p-form electromagnetism on discrete spacetimes*

INSTITUTE OF PHYSICS PUBLISHING CLASSICAL AND QUANTUM GRAVITY

Class. Quantum Grav. **23** (2006) 5129–5176 doi:10.1088/0264-9381/23/17/004

p-form electromagnetism on discrete spacetimes*

Derek K Wise

Department of Mathematics, University of California, Riverside, CA 92521, USA

E-mail: derek@math.ucr.edu

Received 25 February 2006, in final form 21 June 2006
Published 7 August 2006
Online at stacks.iop.org/CQG/23/5129

Abstract

We investigate p-form electromagnetism—with the Maxwell and Kalb–Ramond fields as lowest-order cases—on discrete spacetimes, including not only the regular lattices commonly used in lattice gauge theory, but also more general examples. After constructing a maximally general model of discrete spacetime suitable for our purpose—a chain complex equipped with an inner product on $(p+1)$-cochains—we study both the classical and quantum versions of the theory, with either $\mathbb{R}$ or $U(1)$ as gauge group. We find results—such as a 'p-form Bohm–Aharonov effect'—that depend in interesting ways on the cohomology of spacetime. We quantize the theory via the Euclidean path integral formalism, where the natural kernels in the $U(1)$ theory are not Gaussians but theta functions. As a special case of the general theory, we show that p-form electromagnetism in $p+1$ dimensions has an exact solution which reduces when $p = 1$ to the Abelian case of 2D Yang–Mills theory as studied by Migdal and Witten. Our main result describes p-form electromagnetism as a 'chain field theory'—a theory analogous to a topological quantum field theory, but with chain complexes replacing manifolds. This makes precise a notion of time evolution in the context of discrete spacetimes of arbitrary topology.

PACS numbers: 04.60.Nc, 01.15.Ha, 02.10.Ws

1. Introduction

For the relativist, describing electromagnetism as a gauge theory has obvious appeal. The language of fibre bundles in which gauge theory is written provides a coherent framework for doing field theory in curved spacetimes of arbitrary dimension and with very general global topology. The bundle formalism emphasizes the idea that there is no canonical way to compare 'states at x' with 'states at y' when x and y are distinct points of spacetime—the notion of a connection is central. Briefly, gauge theory has much of the sort of flexibility that relativists like to demand.

* A somewhat less technical, more expository version of this material appeared as gr-qc/0510033 [38].

In contrast, *discrete* analogues of gauge theory tend to be much more rigid. Field theories are often discretized using lattice methods [26, 32], in which Euclidean spacetime is approximated by a regular hypercubical lattice $a\mathbb{Z}^4$, where a is some fixed length called the lattice spacing. The lattice way of doing field theory has certain advantages, both as a means of regularizing path integrals and because numerical computations necessitate discretization of fields. But lattice gauge theory also represents a certain theoretical compromise. As Zee has pointed out (see chapter VII.1 of his textbook [41]), lattice gauge theory rescues gauge invariance from the 'mangling' it suffers in perturbative field theory—including the introduction of unphysical 'ghost fields' necessary to make sense of the results—but does so at the expense of Lorentz invariance. The discretization shatters the symmetry of spacetime. The lattice picks out preferred directions. And the situation is worse for those interested in spacetimes more general than affine Minkowski space: there is no such thing as a diffeomorphism-invariant lattice.

Of course, lattice gauge theory is often considered merely as a computational tool for doing calculations in ordinary continuum gauge theory. One expects to recover Lorentz invariance in the limit $a \to 0$, after Wick rotation. If we are really interested in *manifestly* discrete theories of physics, however, the success of lattice gauge theory in its usual form is not very satisfying: we would like to have 'unmangled' gauge invariance without doing such drastic damage to spacetime's symmetry. In the context of quantum gravity, where spacetime itself is a dynamical variable, the 'spin foam' approach avoids picking out preferred directions in spacetime by integrating over possible discretizations; in lattice electromagnetism we should at least like to have a theory which allows discretizations more general than the hypercubical lattices of lattice gauge theory, with the hope of recovering something that looks like diffeomorphism invariance at large scales.

More generally, we would like such a flexible discretization of *p-form electromagnetism*, the generalization of Maxwell's theory one obtains by replacing the gauge field A by a p-form [1]. More precisely, the gauge field in electromagnetism is a connection on a $U(1)$-bundle, which may be thought of in local coordinates as a 1-form

$$A = A_\mu dx^\mu,$$

and *this* may be generalized to a p-form:

$$A = A_{\mu_1 \ldots \mu_p} \frac{1}{p!} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_p}.$$

The electromagnetic field, locally the 2-form $F = dA$, is globally just the curvature of the connection A. In direct analogy, the p-form electromagnetic field is described locally by the $(p+1)$-form

$$F = dA.$$

The classical equations of motion for p-form electromagnetism, which we might call the *p-form Maxwell equations*, look formally identical to the 1-form case [6]:

$$dF = 0, \qquad \star d \star F = J.$$

But here, since F is a $(p+1)$-form, $\star d \star F$ is a p-form, so the *current* J must also be a p-form. One sees how the dimensions get boosted: whereas the current in ordinary electromagnetism is fundamentally one dimensional, corresponding to zero-dimensional point particles tracing out one-dimensional curves, 2-form electromagnetism has a two-dimensional current. One might expect such a theory to describe one-dimensional charges sweeping out surfaces in spacetime. It is thus not surprising that something like this shows up in string theory, where it is called the 'Kalb–Ramond field' [20, 25]. Integrating the Kalb–Ramond field over the

string's worldsheet gives a term in the action, analogous to the term in the action of a charged particle given by integrating the vector potential A along the particle's worldline.

A natural question which has been pursued by Baez and others is how to describe p-form electromagnetism, and other 'higher gauge theories' [7, 34] with more general gauge group, *globally*. In fact, putting 'p-form gauge theory' on the same global footing as ordinary gauge theory requires some higher-dimensional generalization of the notion of fibre bundle [7, 9, 11, 27]. In particular, one needs to know what is meant by parallel transporting a $(p-1)$-dimensional extended object along a p-dimensional submanifold, rather than merely translating point particles along paths.

Remarkably, many of the technical difficulties of defining 'higher gauge theory' melt away in a discrete setting when the gauge group is Abelian, that is, in the case of discrete p-form electromagnetism! The need for the bundle formalism in gauge theory arises from the possibility that the gauge field might be 'twisted' in a global way. But in the lattice context, changes in the gauge field occur not continuously but in discrete steps. It therefore becomes impossible to decide if the field is 'twisted' around some noncontractible loop. Moreover, the Abelian nature of electromagnetism allows higher-dimensional generalization to be carried out in a straightforward way. Rather than increasingly rich algebraic structures necessary for boosting the dimension in non-Abelian gauge theory, higher-dimensional electromagnetism can always be described in terms of chain complexes of Abelian groups. In fact, p-form electromagnetism on discrete spacetime turns out to be no more difficult than ordinary electromagnetism.

The author hopes that this detailed and pedagogical study of discrete p-form electromagnetism will serve as a starting point for further study in both lattice gauge theory and higher-dimensional analogues of gauge theories, both fields of active research, as well as theories at their interface. A subsidiary motivation is to smooth the reader's way towards an understanding of discrete models of spacetime which arise in the study of quantum gravity, especially 'spin foams'. It is also hoped that the expository flavour of this paper will make it accessible to a wider variety of readers, including physicists wishing to see an elegant treatment of lattice gauge theory and mathematicians wishing to see applications of homological algebra and category theory in physics.

The plan of the paper is as follows. Section 2 serves as an invitation to much of the material that follows it, with the intent of explaining as quickly as possible the path integral approach to discrete quantum electromagnetism. As typical in lattice gauge theory, we work in this section and throughout the paper in Euclidean (i.e. 'Wick rotated') spacetime. In this section, and until section 7, we consider p-form electromagnetism with gauge group $\mathbb{R}$, as opposed to the usual $U(1)$. Readers already well acquainted with path integrals in lattice gauge theory may wish to skim through this to fix notation and then move on to section 3.

In section 3, we begin a more careful approach, considering first which mathematical model of discrete spacetime is appropriate for the description of an Abelian p-form gauge theory. The essential ingredient is a chain complex of free Abelian groups, with bases corresponding to the 'atomic' cells of discrete spacetime.

In section 4, we first define the fields of p-form electromagnetism on the spacetime model developed in the previous section. We then draw a comparison to the continuum theory and use this to construct an appropriate form for the action in the discrete context. In the final subsection, it shown that, just as ordinary gauge theories are particularly simple in two dimensions, p-form electromagnetism simplifies greatly in $p+1$ dimensions. In fact, the natural choice for the action leads to a theory which is 'almost' a topological quantum field theory—it is what we call a 'volumetric field theory'—where the volume of spacetime is the only nontopological input.

Section 5 considers the classical theory arising from the action derived in section 4. It is shown that, quite surprisingly, under certain conditions classical p-form electromagnetism has a strongly topological character. More precisely, when these conditions are met, the space of classical solutions in the absence of charged matter depends only on the topology of spacetime.

In section 6, a topological criterion is found for the convergence of the path integral when the gauge group is $\mathbb{R}$, and path integrals are computed explicitly in examples. The p-form analogue of the Bohm–Aharonov effect is also discussed.

