simple if H is a normal subgroup of G. In that case, a given G conjugacy class is either entirely contained in H or is disjoint from H. By studying B_a for various class representatives, we can determine which classes are contained in H, and so reconstruct H.

The general case is somewhat more complicated, and B_a by itself is not sufficient to completely determine the unbroken subgroup H. Instead, one way to proceed is the following: If $G = \{a_1, a_2, \ldots, a_{n_G}\}$ is a finite group of order n_G, consider

$$\langle B_{a_1,a_2,\ldots,a_{n_G}}(C_1, \Sigma_1, C_2, \Sigma_2, \ldots, C_{n_G}, \Sigma_{n_G}, x_0)\rangle \ , \tag{2.75}$$

in which all group elements appear. Choose the loops to be large and widely separated, with their sizes ordered according to

$$A(C_1) >> A(C_2) >> \cdots >> A(C_{n_G}) \ . \tag{2.76}$$

Now $B_{a_1,a_2,\ldots}$ is effectively averaged over gauge transformations at the basepoint, and so may be replaced by

$$\frac{1}{n_G}\sum_{g\in G} B_{ga_1g^{-1},ga_2g^{-1},\ldots,ga_{n_G}g^{-1}}(C_1, \Sigma_1, C_2, \Sigma_2, \ldots, C_{n_G}, \Sigma_{n_G}, x_0) \ . \tag{2.77}$$

The expectation value is dominated by the configurations that minimize

$$\kappa^{(\mathrm{ren})}_{ga_1g^{-1}} A(\Sigma_1) + \kappa^{(\mathrm{dyn})}_{ga_1g^{-1}} A(C_1) \tag{2.78}$$

(where $\kappa^{(\mathrm{dyn})}_a = 0$ for $a \notin H$). This condition may not determine g uniquely. Among those g that minimize eq. (2.78), the dominant configurations are such that $\kappa^{(\mathrm{ren})}_{a_2} A(\Sigma_2) + \kappa^{(\mathrm{dyn})}_{a_2} A(C_2)$ is also minimized. And so on. Now by varying $A(C)$ and $A(\Sigma)$ independently, we find the group elements a for which $\kappa^{(\mathrm{dyn})}_{gag^{-1}} = 0$. We thus determine the unbroken subgroup H up to one overall conjugacy $H \to gHg^{-1}$. This ambiguity is expected; it corresponds to the freedom to change the embedding of H in G by performing a gauge transformation. (Of course, the calculation of eq. (2.75) involves an average with respect to this embedding.)

In passing, we have found the tension of all of the stable strings associated with the various elements of H.

If G is Abelian, as in the discussion in Section 3, then we can overcome the second problem (that B_a may create a stable string even for $a \notin G$) easily enough. By measuring $A_a^{(\nu)}$, we can determine the Aharonov-Bohm interactions of the string created by B_a, and so identify the flux of the string as belonging to H. If G is non-Abelian, though, life is more complicated. For as we noted above, we have no *a priori* knowledge of what H representation dominates the asymptotic decay of $W^{(\nu)}$. While it seems altogether physically reasonable that the stability and Aharonov-Bohm interactions of strings can be used to identify an unbroken gauge group H, it is not so easy to specify how this should be done with gauge-invariant correlation functions.

The VOOP Another promising probe of charge screening in a gauge theory is the "vacuum overlap order parameter" (VOOP) proposed by Marcu and Fredenhagen.[17,20,18] Let us compare and contrast the VOOP with the Aharonov-Bohm order parameter that has been discussed in this chapter.

Suppose that an H gauge theory contains a matter field $\Phi^{(\mu)}$ that transforms as the irreducible representation (μ) of H. If the local H symmetry is unbroken, and the representation (μ) is not confined or screened, then (loosely speaking) the field $\Phi^\dagger$ should create a stable particle. We can express this in gauge-invariant language. If x and y are distantly separated points, and $P_{x,y}$ is a path from y to x, consider the non-local gauge-invariant operator

$$K^{(\mu)}(x,y,P_{x,y}) = \Phi_x^{(\mu)\dagger} D^{(\mu)} \left(\left(\prod_{l \in P_{x,y}} U_l \right) \right) \Phi_y^{(\mu)} \ . \tag{2.79}$$

If $\Phi^{(\mu)\dagger}$ creates a free charge, then this charge must propagate between x and y, as in Fig. 25. Thus, we have

$$\left\langle K^{(\mu)}(x,y,P) \right\rangle \sim \exp\left(-M_{\text{ren}}^{(\mu)} \, L(P) \right) \exp\left(-M_{\text{dyn}}^{(\mu)} \, |x-y| \right) \ , \tag{2.80}$$

where $M_{\text{dyn}}^{(\mu)}$ is the mass of the stable particle created by $\Phi^{(\mu)\dagger}$, and $M_{\text{ren}}^{(\mu)}$ is the renormalization of the mass of the classical source propagating along P. (Here, $L(P)$ is the length of P, and $|x-y|$ is the distance from x to y.) Since $M_{\text{ren}}^{(\mu)}$ can be determined independently by measuring $\left\langle W^{(\mu)}(C) \right\rangle$ (or by varying $L(P)$ with $|x-y|$

fixed), eq. (2.80) can be used to find $M_{\rm dyn}^{(\mu)}$. But if the representation (μ) is screened (or confined), then $\left\langle K^{(\mu)}(x,y,P)\right\rangle$ will become independent of $|x-y|$ for large separation; in effect, we have $M_{\rm dyn}^{(\mu)} = 0$, with $M_{\rm dyn}^{(\mu)}$ defined by eq. (2.80). Thus, Marcu and Fredenhagen suggested that $M_{\rm dyn}^{(\mu)} > 0$ if and only if $\Phi^{(\mu)\dagger}$ creates a free charge.

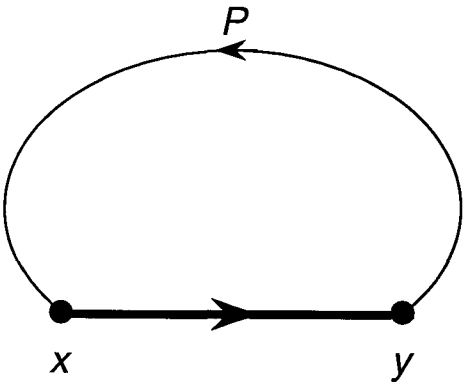

Fig. 25: The path P from y to x that is used to construct the gauge-invariant correlation function $K^{(\mu)}(x,y,P)$. A classical charged particle propagates along P, and a dynamical charged particle propagates from x to y.

This construction is obviously closely similar to our discussion in Section 6 of string stability, and it suffers from related problems. Suppose that, in a model with gauge group G, spontaneous breakdown to a subgroup H occurs. Then, an irreducible representation (μ) of G contains various irreducible representations of H. It may be that among these representations are some that can exist as free charges, and others that are screened. We might expect, then, that the screened charges dominate $\left\langle K^{(\mu)}\right\rangle$, so that $M_{\rm dyn}^{(\mu)} = 0$ even though (μ) contains some unscreened representations of H. Even this is not clear; because of the competition between $M_{\rm dyn}$ and $M_{\rm ren}$ in eq. (2.80), the free charges may actually dominate.

Because of these difficulties, it is not at all easy, in general, to identify the unbroken gauge group H based on the behavior of $\left\langle K^{(\mu)}(x,y,P)\right\rangle$.

Order parameters: Some concluding remarks The existence of stable cosmic strings (or vortices), and of Aharonov-Bohm interactions between strings and free charges, can be used to identify and classify the various phases of a gauge theory. Yet, because of the problems discussed above, it proves difficult to formulate a general procedure

that unambiguously specifies the realization of the gauge symmetry, *e.g.*, the "unbroken" subgroup. This is surprising (to us), but we are reluctant to attach any fundamental significance to it.

In fact, though, just because of these problems, the phase structure of certain gauge theories may be richer than one might naively expect. For example, if the gauge group G is "spontaneously broken" to H, then, as we have remarked, just one of the irreducible H representations contained in the G representation (ν) will dominate the asymptotic behavior of $\left\langle W^{(\nu)}(C)\right\rangle$. On some surface in the parameter space of the theory, a "crossover" may occur, where this H representation changes. Thus, the order parameter $A_a^{(\nu)}$ might be non-analytic on this surface—the surface would be a sort of phase boundary, even though no change in symmetry occurs there.

For most of the discussion in this chapter, we have taken the gauge group G to be a discrete group. There is no obstacle, however, to generalizing our results to the case where G is a continuous group.

2.9. Electric flux tubes and dynamical monopoles

In this chapter, we have systematically discussed the Aharonov-Bohm interactions between magnetic flux tubes and electric charges, which can occur in the Higgs phase of a gauge theory. Central to the discussion has been the effect of quantum-mechanical electric charge fluctuations on the interaction.

There is another type of Aharonov-Bohm interaction, which can occur in the *confining* phase of a gauge theory—the interaction between an *electric* flux tube and a *magnetic* charge. The existence of such interactions has been noted previously,[22,23,5,6] as have the implications concerning the "magnetic hair" carried by black holes.[5,9] But the effects of quantum-mechanical magnetic charge fluctuations on this incarnation of the Aharonov-Bohm phenomenon have not been analyzed before.

In this section, we will develop a quantum field-theoretic treatment of electric flux tubes in a confining gauge theory that contains *dynamical* magnetic monopoles, and will investigate the interactions of flux tubes with monopoles. This treatment, of course, will be closely similar to our theory of magnetic flux tubes, in a Higgs phase.

First, we will describe (following closely the work of Srednicki and Susskind[23]) how dynamical magnetic monopoles are introduced into a lattice gauge theory. Then we will construct a generalized Wilson loop operator that creates an electric flux tube that has Aharonov-Bohm interactions with the monopoles. By studying the properties of the correlation functions of this operator, we will investigate the effect of magnetic charge fluctuations on the Aharonov-Bohm effect.

Magnetic Monopoles on the Lattice

For definiteness, we will consider a gauge theory with gauge group $SU(N)$. A pure $SU(N)$ gauge theory (no matter), or a theory with the matter fields transforming trivially under the center Z_N of $SU(N)$, admits magnetic monopoles with Z_N magnetic charges. (Fields that transform under Z_N would be able to see the Dirac string of such a monopole; that is, the monopole with minimal Z_N charge would not satisfy the Dirac quantization condition.) If $SU(N)/Z_N$ is actually the unbroken gauge symmetry of an underlying theory with simply connected gauge group $G \supset SU(N)/Z_N$, where G is broken via the Higgs mechanism, then such monopoles arise as topological solitons. For example, there are models with $G = Spin(N^2-1)$ broken to $SU(N)/Z_N$ that contain Z_N monopoles.

The (usual) 't Hooft loop operator $B(C)$ inserts a world line of a Z_N monopole along the closed loop C. But the monopole introduced by an 't Hooft loop is a classical source, not a dynamical object. We wish to elevate the monopoles to the status of dynamical degrees of freedom, and introduce a coupling constant that controls the effects of virtual monopoles.

It is most convenient to choose the dynamical variable to be a "Dirac string field" that is summed over in the path integral. On the lattice, this field associates with each plaquette P a quantity $\eta_P \in Z_N$, which may be regarded as the Z_N magnetic flux carried by a Dirac string that pierces that plaquette. If the total magnetic flux entering a cube of the lattice is non-trivial, then a magnetic monopole resides in that cube. Of course, since the Dirac strings themselves must be unobservable, this theory should respect, as well as the usual $SU(N)$ local symmetry, an additional Z_N local symmetry that deforms the Dirac strings (without, of course, changing where the monopoles are).

The construction of this theory was described by Srednicki and Susskind.[23] In the absence of matter fields, it has the action

$$S = -\beta \sum_{\text{plaq}} \text{tr}(\eta_P U_P) - \lambda \sum_{\text{cubes}} \eta_c + c.c. \ , \tag{2.81}$$

where

$$\eta_c = \prod_{P \in c} \eta_P \tag{2.82}$$

is the product of the six η_P's associated with the oriented faces of the cube c; in other words, η_c is the Z_N magnetic charge inside c.

The extra Z_N local symmetry respected by this theory is defined on links; it acts according to

$$\begin{aligned} n: \ & U_l \to e^{2\pi i n/N} U_l \ , \\ & \eta_P \to e^{-2\pi i n/N} \eta_P \ , \quad P \ni l \ , \end{aligned} \tag{2.83}$$

where $n = 0, 1, 2, \ldots, N-1$. This transformation moves the Dirac strings without changing the magnetic charge η_c or the magnetic flux $\eta_P U_P$ that appears in the gauge field plaquette action. (The quantity U_P is not invariant under the extra local symmetry, and so is unphysical; it can be interpreted as a fictitious magnetic flux that includes a contribution from the (unobservable) Dirac string that crosses P.)

The coupling constant λ controls the strength of the effects of virtual monopoles. For $\lambda << 1$, magnetic charge fluctuations occur copiously, but for $\lambda >> 1$, magnetic charge fluctuations are strongly suppressed. In the limit $\lambda \to \infty$, the monopoles freeze out, and eq. (2.81) becomes the usual Wilson action.

Matter fields can be coupled to the gauge theory in the usual way. For example, if $\phi_i \in SU(N)$ is defined on sites (labeled by i), we may define

$$S_{\text{matter}} = -\gamma \sum_{\text{links}} \chi^{(R)} \left((\phi^{-1} U \phi)_l \right) \ + c.c. \ . \tag{2.84}$$

This is invariant under the local symmetry eq. (2.83) only if Z_N is contained in the kernel of the representation (R). Of course, this is just the Dirac quantization condition—the matter fields must be chosen so that the string of a monopole is invisible. We may introduce matter fields that are only invariant under some subgroup of Z_N, but then we must restrict the η_P's to take values in that subgroup.

Wilson Loop Operator Now we want to define a Wilson loop operator $W^{(\nu)}(C)$ that creates an electric flux tube on the loop C. But the usual construction

$$W^{(\nu)}(C) = \frac{1}{n_\nu}\chi^{(\nu)}\left(\prod_{l\in C} U_l\right) \tag{2.85}$$

is not invariant under the local symmetry eq. (2.83) unless Z_N is contained in the kernel of the representation (ν). The operator $W^{(\nu)}(C)$ is physically sensible, then, only if the string created by it has no Aharonov-Bohm interaction with a Z_N monopole. In fact, in the $SU(N)$ theory defined by eq. (2.81), this operator does not create a stable flux tube at all—even at strong coupling ($\beta << 1$), glue fluctuations cause the string to break.

An operator that creates a stable electric flux tube that *does* have Aharonov-Bohm interactions with monopoles cannot depend on the loop C alone; it must also depend on a surface Σ that is bounded by C. After our discussion of the 't Hooft operator in Section 4, this comes as no surprise. A (naive) Wilson loop operator, in the presence of monopoles, is a multi-valued object, for it acquires a non-trivial Aharonov-Bohm phase when the loop winds around a magnetic charge. To construct a single-valued object, we introduce a branch cut on the surface Σ, so that the operator jumps discontinuously when a monopole crosses Σ. We augment the naive Wilson loop operator, then, by a factor that counts the total magnetic flux of the Dirac strings that cross Σ, obtaining

$$W^{(\nu)}(C,\Sigma) \equiv \frac{1}{n_\nu}\chi^{(\nu)}\Big(\Big(\prod_{l\in C} U_l\Big)\Big(\prod_{P\in\Sigma}\eta_P\Big)\Big)\ . \tag{2.86}$$

This operator *is* invariant under eq. (2.83), for any representation (ν) of $SU(N)$, and of course it reduces to $W^{(\nu)}(C)$ if (ν) represents Z_N trivially.

We may also consider the degenerate case in which the loop C shrinks to a point. If (ν) represents Z_N faithfully, then the operator

$$G^{(\nu)}(\Sigma) = \frac{1}{n_\nu}\chi^{(\nu)}\left(\prod_{P\in\Sigma}\eta_P\right) \tag{2.87}$$

inserts the world sheet of a tube with minimal Z_N electric flux, as a classical source, on the surface Σ. (Of course, if (ν) represents Z_N trivially, then the flux tube is

invisible, and the operator is trivial.) If we think of the surface Σ as lying in a time slice, then $G(\Sigma)$ has another interpretation; it is a magnetic charge operator that measures the total Z_N magnetic flux through Σ. Obviously, $G(\Sigma)$ is the magnetic analog of the operator that we called $F(\Sigma)$ in Section 3.

Note also that if matter is introduced as in eq. (2.84), an operator can be constructed that creates a separated particle-antiparticle pair; it is

$$\chi^{(R)}\left(\phi_i^{-1}\left(\prod_{l\in P_{i,j}} U_l\right)\phi_j\right) , \tag{2.88}$$

where $P_{i,j}$ is a path from the site i to the site j. This operator is invariant under the local symmetry, if (R) represents Z_N trivially.

't Hooft Loop Operator We can also construct an 't Hooft loop operator $B_n(C)$; it inserts on the world line C a classical monopole source with Z_N magnetic charge $n = 1, 2, \ldots, N-1$. Alternatively, we may interpret $B_n(C)$, acting in a time slice, as an operator that creates a loop of magnetic flux tube on C.

The construction of this 't Hooft operator may be carried out in much the same way as in a gauge theory without dynamical monopoles. We regard C as a closed loop composed of links of the *dual* lattice, and we chose an arbitrary surface Σ (composed of plaquettes of the dual lattice) whose boundary is C. Dual to the plaquettes of Σ is a set Σ^* of plaquettes of the original lattice. Now, if there are no dynamical monopoles, we regard the η_P's as classical (non-fluctuating) variables. Then, to evaluate a Green function with an insertion of $B_n(C)$, we modify the plaquette action on the plaquettes in Σ^*, by choosing

$$\begin{aligned} \eta_P &= e^{2\pi i n/N} , \quad P \in \Sigma^* . \\ \eta_P &= \quad 1 \quad , \quad P \notin \Sigma^* . \end{aligned} \tag{2.89}$$

When there are dynamical monopoles, however, and the η_P's are in the configuration eq. (2.89), the cubes that are dual to the links of C are frustrated. Thus, this configuration represents a *dynamical* monopole propagating on the world line C. Since we want the operator $B_n(C)$ to introduce a *classical* monopole source, we should modify the cube action so that the configuration eq. (2.89) does not frustrate any

cubes. Thus, we evaluate a Green function with an insertion of $B_n(C)$ by changing the cube action according to

$$\eta_c \to e^{2\pi i n/N}\eta_c \ , \quad c \in C^* \ . \tag{2.90}$$

Equivalently, we have

$$B_n(C) = \prod_{c\in C^*} \exp\left(\lambda(e^{2\pi i n/N}\eta_c - \eta_c + c.c)\right) \ . \tag{2.91}$$

This operator is the magnetic analog of the Wilson loop operator.

In the weakly-coupled gauge theory without dynamical monopoles, the 't Hooft loop operator creates a stable magnetic flux tube, and $\langle B_n(C)\rangle$ exhibits area-law decay. But when there are dynamical monopoles, $B_n(C)$ always exhibits perimeter-law decay. The interpretation is clear. For any finite λ, a Z_N magnetic flux tube is unstable, for the tube can break via nucleation of a monopole-antimonopole pair.

Note that the operator defined by eq. (2.91) makes sense even if C is an open path rather than a closed loop. That is, we may define in like fashion an operator

$$B_n(P_{i,j}) = \prod_{c\in P^*_{i,j}} \exp\left(\lambda(e^{2\pi i n/N}\eta_c - \eta_c + c.c)\right) \ , \tag{2.92}$$

where i and j are sites of the dual lattice, and $P_{i,j}$ is a path connecting these sites. This operator creates a monopole-antimonopole pair, connected by a Dirac string; it can be used to compute the mass of a dynamical monopole. Obviously, it is closely analogous to the operator eq. (2.88).

Having now in hand the operator $B_n(C)$ that introduces a classical Z_N monopole on the world line C, and the operator $G^{(\nu)}(\Sigma)$ that introduces a classical Z_N electric flux tube on the world sheet Σ, we are ready to construct the operator, analogous to $A_n^{(\nu)}(\Sigma, C)$, that probes the Aharonov-Bohm interaction between monopoles and electric flux tubes; it is

$$E_n^{(\nu)}(\Sigma, C) = \frac{G^{(\nu)}(\Sigma)\ B_n(C)}{\langle G^{(\nu)}(\Sigma)\rangle\ \langle B_n(C)\rangle} \ . \tag{2.93}$$

If there is an infinite range Aharonov-Bohm interaction, the expectation value of this operator will have the asymptotic behavior

$$\lim \left\langle E_n^{(\nu)}\right\rangle = \frac{1}{n_\nu}\chi^{(\nu)}\left(\left(e^{2\pi i n/N}\right)^{k(\Sigma,C)}\right) \ , \tag{2.94}$$

where $k(\Sigma, C)$ is the linking number of Σ and C.

Finally, we remark that it is straightforward to generalize the construction in Section 4, and define an operator that creates a loop of *non-Abelian* cosmic string, in a theory that contains dynamical monopoles. We need only be cognizant of the change in the plaquette action that occurs when dynamical monopoles are included; a string that carries flux $a \in SU(N)$ is now created by

$$B_a(C,\Sigma,x_0) = \prod_{P\in\Sigma^*} \exp\Big\{\beta\Big(\mathrm{tr}(V_{l_P} a V_{l_P}^{-1}\eta_P U_P) - \mathrm{tr}(\eta_P U_P)\Big) + c.c.\Big\}\,, \qquad (2.95)$$

where Σ is a surface on the dual lattice, bounded by the loop C. When C shrinks to a point, we obtain the operator $F_a(\Sigma,x_0)$ that introduces a classical string source on the closed world sheet Σ.

Monopole Condensation A pure $SU(N)$ gauge theory, without dynamical monopoles, is confining at strong coupling. For $\beta << 1$, gauge field fluctuations are unsuppressed, and the resulting magnetic disorder gives rise to stable electric flux tubes. We want to explore how dynamical magnetic monopoles modify the physics of this theory.

First, we consider the parameter regime $\beta << 1$ and $\lambda << 1$, so that virtual monopoles are unsuppressed. It is easy to anticipate what will happen. A "monopole condensate" will form, which, in effect, will spontaneously break the local Z_N symmetry of the theory. Thus, the electric flux tube will become the boundary of a domain wall. As usual, this domain wall will decay by quantum tunneling—an electric flux tube will spontaneously nucleate, and expand, consuming the wall. Thus, there will be no stable electric flux tubes, and no infinite range Aharonov-Bohm interaction between flux tubes and monopoles.

We can check whether this expectation is correct in strong-coupling perturbation theory. We proceed by expanding $e^{-S_{\rm plaq}}$ in powers of β at each plaquette, and $e^{-S_{\rm cube}}$ in powers of λ at each cube. Roughly speaking, the terms that survive when the U_l's and η_P's are summed are ones such that a set of "tiled" cubes forms a closed (three-dimensional) hypersurface, or else the tiled cubes form an open hypersurface that is bounded by a (two-dimensional) surface of tiled plaquettes. In other words, strong-coupling perturbation theory can be interpreted as a sum over histories for (heavily suppressed) domain walls bounded by electric flux tubes.

Consider, now, the behavior of $G^{(f)}(\Sigma)$, where (f) denotes the defining representation of $SU(N)$. If $\beta = 0$, then the leading contribution to $\left\langle G^{(f)}(\Sigma)\right\rangle$, for $\lambda << 1$, is obtained by tiling the minimal hypersurface that is bounded by Σ; thus we have

$$\left\langle G^{(f)}(\Sigma)\right\rangle \sim (\lambda)^{\text{Volume}(\Sigma)} . \tag{2.96}$$

The interpretation is that $G^{(f)}(\Sigma)$ is the boundary of a domain wall, where the wall tension is $\sigma \sim -\epsilon^{-3} \ln(1/\lambda)$ (ϵ is the lattice spacing). But for $\beta > 0$, this domain wall is unstable. When the surface Σ is very large, a much larger contribution to $\left\langle G^{(f)}(\Sigma)\right\rangle$ is obtained by tiling the plaquettes of Σ; this contribution is

$$\left\langle G^{(f)}(\Sigma)\right\rangle \sim (\beta/N)^{\text{Area}(\Sigma)} . \tag{2.97}$$

(As $\beta \to 0$, the electric flux tube becomes arbitrarily heavy, and the domain wall is arbitrarily long-lived.)

Accordingly, the operator $W^{(f)}(C,\Sigma)$ does not create a stable electric flux tube. The leading behavior of its expectation value, too, is found by tiling the plaquettes of Σ, so that

$$\left\langle W^{(f)}(C,\Sigma)\right\rangle \sim (\beta/N)^{\text{Area}(\Sigma)} . \tag{2.98}$$

There is no dependence on the area of the minimal surface bounded by C (compare eq. (2.59)), signifying that the dynamical flux tube created by $W^{(f)}(C,\Sigma)$ has vanishing tension.

It is obvious that the leading contribution to $\left\langle G^{(f)}(\Sigma)\right\rangle$ is unaffected by an insertion of $B_n(C)$, even if C and Σ link. So we have

$$\lim \left\langle E_n^{(f)}(\Sigma, C)\right\rangle = 1 ; \tag{2.99}$$

there is no long-range Aharonov-Bohm interaction.

Incidentally, we could have chosen the cube action to be

$$S_c = -\lambda(\eta_c)^m + c.c . \tag{2.100}$$

Then, in effect, for $\lambda << 1$, charge-m monopoles condense; this breaks the local symmetry to Z_M, where M is the greatest common factor of N and m (compare Section 3). In other words, $W^{(\nu)}(C,\Sigma)$ *does* create a stable flux tube, if a source transforming as the representation (ν) has no Aharonov-Bohm interaction with a charge-m monopole, and if (ν) represents Z_M non-trivially.

Magnetic Hair Now we consider the regime $\beta << 1$ and $\lambda >> 1$. In this limit, virtual monopoles are heavily suppressed. We anticipate that quantum-mechanical magnetic charge fluctuations will *not* wipe out the infinite-range Aharonov-Bohm interaction between monopoles and electric flux tubes.

Again, we can check this expectation against perturbation theory. Weak-coupling perturbation theory, for $\lambda >> 1$ is an expansion in the number of frustrated cubes. The frustrated cubes form closed loops that we may interpret as the world lines of virtual monopoles. The expansion in β, as before, is an expansion in the number of tiled plaquettes. The tiled plaquettes form closed surfaces that we may interpret as the world sheets of electric flux tubes. Perturbation theory, then, is a sum over histories for (heavily suppressed) magnetic monopoles and electric flux tubes.

For example, consider $\left\langle G^{(f)}(\Sigma)\right\rangle$. The leading non-trivial contribution arises from a configuration such that (in a particular gauge) $\eta_P = e^{\pm 2\pi i/N}$ for a single plaquette P contained in Σ^*, while $\eta_P = 1$ for all other plaquettes. Flipping one plaquette frustrates four cubes, so this contribution gives

$$\left\langle G^{(f)}(\Sigma)\right\rangle \sim 1 + \sum_{n=\pm 1} \left(e^{2\pi i n/N} - 1\right) \left(\exp\left[-2\lambda\left(1 - \cos\left(\frac{2\pi}{N}\right)\right)\right]\right)^4 . \quad (2.101)$$

Summing disconnected contributions causes the result to exponentiate; we find

$$\left\langle G^{(f)}(\Sigma)\right\rangle \sim \exp\left(-\kappa^{(\mathrm{ren})}\mathrm{Area}(\Sigma)\right) , \quad (2.102)$$

where

$$\epsilon^2\kappa^{(\mathrm{ren})} \sim 2\left(1 - \cos(2\pi/N)\right) \exp\left(-8\lambda\left[1 - \cos\left(\frac{2\pi}{N}\right)\right]\right) \quad (2.103)$$

(and ϵ is the lattice spacing). The interpretation is that, because of the Aharonov-Bohm interaction between monopoles and flux tubes, inserting $G^{(f)}(\Sigma)$ modifies the contribution to the vacuum energy due to virtual monopole pairs that wind around Σ, resulting in a renormalization of the tension of the classical string source.

When we compute $\left\langle W^{(f)}(C,\Sigma)\right\rangle$, a similar renormalization of the tension of the classical string on Σ occurs. But in addition, for $\beta << 1$, the configurations that

contribute have a surface of tiled plaquettes bounded by C. The leading behavior, then, is

$$\left\langle W^{(f)}(C,\Sigma)\right\rangle \sim (\beta/N)^{\text{Area}(C)}\ \exp\left(-\kappa^{(\text{ren})}\text{Area}(\Sigma)\right)\ . \tag{2.104}$$

We conclude that $W^{(f)}(C,\Sigma)$ creates a stable electric flux tube with string tension

$$\epsilon^2\kappa^{(\text{dyn})} \sim -\ln(\beta/N)\ . \tag{2.105}$$

When the operator $B_n(C)$ is inserted, it tends to frustrate the cubes in C^*. But, as we already noted in subsection of 't Hooft loop operator, frustrated cubes can be avoided if the η_P's assume a suitable configuration. We may choose an arbitrary surface T on the dual lattice whose boundary is C. Dual to the plaquettes of T is a set of plaquettes T* of the original lattice. The desired configuration (in a particular gauge) is

$$\begin{aligned}\eta_P &= e^{2\pi i n/N}\ , \quad P\in \text{T}^*\ ,\\ \eta_P &= \ \ 1\ \ \ \ , \quad P\notin \text{T}^*\ .\end{aligned} \tag{2.106}$$

(The local symmetry transformation eq. (2.83) deforms the surface T, but leaves its boundary intact.)

By summing over gauge field fluctuations about the configuration eq. (2.106), we find the leading behavior

$$\langle B_n(C)\rangle \sim \exp\left(-M_n^{(\text{ren})}\ \text{Perimeter}(C)\right)\ , \tag{2.107}$$

where

$$\epsilon M_n^{(\text{ren})} \sim 2N^2(\beta/N)^6\left(1-\cos\left(\frac{2\pi n}{N}\right)\right) \tag{2.108}$$

is the renormalization of the mass of the classical monopole source. This renormalization is associated with virtual electric flux tubes whose world sheets link the world line of the monopole, and arises because of the Aharonov-Bohm interaction between monopole and flux tube.

When the operator $B_n(P_{i,j})$ is inserted, a line of frustrated cubes connecting i and j cannot be avoided, and so we find the leading behavior

$$\langle B_n(P_{i,j})\rangle \sim \exp\left(-M_n^{(\mathrm{dyn})}\ \mathrm{Distance}(i,j)\right)\ \exp\left(-M_n^{(\mathrm{ren})}\ \mathrm{Length}(P)\right)\ , \quad (2.109)$$

where

$$\epsilon M_n^{(\mathrm{dyn})} \sim 2\lambda\left(1-\cos\left(\frac{2\pi n}{N}\right)\right)\ . \quad (2.110)$$

Evidently, for $\lambda >> 1$ and $\beta << 1$, we have $M_n^{(\mathrm{dyn})} >> M_n^{(\mathrm{ren})}$. Indeed, $\langle B_n(C)\rangle$ is dominated by small fluctuations about the configuration eq. (2.106) for precisely this reason—the renormalization of the classical source is much less costly than screening the source with dynamical monopoles.

With the η_P's in the configuration eq. (2.106), the operator $G^{(f)}(\Sigma)$ assumes the value $\exp(2\pi ink/N)$, where k is the linking number of Σ and C. Furthermore, except on the loop C, this configuration is locally equivalent to the trivial configuration with $\eta_P = 1$ everywhere. Thus, as we expand in the small fluctuations about eq. (2.106), we find, to each order of the expansion,

$$\lim\ \left\langle E_n^{(f)}\right\rangle = \left(e^{2\pi in/N}\right)^{k(\Sigma,C)}\ . \quad (2.111)$$

We see that, at least to each order of perturbation theory, our expectation is confirmed. In a confining theory that contains weakly coupled dynamical magnetic monopoles, there is an infinite-range Aharonov-Bohm interaction between monopoles and electric flux tubes.

2.10. Appendix

In this appendix we consider in more detail some of the lattice perturbation theory calculations mentioned in the body of the text. Specifically, we calculate, in a pure gauge theory, the behavior of $\langle F_a(\Sigma, x_0)\rangle$ and $\langle B_a(\Sigma, C, x_0)\rangle$ in leading order in both strong and weak coupling perturbation theory. We demonstrate that $\langle F_a(\Sigma, x_0)\rangle$ exhibits an area law decay in both limits, and that only in the weak coupling limit does $\langle B_a(\Sigma, C, x_0)\rangle$ create a dynamical string on C. We also consider the problem that arises when one attempts to use the untraced Wilson loop operator eq. (2.48) to construct an Aharonov-Bohm order parameter.

Pure gauge theory: strong coupling We start with the calculation of $\langle F_a(\Sigma, x_0)\rangle$ and $\langle B_a(\Sigma, C, x_0)\rangle$ in the strong coupling regime ($\beta << 1$) of the pure gauge theory defined by the plaquette action eq. (2.44), namely

$$S^{(R)}_{\text{gauge},P} = -\beta\chi^{(R)}(U_P) + c.c. \tag{2.112}$$

For definiteness we will assume that the representation (R) that defines the theory is irreducible, and that it satisfies the constraints that $R \otimes R$ does not contain the trivial representation, while $R \otimes R^*$ contains it exactly once.

The expectation value $\langle F_a(\Sigma, x_0)\rangle$ is given by

$$\langle F_a(\Sigma, x_0)\rangle = \frac{\langle\langle F_a(\Sigma, x_0)\rangle\rangle}{\langle\langle 1\rangle\rangle}, \tag{2.113}$$

where the unnormalized expectation value of an operator X is defined by

$$\langle\langle X\rangle\rangle = \prod_{\langle ij\rangle}\int dU_{\langle ij\rangle}\Big(X\prod_P \exp(-S_P)\Big). \tag{2.114}$$

To find the renormalized string tension we need to calculate the leading behavior of both $\langle\langle F_a(\Sigma, x_0)\rangle\rangle$ and $\langle\langle 1\rangle\rangle$. The perturbation expansion in the strong coupling regime is of the form of a sum over closed surfaces, in which surfaces of greater area are suppressed by powers of β compared to smaller ones. Formally, the strong coupling expansion proceeds by performing a character expansion on the exponentiated Wilson action,

$$\exp\big(\beta(\chi^{(R)}(U_P) + \chi^{(R^*)}(U_P))\big) = N(\beta)\Big(1 + \sum_{(\nu)\neq 1} C^{(\nu)}(\beta)\chi^{(\nu)}(U_P)\Big). \tag{2.115}$$

The link integrations then select out closed surfaces S, each formed by "tiling" S with factors of $\chi^{(\nu)}(U_P)$ for each plaquette $P \in S$. The leading non-trivial contribution will be from the smallest surfaces, which furthermore are tiled with characters whose associated factors of $C^{(\nu)}(\beta)$ have the lowest non-trivial dependence on β. The factors $C^{(\nu)}(\beta)$ are found by multiplying eq. (2.115) by $\chi^{(\nu^*)}(U_P)$, summing U_P over the group, and using the character orthogonality relations along with the assumed

properties of the representation (R), giving

$$C^{(R)}(\beta) = 1 + \beta + O(\beta^2), \quad C^{(R^*)}(\beta) = 1 + \beta + O(\beta^2), \tag{2.116}$$

with all other irreducible representations acquiring their first non-trivial β dependence at $O(\beta^2)$. Therefore, the leading contribution will come from surfaces tiled with characters in the representations (R) and (R^*).

In the evaluation of $\langle F_a \rangle$, we must, in addition, identify the leading contribution that is sensitive to the presence of F_a, and hence does not cancel between $\langle\langle F_a \rangle\rangle$ and $\langle\langle 1 \rangle\rangle$. All closed surfaces that do not intersect Σ are unaffected by the presence of F_a. Therefore, we factorize both $\langle\langle F_a \rangle\rangle$ and $\langle\langle 1 \rangle\rangle$ in eq. (2.113) into contributions from surfaces "on- Σ" and "off-Σ," the off-Σ dependence canceling between them. The dominant on-Σ contribution is shown in Fig. 26, where all six faces are tiled with factors of either $\chi^{(R)}$, or $\chi^{(R^*)}$. This results in a contribution to $\langle\langle F_a \rangle\rangle$ from Fig. 26 of

$$2(\beta/n_R)^6 \left|\chi^{(R)}(a)\right|^2, \tag{2.117}$$

where n_R is the dimension of the representation (R). The factor of $(n_R)^{-6}$ comes from the six non-trivial link integrations, and the factor of 2 from the two equal contributions of either $\chi^{(R)}$ or $\chi^{(R^*)}$. Since the number of possible positions of this elementary cube on the surface Σ is given by the area $A(\Sigma)$ of Σ (measured in terms of the size of the elementary plaquette ϵ^2), to get the contribution of a single elementary cube located anywhere on Σ, we should further multiply (2.117) by $A(\Sigma)$.

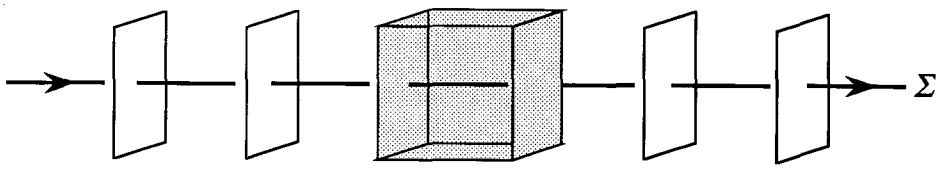

Fig. 26: The leading contribution to $\langle F_a(\Sigma, x_0) \rangle$, in the strong-coupling limit of the pure gauge theory, arises from tiling a cube (shaded) that contains two plaquettes of Σ^*.

The area law decay of $\langle F_a \rangle$ arises from considering an arbitrary number of such elementary tiled cubes on the surface Σ, rather than just a single contribution. In

the "dilute gas" approximation where we can ignore correlations between the various elementary cubes, this leads to an exponentiation of the above result with an overall factor of $A(\Sigma)$. Finally, the "on Σ" contributions to $\langle\langle 1\rangle\rangle$ lead to a similar exponential dependence on $A(\Sigma)$, except that the group element a is replaced by the identity element. Putting these together yields the result

$$\langle F_a(\Sigma, x_0)\rangle = \exp\left(2(\beta/n_R)^6\left(|\chi^{(R)}(a)|^2 - n_R^2\right)A(\Sigma)\right). \tag{2.118}$$

In other words, the renormalized classical string tension defined in eq. (2.54) is given by

$$\epsilon^2\kappa_a^{(ren)} \sim 2\left(\frac{\beta}{n_R}\right)^6\left(n_R^2 - \left|\chi^{(R)}(a)\right|^2\right) \tag{2.119}$$

to lowest order in β.

The calculation of $\langle B_a(\Sigma, C, x_0)\rangle$ proceeds in a very similar way; the only question concerns the behavior at the boundary curve C. Here there exist contributions to $\langle B_a(\Sigma, C, x_0)\rangle$ similar to those of Fig. 26 except only one plaquette $P \in \Sigma^*$ is contained in the cube. This leads to an additional decay of $\langle B_a(\Sigma, C, x_0)\rangle$ depending on the length $P(C)$ of the perimeter (eq. (2.60)),

$$\langle B_a(\Sigma, C, x_0)\rangle \sim \exp\left(-\kappa_a^{(ren)}A(\Sigma)\right)\exp\left(-m_a^{(ren)}P(C)\right), \tag{2.120}$$

with

$$\epsilon m_a^{(ren)} \sim 2\left(\frac{\beta}{n_R}\right)^6\left(n_R^2 - n_R\mathrm{Re}\,\chi^{(R)}(a)\right). \tag{2.121}$$

Since the configurations that dominate the expectation values of F_a and B_a do not extend deep into the volume enclosed by Σ, we find that there are no stable dynamical strings (or vortices) in this regime and, therefore, that there is no long-distance Aharonov-Bohm interaction between flux tubes and charges.

Weak coupling

This is the regime in which we expect to find $A_a^{(\nu)}(\Sigma, x_0; C)$ displaying a non-trivial Aharonov-Bohm interaction. To start we need to discover which configurations dominate the behavior of $\langle F_a(\Sigma, x_0)\rangle$. When $\beta >> 1$, frustrated plaquettes (ones

with non-minimal gauge action) are highly suppressed. However, as described in Section 5, an insertion of $F_a(\Sigma, x_0)$ tends to frustrate the plaquettes in Σ^*, unless the link variables assume a suitable configuration. The configuration that leaves no frustrated plaquettes, and thus acts as the "ground state" in the presence of F_a is illustrated in Fig. 9. It consists of a "forest" of links inside Σ that have value $U_l = a$. Specifically, we choose a three-dimensional hypersurface Ω made up of cubes of the dual lattice that has boundary Σ. Dual to these cubes is a set of links Ω^* on the original lattice. The configuration with no excited plaquettes is then given by eq. (2.57), namely $U_l = a$ for $l \in \Omega^*$, and $U_l = e$ otherwise. The area law decay of $\langle F_a(\Sigma, x_0)\rangle$ in the weak coupling regime arises from fluctuations of the link variables around this state, which are sensitive to the presence of the non-trivial background.

The leading such contribution to the decay of $\langle F_a(\Sigma, x_0)\rangle$ in the pure gauge theory is shown in Fig. 27a. (For simplicity we discuss the situation for vortices in three Euclidean dimensions.) It consists of a configuration that has nine excited plaquettes, and is constructed by considering two neighboring plaquettes in Σ^*. These plaquettes are connected with four links with $U_l = b$ to make a cube, with b summed over the group. In this configuration, there are eight excited plaquettes with plaquette action proportional to $\beta \mathrm{Re}\,\chi(b)$ (the unshaded plaquettes in Fig. 27b). But, in addition, there is an excited plaquette (if a and b do not commute) with action proportional to $\beta \mathrm{Re}\,\chi(aba^{-1}b^{-1})$ (the shaded plaquette in Fig. 27b). This extra excitation distinguishes the contributions of Fig. 27 to $\langle\langle F_a(\Sigma, x_0)\rangle\rangle$ and $\langle\langle 1\rangle\rangle$. Changes of variable can move the ninth excited plaquette around, but cannot get rid of it.

Of course, the process illustrated in Fig. 27 is only one of a set of similar processes where instead of summing over the same value b for each of the four connecting links, we sum independently over their values b_i. These differ from the contribution of Fig. 27 in two ways: there will either be a greater number of excited plaquettes leading to a subdominant contribution, or, if some of the b_i are taken to be the indentity, changes of variable on the remaining, independent b_i show that the a-links have no physical effect. This is why, for instance, we do not consider the less suppressed, and apparently non-trivial, contribution arising from the excitation of just two joining links.

We can understand this in physical terms. The contribution to $\langle F_a(\Sigma, x_0)\rangle$ reflects

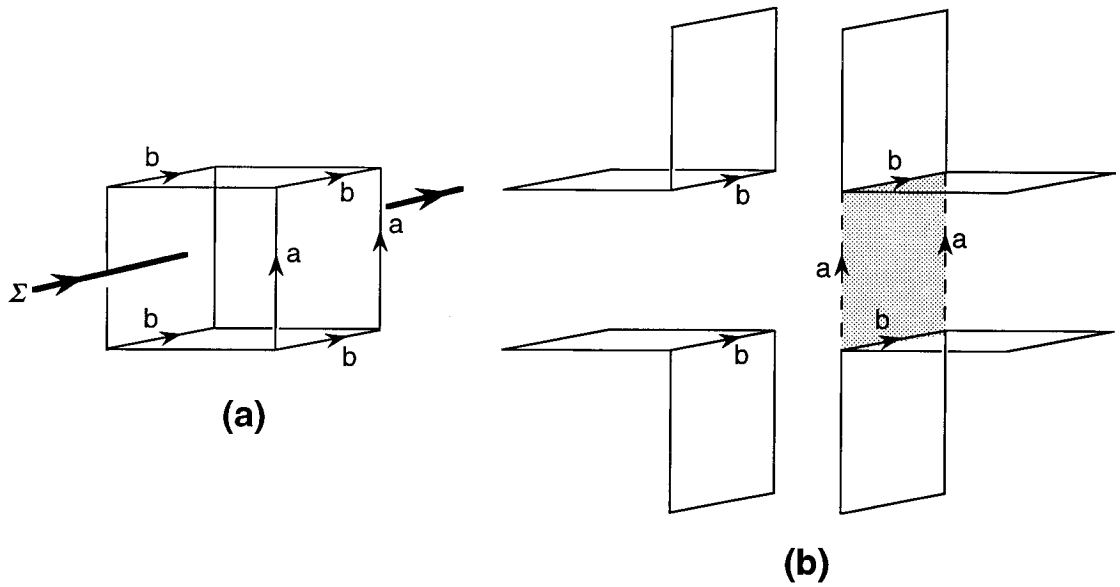

Fig. 27: The leading contribution to $\langle F_a(\Sigma, x_0)\rangle$, in the weak-coupling limit of the pure gauge theory. The cube shown contains two plaquettes that are in Σ^*, and the link variables assume the indicated values. (Unmarked links have the value $U_l = 1$.) There are altogether nine excited plaquettes—eight (unshaded) plaquettes with $U_P = b$ and one (shaded) plaquette with $U_P = aba^{-1}b^{-1}$.

how the non-trivial boundary condition introduced by the vortex (or string) affects the quantum fluctuations of the gauge fields. To see the effect, we must consider processes in which virtual gauge field excitations (with gauge quantum numbers that Aharonov-Bohm scatter off the inserted string of flux a) propagate around the string. The process illustrated in Fig. 27 is the leading one of this type.

The end result of these considerations is a renormalization of the classical string tension (or vortex mass in three dimensions) leading to an area law (respectively, perimeter law) decay of $\langle F_a(\Sigma, x_0)\rangle$ as in eq. (2.54). An estimate of the renormalized vortex mass in the weak-coupling regime arising from Fig. 27 is

$$\epsilon m_a^{(ren)} \sim \frac{1}{n_G}\sum_{b\neq 1}\Big(\exp\{(-16\beta)[n_R - \mathrm{Re}\,\chi^{(R)}(b)]\}\times \Big(1-\exp\{(-2\beta)[n_R - \mathrm{Re}\,\chi^{(R)}(bab^{-1}a^{-1})]\}\Big)\Big), \tag{2.122}$$

where we require more information about the group G to explicitly evaluate the sum.

Also, note that we have found area law decay of $\langle F_a(\Sigma, x_0)\rangle$ in both regimes of the pure non-Abelian gauge theory described by eq. (2.112) (when a is not in the

center of G). This differs from the situation in a pure Abelian gauge theory where a change of variables shows that $\langle F_a(\Sigma, x_0)\rangle \equiv 1$.[5] This is to be expected since the pure non-Abelian gauge theory contains excitations that, via the Aharonov-Bohm effect, are sensitive to the introduction of flux on Σ.

To see that there exist stable dynamical strings in this regime of the pure gauge theory, we must consider $\langle B_a(\Sigma, C, x_0)\rangle$, rather than $\langle F_a(\Sigma, x_0)\rangle$. Since the surface Σ now has a boundary, the three-dimensional hypersurface Ω must also end on a two-dimensional surface S. This surface is dual to a set of plaquettes S^* on the original lattice that must be frustrated, since only one of their links is contained within Ω^*. Clearly the dominant configuration is the one in which S is the minimal area surface with boundary C. Each such plaquette is suppressed relative to the corresponding vacuum configuration that dominates $\langle\langle 1\rangle\rangle$ by an amount

$$\exp\bigl(2\beta(\mathrm{Re}\,\chi^{(R)}(a) - n_R)\bigr). \tag{2.123}$$

In other words $\langle B_a(\Sigma, C, x_0)\rangle$ behaves as in eq. (2.59)

$$\langle B_a(\Sigma, C, x_0)\rangle = \exp\bigl(-\kappa_a^{(\mathrm{ren})} A(\Sigma)\bigr)\exp\bigl(-\kappa_a^{(\mathrm{dyn})} A(C)\bigr), \tag{2.124}$$

where $A(C)$ is the area of the minimal surface S, and $\kappa_a^{(\mathrm{dyn})}$ is the dynamical string tension of the stable string with flux in the a conjugacy class,

$$\epsilon^2 \kappa_a^{(\mathrm{dyn})} \sim 2\beta\left(n_R - \mathrm{Re}\,\chi^{(R)}(a)\right). \tag{2.125}$$

In some situations it is possible that $B_a(\Sigma, C, x_0)$ creates not just one dynamical string, of flux a, on C, but instead creates a number of strings of varying fluxes $a_1, \ldots, a_k$, where $a_1 a_2 \ldots a_k = a$. The corresponding link configuration would be a "forest state" of links of value $U_l = a$ that terminates gradually in a number of steps rather than dropping to the indentity across a single plaquette. This is the more favorable configuration if the total area law suppression of the multiple strings is less than that for the single string. This translates into the condition (in the pure gauge case, and ignoring perimeter corrections that vanish in the infinite size limit)

$$(k-1)n_R + \mathrm{Re}\,\chi^{(R)}(a) - \sum_{i=1}^{k} \mathrm{Re}\,\chi^{(R)}(a_i) < 0\ . \tag{2.126}$$

As an example of this, consider the expectation value, in a pure Z_6 gauge theory, of the operator $B_3(\Sigma, C, x_0)$ that attempts to create a dynamical string with flux $a = \omega^3$

on C (where $\omega = \exp(2\pi i/6)$). The dominant process in this case is not the creation of an ω^3 string, but instead that of three ω strings, since, substituting into eq. (2.126), we find $2 + (-1) - 3(1/2) < 0$.