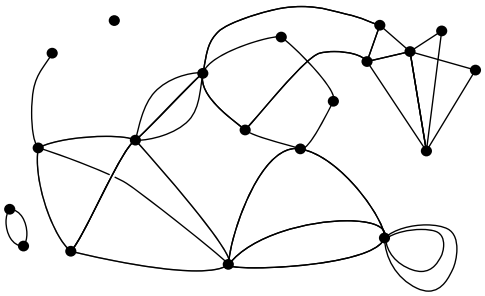
In section 7, we examine the relationship between the groups $\mathbb{R}$ and $U(1)$ as gauge groups for electromagnetism and use the formalism developed in previous sections for the $\mathbb{R}$ theory to set up path integrals in the more usual $U(1)$ case. Since with the latter gauge group the configuration space is a torus, these path integrals turn out to involve not Gaussian integrals in the ordinary sense, but the analogous integrals on tori, which are integrals of theta functions. We consider these theta function integrals in detail in section 7.2 before setting up the path integral for the $U(1)$ theory in section 7.3.

Before concluding section 7, we return in 7.4 to the case of p-form electromagnetism in $p+1$ dimensions, this time in the $U(1)$ setting, and make explicit the connection to 2D Yang–Mills theory. We compute the partition function for lattice p-form electromagnetism on any discretization of an arbitrary compact oriented $(p+1)$-manifold.

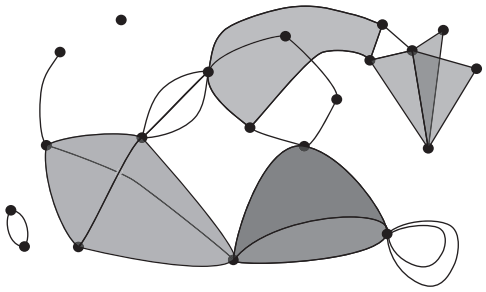
The final section of the paper, section 8, presents the main result of this work: a description of lattice p-form electromagnetism that parallels the category-theoretic description of topological field theory. The importance of this result is that it gives a rigorous description of *time evolution* in lattice p-form electromagnetism. Namely, it allows an n-dimensional discrete spacetime to be split into $(n-1)$-dimensional slices representing space and provides time evolution operators between the Hilbert spaces of states on these. This final section is somewhat more technically demanding than the rest of the paper. In particular, it assumes at least a passing knowledge of categories and functors. For those readers less familiar with category theory, a highly readable introduction which also emphasizes applications to physics is the book by Geroch [18]. The more technical aspects of category theory needed in this section are all covered in Mac Lane's textbook [28].

2. Survey: from *n*-graphs to path integrals

A discrete model of spacetime might begin with a simple discrete set of points, the 'events'. Soon however, if we want to write theories that look like their desired continuum limits, we are forced to equip our model with additional data. In the context of gauge theory, for example, we would like to be able to define an analogue of the connection and take holonomies along paths, and this requires that we specify ways of getting from one event to another. This has the effect of turning our model into a *graph*, with events as vertices and paths between events as finite strings of edges:



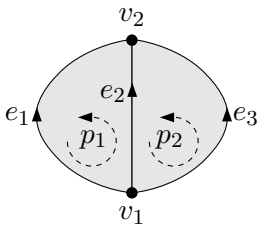
Likewise, the desire to give meaning to notions such as area and curvature makes it natural to fill in empty spaces bounded by edges with two-dimensional faces, or *plaquettes*. Thus, our model becomes a '2-graph':



In a similar way, we are led to adjoin three-dimensional cells bounded by plaquettes, and so on, all the way up to the highest dimension of the spacetime. In general, an n-dimensional discrete spacetime should be constructed from $n + 1$ distinct classes of geometric objects: vertices, edges between vertices, plaquettes bounded by edges, 3-cells bounded by plaquettes, ... and n-cells bounded by $(n - 1)$-cells. Spacetime thus becomes an 'n-graph'.

The problem is that nobody knows what an n-graph is! Of course, we have some solid ideas of what an n-graph should be like, but there seems to be no accepted intrinsic definition that is flexible enough to encompass every would-be example. We take up this issue in section 3.

For now, let us rely on our intuitive notion of a 2-graph as a combinatorial object consisting of finite sets of vertices, edges and plaquettes, and begin with a naive attempt at discretizing quantum electromagnetism in two dimensions. Let V be the set of *vertices*, E the set of *edges* and P the set of *plaquettes* or *faces*. An edge is just a line segment or curve connecting one vertex to another, while a plaquette is a two-dimensional surface whose boundary consists of one or more edges. A simple example, with two vertices, three edges and two plaquettes, is depicted below:



where we label the elements of V, E and P using lower-case letters v, e and p, with subscripts to distinguish between them. As a matter of convenience in computations, we have also assigned an arbitrary orientation to each cell.

The usual gauge group for electromagnetism is $U(1)$, but there is another obvious choice, namely $\mathbb{R}$. We will follow the latter alternative in the first part of this paper, both because a noncompact gauge group in electromagnetism is interesting in its own right and because the construction of the $U(1)$ path integral in section 7 is made easier having established the formalism with gauge group $\mathbb{R}$. In lattice gauge theory, the *gauge field* or *connection* A assigns to each edge in the lattice an element of the gauge group:

$$A\colon E \to \mathbb{R},$$

so the space of connections is just

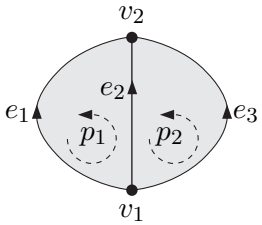
$$\mathcal{A} = \mathbb{R}^E.$$

Suppose γ is an *edge path* through the spacetime lattice—a concatenation $\gamma = e_1^{\varepsilon_1} e_2^{\varepsilon_2} \cdots e_n^{\varepsilon_n}$ of oriented edges, where $\varepsilon_i = \pm 1$ with e_i^{-1} denoting the edge $e_i = e_i^{+1}$ with its orientation reversed. We define the *holonomy* of A along the path γ to be the sum

$$H(A, \gamma) = \varepsilon_1 A(e_1) + \varepsilon_2 A(e_2) + \cdots + \varepsilon_n A(e_n),$$

where we use additive notation since we are using $\mathbb{R}$ as the gauge group. For gauge theories with multiplicative gauge groups, the holonomy of course becomes $H(A, \gamma) = A(e_1)^{\varepsilon_1} A(e_2)^{\varepsilon_2} \cdots A(e_n)^{\varepsilon_n}$.

A discrete connection is *flat* if its holonomy around any contractible loop is the identity. We therefore define the *curvature* or *field strength* $F(p)$ on a plaquette $p \in P$ to be the holonomy of the connection A around the loop that forms its boundary ∂p. For example, in the lattice



the curvature on plaquette p_1 is $A_2 - A_1$, while the curvature on p_2 is $A_3 - A_2$, where $A_i := A(e_i)$. Note that since the gauge group is Abelian, we need not pick a starting vertex for the loop—we may traverse the edges in any order, so long as we take orientations into account. This fact will be of great use to us in section 3.2.

As is standard practice in lattice gauge theory, we quantize this theory using the Euclidean path integral approach. We must therefore specify the form of the action for the gauge field. Roughly, when there are no sources present, the action should be a positive quadratic function of the curvature, so that the connections with zero curvature have the least action, hence are most probable. In section 4, we discuss in more detail what the action for a given connection should look like. In the case of lattice $\mathbb{R}$ electromagnetism in two dimensions, we find that the best choice for the action is

$$S(A) = \frac{1}{2e^2} \sum_{p_i \in P} \frac{{F_i}^2}{V_i},$$

where $F_i := F(p_i)$ is the curvature induced by the connection on the ith plaquette, V_i is the area of the ith plaquette and e^2 is the square of the charge on the electron (or if you wish, the fine structure constant $\alpha = e^2/\hbar c$, since we are using units with $\hbar = c = 1$).

An observable O in our theory is a real-valued function of the gauge field:

$$O\colon \mathcal{A} \to \mathbb{R},$$

so we can try to calculate its expected value, using our action $S(A)$, as

$$\langle O \rangle = \frac{\int_{\mathcal{A}} O(A)\, \mathrm{e}^{-S(A)} \mathcal{D}A}{\int_{\mathcal{A}} \mathrm{e}^{-S(A)} \mathcal{D}A}.$$

In certain cases, it might be difficult to define exactly what this means. For example, if our lattice is infinite, then we must deal with the difficulties of defining measures on infinite-dimensional Euclidean space. However, even in the finite-dimensional case, where $\mathcal{D}A = d^E A$ is just Lebesgue measure on $\mathcal{A} = \mathbb{R}^E$, path integrals like that for $\langle O \rangle$ above are almost always ill defined. To see this, imagine calculating the partition function

$$Z := \int_{\mathcal{A}} \mathrm{e}^{-S(A)} \mathcal{D}A,$$

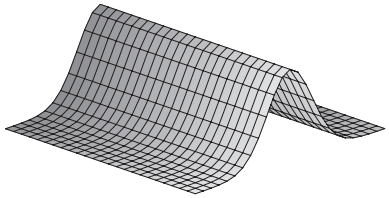
say for the little two-plaquette complex above, to be definite. If the plaquette areas are V_1 and V_2, then the action for this example becomes

$$S(A) = \frac{1}{2e^2}\left(\frac{(A_1 - A_2)^2}{V_1} + \frac{(A_2 - A_3)^2}{V_2}\right).$$

Note that this action is unaffected by the gauge transformation

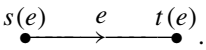
$$(A_1, A_2, A_3) \mapsto (A_1 + d\phi, A_2 + d\phi, A_3 + d\phi), \qquad d\phi \in \mathbb{R},$$

so that one cross section of the integrand $\exp(-S)$ looks like this:



where the direction of gauge freedom is obvious. In particular, $\exp(-S)$ equals 1 along an entire line through the origin of $\mathcal{A} = \mathbb{R}^E$. Because the gauge group $\mathbb{R}$ is noncompact, the partition function therefore diverges.