We now turn to the order parameter $A_a^{(\nu)}(\Sigma, x_0; C)$, which is defined (eq. (2.36)) in terms of $F_a(\Sigma)$ and the Wilson loop operator $W^{(\nu)}(C)$. $F_a(\Sigma)$ has been described above, so we will now discuss the Wilson loop operator, and demonstrate eq. (2.37).

First, let us consider the untraced Wilson operator $U^{(\nu)}(C, x_0)$ (eq. (2.48)). It is easy to calculate the leading behavior of the expectation value of $U^{(\nu)}(C, x_0)$ at weak coupling: as above we factorize the contributions from links on-C and off-C, and cancel the off-C contributions,

$$\frac{\langle\langle U^{(\nu)}(C, x_0)\rangle\rangle}{\langle\langle 1\rangle\rangle} = \frac{\sum_{U_1}\cdots\sum_{U_L} D^{(\nu)}(U_1\cdots U_L)\prod_{i=1}^{L}\exp\Big(4D\beta\mathrm{Re}\,\chi^{(R)}(U_i)\Big)}{\sum_{U_1}\cdots\sum_{U_L}\prod_{i=1}^{L}\exp\Big(4D\beta\mathrm{Re}\,\chi^{(R)}(U_i)\Big)}\,, \tag{2.127}$$

where we work in $D+1$ Euclidean dimensions, and U_i for $i = 1\ldots L$ are the links in C. (In eq. (2.127), we sum over all values of the link variables contained in C, but keep all other link variables fixed at $U_l = 1$.) Because we have not traced the Wilson loop, the multiple summations factorize. We then use

$$\frac{1}{n_G}\sum_{g\in G} D^{(\nu^*)}(g)\chi^{(\mu)}(g) = \frac{1}{n_\nu}\delta^{\mu,\nu} I^{(\nu)}, \tag{2.128}$$

where $I^{(\nu)}$ is the identity matrix. Along with eq. (2.115) this implies that

$$\langle U^{(\nu)}(C, x_0)\rangle = \left(\frac{C^{(\nu^*)}(2D\beta)}{n_\nu}\right)^L I^{(\nu)}. \tag{2.129}$$

If we repeat the calculation in the background created by a string insertion operator $F_a(\Sigma, x_0)$ then the only difference is that for each time C links Σ there is a link in C that is multiplied by a. Neither $F_a(\Sigma, x_0)$ nor $U^{(\nu)}(C, x_0)$ is gauge-invariant, but a gauge transformation conjugates both a and $U^{(\nu)}$ by the value of the transformation at x_0, so these conjugations cancel, and the gauge-invariant result is that $I^{(\nu)}$ is replaced by $D^{(\nu)}(a^{k(\Sigma,C)})$, where $k(\Sigma, C)$ is the linking number of the two curves. This proves eq. (2.49), and by taking a trace it proves eq. (2.37).

However, there is a major caveat that limits the usefulness of the untraced Wilson loop: there are corrections to eq. (2.49) that do not go to zero as C and Σ become large and well separated. Recall that $F_a(\Sigma, x_0)$ corresponds to a thought experiment in which a specific group element a (not just a conjugacy class) was associated with a string, using an arbitrary basis at a point x_0. As described in Section 4 this means that in defining $F_a(\Sigma, x_0)$ we give each plaquette in Σ^* a long "tail" of links that stretches back to x_0. The problem manifests itself (in a particular gauge) when we take into account configurations in which some (but not all) of the links coming out of x_0 are excited. In these configurations the inserted flux may be conjugated, but not the measured group element (or vice versa).

Concretely, the leading contribution, in weak coupling perturbation theory, that causes problems for the untraced Wilson loop is simply the excitation of a single link $U_l = b$ on the path that connects x_0 to Σ. In the three-dimensional case this leads to the excitation of four plaquettes, and is thus suppressed relative to the ground state configuration by a factor of $\exp(8\beta(\mathrm{Re}\,\chi^{(R)}(b) - n_R))$. However, the action of these configurations, though large at weak coupling, is completely independent of the size and separation of Σ and C, and so remains finite as Σ and C become large and far apart. Their effect on the untraced order parameter is to conjugate the inserted flux (and thus the flux measured by $U^{(\nu)}(C, x_0)$). Therefore, there are corrections to eq. (2.49) that occur at finite order in weak coupling perturbation theory that render the measured flux uncertain up to conjugation.

The physical interpretation of these configurations is that the path connecting x_0 to Σ is linked by a virtual vortex-antivortex pair of flux b. In four dimensions there exist analogous processes where a virtual cosmic string links the connecting path. These effects do not correct the leading behavior of the ABOP given by eq. (2.37), because the traced Wilson loop is gauge-invariant, and so is not sensitive to conjugation of the inserted flux.

REFERENCES

1. See, for example: J. Preskill, "Vortices and Monopoles," in *Architecture of the Fundamental Interactions at Short Distances*, ed. P. Ramond and R. Stora (North-Holland, Amsterdam, 1987).
2. Aharonov, Y. and Bohm, D., Phys. Rev. **119** (1959) 485.
3. R. Rohm, Princeton University Ph.D. Thesis (unpublished) (1985); M. G. Alford and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1071.
4. L. Krauss and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1221.
5. J. Preskill and L. Krauss, Nucl. Phys. **B341** (1990) 50.
6. M. G. Alford, J. March-Russell and F. Wilczek, Nucl. Phys. **B337** (1990) 695.
7. M. J. Bowick *et al.*, Phys. Rev. Lett. **61** (1988) 2823.
8. J. Preskill, Phys. Scrip. **T36** (1991) 258.
9. S. Coleman, J. Preskill, and F. Wilczek, Mod. Phys. Lett. **A6** (1991) 1631; Phys. Rev. Lett. **67** (1991) 1975; Nucl. Phys. **B378** (1992) 175.
10. Early discussions of non-Abelian strings appear in F. Bais, Nucl. Phys. **B170**(FS1) (1980) 32; T. W. B. Kibble, Phys. Rep. **67** (1980) 183.
11. A.S. Schwarz, Nucl. Phys. **B208** (1982) 141.
12. S. Coleman and P. Ginsparg, (1983), unpublished.
13. M. Alford *et al.*, Phys. Rev. Lett. **64** (1990) 1632; **65** (1990) 668 (E); Nucl. Phys. **B349** (1991) 414.
14. An earlier discussion is contained in M. G. Alford and J. March-Russell, Nucl. Phys. **B369** (1992) 276.
15. K. Wilson, Phys. Rev. **D10** (1974) 2445.
16. G. 't Hooft, Nucl. Phys. **B138** (1978) 1.
17. K. Fredenhagen and M. Marcu, Comm. Math. Phys. **92** (1983) 81.
18. K. Fredenhagen, in *Fundamental Problems of Gauge Field Theory*, ed. G. Velo and A. S. Wightman (Plenum, New York, 1986).

19. E. Fradkin and S. Shenker, Phys. Rev. **D19** (1979) 3682; T. Banks and E. Rabinovici, Nucl. Phys. **B160** (1979) 349; S. Dimopoulos, S. Raby, and L. Susskind, Nucl. Phys. **B173** (1980) 208.

20. K. Fredenhagen and M. Marcu, Phys. Rev. Lett. **56** (1986) 223.

21. M. Alford, S. Coleman and J. March-Russell, Nucl. Phys. **B351** (1991) 735.

22. S. Coleman, "The Magnetic Monopole Fifty Years Later," in *The Unity of the Fundamental Interactions*, ed. A. Zichichi (Plenum, New York, 1983).

23. M. Srednicki and L. Susskind, Nucl. Phys. **B179** (1981) 239.

24. M. Bucher, Nucl. Phys. **B350** (1991) 163.

25. F. Wilczek and Y.-S. Wu, Phys. Rev. Lett. **65** (1990) 13.

26. The first formulation of local discrete symmetries in the continuum was given by J. Kiskis, Phys. Rev. **D17** (1978) 3196, although not in the context of spontaneously broken gauge theories.

27. T. Banks, Nucl. Phys. **B323** (1989) 90.

28. K. Li, Nucl. Phys. **B361** (1991) 437.

29. M. G. Alford and J. March-Russell, Int. J. Mod. Phys. **B5** (1991) 2641.

30. M. Bucher, K.-M. Lee, and J. Preskill, Nucl. Phys. **B386** (1992) 27.

31. V. Poénaru and G. Toulouse, J. Phys. (Paris) **38** (1977) 887; N. D. Mermin, Rev. Mod. Phys. **51** (1979) 591.

32. T. W. B. Kibble, G. Lazerides, and Q. Shafi, Phys. Rev. **D26** (1982) 435.

33. G. Mack and V. Petkova, Ann. Phys. **123** (1979) 442.

34. A. Ukawa, P. Windey, and A. Guth, Phys. Rev. **D21** (1980) 1013.

Chapter 3: Detecting Cheshire Charge

In a spontaneously broken gauge theory, if the unbroken gauge group H is a *discrete* subgroup of the underlying continuous gauge group G, then the theory will contain topologically stable strings (in 3+1 dimensions) or vortices (in 2+1 dimensions). If H is non-Abelian, the strings have remarkable properties. In particular, a closed loop of string can carry a non-trivial H charge. Oddly, this charge is a global property of the string that cannot be attributed to any locally defined charge density. Yet the charge is physically detectable, for the charged string loop has an infinite range Aharonov-Bohm interaction with other strings. Furthermore, if a pointlike particle carrying H charge winds through a string loop, the particle and the loop can exchange charge.

Charge with no localized source has been called "Cheshire charge." [1] It was first discussed for the case of the "Alice" string.[2] A loop of Alice string can carry electric charge, and have a long-range electric field, even though the electric charge density vanishes everywhere.[3,1,4] Processes in which electric (or magnetic) charge is exchanged between string loops and point particles were discussed in Ref. 1, 4-6.

In this chapter, we analyze the purely quantum-mechanical version of Cheshire charge[7,4] that arises in a theory with a non-Abelian discrete local H symmetry.[8,9] The semiclassical theory of discrete Cheshire charge was formulated in Ref. 4 and elaborated in Ref. 10, 11. Here we extend the theory further, in several respects. We describe how a charge operator can be constructed, such that the expectation value of the operator in a state specifies the transformation properties of the state under global H transformations. We then study processes in which charge is exchanged between string loops and point particles, and derive general formulas for how the expectation value of the charge of the loop is altered by the exchange. Finally, we explain how the charge exchange processes can be probed using gauge-invariant correlation functions.

The charge operator and correlation functions are also treated in Ref. 12, where lattice realizations of operators and correlators are extensively discussed.

The rest of this chapter is organized as follows: In Section 1, we briefly review the basic properties of non-Abelian strings and the concept of Cheshire charge. We con-

struct the non-Abelian charge operator in Section 2, and analyze the charge exchange process in Section 3. Section 4 contains a final comment.

3.1. Non-Abelian strings

Let us briefly recall some of the properties of non-Abelian strings in three spatial dimensions (and vortices in two spatial dimensions).

If a simply connected gauge group G is broken to a discrete subgroup H, then strings are classified by elements of H. To assign a group element to a loop of string, we fix an (arbitrary) basepoint x_0, and specify a path C, beginning and ending at x_0, that winds once through the string loop. (See Fig. 1.) The assigned group element is then

$$a(C, x_0) = P \ \exp\left(i \int\limits_{C, x_0} A \cdot dx \right) \ . \tag{3.1}$$

We refer to $a(C, x_0)$ as the "flux" of the string; it encodes the effect of parallel transport around the path C. The flux takes values in $H(x_0)$, the subgroup of G that stabilizes the Higgs condensate at the point x_0 (since parallel transport around C must return the condensate to its original value). Since H is discrete, the flux $a(C, x_0)$ is unchanged by deformations of C that leave x_0 fixed, as long as C never crosses the core of the string.

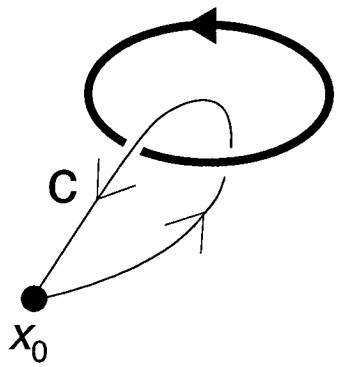

Fig. 1: The path C, starting and ending at the point x_0, encircles a loop of string.

For a configuration of many string loops, we specify a standard path for each loop, where all paths have the same basepoint. Evidently, the flux associated with the product path $C_2 \circ C_1$ obtained by traversing first C_1 and then C_2 is just the product $a(C_2, x_0) \cdot a(C_1, x_0)$ of the two fluxes associated with C_1 and C_2. Thus, $a(C, x_0)$ defines a homomorphism that maps $\pi_1(\mathcal{M}, x_0)$ to H, where $\mathcal{M}$ is the manifold that is obtained when the cores of all strings are removed from R^3.

The flux assigned to a path is not gauge invariant. The gauge transformations at the basepoint x_0 that preserve the condensate at the basepoint, and so preserve the embedding of H in G, take values in $H(x_0)$. Under such a gauge transformation $h \in H(x_0)$, the flux transforms as

$$a(C, x_0) \to h\ a(C, x_0)\ h^{-1}\ . \tag{3.2}$$

In a many-string configuration, the flux of each string becomes conjugated by h.

In the presence of strings, the embedding of the unbroken group H in G necessarily depends on the spatial position x. If the strings are non-Abelian, this position dependence is described by a non-trivial fiber bundle. The base space of the bundle is the spatial manifold $\mathcal{M}$, the fiber is H, and the structure group is also H, which acts on the fiber by conjugation. The bundle is twisted: Upon transport around the path C, the group element $h \in H(x_0)$ becomes conjugated by $a(C, x_0)$. This twist prevents the bundle from being smoothly deformed to the trivial bundle $\mathcal{M} \times \mathcal{H}$. One thus says that the unbroken H symmetry is not "globally realizable";[13,1,4] there is no smooth function of position that describes how the unbroken group is embedded in G. Only the subgroup of H that commutes with the flux of all strings is globally realizable on $\mathcal{M}$.

To define the H-charge of a state, we will want to consider how the state transforms under global H transformations. Fortunately, these global gauge transformations can be implemented, even though a topological obstruction prevents H from being globally realized. The point is that it is sufficient to be able to define an H transformation on and outside a large surface Σ (homeomorphic to S^2) that encloses all of the string loops. The transformation cannot be smoothly extended inside the sphere if it is required to take values in $H(x)$. However, one may relax this requirement and allow the gauge transformation to take values in G inside of Σ; then a

smooth extension is possible. It makes no difference what extension is chosen, for gauge transformations of compact support act trivially on physical states. (In two spatial dimensions, the only global H transformations that can be implemented are those that commute with the *total* flux; *i.e.*, the flux associated with a path that encloses all of the vortices.)

If the basepoint x_0 lies outside the surface Σ, then, under the global gauge transformation $h \in H$, the flux of a string transforms as in eq. (3.2). Thus, the H representations mix up the string loop state labeled by $a \in H$ with string loop states labeled by other group elements in the same conjugacy class as a. Let $[a]$ denote the conjugacy class that contains a. The action of H on the members of the class $[a]$ defines a (reducible) representation that we denote as $D^{([a])}$. In $D^{([a])}$, each element of H is represented by a permutation of the class, according to

$$D^{([a])}(h): \ |a'\rangle \to |ha'h^{-1}\rangle \ , \quad a' \in [a] \ . \tag{3.3}$$

This representation can be decomposed into irreducible representations of H. For each class $[a]$ there is a unique state that can be constructed that transforms trivially under H; it is the superposition of flux eigenstates

$$|0;[a]\rangle = \frac{1}{\sqrt{n_{[a]}}} \sum_{a' \in [a]} |a'\rangle \ , \tag{3.4}$$

where $n_{[a]}$ denotes the order of the class. The other states contained in the decomposition of $D^{([a])}$ carry H-charge. This is "discrete Cheshire charge," for the charge of the loop has no localized source. (Note that the charged string states transform trivially under the *center* of H, since $D^{([a])}$ represents the center trivially.)

The splitting between the charge-0 string state eq. (3.4) and the lowest charge excitation of the string is of order $\exp(-\kappa A)$, where κ is a string tension, and A is the area of the string loop.[11,12] It is a remarkable property of Cheshire charge that, in the presence of a large string loop, the gap between the ground state and the first charged excitation is much less than the corresponding gap when the string is absent. Indeed, the gap approaches zero very rapidly as the size of the loop increases.

The above discussion of the discrete Cheshire charge carried by a string loop also applies, in two spatial dimensions, to the charge carried by a vortex pair, if the total

flux of the pair is trivial. In fact, in two dimensions, there is a somewhat more general notion of Cheshire charge. We may consider a pair of vortices with flux a_1 and a_2, such that the total flux $a_{\rm tot} = a_1 a_2$ is a non-trivial element of H. Then the global gauge transformations that can be implemented belong to the normalizer $N(a_{\rm tot})$, the subgroup of H that commutes with the total flux. Unless a_1 and a_2 commute with all elements of $N(a_{\rm tot})$, the two-vortex states will transform as a non-trivial representation of $N(a_{\rm tot})$, which can be decomposed into its irreducible components. Thus, there are two-vortex states that carry non-trivial $N(a_{\rm tot})$-charge.

3.2. Charge operator

The discrete charge of an object, including a charged string loop, can be detected at long range by means of the Aharonov-Bohm effect.[14] Let $|u\rangle$ denote the wave-function in internal-symmetry space of an object located at x_0 that transforms as the irreducible representation $D^{(\nu)}$ of H. Then when the particle is transported around the closed path C that begins and ends at x_0, the wave-function is modified according to

$$|u\rangle \to D^{(\nu)}\left[a(C,x_0)\right]\ |u\rangle\ ; \tag{3.5}$$

if the string is in the flux eigenstate $|a\rangle$, then the Aharonov-Bohm phase that can be measured in an interference experiment is

$$\langle u|\, D^{(\nu)}(a)\, |u\rangle\ . \tag{3.6}$$

But if the string is in the charge-zero eigenstate $|0;[a]\rangle$ given by eq. (3.4), then the expectation value of the "phase" $D^{(\nu)}(a)$ becomes

$$\frac{1}{n_{[a]}}\sum_{a'\in[a]} D^{(\nu)}(a') = \frac{1}{n_H}\sum_{h\in H} D^{(\nu)}(hah^{-1}) = \frac{1}{n_\nu}\chi^{(\nu)}(a)\ \mathbf{1}\ , \tag{3.7}$$

where n_H is the order of the group, n_ν is the dimension of $D^{(\nu)}$, and $\chi^{(\nu)}$ is the character of the representation. The second equality follows from Schur's lemma.

In principle, the charge inside a large region can be measured by means of a process in which the world sheet of a string sweeps over the boundary of the region.

If the string is in the charge-zero eigenstate $|0;[a]\rangle$, and the object enclosed by the world sheet transforms as the irreducible representation (ν) of H, then the amplitude for this process will be weighted by the Aharonov-Bohm factor $(1/n_\nu)\chi^{(\nu)}(a)$. The charge (ν) of an unidentified object can be determined by measuring this factor for each class $[a]$.

A gauge-invariant operator $F_{[a]}(\Sigma)$ can be constructed that inserts, as a classical source, a string world sheet in the state $|0;[a]\rangle$ on the closed surface Σ. The realization of this operator in a Euclidean lattice gauge theory was described in Ref. 4 in the case where H is Abelian (see also Ref. 15), and in Ref. 16, 12 for H non-Abelian. (It is closely related to the 't Hooft loop operator.[17]) If the surface Σ is chosen to lie in a time slice, then the operator $F_{[a]}(\Sigma)$ measures the charge enclosed by Σ. To define the charge of an isolated object, we consider a surface Σ that encloses the object, and whose closest approach to the object is large compared to the correlation length of the theory. Let $|\psi\rangle$ denote the quantum state of the object. Then we have

$$\frac{\langle\psi|\, F_{[a]}(\Sigma)\,|\psi\rangle}{\langle F_{[a]}(\Sigma)\rangle_0} = \sum_\nu p^{(\nu)}(\psi;\Sigma)\,\frac{1}{n_\nu}\chi^{(\nu)}(a)\ , \tag{3.8}$$

where $p^{(\nu)}(\psi;\Sigma)$ is the probability that the object carries charge (ν). By measuring $F_{[a]}(\Sigma)$ for each class, we can determine all of the $p^{(\nu)}$'s. (It is necessary to divide by the vacuum expectation value $\langle F_{[a]}(\Sigma)\rangle_0$ to remove the effects of quantum-mechanical vacuum charge fluctuations near the surface Σ.[4])

The Aharonov-Bohm interaction makes it possible to detect H-charge at arbitrarily long range. Thus, a theory with discrete local H symmetry obeys a charge superselection rule—no gauge-invariant local operator can create or destroy H-charge. We have

$$\langle(\mu)|\,\mathcal{O}\,|(\nu)\rangle = 0\ , \quad (\mu)\neq(\nu)\ , \tag{3.9}$$

where $\mathcal{O}$ is any local observable, and $|(\nu)\rangle$ denotes a state that transforms as the irreducible representation (ν) of H. We can construct a projection operator that projects out a given superselection sector of the Hilbert space. It is

$$P^{(\nu)} = \frac{n_\nu}{n_H}\sum_{a\in H}\chi^{(\nu)}(a)^*\,U(a)\ , \tag{3.10}$$

where $U(a)$ represents the global H transformation $a\in H$ acting on the Hilbert space.

This projection operator can be expressed in terms of the operators $F_{[a]}(\Sigma)$, for it follows from eq. (3.8) that

$$\frac{F_{[a]}(\Sigma)}{\langle F_{[a]}(\Sigma)\rangle_0} \longrightarrow \frac{1}{n_{[a]}} \sum_{a' \in [a]} U(a') \ , \tag{3.11}$$

as the surface Σ approaches the surface at spatial infinity.

We can also use the operator $F_{[a]}$ to construct an "Aharonov-Bohm Order Parameter" (ABOP) that probes whether non-trivial superselection sectors actually exist. Let

$$W^{(\nu)}(C) \equiv \frac{1}{n_\nu} \chi^{(\nu)} \left[P \ \exp \left(i \int_C A \cdot dx \right) \right] \tag{3.12}$$

denote the Wilson loop operator in the irreducible representation (ν). This operator introduces a classical source with charge (ν) propagating on the world line C. The ABOP is defined by

$$A^{(\nu)}_{[a]}(\Sigma, C) \equiv \frac{F_{[a]}(\Sigma)\ W^{(\nu)}(C)}{\langle F_{[a]}(\Sigma)\rangle_0 \langle W^{(\nu)}(C)\rangle_0} \ . \tag{3.13}$$

If H quantum numbers can indeed be detected at infinite range, then we expect that

$$\left\langle A^{(\nu)}_{[a]}(\Sigma, C) \right\rangle_0 \longrightarrow \frac{1}{n_\nu} \chi^{(\nu)}(a^{k(\Sigma,C)}) \ , \tag{3.14}$$

in the limit in which Σ and C increase to infinite size, with the closest approach of Σ to C also approaching infinity. Here $k(\Sigma, C)$ is the linking number of the surface Σ and the loop C. (In the Abelian case, the ABOP was first described in Ref. 15, and was further elaborated in Ref. 4. The non-Abelian generalization was introduced in Ref. 16, and its properties were extensively discussed in Ref. 12.)