Let us recall how gauge freedom manifests itself in the lattice context. First, for each directed edge $e \in E$ in the lattice, let $s(e), t(e) \in V$ denote the *source* and *target* vertices of e, respectively,



Then let ϕ be any function of the vertices of spacetime valued in the gauge group, in our case

$$\phi: V \to \mathbb{R},$$

and define the *differential* of ϕ by

$$d\phi: E \to \mathbb{R}, \qquad d\phi(e) = \phi(t(e)) - \phi(s(e)).$$

It is easy to see that a gauge transformation

$$A \mapsto A + d\phi$$

does not affect the holonomy of A around a loop, since the contributions from $d\phi$ all cancel in pairs. It is thus convenient to think of a gauge transformation as being a map from the set of vertices to the gauge group, in our case an element of

$$\mathcal{G} = \mathbb{R}^V.$$

A common way of eliminating divergences caused by gauge freedom is *gauge fixing*. Fixing a gauge means choosing some method to pick one representative of each gauge-equivalence class $[A]$ and doing path integrals over just these. There are problems with this approach, however. In a general gauge theory, it might not even be possible to fix a smooth, global gauge over which to integrate. But even when we can do this, the arbitrary choice involved in fixing a gauge is undesirable, philosophically. Gauge fixing amounts to pretending that a *quotient space* of $\mathcal{A}$, the space of *connections modulo gauge transformations*, is a *subspace*. A better approach is to use the quotient space directly. Namely, $\mathcal{G}$ is a group which acts on the space $\mathcal{A}$ of connections by

$$\mathcal{G} \times \mathcal{A} \to \mathcal{A}, \qquad (\phi, A) \mapsto A + d\phi.$$

Modding out by the action of $\mathcal{G}$ on $\mathcal{A}$ gives the quotient space $\mathcal{A}/\mathcal{G}$ consisting of gauge-equivalence classes of connections, and we do path integrals like

$$Z = \int_{\mathcal{A}/\mathcal{G}} \mathrm{e}^{-S([A])} d[A].$$

When gauge fixing works, path integrals over $\mathcal{A}/\mathcal{G}$ give the same results as gauge fixing. But integrating over $\mathcal{A}/\mathcal{G}$ is more general and involves no arbitrary choices. Since a physical observable in gauge theory is required to be gauge invariant, it can be considered a function on the physical configuration space $\mathcal{A}/\mathcal{G}$. We calculate path integrals on $\mathcal{A}/\mathcal{G}$ explicitly for the $\mathbb{R}$ theory in section 6.

In the case of noncompact gauge group, even factoring out all of the gauge freedom may be insufficient to regularize our path integrals. In particular, there are certain topological conditions our spacetime must meet for this programme to give convergent path integrals. As a consequence, in the case where the gauge group is $\mathbb{R}$, we must sometimes take more drastic measures if we want finite path integrals, and this leads to some interesting differences between the cases $G = \mathbb{R}$ and $G = U(1)$ as gauge groups for electromagnetism. These differences are related to the Bohm–Aharonov effect, which has higher-dimensional generalizations in the case of p-form electromagnetism. We discuss this issue in section 7, where we also study path integrals in the $U(1)$ theory.

3. Discrete spacetime as an n-complex

Just as the concept of a manifold is fundamental to an elegant treatment of general relativity, lattice gauge theory must give due attention to the spacetime lattice itself. In particular, we should say with some precision what we mean when we describe discrete spacetime as an 'n-graph'. We make no attempt in this section at giving the most general definition of n-graph suitable for describing arbitrary discrete gauge theories. Rather, we show that for *Abelian* gauge theory, we really only need a weaker notion—what we shall call an 'n-complex'.

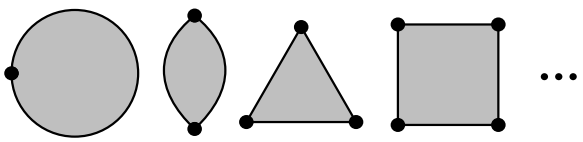
3.1. The n-graph problem

Formulating a precise definition of n-graph is not an easy task. Ideally, we should be able to describe n-graphs in a purely combinatorial way: an n-graph should consist of sets of various kinds of cells, together with maps telling how the cells are linked together. An ordinary (directed) *graph*, for example, is specified by a set V of vertices and a set E of edges, together with *source* and *target* maps

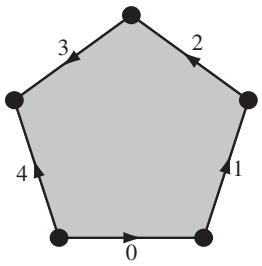
$$s, t: E \to V$$

telling at which vertex each edge begins and ends.

If one tries copying this definition to define a notion of n-graph, one quickly realizes the immensity of the combinatorial problem at hand! Even if we only wish to add two-dimensional faces, or plaquettes, complete generality requires an infinite number of new sets and set maps; whereas an edge has only two ends, a 2D face can have arbitrarily many sides:



What's more, the specified maps must satisfy certain equations. For example, in the pentagonal plaquette



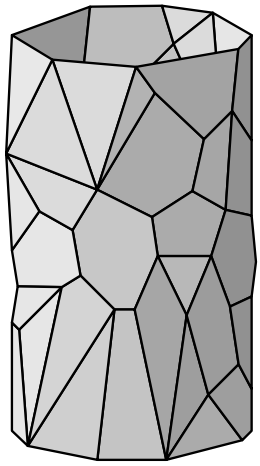
we must impose equations stating that the zeroth and fourth edges have the same source vertex, that the target of the second edge is the source of the third, and so on (three more equations). These equations are called *incidence relations*.

And it gets worse. A 3-cell may have arbitrarily many faces, each of which may have any number of sides . . .

There are various attempts at formalizing the notion of n-graph in a fully satisfactory way, most relying heavily on category theory, or rather n-category theory, which itself is still under development. At the present state of the art, however, we are forced to make compromises of one sort or another. Each of these compromises carries with it its own weaknesses.

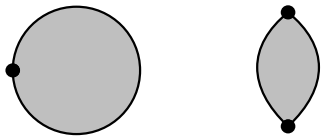
We could, for example, restrict the allowable shapes of cells. Allowing only simplices as cells, for example, lets us define 'simplicial n-graphs' as simplicial sets [29], where we simply remove any cells of dimension greater than n. Similarly, we could define 'cubical n-graphs' using cubical sets [19]. If we are only interested in lattice gauge theory as a computational tool, these should be sufficient, and indeed the study of lattice gauge theory began with regular cubical lattices. However, if we really believe spacetime is discrete at some scale, then there seems to be no *a priori* physical justification for imposing such strong conditions on the shapes of cells!

For this reason, piecewise-linear CW complexes (or PLCW complexes, for short) [23, 24] have been used as discrete models of spacetime in loop quantum gravity [4, 5]. These allow a wider variety of cell shapes and are quite handy if, for example, we want to chop up a manifold with boundary into polyhedra:

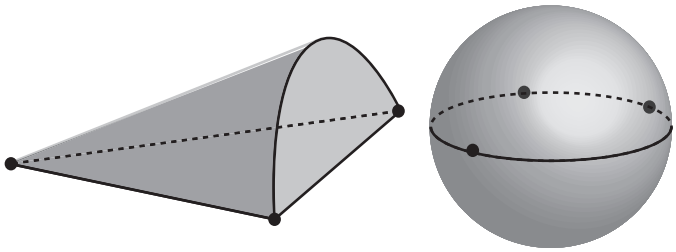


Unlike simplicial sets, however, we do not know how to define PLCW complexes *intrinsically*—they rely on an ambient space for their construction. This too is undesirable for

physics, analogous to defining the spacetime manifold in relativity as a submanifold of some higher-dimensional Euclidean space. In addition, the polyhedral cells in PLCW complexes would not allow, for example, plaquettes bounded by only one or two edges:



or 3-cells bounded by fewer than four plaquettes, like these:



and so on. The problem with all of these is that they involve curved sides; cells in PLCW complexes must be constructible in Euclidean space using only polyhedra with straight sides. So in fact, even our simple two-plaquette example from section 2 is not a PLCW complex. Worse, a typical *graph* is not a one-dimensional PLCW complex, so PLCW complexes are hardly a candidate for general n-graphs.