In two spatial dimensions, the classification of superselection sectors is richer than in three dimensions.[4,12] The Aharonov-Bohm interaction makes it possible to measure the magnetic flux of a vortex (up to conjugation) at arbitrarily long range; therefore, no local observable can change the conjugacy class of the total magnetic flux. The magnetic flux (up to conjugation) of an object can be measured by the Wilson loop

operator $W^{(\nu)}(C)$. We choose the loop C to lie in a time slice, such that its closest approach to the object is large compared to the correlation length of the theory. Then

$$\frac{\langle\psi|\, W^{(\nu)}(C)\,|\psi\rangle}{\langle W^{(\nu)}(C)\rangle_0} = \sum_{[a]} p_{[a]}(\psi;C)\,\frac{1}{n_\nu}\chi^{(\nu)}(a)\ , \tag{3.15}$$

where $p_{[a]}(\psi;C)$ is the probability that the state $|\psi\rangle$ has flux $[a]$. By measuring $W^{(\nu)}(C)$ for each irreducible representation (ν), we can determine all of the $p_{[a]}$'s.

As we noted in the beginning of this chapter, if the total flux of a state is $a_{\rm tot}$, then the charge of the state is specified by an irreducible representation of the normalizer $N(a_{\rm tot})$. This is because the only global gauge transformations that can be implemented acting on this state are elements of the normalizer. In more physical terms, the charge of an object is measured by means of its Aharonov-Bohm interaction with a distant test-vortex. Such a measurement can be carried out only if the flux of the test-vortex commutes with the flux of the object. Otherwise, the flux of the test vortex changes when it circumnavigates the object, due to the holonomy interaction,[18,10] and this obscures the interference pattern. Thus, superselection sectors are labeled by the class $[a_{\rm tot}]$ of the total flux, and an irreducible representation of the normalizer $N(a_{\rm tot})$. (The abstract group $N(a_{\rm tot})$ is independent of how the class representative $a_{\rm tot}$ is chosen.)

The operators $F_{[a]}(\Sigma)$ and $A^{(\nu)}_{[a]}(\Sigma,C)$ can also be constructed in two spatial dimensions. Then Σ becomes a closed curve that can be interpreted as the world line of a vortex-antivortex pair. But for the purpose of measuring the charge of a state in two dimensions, it is convenient to generalize the construction of the operator $F_{[a]}(\Sigma)$. If the state has non-trivial flux, the charge is a representation of the normalizer of the flux, and there is a charge-zero vortex pair associated with each class of the *normalizer*. There are gauge-invariant operators that, in effect, insert such charge-zero pairs (as classical sources) on the world line Σ.

To construct these operators, we must specify a basepoint x_0. Then a projection operator $P_{(a)}(\Sigma,x_0)$ can be constructed that projects out states that have flux a associated with the path Σ that begins and ends at x_0. There is also an operator $\tilde{F}_{(b)}(\Sigma,x_0)$ that introduces a *flux eigenstate* vortex pair on Σ, where the flux b is defined with respect to paths that begin and end at x_0. These operators are not

invariant under H gauge transformations; under a gauge transformation $h \in H(x_0)$, they transform as

$$h: \quad P_{(a)} \to P_{(hah^{-1})} \ , \quad \tilde{F}_{(b)} \to \tilde{F}_{(hbh^{-1})} \ . \tag{3.16}$$

But the combination

$$F_{[a,b]}(\Sigma, x_0) \equiv \frac{1}{n_H} \sum_{h \in H} P_{(hah^{-1})}(\Sigma, x_0) \tilde{F}_{(hbh^{-1})}(\Sigma, x_0) \tag{3.17}$$

is gauge invariant. (It is unchanged if a and b are conjugated by the same element of H, and is also independent of how the basepoint x_0 is chosen.) Were it not for the flux projection operators, the left-hand side of eq. (3.17) would be just the operator $F_{[b]}$. And, indeed, if a is the identity, we have

$$F_{[e,b]} = F_{[b]} P_{(e)} \ ; \tag{3.18}$$

just our old F operator times a projection onto the trivial flux sector.

Now suppose that Σ lies in a time slice, and that we evaluate the expectation value of $F_{[a,b]}(\Sigma)$ in a state with non-trivial flux. Suppose, to be specific, that the flux contained inside Σ is in the same H conjugacy class as a. Then the projection operator will restrict the sum over h to the elements of the normalizer of the flux. Furthermore, the contribution to the expectation value due to terms in which hbh^{-1} does not commute with the flux will be heavily suppressed when Σ is large and far from the object that is being measured. To see this, note that, if a vortex-antivortex pair with flux b' winds around a vortex with flux a', and a' and b' do not commute, then the pair cannot reannihilate unless flux is exchanged between the pair and the a' vortex. (This consequence of the holonomy interaction was discussed in Ref. 12.) In effect, then, hbh^{-1} is restricted to the normalizer of the flux of the state (because the other contributions are suppressed by $\tilde{F}$), and ranges over precisely one class of the normalizer (because of the projection operator P). Therefore, if the state $|[a],(\nu_a)\rangle$ transforms as the irreducible representation (ν_a) of the *normalizer* $N(a)$, and if a and b commute, we find

$$\frac{\langle [a],(\nu_a)|\, F_{[a,b]}(\Sigma)\, |[a],(\nu_a)\rangle}{\langle F_{[a,b]}(\Sigma)\rangle_0} = \frac{1}{n_{(\nu_a)}} \chi^{(\nu_a)}(b) \ , \quad b \in N(a) \ . \tag{3.19}$$

But if a and b do not commute, or if the flux of the state is not in the class $[a]$, then

$$\frac{\langle [c],(\nu_c)|\, F_{[a,b]}(\Sigma)\, |[c],(\nu_c)\rangle}{\langle F_{[a,b]}(\Sigma)\rangle_0} = 0\ , \quad c \notin [a] \ \text{ or } \ b \notin N(a) \tag{3.20}$$

(in the limit where Σ becomes very large). By measuring the expectation value of $F_{[a,b]}$ for all commuting a and b, we can determine the flux and charge of a state.

3.3. Charge transfer

We will now consider the non-Abelian Aharonov-Bohm interactions between string loops and point particles, and demonstrate that exchange of H-charge can occur.

The total H-charge of a composite system consisting of a string loop and a charged particle can be measured by studying the Aharonov-Bohm interaction of the composite with other, much larger, string loops. Obviously, then, the total H-charge of the composite must be conserved; it cannot change when the particle winds through the loop. Charge exchange is an inevitable consequence of charge conservation.

To see this, it is convenient to imagine a composite of a string loop and a particle-antiparticle pair, where, initially, both the loop and the pair have zero charge (transform trivially under H). Suppose that the particle transforms as the irreducible representation $D^{(\nu)}$ of H; the antiparticle transforms as the conjugate representation. Let $\{e_i^{(\nu)} \mid i = 1, 2 \ldots n_\nu\}$ denote an orthonormal basis for the vector space on which $D^{(\nu)}$ acts. Then the initial state of the pair has the group-theoretic structure

$$\left|\psi_{\text{in}}^{(\nu)}\right\rangle = \frac{1}{\sqrt{n_\nu}}\ \left|e_i^{(\nu)*} \otimes e_i^{(\nu)}\right\rangle \tag{3.21}$$

(summed over i). The initial state of the loop is the state $|0;[a]\rangle$ defined in eq. (3.4).

Suppose that the particle and antiparticle are initially at the point x_0. Then the particle traverses a path C that winds through the string loop and returns to x_0. After this traversal, the state of the pair and the state of the loop are correlated. The total charge is still zero, but in general the pair and the loop both have a non-trivial charge. We can infer the final charge on the loop by calculating the final charge carried by the pair. In fact, the final charge of the pair is actually independent of

the initial charge of the loop; it depends only on the class $[a]$. Thus, to calculate the final charge of the pair, we may take the state of the loop to be the flux eigenstate $|a\rangle$ (where the flux is defined in terms of the path C as in eq. (3.1)). It does not matter how the class representative a is chosen.

Using eq. (3.5), we find that the state of the pair after the traversal is

$$\left|\psi^{(\nu)}_{\rm fin}, a\right\rangle = \frac{1}{\sqrt{n_\nu}} \left|e^{(\nu)*}_i \otimes e^{(\nu)}_j\right\rangle D^{(\nu)}_{ji}(a) \ . \tag{3.22}$$

This state $\left|\psi^{(\nu)}_{\rm fin}, a\right\rangle$ does not transform as a definite irreducible representation of H, but it can, of course, be decomposed into states of definite H-charge. The probability $p^{(\mu)}$ that the H-charge is (μ) can be extracted by using the projection operator $P^{(\mu)}$ defined by eq. (3.10). We find

$$\begin{aligned} p^{(\mu)}_{\rm pair}(\nu,[a]) &= \left\langle \psi^{(\nu)}_{\rm fin}, a\right| P^{(\mu)} \left|\psi^{(\nu)}_{\rm fin}, a\right\rangle \\ &= \frac{n_\mu}{n_\nu n_H} \sum_{b\in H} \chi^{(\mu)}(b^{-1}) D^{(\nu)*}_{nm}(a) \left\langle e^{(\nu)*}_m \otimes e^{(\nu)}_n \right| \left. e^{(\nu)*}_k \otimes e^{(\nu)}_l \right\rangle D^{(\nu)*}_{ki}(b) D^{(\nu)}_{lj}(b) D^{(\nu)}_{ji}(a) \\ &= \frac{n_\mu}{n_\nu n_H} \sum_{b\in H} \chi^{(\mu)}(b^{-1}) \chi^{(\nu)}(bab^{-1}a^{-1}) \ . \end{aligned} \tag{3.23}$$

As we anticipated, this result is unchanged if a is replaced by $a' \in [a]$.

If the total H-charge is zero, then the composite of string loop and pair has a wavefunction of the form

$$\left|\psi^{(\nu)}_{[a]}\right\rangle = \sum_\mu \sqrt{p^{(\mu)}_{\rm pair}} \ |{\rm loop}, \mu*\rangle \otimes |{\rm pair}, \mu\rangle \ . \tag{3.24}$$

Thus, the probability that the loop carries charge (μ) is given by

$$p^{(\mu)}_{\rm loop}(\nu,[a]) = p^{(\mu*)}_{\rm pair}(\nu,[a]) = p^{(\mu*)}_{\rm loop}(\nu*,[a]) = p^{(\mu*)}_{\rm loop}(\nu,[a^{-1}]) \ . \tag{3.25}$$

Of course, this probability is non-vanishing only if $D^{(\mu)}$ is contained in $D^{(\nu)*} \otimes D^{(\nu)}$ and represents the center of H trivially.

We can directly verify that detectable Cheshire charge now resides on the string loop by studying an appropriate gauge-invariant correlation function. Consider the

process depicted in Fig. 2. This process is shown in 2+1 dimensions for ease of visualization, but the generalization to 3+1 dimensions is straightforward. At time t_1, a vortex-antivortex pair is created. The flux of the vortex lies in the class $[a]$, and the (initial) H-charge of the vortex pair is trivial. At time t_2, a particle-antiparticle pair is created. The particle has H-charge (ν), and the pair is (initially) uncharged. Then the particle winds counterclockwise around the $[a]$ vortex, transferring charge to the vortex pair. Next, another vortex-antivortex pair, with flux lying in the class $[b]$, winds around the (now charged) $[a]$ vortex pair, acquiring an Aharonov-Bohm phase that is sensitive to the charge of the $[a]$ pair. Then the charge-(ν) particle winds clockwise around the $[a]$ vortex, discharging the $[a]$ pair. Finally, the particle-antiparticle pair is annihilated at time t_3, and the $[a]$ vortex-antivortex pair is annihilated at time t_4.

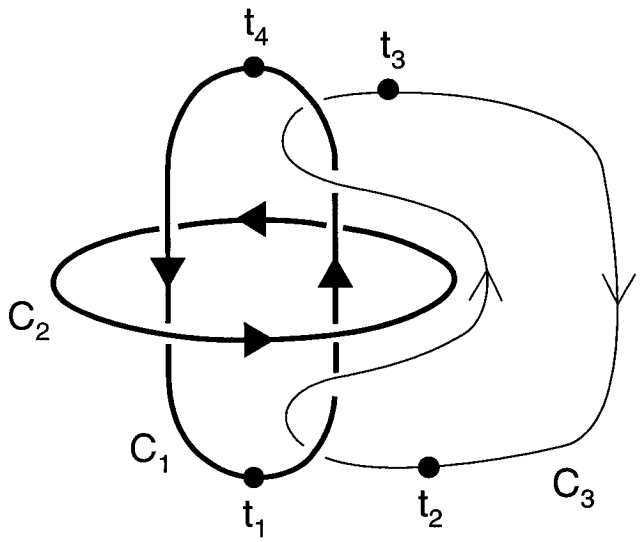

Fig. 2: The Borromean rings. C_1 is the world line of an $[a]$ vortex, C_2 is the world line of a $[b]$ vortex, and C_3 is the world line of a charged particle that transforms as the representation (ν). The charged particle transfers charge to the $[a]$ vortex-antivortex pair, and the charge is subsequently detected via the Aharonov-Bohm interaction of the pair with the $[b]$ vortex.

If the vortices and charged particles are treated as classical sources, this process is described by the correlation function

$$\left\langle F_{[a]}(C_1)\; F_{[b]}(C_2)\; W^{(\nu)}(C_3)\right\rangle_0 \;, \tag{3.26}$$

where C_1 is the world line of the $[a]$ vortex, C_2 is the world line of the $[b]$ vortex,

and C_3 is the world line of the charged particle. As shown in Fig. 2, the three loops C_1, C_2, and C_3 are joined in a topologically non-trivial configuration known as the "Borromean rings";[19] no two loops are linked, yet the loops cannot be separated without crossing.

By considering the case where the loops are large and far apart, and comparing with the case where the loops are unjoined, we can isolate the Aharonov-Bohm factor acquired by the $[b]$ vortex pair that winds around the charged $[a]$ vortex pair. The calculation of eq. (3.26), using weak-coupling perturbation theory on the lattice, is described in Ref. 12. We will not repeat the details of the calculation here, but it is easy to explain the main idea. Loosely speaking, inserting a classical vortex with flux a on the closed path C_1 is equivalent to performing a singular gauge transformation on a surface S_1 that is bounded by C_1. The path has an orientation, which induces an orientation of the surface. The effect of the singular gauge transformation on the Wilson loop $W^{(\nu)}(C_3)$ is to insert the factor $D^{(\nu)}(a)$ where C_3 crosses S_1 in a positive sense, and to insert the factor $D^{(\nu)}(a^{-1})$ where C_3 crosses S_1 in a negative sense. In Fig. 3, we see that the loop C_3 successively crosses S_2 in a negative sense, S_1 in a negative sense, S_2 in a positive sense, and S_1 in a positive sense, before closing. Due to the path ordering of the Wilson loop, the factor due to a later crossing appears to the left of the factor due to an earlier crossing. These crossings therefore modify $\left\langle W^{(\nu)}(C_3)\right\rangle_0$ by the factor $(1/n_\nu)\chi^{(\nu)}(aba^{-1}b^{-1})$ compared to the case where C_3 is unjoined with C_1 and C_2. Recalling that a and b are averaged over a class when $F_{[a]}$ and $F_{[b]}$ are inserted, we find that

$$\frac{\left\langle F_{[a]}(C_1)\; F_{[b]}(C_2)\; W^{(\nu)}(C_3)\right\rangle_0}{\left\langle F_{[a]}(C_1)\right\rangle_0 \left\langle F_{[b]}(C_2)\right\rangle_0 \left\langle W^{(\nu)}(C_3)\right\rangle_0} \longrightarrow \frac{1}{n_H}\sum_{h\in H}\frac{1}{n_\nu}\chi^{(\nu)}(hah^{-1}bha^{-1}h^{-1}b^{-1}) \tag{3.27}$$

when the loops are large, far apart, and joined.

In 3+1 dimensions, there is an analog of the Borromean ring configuration, in which two disjoint closed surfaces Σ_1 and Σ_2 are joined by a closed loop C_3 that does not link with either surface. For this configuration, eq. (3.27) still applies, with C_1 and C_2 replaced by Σ_1 and Σ_2. We can decompose the right-hand-side of eq. (3.27)

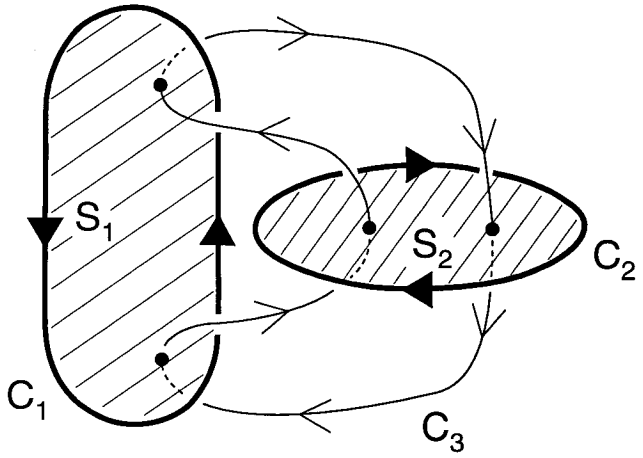

Fig. 3: A deformation of the rings shown in Fig. 2. The gauge field is singular on the surfaces S_1 and S_2 that are bounded by the loops C_1 and C_2.

into characters as

$$\sum_{\mu} p^{(\mu)}_{\text{loop}}(\nu,[a])\ \frac{1}{n_\mu}\chi^{(\mu)}(b)\ , \tag{3.28}$$

where $p^{(\mu)}_{\text{loop}}(\nu,[a])$ is the probability that the charge carried by the $[a]$ string loop, and detected by the $[b]$ string loop, is (μ). (Compare eq. (3.8).) Using the orthogonality of the characters, we find from eq. (3.27) and (3.28) that

$$p^{(\mu)}_{\text{loop}}(\nu,[a]) = \frac{n_\mu}{n_\nu n_H}\sum_{b\in H}\chi^{(\mu)}(b^{-1})\chi^{(\nu)}(aba^{-1}b^{-1})\ , \tag{3.29}$$

in agreement with eq. (3.25) and (3.23). Thus, the charge lost by the particle pair has indeed been transferred to the $[a]$ string loop. (Note that, in order to get the right answer, it is important to choose consistent orientations for the world sheets Σ_1 and Σ_2—the $[a]$ string must pass through the $[b]$ string in the same sense that the Wilson loop passes through the $[a]$ string. Otherwise, we would in effect be measuring the charge of the $[a]$ string with a $[b^{-1}]$ string, rather than a $[b]$ string.)

We will now derive eq. (3.27) by a different method that invokes the "holonomy interaction" between string loops. Consider two flux-eigenstate string loops that initially carry flux a and b. Now suppose that the b loop sweeps around the a loop and returns to its original position. After this process, the flux of the b loop is unchanged, but the flux of the a loop has been altered; it has become a bab^{-1} loop.[18,10,6] (Here

again, we must be careful about the orientations of the string loops. The a loop becomes a bab^{-1} loop if it passes through the b loop in the same sense as the path C that is used to define the flux of the b loop. If it passes through the b loop in the opposite sense, it becomes a $b^{-1}ab$ loop.)

Return now to the Borromean ring process. Suppose that two string loops are initially in the flux eigenstate $|a,b\rangle$. Then a particle-antiparticle pair is created, and the particle winds through the a loop; the new state of the string loops and the particle-antiparticle pair can be expressed as

$$\frac{1}{\sqrt{n_\nu}} \left| a, b, e_i^{(\nu)*} \otimes e_j^{(\nu)} \right\rangle D_{ji}^{(\nu)}(a) \ . \tag{3.30}$$

(Compare eq. (3.22).) When the b loop sweeps around the a loop, the state becomes

$$\frac{1}{\sqrt{n_\nu}} \left| bab^{-1}, b, e_i^{(\nu)*} \otimes e_j^{(\nu)} \right\rangle D_{ji}^{(\nu)}(a) \ , \tag{3.31}$$

due to the holonomy interaction. Now the particle winds back through the bab^{-1} loop (in the opposite sense), and the state becomes

$$\frac{1}{\sqrt{n_\nu}} \left| bab^{-1}, b, e_i^{(\nu)*} \otimes e_k^{(\nu)} \right\rangle D_{kj}^{(\nu)}(ba^{-1}b^{-1}) D_{ji}^{(\nu)}(a) \ . \tag{3.32}$$

Finally, the particle-antiparticle pair annihilates, and we have

$$\frac{1}{n_\nu} \chi^{(\nu)}(aba^{-1}b^{-1}) \left| b^{-1}ab, b \right\rangle \ . \tag{3.33}$$

To reproduce eq. (3.27), we must take the initial string state to be $|0;[a],\ 0;[b]\rangle$, in which the $[a]$ and $[b]$ loops are both uncharged. Thus, we average both a and b over a class. We find that the effect of the particle-antiparticle pair on the string state is

$$|0;[a],\ 0;[b]\rangle \longrightarrow \left(\frac{1}{n_H} \sum_{h \in H} \frac{1}{n_\nu} \chi^{(\nu)}(hah^{-1}bha^{-1}h^{-1}b^{-1}) \right) \ |0;[a],\ 0;[b]\rangle \ . \tag{3.34}$$

By creating the initial string state and annihilating the final string state, we obtain eq. (3.27).

3.4. Some final comments

We described in Section 2 how a charge-zero string loop can be used in an Aharonov-Bohm interference experiment to measure the charge of an object. It seems that charge could also be measured using strings that are in flux eigenstates rather than charge eigenstates. This alternative measurement process was discussed in detail in Ref. 11, and in some respects it is simpler than the process that we described. So our insistence on using charge-zero string loops requires a word of explanation.

Part of the answer is that the gauge-invariant charge operator that we constructed in Section 2 can be interpreted in terms of a process involving a charge-zero loop. In this process, a tiny string loop is created, and is stretched to a large size. The loop then winds around a region, shrinks, and reannihilates. The Aharonov-Bohm factor acquired by the loop is sensitive to the total charge inside the region. This string loop must be in a charge-zero state, rather than a flux eigenstate, because otherwise charge conservation would forbid the creation of the loop and its subsequent annihilation.

Another point is that the notion of a flux eigenstate string suffers from an ambiguity. Of course, flux eigenstates are not energy eigenstates, so there is quantum-mechanical mixing between flux eigenstates that are in the same H conjugacy class. The authors of Ref. 11 note that the time scale for this mixing is of order $\exp(\kappa A)$, where κ is the string tension and A is the area of the loop. Therefore, the mixing can be easily turned off by choosing the string loop to be sufficiently large.

But this is not quite the whole story, for there is another type of "mixing" that should be considered.[12] Recall that to define the flux of a string, we must choose an arbitrary "basepoint" x_0, and a path C, beginning and ending at x_0, that winds around the string. Suppose that a virtual string loop nucleates, lassoes the basepoint, and then reannihilates. If the string is initially assigned the flux a, and the virtual string has flux b, then this process changes the flux to bab^{-1}, due to the holonomy interaction. In physical terms, we may prepare a string in a flux eigenstate, and then measure its flux later by doing an Aharonov-Bohm interference experiment. We may find that the flux has been altered, not because of the mixing of flux eigenstates, but rather because of quantum fluctuations that are independent of the size of the loop. Of course, virtual string loops are strongly suppressed at weak coupling; the

amplitude for such a process is of order $\exp(-C/e^2)$, where e is the gauge coupling. But it is a significant point of principle that quantum fluctuations render ambiguous the notion of the flux of the string. No such ambiguity afflicts the conjugacy class of the flux, or the charge carried by a loop.

For these reasons, we have used charge-zero strings in our discussion of the charge measurement process. We can imagine doing a double-slit experiment with a beam of particles of unknown charge, where a string loop in the state $|0;[a]\rangle$ surrounds one of the slits. By observing how the shift in the interference pattern depends on the class $[a]$, we can determine the character of the representation according to which the particles in the beam transform, and so infer their charge.

However, the phenomenon of charge transfer raises a puzzle. If a particle passes through the slit that is surrounded by the string, it transfers charge to the string. By measuring the charge on the string loop later, we can find out which slit the particle passed through. Thus, no interference pattern should be seen.

The resolution of this puzzle is that there is a non-vanishing probability, in general, that no charge transfer takes place. This probability is given by eq. (3.29) in the case where (μ) is the trivial representation (0); we then have

$$p^{(0)}_{\text{loop}}(\nu,[a]) = \left|\frac{1}{n_\nu}\chi^{(\nu)}(a)\right|^2 \quad . \tag{3.35}$$

Therefore, as long as the character does not vanish, it is possible for the particle to slip through the string loop without being detected, and an interference pattern is observed. From the interference pattern, the phase of the character, as well as its modulus, can be deduced.

REFERENCES

1. M. Alford, K. Benson, S. Coleman, J. March-Russell, and F. Wilczek, Phys. Rev. Lett. **64** (1990) 1632; **65** (1990) 668 (E).
2. A.S. Schwarz, Nucl. Phys. **B208** (1982) 141; A.S. Schwarz and Y.S. Tyupkin, Nucl. Phys. **B209** (1982) 427.
3. P. Ginsparg and S. Coleman, unpublished (1982).
4. J. Preskill and L. Krauss, Nucl. Phys. **B341** (1990) 50.
5. M. Alford, K. Benson, S. Coleman, J. March-Russell, and F. Wilczek, Nucl. Phys. **B349** (1991) 414.
6. M. Bucher, H.-K. Lo, and J. Preskill, Nucl. Phys. **B386** (1992) 3.
7. M. Alford, J. March-Russell, and F. Wilczek, Nucl. Phys. **B337** (1990) 695.
8. F. Bais, Nucl. Phys. **B170** (1980) 32.
9. T. W. B. Kibble, Phys. Rep. **67** (1980) 183.
10. M. Bucher, Nucl. Phys. **B350** (1991) 163.
11. M. Alford, S. Coleman and J. March-Russell, Nucl. Phys. **B351** (1991) 735.
12. M. Alford, K.-M. Lee, J. March-Russell, and J. Preskill, Nucl. Phys. **B384** (1992) 251.
13. A. Balachandran, F. Lizzi, and V. Rogers, Phys. Rev. Lett. **52** (1984) 1818.
14. L. Krauss and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1221.
15. K. Fredenhagen and M. Marcu, Comm. Math. Phys. **92** (1983) 81; K. Fredenhagen, in *Fundamental Problems of Gauge Field Theory*, ed. G. Velo and A. S. Wightman (Plenum, New York, 1986).
16. M. Alford and J. March-Russell, Nucl. Phys. **B369** (1992) 276.
17. G. 't Hooft, Nucl. Phys. **B138** (1978) 1; Nucl. Phys. **B153** (1979) 141.
18. F. Wilczek and Y.-S. Wu, Phys. Rev. Lett. **65** (1990) 13.
19. See, for example, D. Rolfsen, *Knots and Links* (Publish or Perish, Wilmington, 1976).

Chapter 4: Vortices on Higher Genus Surfaces

When a gauge symmetry is spontaneously broken, in general there will be stable topological defects.[1] What types of defects will be created depend on the spacetime dimension and the topology of the vacuum manifold. In two spatial dimensions, if the fundamental group of the vacuum manifold is non-trivial, there will be point defects which are called vortices. A charged particle winding around a vortex will be transformed by an element of the unbroken gauge group. This is the (non-Abelian) Aharonov-Bohm effect.[2–16] It is long range and topological. This means that the gauge transformation will not depend either on how far apart the particle and the vortex are or on the exact loops the charged particle travels along as long as their linking numbers with the vortex are the same. We will say that the vortex carries (non-Abelian) magnetic flux.

Another way to look at it is that in the presence of the vortices, the fundamental group of the surface is non-trivial.[7] After a charged particle travels along a non-contractible loop around a vortex, it will remain the same only up to a gauge transformation. The element of the unbroken gauge group associated with that transformation is the magnetic flux carried by the vortex. However, the fundamental group of the surface may be non-trivial even without any vortices. For example, there may be handles on the surface. There are two non-equivalent non-contractible loops associated to each handle. Then, by the same argument, we expect we can assign group elements to the two loops and the handle can carry magnetic flux; therefore, the handles will have topological interaction with the vortices and the charged particles. If we interchange two vortices or let a vortex go along a non-contractible loop, the magnetic flux carried by the vortices and the loop will be changed. This kind of motion of the vortices can be described by the braid group of the surface.[17,18] We then have a natural action of the elements of the braid group on the states of the vortices and the surface.

If the surface is compact, there is one relation between the generators of the fundamental group of the surface and the group elements assigned to the vortices, and the non-contractible loops must satisfy a relation induced from that relation. This restricts the possible magnetic flux carried by the vortices and handles on any

compact surfaces. The simplest example is that we cannot put a single non-trivial vortex on a sphere.

In semiclassical approximation, a pair of vortex and anti-vortex may carry electric charge, "Cheshire charge."[5,6,10] It turns out that the properties of a handle are similar (but not equal) to the properties of two vortex-antivortex pairs. In particular, a handle can carry Cheshire charge. If the size of the handle is very small, an observer outside the handle will see a "particle" that carries both magnetic flux and electric charge, a dyon.[1] (The term "dyon" is originally for a particle that carries both electric and magnetic charge in $3+1$ dimensions. We stretch its meaning to $2+1$ dimensional spacetime.) In fact, any particle that carries magnetic flux and/or electric charge falls into the representations of the quantum double associated with the gauge group.[19,14] In the language of quantum double, we have a unified treatment of the magnetic flux and electric charge. There is also a restriction on the configurations of the dyons on any compact surfaces.

In Section 1, the basic properties of vortices will be briefly reviewed. The purpose of this section is to establish conventions. In Section 2, the braid group of a surface will be described and the topological interaction between vortices and handles will be analyzed. We will find out that locally, there is no restriction on the assignment of group elements to the non-trivial loops of a handle. In Section 3, we will give a semiclassical analysis of the theory. The argument that the handle can carry Cheshire charge is given. We also explain what a quantum double is and why it is relevant. In Section 4, the most general formulation of dyons on a surface is given. (The analysis of the previous sections is a special case in this formulation.) We give the conclusions and some comments in Section 5. There is also an appendix to explain how to measure the flux of vortices using charged particles and how to measure the charge of particles using vortices.

4.1. NON-ABELIAN VORTICES

We assume that for our theory, in high energy regime, the gauge group is a simply connected Lie group. In low energy regime, the symmetry is spontaneously broken by the Higgs mechanism, say, to a finite group, G. Then in $2+1$ dimensions, the point defects will be classified by $\pi_0(G) \cong G$.[1] From now on, we only consider the unbroken gauge group and its representations. The original broken gauge group plays no role in the following discussion. If the energy scale of the symmetry breaking is very high, the size of the vortices will be very small. Low energy experiment usually cannot probe the core of vortices. Then the space that a low energy particle sees is the original space with the points where the vortices are removed.

We can assign a group element to any isolated vortex to label the flux by the following method. Let us consider vortex 1 in Fig. 1. Choose a fixed but arbitrary base point, x_0 (away from the vortex), and a loop around it. Then, calculate the untraced Wilson loop,

$$a(C_1, x_0) = P \exp\left(i \int_{C_1, x_0} A \cdot dx \right) , \tag{4.1}$$

where P denotes the path ordering. The orientation of the loop, C_1, is only a convention. We adopt the convention indicated in the figure. Then, if a charged particle in representation (ν) of G is transported along the loop C_1, it will be transformed by $D^{(\nu)}(a(C_1, x_0))$:[4,20]

$$v^{(\nu)} \to D^{(\nu)}(a(C_1, x_0)) v^{(\nu)} \tag{4.2}$$

Since the unbroken gauge group is discrete, there is no local low energy gauge excitation. The group element $a(C_1, x_0)$ is invariant under continuous deformation of the loop C. This is how the fundamental group of the space comes in. We have assigned a group element to a generator of the fundamental group of the space with punctures.

If there are two or more vortices, we have to choose a standard loop for each vortex as in Fig. 1.[7,13] Then we can assign group elements to the loops. The combined

magnetic flux, for example, of vortex 1 and vortex 2 is the product of the group elements associated to them. Here we adopt the convention that the product of two loops, C_1C_2, in the fundamental group means that the particle will travel C_2 first and then C_1. So, the combined magnetic flux, in this convention, is $a(C_1C_2, x_0) = a(C_1, x_0)a(C_2, x_0)$.

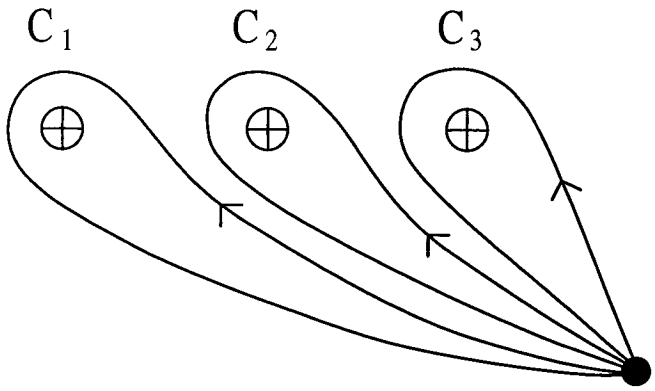

Fig. 1: The vortices in this figure are represented by circles and crosses. We choose a base point x_0 and a standard path C_i around each vortex to measure its flux.

Let us consider what will happen if we interchange two vortices. Let the flux of vortex 1 and 2 be h_1 and h_2 respectively. If we interchange the vortices counterclockwise, Fig. 2a, the magnetic flux of them will change. We have to find two loops such that *after* the interchange, they will deform to the standard loops. Then the group elements associated to them are the magnetic flux of the vortices after the interchange. From Fig. 2b, we find that

$$\begin{aligned} h_1 &\rightarrow h_1 h_2 h_1^{-1} \\ h_2 &\rightarrow h_1 \ . \end{aligned} \tag{4.3}$$

We will rely on this kind of loop tracing method to calculate various processes in the coming sections.

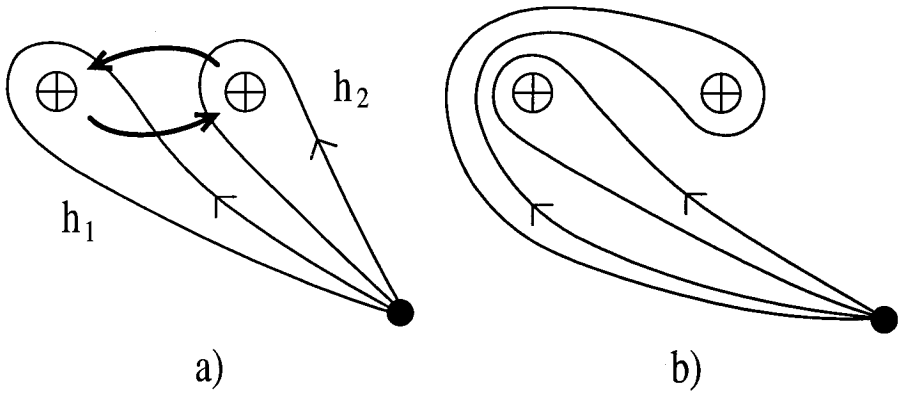

Fig. 2: In a), the two paths are the standard paths based on the same based point. The dark curves with arrows represent the interchange of the two vortices. In b), the two paths will deform to the standard paths after the interchange of the two vortices. So the flux associated with them are the flux of the two vortices after the interchange.

4.2. VORTICES ON HIGHER GENUS SURFACES

The basic element of a surface with genus greater than zero is a handle.[21] All compact surfaces can be classified according to the number of handles they have. For a single handle, there are two generators in the fundamental group. We can choose the generators to be the loops α and β in Fig. 3. Then it is easy to see that the loop in Fig. 4 is equal to $\alpha\beta\alpha^{-1}\beta^{-1}$.

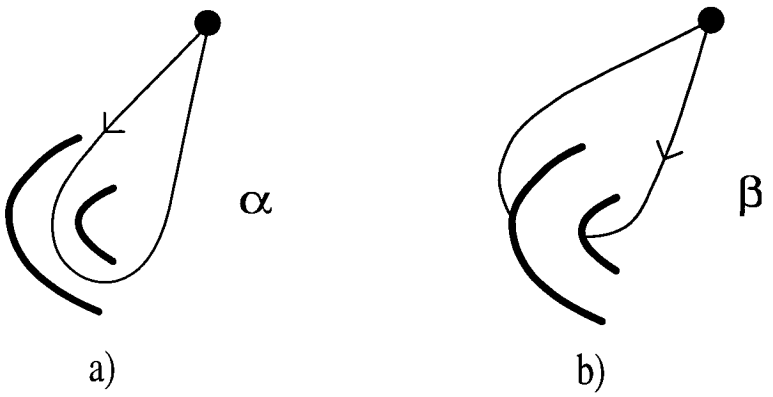

Fig. 3: The wide curves represent a handle that stands out of the paper. The two standard paths of the handle are the α in a) and β in b). Part of β is under the handle.

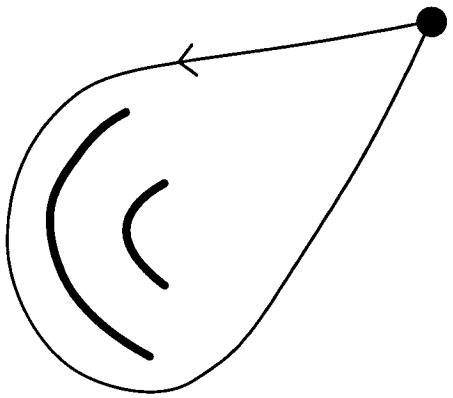

Fig. 4: We can use this path to calculate the flux of a handle. The path is equal to $\alpha\beta\alpha^{-1}\beta^{-1}$.

Since α and β are non-contractible, a charged particle transported along them may be transformed by an element of the gauge group. By sending charged particles along the loops, we can measure the group elements associated to them; for example,

$$\begin{aligned} \alpha &\mapsto a \\ \beta &\mapsto b \end{aligned} \tag{4.4}$$

where a, $b \in G$.

What will happen if a vortex winds around the loops? We expect that the magnetic flux of the vortex and group elements associated with the loops will change. We can calculate the changes by the loop tracing method as in section 1. This means that we have to find loops such that after the traveling of the vortex, these loops deform to the standard loops we used to measure the flux.

If the vortex with flux h winds around the loop in Fig. 5a, we will say that it winds around α. It implicitly means that we have chosen a path (in this case, the path can be a straight line segement) from the position of the vortex to the base point and the vortex goes along this path, then the α defined in Fig. 3 and finally that path again in reverse. If also the elements associated with α and β are a and b respectively, from Fig. 5b, c and d, we find that they will change to

$$\begin{aligned} h &\to aha^{-1} \\ a &\to ahah^{-1}a^{-1} \\ b &\to aha^{-1}h^{-1}bah^{-1}a^{-1} \ . \end{aligned} \tag{4.5}$$

If the vortex winds around β in a similar sense, from Fig. 6, they will change to

$$\begin{aligned} h &\to h^{-1}bhb^{-1}h \\ a &\to ah \\ b &\to h^{-1}bh \ . \end{aligned} \tag{4.6}$$

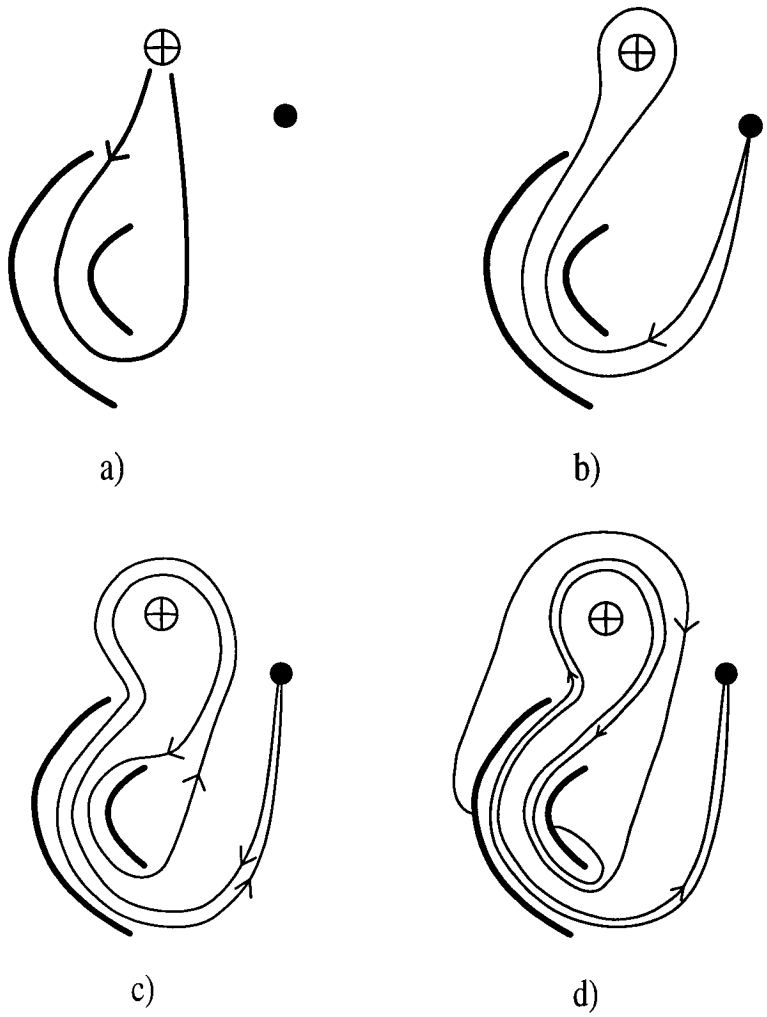

Fig. 5: The motion of the vortex is represented by the loop in a). It is also called the α loop. After the motion of the vortex as in a), the path in b) will deform to the standard path in Fig. 1. The path in c) will deform to the standard path α in Fig. 3a. The path in d) will deform to the standard path β in Fig. 3b.

Is it possible to assign arbitrary group elements to α and β? The answer is yes, at least locally. From (4.5) and (4.6), it is easy to see that even if the group is Abelian

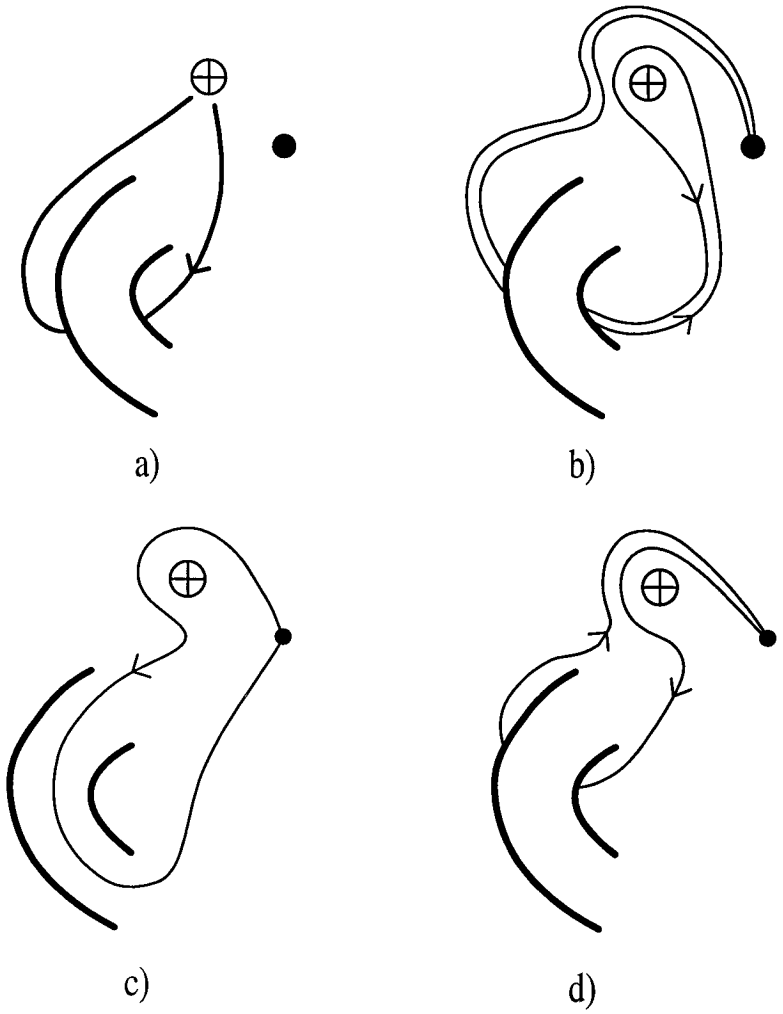

Fig. 6: Similar to Fig. 5, the loop in a) represents the motion of the vortex. Paths in b), c) and d) will deform to the standard paths in Fig. 1, α and β in Fig. 3 respectively.

and initially the elements associated with the loops are identity, after the winding of a non-trivial vortex, the group elements are no longer trivial. So, we can transfer the magnetic flux from a vortex to the handle by sending the vortex to go along α or β. To excite the handle to a state with $\alpha \mapsto a$ and $\beta \mapsto b$, consider the following: We assume that vortices with arbitrary flux exist and initially the group elements associated with α and β are identity. And we send an $ab^{-1}a^{-1}$ vortex to go along α. Then, we send an a vortex to go along β. By (4.5) and (4.6), we have

$$\begin{array}{lllllll} \alpha & \mapsto & e & \stackrel{\alpha}{\longrightarrow} & e & \stackrel{\beta}{\longrightarrow} & a \\ \beta & \mapsto & e & \longrightarrow & aba^{-1} & \longrightarrow & b \quad . \end{array} \tag{4.7}$$

Thus, after the vortices execute the prescribed motion, the loops α and β are associated with the desired group elements a and b, respectively.

If the throat of the handles is small or we ignore the internal structure of the handle and there are no pointlike vortices hiding inside the handle, we can measure the flux of the "particle" by the loop in Fig. 4. The flux is $aba^{-1}b^{-1}$; the flux carried by a handle must be in this form.

Now let us formulate the theory in precise mathematical terms. Let the space be an orientable connected surface, Σ. If there are n vortices on it, we have to consider the fundamental group of the surface with n punctures and a base point, $\pi_1(\Sigma(n), x_0)$. The combined magnetic flux (or group elements) of vortices or handles follows from the multiplication rule of the fundamental group. Any classical state of the vortices and the surface is a homomorphism from $\pi_1(\Sigma(n), x_0)$ to G,

$$\rho : \pi_1(\Sigma(n), x_0) \to G \ . \tag{4.8}$$

If the surface is also compact, there is one relation between the generators of the fundamental group. For example, the relation for a surface of genus g, Σ_g, is[21]

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\ldots\alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1} = e \ . \tag{4.9}$$

For the compact surface of genus g with n punctures, $\Sigma_g(n)$, the relation is

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\ldots\alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1}C_1\ldots C_n = e \tag{4.10}$$

where our convention of the loops is shown in Fig. 7.

The flux of the vortices and handles must satisfy this relation. For example, the relation associated with a single vortex on a sphere is

$$C_1 = e \ . \tag{4.11}$$

It is inconsistent to put a single non-trivial vortex on a sphere. Also, the relation of a torus without any vortex is

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1} = e \ . \tag{4.12}$$

Therefore, the group elements associated with the two loops must commute.

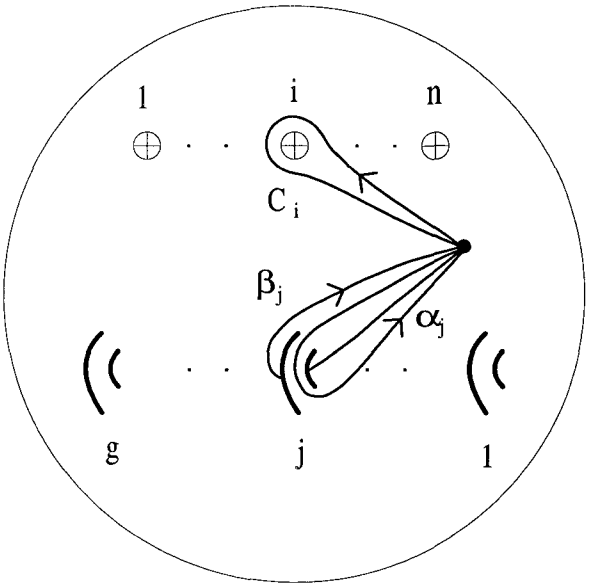

Fig. 7: A compact surface is represented by a large sphere with handles here. We choose the standard positions of the n vortices and the g handles as in this figure. We also choose a standard path for each vortex and the two standard paths α and β for each handle. The positions and the paths are consistent with Fig. 1 and Fig. 3.