3.2. Homological n-complexes

Luckily, it turns out that in gauge theory, and even 'higher gauge theory', we can avoid many of the technical issues with n-graphs when the gauge group is *Abelian*—that is, in the case of p-form electromagnetism. To see why, note that taking the holonomy of a connection along some path through a graph, we multiply group elements of the edges composing the path *in order*. In a non-Abelian gauge theory, this requires that we keep close track of how the edges are interconnected. This is manageable for one-dimensional holonomies, but in 'p-form gauge theory', where holonomy involves multiplying group elements labelling p-cells, it is difficult to know what order to use. In general, the holonomy depends intimately on the incidence relations of the n-graph—that is, on the details of how cells are linked up—and keeping track of this requires a fairly sophisticated notion of n-graph. If the gauge group is Abelian, however, then the order in which we apply the group operation is unimportant, and all that matters is the orientation of each cell!

For us, spacetime will be a (homological) 'cell complex' M, which we define as follows. First, M consists of a list of sets

$$X_0, X_1, X_2, \ldots, X_k, \ldots,$$

where we call $X_k = X_k(M)$ the set of *k-cells* in M. Now intuitively, the boundary of a k-cell should be a 'sum' of $(k-1)$-cells. To formalize such sums, for each k we let $C_k = C_k(M)$ be the free Abelian group on the set X_k. In other words, C_k just consists of all formal linear combinations of k-cells, with integer coefficients. We call the elements of C_k the *k-chains* in M. We then hypothesize *boundary maps* $\partial_k \colon C_k \to C_{k-1}$ and require that these be linear over $\mathbb{Z}$. If we let $\partial_0 \colon C_0 \to 0$ be the unique map from C_0 to the trivial group, then the boundary maps fit together like this:

$$0 \xleftarrow{\partial_0} C_0 \xleftarrow{\partial_1} C_1 \xleftarrow{\partial_2} C_2 \xleftarrow{\partial_3} \cdots \xleftarrow{\partial_k} C_k \xleftarrow{\partial_{k+1}} \cdots,$$

and the topological maxim 'the boundary of a boundary is zero', popularized by Wheeler [13, 31], demands that applying two boundary maps in succession give the zero map. In homological algebra, such a sequence of maps $\partial\colon C_k \to C_{k-1}$ satisfying $\partial\partial = 0$ is called a *chain complex*.

We will say that the cell complex is *n dimensional*, or that it is an *n-complex*, provided $X_k = \emptyset$ for all $k > n$.[1] When the cell complex is *n* dimensional, we might as well truncate the chain and write simply

$$0 \xleftarrow{\partial_0} C_0 \xleftarrow{\partial_1} C_1 \xleftarrow{\partial_2} C_2 \xleftarrow{\partial_3} \cdots \xleftarrow{\partial_n} C_n.$$

We may now summarize the definition in a concise way.

Definition 1. *An n-dimensional homological cell complex, or n-complex, M is a chain complex*

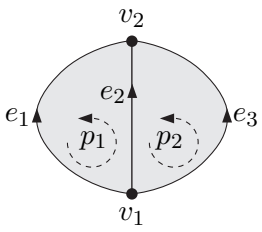
$$0 \xleftarrow{\partial} C_0 \xleftarrow{\partial} C_1 \xleftarrow{\partial} C_2 \xleftarrow{\partial} \cdots \xleftarrow{\partial} C_n$$

of free Abelian groups $C_k(M)$, each equipped with a preferred basis $X_k(M)$.

In a certain sense, this definition is perhaps too inclusive. We could take, for instance, the sets X_k of cells to be any sets whatsoever and let each boundary map ∂ be the zero map. One can also concoct other sorts of perverse examples, all falling squarely within the bounds of the definition, but with seemingly no connection between the boundary maps and the geometric notion of boundary.

The point is that, although the result might be strange or impractical, one actually can write a theory of electromagnetism on such *n*-complexes. Our attitude will be simply that it is up to the user of the definition to decide what additional conditions to impose to guarantee more 'reasonable' *n*-complexes. But to see that our definition of an *n*-complex at least does include the sort of discrete spacetimes we really *are* interested in and to show how the definition might be used in practice, an example or two might be in order.

Example 1. In section 2, we had the following spacetime:



But now, instead of simply talking about the sets $X_0 = V$, $X_1 = E$ and $X_2 = P$ of vertices, edges and plaquettes, we make them into free Abelian groups

$$C_0 \cong \mathbb{Z}^V, \qquad C_1 \cong \mathbb{Z}^E \qquad \text{and} \qquad C_2 \cong \mathbb{Z}^P,$$

with ordered bases V, E and P, respectively. We define the boundary map $\partial\colon C_1 \to C_0$ by assigning to each edge $e \in C_1$ its target minus its source:

$$\partial(e) = t(e) - s(e).$$

In particular, since all three edges in this example go from v_1 to v_2, we have

$$\partial(e_1) = \partial(e_2) = \partial(e_3) = v_2 - v_1.$$

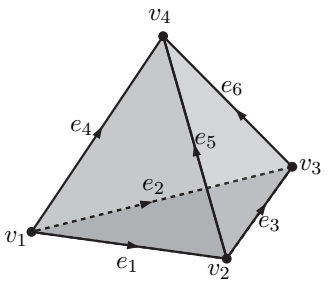
[1] Note that X_n can be empty. Thus, an *n*-complex is trivially also an $(n+1)$-complex.

Similarly, we define the boundary of a plaquette $p \in C_2$ by taking the sum of the edges around its physical boundary, but with opposite sign if an edge points opposite to the plaquette orientation. Thus, in our example,

$$\partial(p_1) = e_1 - e_2, \qquad \partial(p_2) = e_2 - e_3.$$

It is easy to check that $\partial\partial(p_1) = \partial\partial(p_2) = 0$, as required.

Example 2. As a second example, take the solid tetrahedron:



which has four vertices, six edges, four plaquettes and one 3-cell. Here we have left out the labels on plaquettes, to avoid cluttering the diagram, but for convenience we take the ith plaquette p_i to be the triangular face whose vertices include all vertices but v_i. On each plaquette, let us take the orientation such that its three edges are traversed in numerically increasing order (or a cyclic permutation). To specify the orientation of the 3-cell c, we call each of its faces positively oriented if its orientation is counterclockwise when viewed from inside c. The resulting chain complex then looks like this:

$$\begin{array}{ccccccccc} 0 & \xleftarrow{\partial_0} & C_0 & \xleftarrow{\partial_1} & C_1 & \xleftarrow{\partial_2} & C_2 & \xleftarrow{\partial_3} & C_3 \\ & & \| & & \| & & \| & & \| \\ 0 & \longleftarrow & \mathbb{Z}^4 & \longleftarrow & \mathbb{Z}^6 & \longleftarrow & \mathbb{Z}^4 & \longleftarrow & \mathbb{Z} \end{array}$$

with the boundary maps represented by the matrices

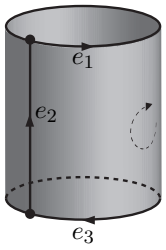
$$\partial_0 = [0 \quad 0 \quad 0 \quad 0], \qquad \partial_1 = \begin{bmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

$$\partial_2 = \begin{bmatrix} 0 & 0 & -1 & -1 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ -1 & -1 & 0 & 0 \end{bmatrix}, \qquad \partial_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

Again, we can easily verify that $\partial_0\partial_1$, $\partial_1\partial_2$ and $\partial_2\partial_3$ are all zero, so we really do get a chain complex.

Note that in passing from the physical model of a spacetime lattice to its corresponding n-complex, one is actually throwing away quite a lot of information! Given a chain complex, it may be impossible to reconstruct every detail of how cells in the original model might have been hooked together. Applying the boundary map to a k-cell only tells us 'how many times'

each $(k-1)$-cell is included in its boundary, with possible cancellation. For example, in the cylinder with one plaquette p:



we have $\partial p = e_1 - e_2 + e_3 + e_2 = e_1 + e_3$, leaving no trace of the fact that e_2 is actually included twice in the physical boundary of p, with opposite orientation.

We must emphasize that we are free to discard all of these incidence data only because electromagnetism is Abelian. As we shall soon see, almost everything we need for electromagnetism is contained in the chain complex of spacetime. We can even think of discrete spacetime as *being* its chain complex, a perspective which we explore more fully in section 8.

4. The discrete gauge field and the action

4.1. p-connections, curvature and gauge transformations

In p-form electromagnetism, the analogue of the connection A is locally a p-form, making the electromagnetic field $F = dA$ a $(p+1)$-form. In *lattice* p-form electromagnetism, we thus define a *p-connection* to be a map assigning to each p-cell an element of the gauge group:

$$A: \quad X_p(M) \to G,$$

where for our purposes $G = \mathbb{R}$ or $U(1)$. But in the framework introduced in the previous section, we can now speak of not only p-cells but also their formal linear combinations. The real beauty of this viewpoint is that we can extend A to a homomorphism of Abelian groups

$$A: \quad C_p(M) \to G$$

in a unique way. That is, we can consider a p-connection to be an element of the dual space

$$C^p(M; G) := \hom(C_p(M), G),$$

which we dub the space of G-valued *p-cochains*. Likewise, the curvature is a linear functional on the space of $(p+1)$-cells or an element of $C^{p+1}(M; G)$.