This is a global constraint on the possible flux. We have seen that we can excite the handle to any state locally. We will show that it is possible to construct the state corresponding to any homomorphism, $\rho : \pi_1(\Sigma(n), x_0) \to G$. We assume that we can create vortex-antivortex pairs with arbitrary flux; to construct the state corresponding to ρ, we create two sets of vortex-antivortex pairs. The first set consists of n pairs. They contain exactly the n vortices that we want. Then we have n antivortices left. Push the antivortices to some simply connected region. The second set of vortex-antivortex pairs contains $2g$ pairs with appropriate flux. By sending them to go along the α's and the β's, we can excite the handles to the desired states. After they go along the loops, their flux will be changed. Now, the combined magnetic flux of the resulting second set of totally $4g$ vortices will not be trivial. Let us push them to the same simply connected region that contains the n antivortices. We claim that the combined magnetic flux of first set of antivortices and the second set of vortices is trivial. Since the surface is compact, a loop wrapped around that region can be deformed to a loop that wraps around all handles and the n vortices. The magnetic flux measured along this loop must be trivial because the flux assigned by ρ satisfies the relation (4.10). This means that the combined magnetic flux of the

leftover vortices is trivial. They can completely annihilate each other and the state of the surface with the n vortices is given by ρ. We conclude that the space of all states is $\mathrm{Hom}(\pi_1(\Sigma(n), x_0), G)$.

When we consider the kinematics of the vortices on the surface, at low energy, they are not allowed to collide with each other. And because the magnetic flux we measure depends on the loops we choose, to determine the flux after any motion, the vortices must be brought back to some standard positions. This kind of motion is exactly described by the braid group of the surface.

If collision of particles is not allowed, the configuration space of n distinguishable particles on a surface, Σ, is $\Sigma^n - \Delta$ where Δ is the subset of Σ^n in which at least two points in the Cartesian product coincide. The permutation group S_n has an obvious action on this configuration space. The configuration space of n indistinguishable particles is then $(\Sigma^n - \Delta)/S_n$. The definition of the braid group of n points on the surface is[17,18,22,23]

$$B_n(\Sigma) = \pi_1((\Sigma^n - \Delta)/S_n) \ . \tag{4.13}$$

Notice that although the definition of the braid group involves n indistinguishable particles, it does not matter if the vortices are distinguishable or not because we assume that they do not collide or after quantization, the position part of their wavefunctions do not overlap during the motion. Of course, if they are not identical, the total wavefunction of the vortices must be a representation of the corresponding color braid groupoid.[24] We use the elements of the braid group to describe the *motion* of our n vortices. They will become operators of the Hilbert space of states of the vortices and the surface after quantization.

If the surface is the plane, R^2, the braid group $B_n(R^2)$ is the usual braid group with $n-1$ generators which interchange adjacent points. In $B_n(\Sigma_g)$, there are $2g$ more generators. They are the α_i and β_i, $1 \leq i \leq g$ as shown in Fig. 8.* (We use the same symbols α and β to denote the loops in the fundamental group and the generators of the braid group as explained above (4.5).)

$\star$ We use a different convention from Ref. 17 because his α's involve all handles.

These generators are not independent. They satisfy, in our convention, the relations

$$
\begin{aligned}
\sigma_i\sigma_j &= \sigma_j\sigma_i \qquad |i-j| \geq 2 \ , \\
\sigma_i\sigma_{i+1}\sigma_i &= \sigma_{i+1}\sigma_i\sigma_{i+1} \qquad 1 \leq i \leq n-2 \ , \\
\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\ldots\alpha_g\beta_g\alpha_g^{-1}\beta_g^{-1}\sigma_{n-1}\ldots\sigma_1^2\ldots\sigma_{n-1} &= e \ , \\
\sigma_i\alpha_l\sigma_i^{-1}\alpha_l^{-1} &= e \qquad 2 \leq i \leq n-1, \quad 1 \leq l \leq g \ , \\
\sigma_i\beta_l\sigma_i^{-1}\beta_l^{-1} &= e \qquad 2 \leq i \leq n-1, \quad 1 \leq l \leq g \ , \\
\sigma_1\alpha_p\sigma_1^{-1}\alpha_l &= \alpha_l\sigma_1\alpha_p\sigma_1^{-1} \qquad 1 \leq p < l \leq g \ , \\
\sigma_1\beta_p\sigma_1^{-1}\beta_l &= \beta_l\sigma_1\beta_p\sigma_1^{-1} \qquad 1 \leq p < l \leq g \ , \\
\sigma_1\alpha_p\sigma_1\alpha_p &= \alpha_p\sigma_1\alpha_p\sigma_1 \qquad 1 \leq p \leq g \ , \\
\sigma_1\beta_p\sigma_1\beta_p &= \beta_p\sigma_1\beta_p\sigma_1 \qquad 1 \leq p \leq g \ , \\
\alpha_p\sigma_1^{-1}\beta_l\sigma_1 &= \sigma_1^{-1}\beta_l\sigma_1\alpha_p \qquad 1 \leq p < l \leq g \ , \\
\beta_p\sigma_1^{-1}\alpha_l\sigma_1 &= \sigma_1^{-1}\alpha_l\sigma_1\beta_p \qquad 1 \leq p < l \leq g \ , \\
\sigma_1\alpha_p\sigma_1\beta_p &= \beta_p\sigma_1\alpha_p\sigma_1^{-1} \qquad 1 \leq p \leq g \ .
\end{aligned}
\tag{4.14}
$$

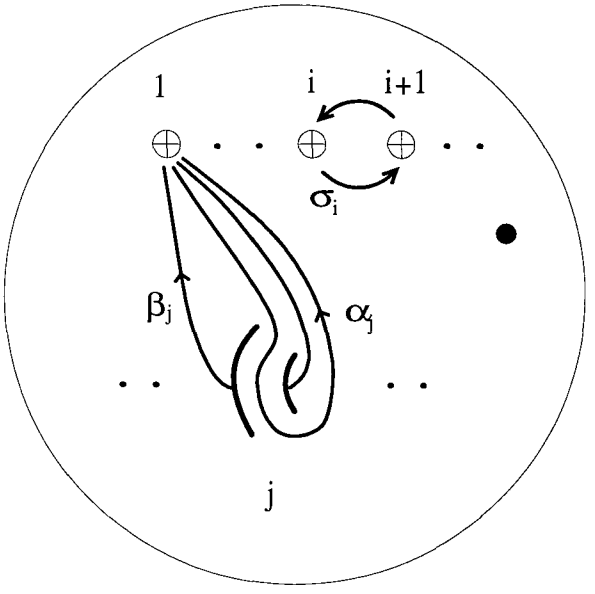

Fig. 8: These are the conventions for the braidings of the vortices with each other and with the handles. They are equal to the braidings in Fig. 2a, Fig. 5a and Fig. 6a.

There is a natural action of the braid group of surface with n points on the fundamental group of the surface with n punctures defined as follows. If $\tau \in B_n(\Sigma)$ and $\gamma \in \pi_1(\Sigma(n), x_0)$, define $\tau(\gamma)$ to be the loop such that *after* the motion of the n points according to τ, the loop will deform to γ. By the calculation in (4.3), (4.5) and (4.6), we find that the non-trivial actions for $B_n(\Sigma_g)$ are

$$\begin{aligned}
\sigma_i(C_i) &= C_{i+1} \\
\sigma_i(C_{i+1}) &= C_{i+1}C_iC_{i+1}^{-1} \\
\alpha_j(C_1) &= \alpha_j C_1 \alpha_j^{-1} \\
\alpha_j(\alpha_j) &= \alpha_j C_1 \alpha_j C_1^{-1} \alpha_j^{-1} \\
\alpha_j(\beta_j) &= \alpha_j C_1 \alpha_j^{-1} C_1^{-1} \beta_j \alpha_j C_1^{-1} \alpha_j^{-1} \\
\beta_j(C_1) &= C_1^{-1} \beta_j C_1 \beta_j^{-1} C_1 \\
\beta_j(\alpha_j) &= \alpha_j C_1 \\
\beta_j(\beta_j) &= C_1^{-1} \beta_j C_1 \ .
\end{aligned} \tag{4.15}$$

It is easy to check that this definition satisfies the relations (4.14). This action induces an action of the braid group on the states of the vortices and surface. And this is exactly how the state will be changed after the motion of the vortices.

4.3. Semi-classical analysis

We are going to argue that if we specify the flux of α and β of a handle, we know the quantum state of the handle completely. (Of course, a general quantum state of the handle could be a linear combination of the flux eigenstates of α and β.) The scheme is as follows. We try to find out a complete set of commuting observables by first choosing an observable, say A, and find out its eigenstates. In general, there are more than one independent eigenvectors with the same eigenvalues. So, we find another observable, B, which commutes with A. Then we can decompose the eigenspaces of A with respect to B. If the dimensions of the simultaneous eigenspaces of A and B are still greater than one, we find yet another observable which commutes with both A and B and decompose the eigenspaces and so on. This process will stop if all the simultaneous eigenspaces are one-dimensional or we run out of observables.

In our case of discrete gauge theory, there are not many observables. First of all, the theory is topological. We don't have any local excitations and the only things that we can measure are the magnetic flux and the electric charge bounded by a loop. For a handle, we can send charged particles to go along α and β to measure the flux of them. The measurement of one loop does not affect the flux of the other; therefore, these two observables commute. Let us denote the state of a handle that α maps to a and β maps to b by $|a, b, X\rangle$ where X specifies any other quantum numbers needed to completely specify the state of a handle. Our objective is to prove that no such X is needed. Now, the only other possible degrees of freedom, X, are the charge bounded by the two loops. It turns out that we cannot measure the charge bounded by α, say, without messing up the flux of β. The charge measurement of α does not commute with the measurement of the flux of β and vice versa. Since these are all the observables in the theory, the flux of α and β form a complete set of commuting observables and we do not need any X.

We now explain why we cannot measure the charge bounded by α without affecting the flux of β. The only way we can measure the charge bounded by a loop is to send vortices along the loop and deduce the charge from the interference pattern.[8] (We explain how to measure flux of a vortex by charged particles and how to measure the charge of a particle by vortices in the appendix.) If the handle is in the state $|a, b, X\rangle$ and the flux of the vortex is h, the initial state is $|h\rangle \otimes |a, b, X\rangle$. Now, suppose that the vortex winds around the loop α, from (4.5), the final state is $|aha^{-1}\rangle \otimes |ahah^{-1}a^{-1}, aha^{-1}h^{-1}bah^{-1}a^{-1}, X'\rangle$ where the quantum number X may change to X' after the winding of the vortex. The interference term is $\langle h|aha^{-1}\rangle\langle a, b, X|ahah^{-1}a^{-1}, aha^{-1}h^{-1}bah^{-1}a^{-1}, X'\rangle$. The first factor is non-zero if a and h commute, but then the second factor is $\langle a, b, X|a, bh^{-1}, X'\rangle$ which is zero for non-trivial h; there is no interference and we cannot know the charge bounded by α.

Notice that in some state of the handle, α can bound some well defined electric charge. For example, in the state $\sum_{b\in G} |e, b\rangle$, the charge bounded by α can be measured and it is trivial. Some other linear combinations will give non-trivial charge; however, none of these α-charge eigenstates are β-flux eignestates.

Recall that if the state of the handle is $|a, b\rangle$, it carries magnetic flux $aba^{-1}b^{-1}$.

In the neighborhood of such a handle (in general, in the neighborhood of a vortex with non-trivial magnetic flux), it is impossible to implement a global gauge transformation h if h does not commute with the flux of the handle. It is because when we try to extend a local gauge transformation h along a loop around the handle, the transformation will be conjugated by the flux of the handle at the end of the loop. There is no way to solve this inconsistent boundary condition. It is called the global color problem.[25] We can only consider global gauge transformations that are in the normalizer of $aba^{-1}b^{-1}$, $N_{aba^{-1}b^{-1}}$. Under such a gauge transformation h, the state is transformed to

$$|a, b\rangle \rightarrow |hah^{-1}, hbh^{-1}\rangle \ . \tag{4.16}$$

Semi-classically, linear combinations of these states are physically attainable. The vector space spanned by all these states of a handle can be decomposed to a direct sum of irreducible representations of $N_{aba^{-1}b^{-1}}$. These irreducible representations are the possible Cheshire charges that a handle can carry. Notice that the mathematical structure of this vector space is equal to the structure of the states of two vortex-antivortex pairs. They can carry the same kinds of Cheshire charge.

It seems possible that a handle as a whole could carry ordinary electric charges belonging to some representations of $N_{aba^{-1}b^{-1}}$ in addition to the Cheshire charge. However, this is not true because in order to measure this electric charge, we could only send vortices around the handle as in Fig. 4. This path is equal to $\alpha\beta\alpha^{-1}\beta^{-1}$ and the measurement is equivalent to the measurement of α and β in the appropriate order. Another way to say this is if an observer does not care about the topology of the space, he or she may think that the handle is "two pairs of vortex-anti-vortex." There is no electric charge associated to any single "vortex." The total electric charge can only be built up from the Cheshire charge and not something else.

Particles that carry both magnetic flux and electric charge are called dyons. The mathematical tool to classify them is the quantum double of a group and its representations. We will give a brief review of the necessary details here. Interested readers can look up the references for a full account.[19,14,26]

The difficulty of classifying dyons is that when a dyon carries magnetic flux a, we can only consider electric charges which fall into the representations of the normalizer

of a, N_a. If there are two dyons with flux a and b, their electric charges will be classified by N_a and N_b respectively. However, when we consider the two dyons as a whole, the total magnetic flux will be ab (in some convention) and the electric charge must be a representation of N_{ab}. We will find out that vectors in the irreducible representations of the quantum double have exactly this property. The irreducible representations of a quantum double are labeled by the conjugacy classes of flux and an irreducible representation of the normalizer of the flux in that conjugacy class. (All normalizers of the flux in the same conjugacy class are isomorphic.) There is a basis in each irreducible representation consisting of vectors that represent states of dyons with definite magnetic flux. A tensor product of two such vectors in two irreducible representations can be decomposed to a direct sum of vectors in irreducible representations of the normalizer of the total flux. There is also an element in (the tensor product of two copies of) the quantum double to implement the braiding operation.

Let us begin by recalling some properties of representations of a group. Any representation of a group G on a vector space V is a homomorphism

$$\phi : G \to \mathrm{End}(V) \ . \tag{4.17}$$

This homomorphism can be extended linearly to the group algebra $\mathcal{C}[G]$ by

$$\phi(\sum k_i h_i) = \sum k_i \phi(h_i) \tag{4.18}$$

where $k_i \in \mathcal{C}$. When we consider the tensor product of two representations, $\phi = \phi_1 \times \phi_2$, we have

$$\phi(h) = \phi_1(h) \otimes \phi_2(h) \tag{4.19}$$

if h is a group element. In order to lift to the group algebra, we define the comultiplication $\Delta : \mathcal{C}[G] \to \mathcal{C}[G] \otimes \mathcal{C}[G]$, by

$$\Delta(\sum k_i h_i) = \sum k_i h_i \otimes h_i \ . \tag{4.20}$$

Then, $\phi(h) = (\phi_1 \otimes \phi_2)\Delta(h)$ where now h can be any element in the group algebra.

The meaning of the comultiplication is that when a system consists of two subsystems, comultiplication bridges between the transformation of the whole system and the individual transformations of the subsystems. In general, if the symmetry transformations of a theory form an algebra, we expect there is a corresponding comultiplication to relate the symmetry transformations of the whole system and the subsystems.

Now, consider the gauge theory of a finite group G in two spatial dimensions. From the above discussion, we expect to have the following operators. For each element, a, of G, there is the gauge transformation operator of a. We can implement this operator by sending an a vortex around the base point (in some convention). The system does not change but the basis we used to measure the flux and charge has changed by a gauge transformation. It is equivalent to relabel everything in the system. For example, if the flux of a vortex is initially labeled by h, after the transformation, it is labeled by aha^{-1}. We denote this operator by the same symbol, a.

An observer far away from the system can also measure the total magnetic flux of the system relative to some fixed gauge choice. We also have a projection operator P_a, for each $a \in G$, to project to the subspace of the total flux, a. The algebra of operators, $D(G)$, is generated by a and P_b where a and $b \in G$.

The multiplication of a, b in $D(G)$ is same as the multiplication in the group. Since P_a is a projection operator,

$$P_a P_b = \delta_{ab} P_a \ . \tag{4.21}$$

After a gauge transformation of a, the magnetic flux of the system changes from b to aba^{-1}, so we have

$$a P_b = P_{aba^{-1}} a \ . \tag{4.22}$$

We have completely determined the algebraic structure of $D(G)$.

The comultiplication, $\Delta : D(G) \to D(G) \otimes D(G)$, of elements of G is the one we discussed before

$$\Delta(a) = a \otimes a \ . \tag{4.23}$$

If the system is composed of two subsystems and the magnetic flux of them is b and

c, then the flux of the whole system is bc. Conversely, if the total flux is a, the flux of the two subsystems can be any b and c as long as $bc = a$, giving us

$$\Delta(P_a) = \sum_{\substack{b,c \\ bc=a}} P_b \otimes P_c \ . \tag{4.24}$$

In this equation, we have implicitly assumed some standard paths are chosen. Then for general elements in $D(G)$, the comultiplication is*

$$\begin{aligned} \Delta(P_a b) &= \Delta(P_a)\Delta(b) \\ &= (\sum_{\substack{c,d \\ cd=a}} P_c \otimes P_d)(b \otimes b) \\ &= \sum_{\substack{c,d \\ cd=a}} (P_c b) \otimes (P_d b) \ . \end{aligned} \tag{4.25}$$

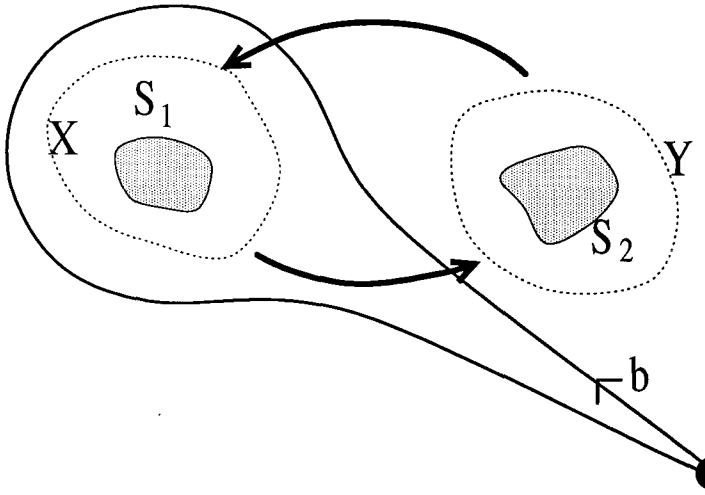

Fig. 9: This is essentially the same as Fig. 2. The shaded areas are the locations of the subsystems. Region X and Y are bounded by dotted lines.

Let us consider the two subsystems, S_1 and S_2, located in region X and region Y respectively. Our convention is that the first factor in $D(G) \otimes D(G)$ acts on the system in region X and the second factor acts on system in region Y. What will happen if the two subsystems interchange positions as in Fig. 9? If the magnetic flux

* $P_a b$ in Ref. 19 and 14 is written as $\begin{smallmatrix} a \\ \llcorner \\ b \end{smallmatrix}$.

of S_1 is b, the effect of the braiding is that apart from the position change, S_1 does not change its state but S_2 will be changed by a gauge transformation b,

$$|S_1\rangle \otimes |S_2\rangle \rightarrow (b|S_2\rangle) \otimes |S_1\rangle \ . \tag{4.26}$$

If S_1 is not in magnetic flux eigenstate, we have

$$|S_1\rangle \otimes |S_2\rangle \rightarrow \sum_{b\in G} (b|S_2\rangle) \otimes (P_b|S_1\rangle) \ . \tag{4.27}$$

If we define τ to be the operator to interchange the two factors in a tensor product and $R = \sum_b P_b \otimes b \in D(G) \otimes D(G)$, then the above action can be described by an operator $\mathcal{R} = \tau \circ R$ because

$$\begin{aligned} \mathcal{R}(|S_1\rangle \otimes |S_2\rangle) &= \tau(\sum_{b\in G} (P_b|S_1\rangle) \otimes (b|S_2\rangle)) \\ &= \sum_{b\in G} (b|S_2\rangle) \otimes (P_b|S_1\rangle) \ . \end{aligned} \tag{4.28}$$

It is easy to show that $R^{-1} = \sum_b P_b \otimes b^{-1}$ and for any $P_a b \in D(G)$,

$$\begin{aligned} R\Delta(P_a b)R^{-1} &= R(\Delta P_a)(\Delta b)R^{-1} \\ &= \sum_{\substack{c,d \\ hk=a}} (P_c \otimes c)(P_h \otimes P_k)(b \otimes b)(P_d \otimes d^{-1}) \\ &= \sum P_c P_h b P_d \otimes c P_k b d^{-1} \\ &= \sum P_c P_h P_{bdb^{-1}} b \otimes c P_k b d^{-1} \\ &= \sum_{hk=a} P_h b \otimes h P_k b (b^{-1} h^{-1} b) \\ &= \sum_{hk=a} P_h b \otimes P_{hkh^{-1}} b \\ &= \sum_{hk=a} P_k b \otimes P_h b \\ &= \tau(\Delta(P_a b)) \ . \end{aligned} \tag{4.29}$$

The meaning of this equation is the following. If an operator $d \in D(G)$ acts on the whole system, we can calculate its effect on the subsystems either by directly applying the comultiplication or by the following procedure. First, interchange the two subsystems in clockwise direction. Then, apply the comultiplication and, finally, interchange the subsystems (counterclockwise) back to their original positions.

With the multiplication, comultiplication and the R operator (and some other structures), the algebra $D(G)$ is called the quantum double associated with the group G.[27] We have seen that quantum double is a generalization of the group algebra, and we expect the states of dyons will fall into representations of the quantum double.

Now we describe the irreducible representations of $D(G)$.[19] Let the set of all conjugacy classes of G be $\{{}^AC\}$. The conjugacy class contained a will be denoted by $[a]$. For each class, fix an ordering of the elements $\{{}^AC\} = \{{}^Ag_1, \ldots, {}^Ag_k\}$. Let ${}^AN$ be the normalizer of ${}^Ag_1$. Choose elements, ${}^Ax_1, \ldots, {}^Ax_k \in G$, such that ${}^Ag_i = {}^Ax_i{}^Ag_1{}^Ax_i^{-1}$. We take ${}^Ax_1 = e$. Consider the vector space, V_ν^A, spanned by the vectors $|{}^Ag_j, {}^\nu v_i\rangle$, $j = 1,\ldots,$k and $i = 1,\ldots,\dim\nu$, where $\{{}^\nu v_i\}$ is a basis of the ν irreducible representation of ${}^AN$. This vector space carries an irreducible representation, Π_ν^A, of $D(G)$ defined by

$$\Pi_\nu^A(P_a b)|{}^Ag_j, {}^\nu v_i\rangle = \delta_{a,b{}^Ag_jb^{-1}}|b{}^Ag_jb^{-1}, D^\nu({}^Ax_l^{-1}b{}^Ax_j){}^\nu v_i\rangle \tag{4.30}$$

where ${}^Ax_l$ is defined by ${}^Ag_l = b{}^Ag_jb^{-1}$. Notice that ${}^Ax_l^{-1}b{}^Ax_j$ is in ${}^AN$. The gauge transformation b is "twisted" into the normalizer of the flux. It can be shown that these representations form a complete set of irreducible representations of $D(G)$. Any representations of $D(G)$ can be decomposed to direct sum of these representations. In $|{}^Ag_j, {}^\nu v_i\rangle$, the conjugacy class labels the magnetic flux of the dyon, the representation of ${}^AN$ labels the electric charge. We can use the comultiplication to define the tensor product of representations of $D(G)$.

The state of an ordinary electrically charged particle is $|e, {}^\nu v\rangle$, where now, ν is an irreducible representation of G. The state of a single vortex in a group eigenstate is $|h, 1\rangle$ relative to some standard path, where the 1 is the trivial representation. It is found that $\Pi_1^{[h]} \otimes \Pi_1^{[h^{-1}]} = \Pi_\nu^{[e]} \oplus \cdots$ where ν is a non-trivial representation of G and this is the Cheshire charge that a pair of vortex-antivortex can carry.[14] If we consider the handle in a state $|a; b\rangle$ as a particle, it has magnetic flux $aba^{-1}b^{-1}$. And the operator h changes its state to $|hah^{-1}, hbh^{-1}\rangle$. The state of the whole handle as a vector of the representation of the quantum double is

$$|a, 1\rangle \otimes |b, 1\rangle \otimes |a^{-1}, 1\rangle \otimes |b^{-1}, 1\rangle \tag{4.31}$$

because they have the same transformation properties under the quantum double. For example, we can calculate the possible Cheshire charge of a handle by decomposing

the tensor product $\Pi_1^{[a]} \otimes \Pi_1^{[b]} \otimes \Pi_1^{[a^{-1}]} \otimes \Pi_1^{[b^{-1}]}$. We must be careful about the meaning of the expression in (4.31). It is originally for the state of four vortices or dyons. In this case, it represents the state of a single handle. For example, we cannot apply the braiding operator to it.