Since the boundary maps form a chain complex,

$$0 \xleftarrow{\partial_0} C_0 \xleftarrow{\partial_1} C_1 \xleftarrow{\partial_2} C_2 \xleftarrow{\partial_3} \cdots \xleftarrow{\partial_n} C_n,$$

we get a chain complex of the dual spaces in the opposite direction, the associated *cochain complex*, by taking the adjoint of each boundary map:

$$C^0 \xrightarrow{d_0} C^1 \xrightarrow{d_1} C^2 \xrightarrow{d_2} \cdots \xrightarrow{d_{n-1}} C^n \xrightarrow{d_n} 0.$$

That is, given $\omega \in C^i$, we define the *coboundary maps* $d_i\omega \in C^{i+1}$ by $d_i\omega(x) = \omega(\partial_{i+1}x)$. Dropping the subscripts on both the boundary and coboundary maps usually causes no confusion and we can write more cleanly

$$df(x) = f(\partial x).$$

This definition justifies the name cochain complex: it is easy to see that $\partial\partial = 0$ implies $dd = 0$.

Using the coboundary maps, we define the *curvature* of a p-connection A to be the $(p+1)$-cochain dA. Likewise, a *gauge transformation* is a $(p-1)$-cochain ϕ which acts on a p-connection A via $A \mapsto A + d\phi$. Thus, one part of the cochain complex looks like this:

$$\underset{\text{gauge transformations}}{C^{p-1}(M;G)} \xrightarrow{d} \underset{\text{p-connections}}{C^{p}(M;G)} \xrightarrow{d} \underset{\text{curvatures}}{C^{p+1}(M;G)}$$

For $p = 1$, these definitions are precisely the definitions in ordinary continuum electromagnetism, except that the complex in question is the cochain complex above, rather than the de Rham complex of differential forms.

When $G = \mathbb{R}$, so that the duals $C^q(M,\mathbb{R}) \cong \mathbb{R}^{X_k}$ are vector spaces, there is an especially strong parallel between our cochain complex and the de Rham complex. In this case, we will slip between the two vocabularies at will, calling cochains 'forms' whenever convenient. Following the language of differential forms, we often call the coboundary maps d_i *differentials* and say that a k-cochain ω is *closed* if $d\omega = 0$ or *exact* if $\omega = d\eta$ for some $(k-1)$-cochain η. The equation $dd = 0$ then has its usual de Rham translation, 'all exact forms are closed'.

4.2. Action for $\mathbb{R}$ p-form electromagnetism

In ordinary continuum electromagnetism, the Lagrangian for the free electromagnetic field $F = dA$ is given by

$$L = \frac{1}{2e^2} F \wedge \star F.$$

This equation holds for free p-form electromagnetism as well, where F is the p-form field $F = dA$. To write a lattice version of this theory, we thus seem to need discrete analogues of the wedge product $\wedge$ and the Hodge star operator $\star$. The Hodge star in differential geometry turns q-forms into $(n-q)$-forms, so we seem to need a map:

$$\star : C^q \to C^{n-q}$$

for each $q = 0, 1, \ldots, n$. Likewise, for an analogue of the wedge product of differential forms we would want

$$\wedge : C^q \times C^r \to C^{q+r}.$$

Defining these maps seems to be the main obstacle to completely formulating a discrete analogue of the theory of differential forms. In certain special cases, it is known what these operations should look like. For example, for simplicial complexes, the best approximation to the wedge product of differential forms is the 'cup product' $\cup$ of cochains [21]. For the Hodge star, a cubical complex is more convenient, since Hodge duality is related to Poincaré duality, and the Poincaré dual of a cubical lattice is also cubical[2].

But we would rather not impose any restrictions on the shapes of cells in our spacetime model. Forcing the lattice to be too regular goes against relativity's lesson that spacetime, at least on the macroscopic level, does not have preferred directions the way a crystal lattice does.

Luckily, if we resist the temptation to dissect the formula for the action

$$S = \frac{1}{2e^2} \int F \wedge \star F,$$

[2] For more on discrete differential geometry, we refer the reader to the papers by Dimakis and Müller-Hoissen [14], Dodziuk [15], Forgy and Schreiber [16] and Forman [17], and references cited therein.

we can instead discretize the whole thing in one fell swoop! The key observation is that

$$\langle F, G\rangle := \int F \wedge \star G$$

defines an inner product, the *Hodge inner product*, on differential $(p+1)$-forms. So in the lattice theory, if we equip our $(p+1)$-cochains with an inner product

$$\langle\,,\rangle : C^{p+1} \times C^{p+1} \to \mathbb{R},$$

then we can define the action corresponding to the field $F = dA$ to be

$$S = \frac{1}{2e^2}\langle F, F\rangle.$$

What this inner product should look like for a given lattice depends on the structure and the dimension of the lattice. We shall not dig too deeply into this issue, but just mention a few 'reasonable' requirements. For this it is useful to write the inner product as

$$\langle F, G\rangle = h^{ij} F(c_i) G(c_j),$$

where the implicit sum is over all pairs of $(p+1)$-cells c_i, c_j. One sensible condition that might be imposed on this inner product is a 'locality condition'. That is, we do not expect the action to have correlations between distant points in spacetime, so we might require that h^{ij} vanish whenever c_i, c_j are 'nonadjacent' $(p+1)$-cells. Also, when the details of lattice geometry are carefully accounted for, the defining equation of the Hodge star operator:

$$\mu \wedge \star\nu = \langle \mu, \nu\rangle \,\mathrm{vol}$$

implies that the off-diagonal entries of h^{ij} should depend on the physical *angle* between the corresponding cells. In particular, on a regular cubical lattice, we expect h^{ij} to be a diagonal matrix. Note that neither the locality condition nor any geometric data used to build the inner product can be derived from the chain complex itself. Both constitute additional structure in the spacetime model.

At this point, we are prepared to formally define the path integral for lattice p-form electromagnetism with gauge group $\mathbb{R}$. Using the action defined above, we define the partition function

$$Z = \sqrt{\det\left(\frac{1}{2\pi} h_{ij}\right)} \int_{\mathcal{A}/\mathcal{G}} \mathrm{e}^{-\frac{1}{2e^2} h^{ij} F_i F_j} \mathcal{D}A. \tag{1}$$

The constant factor in front is purely a matter of convenience, chosen because

$$\sqrt{\det\left(\frac{1}{2\pi} h^{ij}\right)}\, \mathrm{e}^{-\frac{1}{2e^2} h^{ij} F_i F_j}$$

is a normalized Gaussian in the curvature variables $F_i = F(c_i)$ with respect to Lebesgue measure on $\mathcal{F} = C^{p+1}$. We are free to introduce such a factor—or, equivalently, to shift the origin of the action by the log of this factor—since it does not change the physics: expectation values of observables are unaffected. Of course, this also means that we are free to ignore this factor when calculating expectation values.

For the sake of simplicity, many of our examples in this paper involve electromagnetism in two dimensions. For this reason, we next turn briefly away from our more general development of the theory to consider this special case. More generally, we consider the case of p-form electromagnetism in $p+1$ dimensions and find that the requisite inner product on $(p+1)$-cochains has a natural and simple form.

4.3. The case of $p+1$ dimensions and volumetric field theory

Ordinary gauge theories in *two* dimensions are notably well behaved [3, 39]. One of the reasons electromagnetism in two dimensions is especially nice is that the electromagnetic field, being a 2-form, can be written simply as a scalar field times the volume form (or, perhaps more properly, the 'area form'):

$$F = f \text{ vol}.$$

This allows us to choose the best possible inner product on 1-cochains—an exact analogue of the Hodge inner product—without having to define a discrete Hodge star operator in detail. In particular, the only Hodge duals we really need to consider are the trivial ones: $\star\text{vol} = 1$ and $\star 1 = \text{vol}$. More generally, and for the same reason, we can do discrete p-form electromagnetism and avoid any unpleasant entanglements with the Hodge star as long as we work in $(p+1)$-dimensional spacetime.

Since the usual Lagrangian for vacuum p-form electromagnetism is $L = \frac{1}{2e^2} F \wedge \star F$, we find, in the $(p+1)$-dimensional case,

$$\begin{aligned} L &= \frac{1}{2e^2}(f \text{ vol}) \wedge \star(f \text{ vol}) \\ &= \frac{1}{2e^2} f^2 \text{ vol} \wedge 1 \\ &= \frac{f^2}{2e^2} \text{ vol}. \end{aligned}$$

Now, if we want to discretize this theory, we clearly want a fine enough lattice that the value of the scalar curvature field f in the above equations is essentially constant over any $(p+1)$-cell. If we make this assumption, then the action associated with a $(p+1)$-cell c_i whose curvature is F should be

$$\begin{aligned} S(c_i) &= \int_{c_i} \frac{f^2}{2e^2} \text{ vol} \\ &\approx \frac{f^2}{2e^2} \int_{c_i} \text{vol} \\ &= \frac{f^2}{2e^2} V_i, \end{aligned}$$

where V_i is the volume of the $(p+1)$-cell c_i. Similarly, the most obvious way of discretizing the curvature is to assign to each $(p+1)$-cell the volume integral of its continuum curvature:

$$\begin{aligned} F_i &= \int_{c_i} f \text{ vol} \\ &\approx f \int_{c_i} \text{vol} \\ &= f V_i. \end{aligned}$$

Combining these last two results, we find a candidate for the action which now uses only variables that are available to us directly in the lattice model:

$$S_i = \frac{1}{2e^2} \frac{F_i^2}{V_i}.$$

To get the total action, we just sum over $(p+1)$-cells. Recalling from section 4.2 that the action is given by

$$S = \frac{1}{2e^2} \langle F, F \rangle,$$

the corresponding inner product on $(p+1)$-cochains for p-form electromagnetism in $p+1$ dimensions is thus

$$\langle F, G\rangle = \sum_{p_i \in X_{p+1}} \frac{F_i G_i}{V_i}.$$

Here $V_i := \mathrm{vol}(x_i)$, where we now use vol to denote the *discrete volume form*—an $\mathbb{R}$-valued $(p+1)$-cochain

$$\mathrm{vol}{:}C_{p+1} \to \mathbb{R}$$

that assigns to each $(p+1)$-cell its volume.

Specializing the path integral (1) to the action above, we obtain the definition of the partition function for p-form electromagnetism in $p+1$ dimensions:

$$Z = \left(\prod_{c \in X_{p+1}} \frac{1}{\sqrt{2\pi e^2 \,\mathrm{vol}(c)}} \right) \int_{\mathcal{A}/\mathcal{G}} \exp\left(-\frac{1}{2e^2} \sum_{c \in X_{p+1}} \frac{F(c)^2}{\mathrm{vol}(c)} \right) \mathcal{D}A. \tag{2}$$

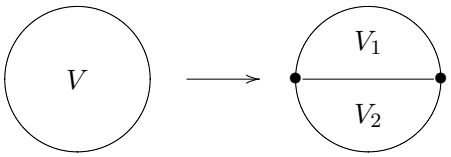
The normalization factor is particularly handy in this case, since one can use the identity

$$\delta(x) = \lim_{\sigma \to 0} \frac{\mathrm{e}^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

to take the limit of Z as the volume of a single $(p+1)$-cell approaches zero; the result is the partition function one would obtain by simply ignoring that cell in the lattice. This is a desirable feature if one wishes to take the continuum limit of the lattice theory. We shall see in section 7.4 that the results of Witten [39] suggest that the properly normalized partition function indeed has the correct continuum limit.

In fact, though our derivation of the action in this section was based on the assumption of a sufficiently fine lattice, the resulting quantum theory does not depend at all on how fine a lattice we choose. Just as Witten has noted for 2D Yang–Mills theory, there is a strong sense in which the quantum theory of p-form electromagnetism in $p+1$ dimensions is 'almost topological'. Namely, the theory requires only one nontopological datum: the volume of spacetime. It is thus what we might call a *volumetric quantum field theory*.

To begin seeing why the p-form theory in $p+1$ dimensions is volumetric, consider splitting a $(p+1)$-cell of volume V in two by slicing through it with a new p-cell, leaving some fraction of the original volume on either side. Schematically,



$$V_1 + V_2 = V$$

We wish to compare the curvature F_0 on the original $(p+1)$-cell to the sum $F = F_1 + F_2$ in the finer discretization. The curvature on the original cell is a Gaussian random variable with probability measure

$$\mu_{F_0} = \frac{\mathrm{e}^{-F_0^2/2e^2V}}{\sqrt{2\pi e^2 V}} dF_0.$$

On the other hand, the joint probability measure for the random variables F_1 and F_2 is

$$\mu_{F_1,F_2} = \frac{\mathrm{e}^{-F_1^2/2e^2V_1}\,\mathrm{e}^{-F_2^2/2e^2V_2}}{2\pi e^2\sqrt{V_1V_2}} dF_1 dF_2.$$

Making the change of variables

$$F = F_1 + F_2, \qquad G = F_1 - F_2,$$

we can rewrite this as a measure for F and G:

$$\mu_{F,G} = \frac{\mathrm{e}^{-(F+G)^2/8e^2V_1}\,\mathrm{e}^{-(F-G)^2/8e^2V_2}}{4\pi e^2\sqrt{V_1V_2}}dF\,dG.$$

To get a measure for $F = F_1 + F_2$, we now have only to integrate out the G-dependence from $\mu_{F,G}$. After completing the square in G and integrating, we find

$$\mu_F = \frac{\mathrm{e}^{-F^2/2e^2V}}{\sqrt{2\pi e^2 V}}dF,$$

which is the same as μ_{F_0}.

What this calculation shows is that we are free to rediscretize our spacetime by knocking down or inserting new p-cells, and that we will always get the same results, as long as we only ask questions that can be asked for either discretization. In the end, the only freedom in the theory besides the topology is the total volume of the spacetime. We take a deeper look at the volumetric nature of p-form electromagnetism in $p+1$ dimensions in section 7.4. For now, we return to the general development and study the classical lattice theory.

5. Classical *p*-form electromagnetism

5.1. Discrete p-form Maxwell equations

Let us return to discrete spacetimes of arbitrary dimension and obtain the classical equations of motion—the discrete analogue of Maxwell's equations, or rather their p-form generalization—for the case where the gauge group is $\mathbb{R}$. Formally, the action in our theory is the map

$$S{:}C^p \to \mathbb{R}$$

given by

$$S(A) = \frac{1}{2e^2}\langle dA, dA\rangle,$$

where $\langle\,,\,\rangle$ is the inner product on $(p+1)$-cochains introduced in the previous section. We get

$$\begin{aligned} 0 = \delta S &= \frac{1}{2e^2}\delta\langle dA, dA\rangle \\ &= \frac{1}{e^2}\langle \delta dA, dA\rangle \\ &= \frac{1}{e^2}\langle d\delta A, F\rangle \\ &= \frac{1}{e^2}\langle \delta A, d^\dagger F\rangle, \end{aligned}$$

where $d^\dagger$ is the Hilbert space adjoint of the linear operator d. We therefore find the equation of motion:

$$d^\dagger F = 0$$

which is our discrete analogue of the vacuum p-form Maxwell equations

$$\star d \star F = 0.$$

It is worth noting the relationship between the continuum and lattice versions of the Maxwell equations. In particular, this means understanding the relationship between the

two kinds of dualities showing up in these equations. In Riemannian signature, the Hodge operator satisfies $\star^2\omega = (-1)^{p(n-p)}\omega$ when acting on any p-form ω. So given a p-form A and $(p+1)$-form G, we have

$$\begin{aligned}\langle dA, G\rangle &= \int dA \wedge \star G \\ &= (-1)^{p+1}\int A \wedge d\star G \\ &= \int A \wedge \star(-1)^{(n-p)p}\star d\star G \\ &= \langle A, (-1)^{(n-p)p}\star d\star G\rangle,\end{aligned}$$

where in the second step we did an integration by parts. Thus,

$$d^\dagger = (-1)^{(n-p)p}\star d\star.$$

Note that while $\star F$ naturally lives on the Poincaré dual lattice, $\star d\star F$ lives on the original lattice, so the p-form Maxwell equations $\star d\star F = J$ make perfect sense on the lattice, without reference to dual lattices, even for nonvanishing p-current J.

5.2. *The topological character of lattice p-form electromagnetism*

In fact, the vacuum p-form Maxwell equations

$$d^\dagger F = 0$$

can be simplified further in the present context. If $F = dA$ satisfies this equation, then we have

$$0 = \langle d^\dagger F, A\rangle = \langle F, dA\rangle = \langle F, F\rangle.$$

In Lorentzian physics, we draw no strong conclusions from this calculation. But in the Riemannian case, the inner product $\langle\,,\rangle$ is *positive definite*, and hence the classical equations of motion for free discrete p-form electromagnetism reduce to

$$F = 0!$$

This is shocking if we are accustomed to electromagnetism on Minkowski spacetime. It means, in particular, that our theory has no analogue of electromagnetic waves propagating through a vacuum! Moreover, the same proof works not only in the discrete case but also, for example, for Riemannian p-form electromagnetism on a compact manifold. The two assumptions leading to the conclusion that $F = 0$ are that (1) we have an inner product defined on *all* $(p+1)$-forms F and (2) this inner product is positive definite. In Minkowski spacetime, the Hodge inner product $\int F \wedge \star G$ is neither defined for all $(p+1)$-forms, nor positive definite, so even compactly supported p-form electromagnetic fields in the vacuum need not vanish.

There are of course other well-known theories whose classical equations of motion amount to stating that some curvature F vanishes, including BF theory [5] (also called 'topological Yang–Mills theory' in 2D or 'topological gravity' in 4D) and Chern–Simons theory [6], with 3D general relativity as a special case of both of these [40]. One major difference between p-form electromagnetism under conditions (1) and (2) and these theories is that p-form electromagnetism is not about ordinary flat connections on bundles, but about flat p-connections on p-bundles [7]. But even in the case $p = 1$, there are important differences.

Most notable from the relativist's perspective is that p-form electromagnetism, unlike general relativity, BF theory or Chern–Simons, relies on a *fixed background structure*. Namely,

one must specify an inner product on $(p+1)$-cochains. This inner product encodes geometrical data, since it corresponds to the Hodge inner product, which is metric dependent. Although the classical equation of motion $F = 0$ is independent of the choice of inner product, the inner product is crucial to write the Lagrangian. Even more importantly, unlike some genuine TQFTs whose Lagrangians are superficially background metric dependent, quantum p-form electromagnetism has expectation values which depend on the background structure. Thus, while we might argue that p-form electromagnetism on a finite lattice is 'topological' as a *classical* field theory, it is not in general a topological *quantum* field theory.