4.4. Dyons on higher genus surfaces

For any surface Σ with n dyons on it, we can specify the state by choosing standard paths for the dyons and the handles and associating a vector in some representation of the quantum double for each path. One may expect that there is a correspondence between the multiplication of paths in the fundamental group and the tensor product of vectors in representations of the quantum double. However, the correspondence does not exist. To illustrate this, consider the product $C_1 {C_1}^{-1}$ in Fig. 1. It is trivial in the fundamental group. The state associated with it must be the vector in the trivial representation, but if the state associated with C_1 is $|h_1, 1\rangle$, the state associated with C_1^{-1} is $|h_1^{-1}, 1\rangle$. The tensor product $|h_1, 1\rangle \otimes |h_1^{-1}, 1\rangle$ transforms as a linear combination of charge eigenstates, not as the vector in the trivial representation. The reason why it does not work is that there is in general no "inverse" of a vector in any representation of the quantum double.

This also occurs in ordinary spacetime. For example, consider QCD in $3+1$ dimensions. When we say that there are two units of red charge inside a closed surface, we mean that we have chosen the *outward normal* of the surface, and after we integrate the color electric field on the surface relative to this normal direction, we get two units of red charge. If we consider the product of the surface and itself with inward normal in the second homotopy group, and the tensor product of the corresponding charge, we run into the same difficulty as described above.

However, the tensor product does give us the combined state of two subsystems. In $3+1$ dimensions, we have to choose the outward normal (or inward normal) for both surfaces and determine the states corresponding to these surfaces. Then the state of the combined system is given by the tensor product. In our case of dyons, the orientations of the standard loops must be in the "same sense." For example, if the states associated with C_1 and C_2 in Fig. 1 are $|h_1, 1\rangle$ and $|h_2, 1\rangle$ respectively, the

state associated with C_1C_2 is $|h_1,1\rangle \otimes |h_2,1\rangle$.

For compact surface Σ_g with n dyons, if we choose the conventions in Fig. 7. The states of the dyons can be measured by charged particles and vortices traveling around C_i (see appendix). We can denote the state where α_i maps to a_i, β_i maps to b_i and C_j maps to $|h_j, {}^{\nu_j}v\rangle$ by

$$|a_1, b_1, \ldots, a_g, b_g; h_1, {}^{\nu_1}v; \ldots; h_n, {}^{\nu_n}v\rangle \ . \tag{4.32}$$

A general state will be linear combination of these. Here we assume that the vortices are distinguishable as explained below (4.13). If some of them are identical, the state must satisfy some other relations; for example, it must be symmetric if they are bosons.

If we send vortices or charged particles along the path on the left-hand side of (4.10), we will conclude that the flux and the electric charge enclosed by the path are trivial for any state. This means that for the state

$$\sum_r k_r |a_i^{(r)}, b_i^{(r)}; h_j^{(r)}, {}^{\nu_j}v_{(r)}\rangle \ , \tag{4.33}$$

where k_r are constants, the following tensor product

$$\begin{aligned} \sum_r k_r |a_1^{(r)}, 1\rangle \otimes |b_1^{(r)}, 1\rangle \otimes |(a_1^{(r)})^{-1}, 1\rangle \otimes |(b_1^{(r)})^{-1}, 1\rangle \\ \otimes \ldots \otimes |a_g^{(r)}, 1\rangle \otimes |b_g^{(r)}, 1\rangle \otimes |(a_g^{(r)})^{-1}, 1\rangle \otimes |(b_g^{(r)})^{-1}, 1\rangle \\ \otimes |h_1^{(r)}, {}^{\nu_1}v_{(r)}\rangle \otimes \ldots \otimes |h_n^{(r)}, {}^{\nu_n}v_{(r)}\rangle \end{aligned} \tag{4.34}$$

must transform as the trivial representation.

We can also consider the motion of the dyons. Similar to the discussion in Section 2, there is an action of the braid group on the states of the surface with dyons. For $B_n(\Sigma_g)$, the action of the σ's are given by the $\mathcal{R}$ operator as discussed above.[14] From (4.15), we also have

$$\begin{aligned} \alpha_j |a_j, b_j; h_1, {}^{\nu_1}v\rangle &= |a_j h_1 a_j h_1^{-1} a_j^{-1}, a_j h_1 a_j^{-1} h_1^{-1} b_j a_j h_1^{-1} a_j^{-1}\rangle \\ &\quad \otimes \Pi_{\nu_1}^{[h_1]}(a_j) |h_1, {}^{\nu_1}v\rangle \\ \beta_j |a_j, b_j; h_1, v^{(\nu_1)}\rangle &= |a_j h_1, h_1^{-1} b_j h_1\rangle \otimes \Pi_{\nu_1}^{[h_1]}(h_1^{-1} b_j) |h_1, {}^{\nu_1}v\rangle \ . \end{aligned} \tag{4.35}$$

One can also check that this definition satisfies the relations (4.14). Notice that the action of the α's and β's cannot be written as the action of some element of the

quantum double because the flux of α and β are not just conjugated. This represents the fact that if we do see the internal structure of the handle, its state is not in the Hilbert space of the states of a particle. If every dyon does not carry electric charge, the representations ν_i are trivial. All formulae here then reduce to the corresponding formulae in section 2.

Let us consider an example. Suppose the group G is the quaternion group $Q = \{\pm 1, \pm i, \pm j, \pm k\}$. There are five conjugacy classes: $\{1\}$, $\{-1\}$, $\{\pm i\}$, $\{\pm j\}$ and $\{\pm k\}$ where, for example, $ij = k$. And there are four one-dimensional irreducible representations: the trivial representation 1 and $1_x, 1_y, 1_z$ where in 1_x, say, $\pm 1, \pm i$ are represented by 1 and the others are represented by -1. There is also a two-dimensional irreducible representation. Notice that the normalizer of -1 is the whole group. So, if a dyon has flux -1, its electric charge can be labeled by representation of the whole group.

Assume the space is a torus and there are two vortices and one charged particle. A possible state is

$$\begin{aligned} |v\rangle = \frac{1}{2}(&|i,j;k,1;k,1;1,1_x\rangle + |i,-j;-k,1;-k,1;1,1_x\rangle \\ &- |-i,j;-k,1;-k,1;1,1_x\rangle - |-i,-j;k,1;k,1;1,1_x\rangle) \ . \end{aligned} \tag{4.36}$$

In each term, the first two factors label the flux carried by α and β ($\pm i$ and $\pm j$). The third and fourth factors are the flux ($\pm k$) and the charge (trivial) of the first vortex. The next two factors have the same meaning. The final two are the trivial flux (1) and the charge (1_x) of the charged particle. The state of the handle and the states of the vortices are entangled, but if we consider them as a whole, they are in the state $|1, 1_x\rangle$, so together with the charged particle, they satisfy (4.34).

If the first vortex winds around β, by (4.35), we have

$$\begin{aligned} \beta|v\rangle =&\frac{1}{2}(|-j,-j;k,1;k,1;1,1_x\rangle + |j,j;-k,1;-k,1;1,1_x\rangle \\ &- |-j,-j;-k,1;-k,1;1,1_x\rangle - |j,j;k,1;k,1;1,1_x\rangle) \\ =&\frac{1}{2}(|-j,-j\rangle - |j,j\rangle) \otimes (|k,1\rangle \otimes |k,1\rangle - |-k,1\rangle \otimes |-k,1\rangle) \otimes |1,1_x\rangle \ . \end{aligned} \tag{4.37}$$

Now, the first factor in the above tensor product is the state of the handle, the second factor is the state of the two vortices and the last factor is the state of the

charged particle. The handle carries magnetic flux -1 and Cheshire charge 1_y. The two vortices together carry flux -1 and charge 1_z. There is magnetic flux transfer between the handle and the pair of vortices.

What will happen if a charged particle winds around a loop of a handle? As we have seen in the beginning of section 3, if the handle is in flux eigenstate, the state of the particle will be transformed by the flux of the loop, and the state of the handle will remain the same. If the handle is in some linear combination of flux eigenstates, something interesting will happen. For example, let the state of the charged particle be $|v\rangle$, and assume the state of the handle is $\sum_{b\in G}|e,b\rangle$. If the charged particle winds around β, then

$$|v\rangle\otimes\sum_{b\in G}|e,b\rangle\rightarrow\sum_{b\in G}|D(b)v\rangle\otimes|e,b\rangle\ . \tag{4.38}$$

If we now introduce an h vortex to measure the charge bounded by α, the state changes to $\sum_{b\in G}|D(b)v\rangle\otimes|e,bh^{-1}\rangle$, then the interference term is proportional to

$$\begin{aligned}&\sum_{b,b'\in G}\langle e,b'|\langle D(b')v|D(b)v\rangle|e,bh^{-1}\rangle\\ =&\sum_{b,b'\in G}\langle v|D(b'^{-1}b)|v\rangle\delta_{b',bh^{-1}}\\ =&\sum_{b\in G}\langle v|D(h)|v\rangle\ .\end{aligned} \tag{4.39}$$

The charge bounded by α is v and the flux of α is identity. We see that the charge of the particle is transferred to the ingoing mouth of the handle. However, the state of the particle entangles with the state of the handle and can no longer be specified by a single vector. This kind of charge transfer between mouths of wormholes or handles and charged particles also occurs in $3+1$ dimensions and for continuous gauge groups[28].

4.5. Conclusion

We have argued that in $2+1$ dimensions, non-trivial topology, the handle, can carry magnetic flux classically. If the unbroken gauge group is finite, we can actually assign arbitrary group elements to the two non-equivalent loops associated to the handle. Semi-classically, the state of a handle can be specified by the flux of the two non-trivial loops. It can also carry Cheshire charge. On the other hand, a general particle will fall into representations of the quantum double, an algebra constructed from the gauge group. If the surface is compact, there is a relation between the generators of the possible flux and charges of the handles and the dyons on that surface. If the surface is non-compact, no such relation exists.

There is topological interaction between the dyons and the handles. The motion of the dyons is described by the braid group of the surface. Then, the topological interaction can be described by an action of the braid group on the states of the handles and the dyons. This action, and hence the topological interaction, can be completely determined by the path tracing method explained in section 1. A similar classical analysis in $3+1$ dimensions for cosmic string has been done by Brekke et al.[29] and the classification of dyons has been generalized to theories with Chern-Simons terms.[15]

4.6. Appendix

In this appendix, we will recall how to measure the flux of a beam of identical vortices by charged particles[8] and how to measure the charge of a beam of identical charged particles by vortices.

Assume that we have a beam of identical vortices with unknown flux, h, and we have charged particles in any desired states. We can send the charged particles in a particular state around the vortices and then observe the interference patterns. If the state of the particles is $|v\rangle$ in some representation ν, the interference gives us

$$\langle v|D^{(\nu)}(h)|v\rangle \ . \tag{4.40}$$

If we replace $|v\rangle$ by $|v\rangle + \lambda|w\rangle$ in the above equation where λ is an arbitrary complex

number and subtract $\langle v|D^{(\nu)}(h)|v\rangle + \langle w|D^{(\nu)}(h)|w\rangle$ from it, we have

$$\lambda\langle v|D^{(\nu)}(h)|w\rangle + \lambda^*\langle w|D^{(\nu)}(h)|v\rangle \ . \tag{4.41}$$

Put λ to be i and 1 in succession, we know the values of two expressions. A linear combination of them gives us

$$\langle v|D^{(\nu)}(h)|w\rangle \ . \tag{4.42}$$

We can determine $\langle v|D^{(\nu)}(h)|w\rangle$ for arbitrary $|v\rangle$, $|w\rangle$ and ν and hence the matrix representation of h. If we choose ν to be some faithful representation, we can determine h.

Now assume that we have a beam of charged particles in some unknown state, $|v\rangle$, in some unknown irreducible representation, ν, and we have vortices with any desired flux. We also assume that $\langle v|v\rangle = 1$. Then a similar interference experiment will give us

$$\langle v|D^{(\nu)}(h)|v\rangle \tag{4.43}$$

for arbitrary h. Because ν is irreducible, the vectors, $D^{(\nu)}(h)|v\rangle$ for $h \in G$, will span the whole representation space. We know the inner products of these vectors because $\langle D^{(\nu)}(h_1)v|D^{(\nu)}(h_2)v\rangle = \langle v|D^{(\nu)}({h_1}^{-1}h_2)|v\rangle$. By Gram-Schmidt's orthogonalization, we can form a basis, $\{|e_i\rangle : i = 1, \ldots d\}$, such that $|e_1\rangle = |v\rangle$ and

$$|e_i\rangle = \sum_{h\in G} c_h^i |D^{(\nu)}(h)v\rangle \ . \tag{4.44}$$

Notice that the coefficients, c_h^i, depend only on the numbers $\langle v|D^{(\nu)}(h')|v\rangle$. We also have $|D^{(\nu)}(h)v\rangle = \sum_i b_i^h |e_i\rangle$ for some coefficients b_i^h.

Now we have a basis, so we can calculate the character of the representation and hence determine the representation itself.

Suppose that there is another vector, $|w\rangle$, in the same representation space such that for all h in G, $\langle w|D^{(\nu)}(h)|w\rangle = \langle v|D^{(\nu)}(h)|v\rangle$. We are going to prove that $|w\rangle$ is equal to $|v\rangle$ up to a phase. If this can be done, we can uniquely determine the state of the beam of charged particles by only sending vortices around them and observing the interference pattern.

Let $|e_i'\rangle = \sum_{h\in G} c_h^i |D^{(\nu)}(h)w\rangle$. Since the coefficients, c_h^i, depend only on the numbers $\langle v|D^{(\nu)}(h')|v\rangle = \langle w|D^{(\nu)}(h')|w\rangle$, $|e_i'\rangle$ form a basis. Then there is an operator L such that $|e_i'\rangle = L|e_i\rangle$. We claim that $LD^{(\nu)}(h) = D^{(\nu)}(h)L$ for all h in G. First of all, we have $|w\rangle = |e_1'\rangle = L|e_1\rangle = |v\rangle$. Then, $D^{(\nu)}(h)L|v\rangle = D^{(\nu)}(h)|w\rangle = \sum_i b_i^h |e_i'\rangle = \sum_i b_i^h L|e_i\rangle = L\sum_i b_i^h |e_i\rangle = LD^{(\nu)}(h)|v\rangle$. We have

$$\begin{aligned} D^{(\nu)}(h)L|e_i\rangle = D^{(\nu)}(h)|e_i'\rangle &= \sum_{h'\in G} c_{h'}^i D^{(\nu)}(hh')|w\rangle \\ &= \sum_{h'\in G} c_{h^{-1}h'}^i D^{(\nu)}(h')|w\rangle = \sum_{h'\in G} c_{h^{-1}h'}^i D^{(\nu)}(h')L|v\rangle \\ &= L\sum_{h'\in G} c_{h^{-1}h'}^i D^{(\nu)}(h')|v\rangle = LD^{(\nu)}(h)|e_i\rangle \end{aligned} \tag{4.45}$$

This proves the claim. Since ν is irreducible, by Schur's lemma, L is the product of a constant and the identity operator and $|w\rangle = e^{i\theta}|v\rangle$ because $\langle w|w\rangle = \langle v|v\rangle = 1$.

For a beam of dyons, we can first measure the magnetic flux by sending charged particles around them. After we know the flux, we can measure their charge by using vortices with flux which commutes with the flux of the dyons. Then, we completely determine the state of the dyons in some representation of the quantum double.

This analysis can be generalized to the measurement of a single particle or particles in reducible representation (with some limitation).[8]

REFERENCES

1. J. Preskill, Vortices and monopoles, *in* Architecture of the fundamental interactions at short distances, ed. P. Ramond and R. Stora (North-Holland, Amsterdam, 1987).
2. Y. Aharonov and D. Bohm, Phys. Rev. **119** (1959) 485.
3. A. S. Schwarz, Nucl. Phys. **B208** (1982) 141.
4. L. Krauss and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1221.
5. M. Alford, K. Benson, S. Coleman, J. March-Russell and F. Wilczek, Phys. Rev. Lett. **64** (1990) 1632; Nucl. Phys. **B349** (1991) 414.
6. J. Preskill and L. Krauss, Nucl. Phys. **B341** (1990) 50.

7. M. Bucher, Nucl. Phys. **B350** (1991) 163.
8. M. Alford, S. Coleman and J. March-Russell, Nucl. Phys. **B351** (1991) 735.
9. M. Alford and J. March-Russell, Nucl. Phys. **B369** (1992) 276.
10. M.G. Alford, K.-M. Lee, J. March-Russell and J. Preskill, Nucl. Phys. **B384** (1992) 251.
11. M. Bucher, H.-K. Lo and J. Preskill, Nucl. Phys. **B386** (1992) 3.
12. M. Bucher, K.-M. Lee and J. Preskill, Nucl. Phys. **B386** (1992) 27.
13. F.A. Bais, Nucl. Phys. **B170 (FSI)** 32 (1980).
14. F.A. Bais, P. van Driel and M. de Wild Propitius, Phys. Lett. **B280** 63 (1992).
15. F.A. Bais, P. van Driel and M. de Wild Propitius, Nucl. Phys. **B393** 547 (1993).
16. E. Verlinde, in *International Colloquium on Modern Quantum Field Theory*, Bombay, India, 1990, edited by S. Das *er al.* (World Scientific, Singapore, 1991).
17. Y. Ladegaillerie, Bull. Sci. Math. **100** (1976) 255.
18. T.D. Imbo and J. March-Russell, Phys. Lett. **B252** (1990) 84.
19. R. Dijkgraaf, V. Pasquier and P. Roche, Nucl. Phys. B (Proc. Suppl.) **18B** (1990) 60.
20. M.G. Alford, J. March-Russell and F. Wilczek, Nucl. Phys. **B337** (1990) 695.
21. See, e.g., H.M. Farkas and I. Kra, Riemann Surfaces, Springer-Verlag 1980; O. Forster, Lectures on Riemann Surfaces, Springer-Verlag 1981; M.W. Hirsh, Differential Topology, Springer-Verlag 1976.
22. M.G.G. Laidlaw and C. De Witt-Morette, Phys. Rev. **D3** (1971) 1375.
23. M. Leinaas and J. Myrheim, Nuovo Cimento **B37** (1977) 1.
24. See, e.g., J. Frohlich, F. Gabbiani and P.A. Marchetti, in *Knots, Topology and Quantum Field Theory*, Proceedings of the 13th Johns Hopkins Workshop on Current Problems in High Energy Particle Theory, Florence, Italy, 1989, edited by L. Lusanna (World Scientific, Singapore, 1989), p. 335.

25. P. Nelson and A. Manohar, Phys. Rev. Lett. **50** (1983) 943; A. P. Balachandran, G. Marmo, N. Mukunda, J. S. Nilsson, E. C. G. Sudarshan and F. Zaccaria, Phys. Rev. Lett. **50** (1983) 1553; P. Nelson and S. Coleman, Nucl. Phys. **B237** (1984) 1; A. P. Balachandran, F. Lizzi and V. G. Rodgers, Phys. Rev. Lett. **52** (1984) 1818; P. A. Horváthy and J. H. Rawnsley, J. Math. Phys. **27** (1986) 982.

26. P. Bantay, Phys. Lett. **B245** 477 (1990); Lett. Math. Phys. **22** (1991) 187; D. Altschuler and A. Coste (Marseille, CPT), CERN-TH-6638-92, Sep 1992. 12pp. To appear in proceedings of 28th Karpacz Winter School of Theoretical Physics, Karpacz, Poland, Feb 17-29, 1992; D. Altschuler and A. Coste, Commun. Math. Phys. **150** (1992) 83.

27. For a review of quantum doubles and quantum groups, see, e.g., V.G. Drinfel'd, Problems of Modern Quantum Field Theory, Proceedings Alushta 1989, Research Reports in Physics, Springer-Verlag, Heidelberg (1989); T. Tjin, Int. J. Mod. Phys. **A7** (1992) 6175.

28. H.-K. Lo, K.-M. Lee and J. Preskill, Phys. Lett. **B318** (1993) 287.

29. L. Brekke, H. Dykstra, S.J. Hughes and T.D. Imbo, Phys. Lett. **B288** (1992) 273.

Chapter 5: Complementarity in Wormhole Chromodynamics

Many years ago, Wheeler[1] and Misner and Wheeler[2] proposed that electric field lines trapped in the topology of a multiply-connected space might explain the origin of electric charge. Consider a three-dimensional space with a handle (or "wormhole") attached to it, where the cross section of the wormhole is a two-sphere. On this space, the source-free Maxwell equations have a solution with electric field lines caught inside the wormhole throat. One mouth of the wormhole, viewed in isolation by an observer who is unable to resolve the small size of the mouth, cannot be distinguished from a pointlike electric charge. Only when the observer inspects the electric field more closely, with higher resolution, does she discover that the electric field is actually source free everywhere.

It is also interesting to consider what happens when a charged particle traverses a wormhole.* (Of course, this "pointlike" charge might actually be one mouth of a smaller wormhole.) Suppose that, initially, the mouths of the wormhole are uncharged (no electric flux is trapped in the wormhole). By following the electric field lines, we see that after an object with electric charge Q traverses the wormhole, the mouth where it entered the wormhole carries charge Q, and the mouth where it exited carries charge $-Q$. Thus, an electric charge that passes through a wormhole transfers charge to the wormhole mouths.

In this chapter, we will address two (closely related) puzzles associated with this type of charge transfer process. Our first puzzle concerns the quantum mechanics of charged particles in the vicinity of a wormhole. We can compute the amplitude for the particle to propagate from an initial position to a final position by performing a sum over histories. Naively, one would expect this sum to include histories that traverse the wormhole, and that the contribution to the path integral due to these histories should be combined coherently with the contribution due to histories that

---

$\star$ Note that we are assuming that the wormhole is traversable. This assumption would be valid for a non-dynamical wormhole three-geometry, but it is in conflict with the "topological censorship" theorem[3] that can be proved in classical general relativity (with suitable assumptions about the positivity of the energy-momentum tensor). The traversability of the wormhole might be enforced by quantum effects. Alternatively, the reader might prefer to envision our space as a thin two-dimensional film, containing objects with Aharonov-Bohm interactions. Such wormholes might actually be fashioned in the laboratory!

do not traverse the wormhole. In fact, the histories can be classified according to their "winding number" around the wormhole, which can take any integer value, and one expects that all of the winding sectors should be combined coherently. Upon further reflection, though, one sees that, for charged particles, this naive expectation must be incorrect. Long after the final position of the particle has been detected, an observer can measure the charge of one of the wormhole mouths. If the mouth was uncharged initially, and carries charge nQ finally, then the observer concludes that the charged particle must have entered that mouth of the wormhole n times. Because the winding sectors are perfectly correlated with the charge transferred to the mouth, the amplitudes associated with different numbers of windings cannot interfere with one another. The puzzle in this case is to understand more clearly the mechanism that destroys the coherence of the different winding sectors.

Our second puzzle arises in a non-Abelian gauge theory, such as quantum chromodynamics. Suppose that a wormhole initially carries no color charge, and consider what happens when a "red" quark traverses the wormhole. (We can give a gauge-invariant meaning to the notion that the quark is red by establishing a "quark bureau of standards" at some preferred location, and carefully preserving a standard red (R) quark, blue (B) quark, and yellow (Y) quark there. When we say that a quark at another location is red, we mean that if it is parallel transported back to the bureau of standards, its color matches that of the standard R quark. This notion is especially simple if we assume that there are no color magnetic fields, so that parallel transport is unaffected by smooth deformations of the path.) An observer who watches the red quark enter one mouth of the wormhole concludes that the mouth becomes a red source of color electric field.† But the other mouth of the wormhole is initially in a color-singlet state, and it cannot suddenly acquire a long-range color electric field as the quark emerges from the mouth. Thus, after the traversal, the quark and mouth must be in the color-singlet state

$$\frac{1}{\sqrt{3}}\left(|R\rangle_{\text{quark}} \otimes |\bar{R}\rangle_{\text{mouth}} + |B\rangle_{\text{quark}} \otimes |\bar{B}\rangle_{\text{mouth}} + |Y\rangle_{\text{quark}} \otimes |\bar{Y}\rangle_{\text{mouth}}\right) . \qquad (5.1)$$

The puzzle in this case is to understand why the quark that emerges from the wormhole is not simply in the color state $|R\rangle$, and how the correlation between the color

† We are assuming that the wormhole is being examined on a sufficiently short distance scale that the effects of color confinement can be neglected.

of the quark and the color of the mouth is established.

The resolution of these puzzles involves some peculiar features of the Aharonov-Bohm effect[4] on non-simply connected manifolds. The essential concept is the magnetic flux "linked" by the wormhole. If a particle with charge Q is carried around a closed path that traverses a wormhole (in a $U(1)$ gauge theory), it in general acquires an Aharonov-Bohm phase $e^{iQ\Phi}$, where Φ is the flux associated with the path. (This flux is defined modulo the flux quantum $\Phi_0 = 2\pi/e$, where e is the charge quantum.) If magnetic field strengths vanish everywhere, this flux is a topological invariant, unchanged by smooth deformations of the path. The crucial point is that the flux Φ and the charge of a wormhole mouth are complementary observables—if the mouth has a definite charge (like zero), then the flux does not take a definite value. Summing over the different possible values of the flux generates the decoherence of the winding sectors described above, and also (in the non-Abelian case) causes the red quark that traverses the wormhole to emerge in the state Eq. (5.1).

5.1. Wormhole complementarity

Let us now analyze these Aharonov-Bohm interactions in greater detail. We will use a notation that is appropriate when the gauge group G is a finite group. This will serve to remind the reader that our analysis applies to the case of a *local discrete symmetry*.[5–7] For the case of a continuous gauge group, one need only replace sums by integrals in some of the expressions below. When the gauge group is discrete (and also when it is continuous), the electric charge of an object, including a wormhole mouth, can be measured in principle by scattering a loop of *cosmic string* (or a closed magnetic solenoid) off of the object. For ease of visualization, we will carry out our explicit analysis for the case of two spatial dimensions, so that charges are measured by scattering magnetic *vortices*. The analysis in three spatial dimensions is similar.

There are actually two types of topological magnetic flux associated with a wormhole in two spatial dimensions, for there are two topologically distinct paths for which Aharonov-Bohm phases can be measured, as shown in Fig. 1. The path α encloses one mouth of the wormhole, and we will denote the group element associated with parallel transport around this path as $a \in G$. The path β passes through both

wormhole mouths, and we denote the associated group element as $b \in G$. We refer to these group elements as the α-flux and β-flux of the wormhole, and denote the corresponding quantum state of the wormhole as $|a, b\rangle_{\text{wormhole}}$.

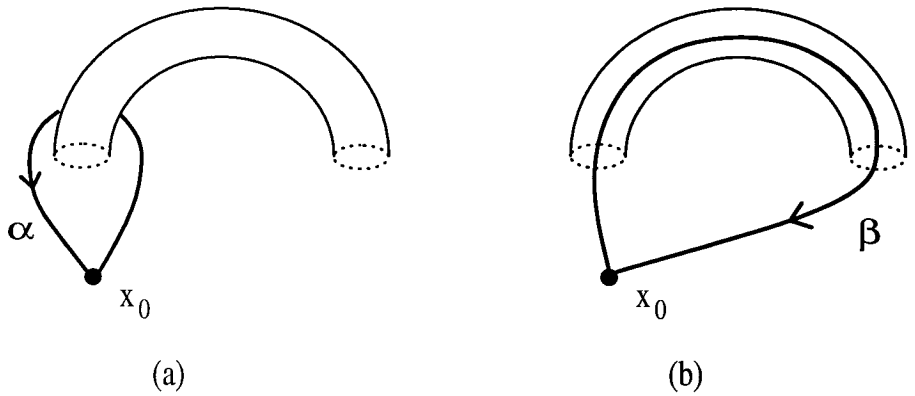

Fig. 1: Two non-contractible paths α and β, beginning and ending at an arbitrarily chosen basepoint x_0, on the wormhole geometry. The group elements associated with parallel transport around these paths are the α-flux and β-flux of the wormhole.

(In three spatial dimensions, we may wish to consider a wormhole that has the topology of $M \times R$, where M is a Riemann surface. In the case most commonly of interest, M is a two sphere; in that case, there is no non-contractible path that "wraps around" one mouth of the wormhole, and hence there is no topological α-flux.)

Now, we can measure the *electric* charge of a wormhole mouth by winding a vortex around the mouth, and observing the Aharonov-Bohm phase acquired by the vortex. However, winding the vortex around the mouth will also change the state $|a, b\rangle$ of the wormhole. For our purposes, it will be sufficient to consider the special case in which $a = e$, the identity. (A more general analysis of non-Abelian Aharonov-Bohm interactions on topologically non-trivial spaces can found in Ref. 8). As shown in Fig. 2, we may enclose the vortex with a closed path γ; we denote the group element associated with transport around γ as $h \in G$, and refer to it as the flux of the vortex. As the vortex winds counterclockwise around the wormhole mouth, the path $\beta\gamma^{-1}$ is deformed to β. (Here, $\beta\gamma^{-1}$ denotes the path that is obtained by tracing γ^{-1} *first*, followed by β.) Thus, when the vortex winds around the mouth, the flux associated with $\beta\gamma^{-1}$ *before* the winding becomes the flux associated with β after the winding;

we conclude that the state of wormhole and vortex is modified according to

$$|e,b\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \rightarrow |e,bh^{-1}\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \ . \tag{5.2}$$

Eq. (5.2) is the centerpiece of our analysis. It says that if the wormhole is in the "flux eigenstate" $|e,b\rangle$, then any attempt to use Aharonov-Bohm interference to measure the electric charge of one mouth is doomed to failure. If we scatter a vortex off of the mouth (with vortex flux $h \neq e$), whether the vortex passed to the left or the right of the mouth is perfectly correlated with the state of the wormhole, and therefore no interference is seen; the probability distribution of the scattered vortex is the incoherent sum of the probability distributions for vortices that pass to the left and pass to the right.

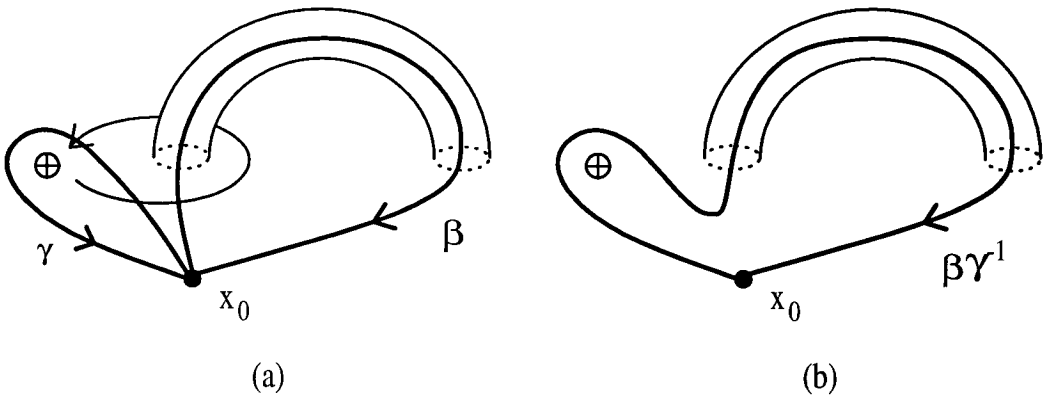

Fig. 2: A vortex winds around one mouth of the wormhole, as shown in (a). If the path $\beta\gamma^{-1}$ shown in (b) is deformed during the winding of the vortex, so that the vortex never crosses the path, $\beta\gamma^{-1}$ evolves to the path β.

However, by superposing the wormhole states of definite β-flux, we can construct states with definite charge. (We need only decompose the regular representation of G into irreducible representations.) In particular, in the state

$$|0\rangle_{\text{wormhole}} = \frac{1}{\sqrt{n_G}} \sum_{b \in G} |e,b\rangle_{\text{wormhole}} \tag{5.3}$$

(where n_G is the order of the group G), each mouth of the wormhole has zero charge. To see this, consider carrying the h-vortex around one mouth of this wormhole. It

is easy to see that the state of the wormhole is unmodified, so that the Aharonov-Bohm phase acquired by the vortex is trivial. On the other hand, suppose that we try to measure the β-flux of the wormhole by carrying a charged particle along the path β. Let us denote the initial state of the particle as $|v\rangle_{\text{particle}}$, and let (ν) be the irreducible representation of G according to which the state transforms. Then if we carry this particle around the path β where the wormhole is initially in the state $|0\rangle_{\text{wormhole}}$, the state of particle and wormhole is modified according to

$$\begin{aligned} &|\text{initial}\rangle \equiv |v\rangle_{\text{particle}} \otimes |0\rangle_{\text{wormhole}} \rightarrow \\ &|\text{final}\rangle \equiv \frac{1}{\sqrt{n_G}} \sum_{b\in G} D^{(\nu)}(b)\,|v\rangle_{\text{particle}} \otimes |e,b\rangle_{\text{wormhole}} \ ; \end{aligned} \tag{5.4}$$

thus the overlap of the final state with the initial state is

$$\langle \text{final}\,|\text{initial}\rangle = \frac{1}{n_G} \sum_{b\in G} \langle v|\, D^{(\nu)}(b)\,|v\rangle = \begin{cases} 1, & \text{if } (\nu) = \text{trivial} \ ; \\ 0, & \text{otherwise} \ . \end{cases} \tag{5.5}$$

Unless (ν) is trivial, the state of the particle that has been carried through the wormhole is orthogonal to the original state. Hence we recover our earlier conclusion that, for charged particles propagating on the wormhole geometry, paths that traverse the wormhole add incoherently with paths that do not.

We see that the wormhole cannot simultaneously have a definite β-flux and a definite charge. We call this property "wormhole complementarity." It is intimately related to the complementary connection between magnetic and electric flux that was first emphasized by 't Hooft,[9] and was generalized to the non-Abelian case in Ref. 10.

By decomposing the regular representation Eq. (5.2) into irreducible representations, we obtain states in which the wormhole *mouth* has a definite charge. The charge of a mouth should not be confused with the "Cheshire charge"[6,11] carried by the whole wormhole. To measure the charge of the whole wormhole, we would wind a vortex around *both* mouths of the wormhole. In this process, the state of vortex and wormhole is modified according to[8]

$$\begin{aligned} &|a,b\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \rightarrow \\ &\left|hah^{-1}, hbh^{-1}\right\rangle_{\text{wormhole}} \otimes \left|h\left(aba^{-1}b^{-1}\right)h\left(aba^{-1}b^{-1}\right)^{-1}h^{-1}\right\rangle_{\text{vortex}} . \end{aligned} \tag{5.6}$$

Note that $aba^{-1}b^{-1}$ is the "total flux" of the wormhole, the flux associated with a path that encloses both mouths. Charge measurement is possible only if the initial and

final vortex states are not orthogonal, so that interference can occur. Therefore, the flux h of the vortex must commute with the total flux of the wormhole—the charge that can be detected is actually a representation of $N(aba^{-1}b^{-1})$, the centralizer of the total flux.[12,6,11] States of definite Cheshire charge are obtained by decomposing the wormhole states $|a,b\rangle$ into states that transform irreducibly under the action Eq. (5.6), where $h \in N(aba^{-1}b^{-1})$.

Of course, to an observer with poor resolution, the wormhole mouths look like pointlike particles, and the Cheshire charge of the wormhole coincides with the Cheshire charge of vortex pairs that has been discussed elsewhere.[10,13,14] For example, if $b = e$ then the mouths appear to be a vortex with flux a and an antivortex with flux a^{-1}. In the case $a = e$ that we have considered, neither wormhole mouth carries any flux, and the states $|e,b\rangle_{\text{wormhole}}$ are transformed as

$$|e,b\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \to \left|e, hbh^{-1}\right\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \tag{5.7}$$

when the vortex winds around the wormhole. The states of definite Cheshire charge are obtained by superposing the flux eigenstates $|e,b\rangle_{\text{wormhole}}$, with b taking values in a particular conjugacy class of G. Specifically, the states

$$|0,[b]\rangle_{\text{wormhole}} = \frac{1}{\sqrt{n_{[b]}}} \sum_{b' \in [b]} \left|e,b'\right\rangle_{\text{wormhole}} \tag{5.8}$$

(where $[b]$ denotes the class containing b, and $n_{[b]}$ is the order of that class) have trivial total charge, although each wormhole mouth carries charge in these states.

The peculiar behavior we found for Aharonov-Bohm scattering off of a wormhole mouth, when the wormhole is in a flux eigenstate, can be given a more conventional interpretation if we think of the wormhole as a pair of charged particles in a particular (correlated) state. For example, the flux eigenstate $|e,e\rangle_{\text{wormhole}}$ can be decomposed as

$$\begin{aligned} |e,e\rangle_{\text{wormhole}} &\equiv |0,[e]\rangle_{\text{wormhole}} \\ &= \sum_\nu C_\nu \sum_i \frac{1}{\sqrt{n_\nu}} |e_i,\nu\rangle \otimes |e_i^*,\nu\rangle \ , \quad \sum_\nu |C_\nu|^2 = 1 \, , \end{aligned} \tag{5.9}$$

where the $|e_i,\nu\rangle$'s are a basis for the space on which the irreducible representation (ν) acts, and n_ν is the dimension of this representation. This is a superposition of

states in which the two particles (the mouths) have non-trivial charges, and are in a combined state of trivial charge. Experiments involving one of the mouths are described by a mixed density matrix of the form

$$\rho = \sum_{\nu} |C_\nu|^2 \frac{1}{n_\nu} \mathbf{1}_\nu \,, \tag{5.10}$$

and Aharonov-Bohm scattering of the h-vortex off the mouth enables us to measure

$$\text{tr } D(h)\rho = \sum_{\nu} |C_\nu|^2 \frac{1}{n_\nu} \chi^{(\nu)}(h) = \begin{cases} 1, & h = e \ ; \\ 0, & \text{otherwise} \ , \end{cases} \tag{5.11}$$

where $\chi^{(\nu)}$ denotes the character of the representation. (The second equality in Eq. (5.11) follows from the property Eq. (5.2).) From the group orthogonality relations, we see that $|C_\nu|^2 = n_\nu^2/n_G$. Thus Aharonov-Bohm scattering enables us to determine the probability that the wormhole mouth carries charge (ν), but does not determine the relative phases of the C_ν's.[13] When we think of it as a point particle, the unusual thing about a wormhole mouth is that it is natural to consider a state such that the mouth is in a superposition of particle states with different gauge charges.

5.2. Charge transfer

Now let us suppose that, after the wormhole in the initial state $|0\rangle_{\text{wormhole}}$ is traversed by the charged particle in the initial state $|v\rangle_{\text{particle}}$, we attempt again to measure the charges of the two mouths. If an h-vortex is carried around the mouth that the charged particle *entered*, then the state of wormhole, particle, and vortex is modified according to

$$\begin{aligned} &\frac{1}{\sqrt{n_G}} \sum_{b\in G} D^{(\nu)}(b) \, |v\rangle_{\text{particle}} \otimes |e, b\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \rightarrow \\ &\frac{1}{\sqrt{n_G}} \sum_{b\in G} D^{(\nu)}(b) \, |v\rangle_{\text{particle}} \otimes \left|e, bh^{-1}\right\rangle_{\text{wormhole}} \otimes |h\rangle_{\text{vortex}} \ , \end{aligned} \tag{5.12}$$

so that the overlap of the initial state with the final state is

$$\text{overlap} = \frac{1}{n_G} \sum_{b,b'\in G} \langle v| \, D^{(\nu)}(b')^{-1} D^{(\nu)}(b) \, |v\rangle \cdot \left\langle e, b' \middle| \, e, bh^{-1} \right\rangle = \langle v| \, D^{(\nu)}(h) \, |v\rangle \ . \tag{5.13}$$

This is exactly the same as the overlap we would have obtained if the vortex had been carried around the initial charged particle. Thus, as we anticipated, the charge of the particle has been transferred to the mouth of the wormhole.

But if we measure instead the charge of the *other* mouth, we obtain a rather different result. It is actually most instructive to consider carrying the h-vortex around *both* the charged particle and the other wormhole mouth. A variant of the argument given earlier shows that carrying the vortex counterclockwise around this mouth changes the wormhole state $|e,b\rangle$ to $|e,hb\rangle$. We thus find that the state of wormhole, particle, and vortex is modified according to

$$\begin{aligned}&\frac{1}{\sqrt{n_G}}\sum_{b\in G} D^{(\nu)}(b)\,|v\rangle_{\text{particle}}\otimes|e,b\rangle_{\text{wormhole}}\otimes|h\rangle_{\text{vortex}}\rightarrow\\&\frac{1}{\sqrt{n_G}}\sum_{b\in G} D^{(\nu)}(h)D^{(\nu)}(b)\,|v\rangle_{\text{particle}}\otimes|e,hb\rangle_{\text{wormhole}}\otimes|h\rangle_{\text{vortex}}\,,\end{aligned}\tag{5.14}$$

and that the overlap of the initial state with the final state is

$$\text{overlap}=\frac{1}{n_G}\sum_{b,b'\in G}\langle v|\,D^{(\nu)}(b')^{-1}D^{(\nu)}(hb)\,|v\rangle\cdot\langle e,b'|\,e,hb\rangle=1\,.\tag{5.15}$$

Thus the Aharonov-Bohm phase is trivial, and we conclude that the charged particle and mouth are combined together into a singlet state, again as anticipated.

Eq. (5.1) is a special case of this result. We now understand that if the wormhole mouth initially carries no color charge, that means that the color holonomy associated with traversing the wormhole does not take a definite value. Thus the red quark emerges from the wormhole mouth carrying indefinite color, but with its color perfectly anti-correlated with the color of the mouth. Furthermore, after the (initially) red quark passes through the wormhole, the wormhole state is a superposition of a color octet and color singlet, so that Cheshire charge has been transferred to the wormhole.[⋆]

In summary, we have seen that the β-flux "linked" by a wormhole and the charge of a wormhole mouth cannot simultaneously have definite values. We call this property "wormhole complementarity." If the β-flux has a definite value, then each wormhole mouth is in an incoherent superposition of charge eigenstates. (The charges of the two mouths are correlated, so that each mouth is described by a mixed density matrix.) There is no Aharonov-Bohm interference when a vortex (or cosmic string)

⋆ $SU(3)_{\text{color}}$ Cheshire charge has also been discussed recently by Bucher and Goldhaber.[15]

scatters off the mouth. If the charge of each mouth has a definite value, then the wormhole is in a coherent superposition of β-flux eigenstates. Thus, after a colored particle traverses the wormhole, its color is correlated with that of the wormhole mouth from which it emerged.

REFERENCES

1. J. A. Wheeler, Ann. Phys. (N. Y.) **2** (1957) 604; J. A. Wheeler, *Geometrodynamics*, Academic Press, New York (1962).
2. C. W. Misner and J. A. Wheeler, Ann. Phys. (N. Y.) **2** (1957) 525.
3. J. Friedman, K. Schleich, and D. Witt, "Topological censorship," ITP preprint NSF-ITP-93-80 (1993).
4. Y. Aharonov and D. Bohm, Phys. Rev. **119** (1959) 485.
5. L. Krauss and F. Wilczek, Phys. Rev. Lett. **62** (1989) 1221.
6. J. Preskill and L. Krauss, Nucl. Phys. **B341** (1990) 50.
7. M. Alford, J. March-Russell, and F. Wilczek, Nucl. Phys. **B337** (1990) 695.
8. K.-M. Lee, Phys. Rev. **D49** (1994) 2030.
9. G. 't Hooft, Nucl. Phys. **B138** (1978) 1; Nucl. Phys. **B153** (1979) 141.
10. M. Alford, K.-M. Lee, J. March-Russell, and J. Preskill, Nucl. Phys. **B384** (1992) 251; M. Bucher, K.-M. Lee, and J. Preskill, Nucl. Phys. **B386** (1992) 27.
11. M. G. Alford, K. Benson, S. Coleman, J. March-Russell and F. Wilczek, Phys. Rev. Lett. **64** (1990) 1632; Nucl. Phys. **B349** (1991) 414.
12. A. P. Balachandran, F. Lizzi, and V. Rogers, Phys. Rev. Lett. **52** (1984) 1818.
13. M. Alford, S. Coleman, and J. March-Russell, Nucl. Phys. **B351** (1991) 735.
14. H.-K. Lo and J. Preskill, Phys. Rev. **D48** (1993) 4821.
15. M. Bucher and A. S. Goldhaber, "$SO(10)$ cosmic strings and $SU(3)_{\text{color}}$ Cheshire charge," Inst. Adv. Study preprint iassns-hep-93-6 (1993).

Chapter 6: Final Remarks

An early motivation for studying discrete gauge theory comes from the black hole paradox. A black hole has very little hair, but it emits thermal radiation. Thus, it seems that any information that goes into a black hole will not come out. Discrete gauge theory tells us that black hole *could* carry quantum hair. We might conclude that discrete gauge theory may have something to do with the paradox, or even solve it.

This expectation did not last very long. There is just not much discrete symmetry in the real world. If we choose any grand unified theory and break it down to the standard model, the number of components of the low energy gauge group is expected to be of order one. However, if the discrete gauge hair did solve the paradox, there would be symmetries for the momentum, the mass and all other quantum numbers of each particle falling into the black hole to encode all the information. This is impossible.

Although discrete gauge theory does not solve the paradox, it is very interesting in its own right. The idea of Cheshire charge is new to us. There is also a magnetic analog of the Cheshire charge.[1] In the "Alice string" model in 3+1 dimensions (which has gauge group equal to the semi-direct product of $U(1)$ and Z_2), the cosmic string loop can carry unlocalized magnetic charge. It can also transfer magnetic charge to a magnetic monopole by means of a topological interaction. It is not difficult to see that a similar analysis to that in chapter 3 will apply to this magnetic case. However, in order to observe these phenomena, there must be some "unusual" non-Abelian gauge symmetries. We have not seen any discrete gauge symmetry in the real world yet. The global analog[3] of the Aharonov-Bohm effect and Cheshire charge point to the possibility of a similar phenomenon. In fact, superfluid ^{3}He-A has a non-Abelian global symmetry and a realization of Cheshire charge may be possible in it.

We do not have any experimental evidence of the gauged or global Cheshire charge yet. However, there is no reason, in principle, why they could not be found in some condensed matter system. Then, a thin film in that system will be a two-dimensional space and the topological interactions discussed in chapter 4 between the topology of the space and dyons could be realized in a lab.

Apart from the possible connections with condensed matter systems, the topological interaction is related to topological quantum field theory.[4] The dyons in chapter 4 are singularities of the gauge fields. They could also be thought of as boundaries of the space manifold. Then, a state in chapter 4 is a state of a topological quantum field theory with finite gauge group.[5] It is thus tempting to speculate that topological quantum field theory could be realized in a lab!

REFERENCES

1. M. Bucher, H.-K. Lo and J. Preskill, Nucl. Phys. **B386** (1992) 3.
2. T. Vachaspati, Phys. Rev. Lett. **68** (1992) 1977.
3. M. V. Khazan, Pis'ma Zh. Eksp. Teor. Fiz **41** (1985) 396; J. March-Russell, J. Preskill and F. Wilczek, Phys. Rev. Lett. **68** (1992) 2567; A. C. Davis and A. P. Martin, "A Global Analog of the Aharonov-Bohm Effect," DAMTP preprint, DAMTP 93-09 (1993), "Global Strings and the Aharonov-Bohm Effect," DAMTP preprint, DAMTP 94-50 (1993); R. Navin, "Global Analogue of the Aharonov-Bohm Effect," Caltech preprint CALT-68-1869; P. McGraw, "A Global Analogue of Cheshire Charge," Caltech preprint CALT-68-1865.
4. For general information of topological quantum field theory, see, e.g., D. Birmingham, M. Blau, M. Rakowski and G. Thompson, Phys. Rep. **209** 129 and references therein.
5. R. Dijkgraaf and E. Witten, Commun. Math. Phys. **129** (1990) 393; D. S. Freed and F. Quinn, Commun. Math. Phys. **156** (1993) 435.