There is one way we might try to weasel out of the conclusion $F = 0$ in cases where conditions (1) and (2) are met. Namely, we can take a pre-gauge theory perspective on electromagnetism and take F as the fundamental field of classical electromagnetism, rather than A. If the field strength F is not necessarily the differential of any gauge potential, then $dF = 0$ is not automatically satisfied, so we must include it explicitly in the p-form Maxwell equations

$$dF = 0 \qquad \text{and} \qquad d^\dagger F = 0.$$

We will see that these equations *do* have solutions other than $F = 0$, but only on spacetimes with nontrivial $(p+1)$th cohomology—roughly, spacetimes with 'holes' of the type which can be enclosed by a surface of dimension $p+1$. We shall soon see that classical p-form electromagnetism from this perspective also has a strongly topological character.

Let us recall some basic constructions from topology. Consider some small portion of a complex of $\mathbb{R}$-valued cochains

$$C^{q-1} \xrightarrow{d_{q-1}} C^q \xrightarrow{d_q} C^{q+1}.$$

Cohomology classifies topological spaces by comparing two subspaces of C^q:

$$Z^q := \ker d_q \qquad \text{(the space of } q\text{-cocycles)},$$
$$B^q := \operatorname{ran} d_{q-1} \qquad \text{(the space of } q\text{-coboundaries)}.$$

For the cochain complex of *any* spacetime lattice we have, by the identity $dd = 0$,

$$B^q \subset Z^q.$$

That is, every q-coboundary is a q-cocycle. Whether the converse of this statement is true depends on the particular topology of the spacetime lattice. If every q-cocycle is a q-coboundary, so that B^q and Z^q are equal, we say that the cochain complex is *exact* at C^q. In topologically interesting spacetimes (or regions of spacetime), exactness may fail, and we measure the failure of exactness by taking the quotient space

$$H^q := Z^q / B^q,$$

called the qth cohomology with real coefficients.

Now, we have just seen that the vacuum discrete p-form Maxwell equations

$$dF = 0 \qquad \text{and} \qquad d^\dagger F = 0$$

have two possible interpretations.

(i) Considering the gauge potential A as fundamental, the positive definiteness of the inner product on $(p+1)$-forms reduces the equations of motion to $F = 0$. From this perspective, the space of solutions of the vacuum discrete p-form Maxwell equations consists of gauge-equivalence classes of flat p-connections. But this is just

$$\frac{\mathcal{A}_0}{\mathcal{G}} = \frac{Z^p}{B^p} = H^p,$$

the pth cohomology.

(ii) Considering the field strength F as fundamental, solutions of the vacuum discrete p-form Maxwell equations are $(p+1)$-forms F satisfying

$$dF = 0 \qquad \text{and} \qquad d^\dagger F = 0.$$

As explained below, by Hodge's theorem we have an isomorphism

$$\ker d_{p+1} \cap \ker d_p^\dagger \cong H^{p+1},$$

so that from this perspective, solutions are classified by the $(p+1)$th cohomology.

From either viewpoint, the classical theory is thus 'topological' in the sense that its space of solutions is determined up to canonical isomorphism by the cohomology of spacetime. The two viewpoints however disagree on which cohomology group determines the solutions.

To fully understand interpretation (ii), it is helpful to consider an alternate viewpoint on cohomology—a viewpoint which is useful whenever we have a cochain complex of (possibly indefinite) inner product spaces. When we have such a complex C, each of the differentials $d_q : C^q \to C^{q+1}$ has an adjoint $d_q^\dagger : C^{q+1} \to C^q$:

$$C^{q-1} \underset{d^\dagger}{\overset{d}{\rightleftarrows}} C^q \underset{d^\dagger}{\overset{d}{\rightleftarrows}} C^{q+1}$$

and $dd = 0$ together with nondegeneracy of the inner product implies $d^\dagger d^\dagger = 0$, so the maps $d^\dagger$ also form a complex. We say that $\omega \in C^q$ is *coclosed* if $d^\dagger \omega = 0$ or *coexact* if $\omega = d^\dagger \alpha$ for some $\alpha \in C^{q+1}$. The defining equation $\langle \omega, d\mu \rangle = \langle d^\dagger \omega, \mu \rangle$ together with nondegeneracy implies that a coclosed q-form is precisely a q-form which is orthogonal to all exact q-forms. Similarly, a closed q-form is one which is orthogonal to all coexact q-forms. These facts may be combined to obtain the orthogonal direct sum

$$C^q = \operatorname{ran} d \oplus (\ker d^\dagger \cap \ker d) \oplus \operatorname{ran} d^\dagger, \tag{3}$$

known as the *Hodge decomposition*. In the case $q = p+1$, the middle summand is just the space of solutions of the p-form Maxwell's equations.

When the inner product on C^q is positive definite, as it is in Riemannian physics, we can go further, since it is then simple to show that $(\ker d^\dagger \cap \ker d) = \ker \Delta$ where

$$\Delta_q := d_q^\dagger d_q + d_{q-1} d_{q-1}^\dagger$$

is the *Laplacian* on C^q, and a cochain in its kernel is said to be *harmonic*. The Hodge decomposition is precisely the statement that any form is uniquely the sum of exact, harmonic and coexact parts. However, even in the case of indefinite inner products, as in Lorentzian physics, we can use (3) to establish an isomorphism between the $(q+1)$th cohomology and the space of closed and coclosed $(q+1)$-forms

$$H^q \cong \left(\ker d_q^\dagger \cap \ker d_{q+1}\right).$$

This result is a version of *Hodge's theorem*[3]. Identifying these two spaces, we can therefore write

$$C^q = dC^{q-1} \oplus H^q \oplus d^\dagger C^{q+1}$$

as an alternate version of the Hodge decomposition of C^q.

Since the free p-form Maxwell equations for F simply state that F is closed and coclosed, when we have an inner product defined on all $(p+1)$-forms, solutions of Maxwell's equations

[3] Here we only need a relatively weak version of the Hodge decomposition and Hodge theorem [38]. More sophisticated versions include the usual one on the de Rham complex of a (compact) manifold [12] and the 'Kodaira decomposition' for complexes of operators that are only densely defined [1]. Álvarez's thesis [1] provides a detailed study of contexts in which Hodge theory is valid.

are classified by the $(p+1)$th cohomology, which may be identified (in Riemannian signature) with the space of harmonic $(p+1)$-forms.

In the following section, we investigate the implications of cohomology to *quantum* p-form electromagnetism. We shall see in particular that cohomology provides a criterion for the convergence of the path integral.

6. Path integrals with gauge group $\mathbb{R}$

6.1. Cohomological criterion for path integral convergence

When $G = \mathbb{R}$, we saw in section 2 that the path integral

$$Z = \int_{\mathcal{A}} \mathrm{e}^{-S(A)} \mathcal{D}A$$

diverges due to gauge freedom. Indeed, in the language of cohomology, the existence of gauge freedom means $\mathcal{G} = B^p$ is nontrivial, and in this case, writing $A = A_0 + A^\perp$ with $A_0 \in B^p$ and $A^\perp \in (B^p)^\perp$, we get

$$\begin{aligned}\int_{\mathcal{A}} \mathrm{e}^{-S(A)} dA &= \int_{B^p} \int_{(B^p)^\perp} \mathrm{e}^{-S(A_0+A^\perp)} dA^\perp dA_0 \\ &= \int_{B^p} dA_0 \int_{(B^p)^\perp} \mathrm{e}^{-S(A^\perp)} dA^\perp = \infty,\end{aligned}$$

where the second equality follows from gauge invariance of the action.

To eliminate these divergences, it is therefore *necessary* that we eliminate any gauge freedom from our space of p-connections. The standard procedure is to pass to the quotient space

$$\frac{\mathcal{A}}{\mathcal{G}} = \frac{C^p}{B^p} = p\text{-connections modulo gauge transformations},$$

effectively declaring *gauge-equivalent* p-connections to be equal.

In many cases, integrating over C^p/B^p instead of just C^p does indeed make the path integral finite. But this does not always work! Whether it works or not depends on a topological condition—the *cohomology* of the lattice. While factoring out gauge freedom is necessary, it is only sufficient if the pth cohomology, H^p, is trivial. Indeed, if $H^p = Z^p/B^p$ is nontrivial, then since

$$\frac{\mathcal{A}}{\mathcal{G}} = \frac{C^p}{B^p} \cong \frac{C^p}{Z^p} \oplus \frac{Z^p}{B^p},$$

we have

$$\begin{aligned}\int_{\mathcal{A}/\mathcal{G}} \mathrm{e}^{-S(A)} dA &= \int_{Z^p/B^p} \int_{C^p/Z^p} \mathrm{e}^{-S(A)} dA_1 dA_2 \\ &= \int_{Z^p/B^p} dA_2 \int_{C^p/Z^p} \mathrm{e}^{-S(A_1)} dA_1 = \infty,\end{aligned}$$

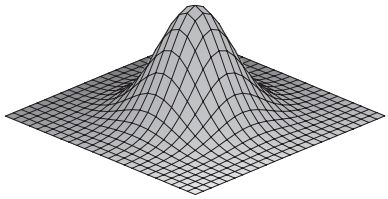
since $S(A)$ really only depends on $dA = d(A_1 + A_2) = dA_1$.

Pondering more carefully what caused the divergence, we realize that to eliminate all divergences from $\mathbb{R}$ electromagnetism we must use the quotient space

$$\frac{\mathcal{A}}{\mathcal{A}_0} = \frac{C^p}{Z^p} = p\text{-connections modulo flat } p\text{-connections}.$$

For any lattice with only finitely many p-cells, integrating over this space *always* gives us a convergent path integral. The point is that, when H^p is nontrivial, the quadratic form S is

still degenerate, even after factoring out all of the gauge freedom. However, when we factor out not just gauge transformations, or in other words 'pure gauge' p-connections, but *all* flat p-connections, S becomes a positive *definite* quadratic form, so that $\int \mathrm{e}^{-S}$ is finite. Visually, on the space of p-connections modulo flat p-connections, every cross section of the graph of e^{-S} looks just how we expect a well-behaved Gaussian to look:



in contrast with the analogous picture from section 2. This gives us a theory in which path integrals are always convergent, but it amounts to ignoring the 'p-form Bohm–Aharonov effect', as we shall soon see.

Whether the pth cohomology is trivial or not, we can calculate the path integral formally on the physical configuration space of connections mod gauge transformations as

$$Z = \int_{\mathcal{A}/\mathcal{G}} \mathrm{e}^{-S(A)} \mathcal{D}A = \int_{C^p/B^p} \mathrm{e}^{-\frac{1}{2e^2}\langle dA, dA\rangle} \mathcal{D}A = \sqrt{\frac{(2\pi e^2)^N}{\det\left(d_p^\dagger d_p\right)}}, \tag{4}$$

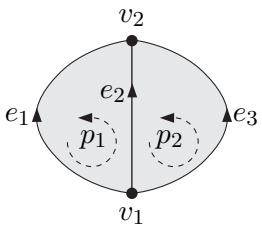
where $N = \dim \mathcal{A}/\mathcal{G}$ and the adjoint $d^\dagger$ is defined relative to the standard dot product on $\mathcal{A}/\mathcal{G} \cong \mathbb{R}^N$:

$$\langle dA, dA\rangle = A \cdot d^\dagger dA.$$

Of course (4) is purely formal unless $H^p = 0$, since otherwise $\det(d^\dagger d) = 0$. One can calculate expected values of observables using standard techniques for dealing with divergent path integrals in quantum field theory, such as imposing a cut-off on the magnitude of the connection and taking a limit.

6.2. *Examples*

Let us consider two examples: one in which modding out by gauge transformations is sufficient to render the path integral finite and other in which it is not. First, let us calculate the partition function for the simple spacetime we considered in section 2:



but this time after factoring out gauge transformations. We note that, while we use the language of 2D electromagnetism, the calculation applies equally well to p-form electromagnetism on a $(p+1)$-ball divided into two $(p+1)$-cells of volumes V_1 and V_2.

We write the connection as

$$A = A_1 e^1 + A_2 e^2 + A_3 e^3 = A_i e^i,$$

where $\{e^i\}$ is the basis of 1-cochains dual to the basis $\{e_i\}$ of edges. A gauge transformation

$$\phi = \phi_1 v^1 + \phi_2 v^2$$

acts on the connection by $A \mapsto A + d\phi$, where

$$d\phi = (\phi_2 - \phi_1)(e^1 + e^2 + e^3).$$

Since the number $\phi_2 - \phi_1 \in \mathbb{R}$ may be chosen arbitrarily, the space of gauge transformations is just the one-dimensional subspace

$$\mathcal{G} = B^1 = \text{span}\{e^1 + e^2 + e^3\} \subset C^1.$$

Similarly, the curvature F is given by

$$F := dA = (A_1 - A_2)p^1 - (A_3 - A_2)p^2,$$

so the space of flat connections is

$$\begin{aligned}\mathcal{A}_0 = Z^1 &= \{A \in C_1 | A_1 = A_2 = A_3\} \\ &= \text{span}\{e^1 + e^2 + e^3\} = B^1.\end{aligned}$$

Thus, due to the simple topology of the model, the flat connections in this case are precisely the pure gauge connections. In other words, the first cohomology is trivial:

$$H^1 = Z^1/B^1 = \{0\}.$$

To evaluate the path integral over the space $C^1/B^1 = C^1/Z^1$, we make use of the canonical isomorphism

$$\frac{C^1}{B^1} \cong (B^1)^\perp$$

given by associating with each gauge-equivalence class $[A] \in C^1/B^1$ its unique member which is also an element of $(B^1)^\perp$. We first define a new orthonormal basis

$$e^{1'} = \tfrac{1}{\sqrt{3}}(e^1 + e^2 + e^3), \qquad e^{2'} = \tfrac{1}{\sqrt{2}}(e^1 - e^2), \qquad e^{3'} = \tfrac{1}{\sqrt{6}}(e^1 + e^2 - 2e^3)$$

of C^1 chosen so that $\{e^{1'}\}$ is a basis for B^1 and $\{e^{2'}, e^{3'}\}$ is a basis for $(B^1)^\perp$.

We can write the connection A in terms of the new dual basis:

$$A = A_{i'}e^{i'} = \frac{A_1 + A_2 + A_3}{\sqrt{3}}e^{1'} + \frac{A_1 - A_2}{\sqrt{2}}e^{2'} + \frac{A_1 + A_2 - 2A_3}{\sqrt{6}}e^{3'}.$$

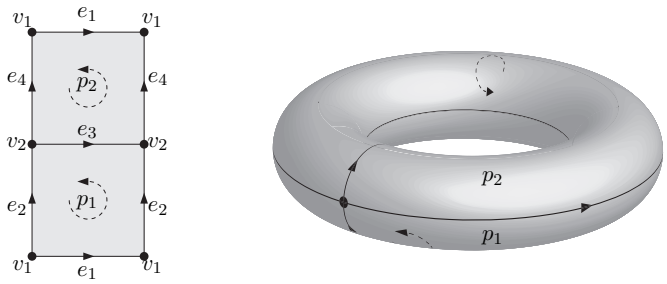
Then the action becomes

$$\begin{aligned}S(A) &= \frac{1}{2e^2}\left(\frac{(A_1 - A_2)^2}{V_1} + \frac{(A_2 - A_3)^2}{V_2}\right) \\ &= \frac{1}{2e^2 V_1 V_2}\left[V_2(\sqrt{2}A_{2'})^2 + V_1\left(-\frac{1}{\sqrt{2}}A_{2'} + \sqrt{\frac{3}{2}}A_{3'}\right)^2\right] \\ &= \frac{1}{2e^2 V_1 V_2}[A_{2'} \quad A_{3'}]\begin{bmatrix} 2V_2 + \frac{1}{2}V_1 & \frac{\sqrt{3}}{2}V_1 \\ \frac{\sqrt{3}}{2}V_1 & \frac{3}{2}V_1 \end{bmatrix}\begin{bmatrix} A_{2'} \\ A_{3'} \end{bmatrix} \\ &= \frac{1}{2e^2 V_1 V_2}A \cdot QA,\end{aligned}$$

where $Q = (1/V_1V_2)d^\dagger d$. Now integrating over $\mathcal{A}/\mathcal{G}$ is a snap! Since the 2×2 matrix in the above expression is nonsingular, the standard Gaussian integral formula now gives a finite result for the path integral (2):

$$Z = \frac{1}{2\pi e^2\sqrt{V_1V_2}}\int_{\mathbb{R}^2} \mathrm{e}^{-S(A)}\, d^2A = \sqrt{\frac{V_1V_2}{\det(Q)}} = \sqrt{\frac{1}{3}}.$$

As a second example, consider again the case of ordinary 1-form electromagnetism in two dimensions, but this time in a spacetime with the topology of a torus:



Note in the diagram that objects with the same label are identified. Let us first calculate the cohomology of this lattice. The cochain complex

$$0 \longrightarrow C^0 \xrightarrow{d_0} C^1 \xrightarrow{d_1} C^2 \longrightarrow 0$$

has differentials represented by the matrices

$$d_0 = \begin{bmatrix} 0 & 0 \\ -1 & 1 \\ 0 & 0 \\ 1 & -1 \end{bmatrix}, \qquad d_1 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}$$

relative to the bases $\{v^1, v^2, v^3\}$ of C^0, $\{e^1, e^2, e^3, e^4\}$ of C^1 and $\{p^0, p^2\}$ of C^2. We are particularly interested in the first cohomology, since it relates flat connections to gauge transformations. Some basic linear algebra shows that

$$\mathcal{A}_0 = Z^1 = \operatorname{span}\{e^1 + e^3, e^2, e^4\} \cong \mathbb{R}^3,$$
$$\mathcal{G} = B^1 = \operatorname{span}\{e^2 - e^4\} \cong \mathbb{R}$$

and therefore

$$H^1 = \frac{Z^1}{B^1} \cong \frac{\mathbb{R}^3}{\mathbb{R}} \cong \mathbb{R}^2.$$

Physically, this says that there are flat connections A on the torus which are not gauge equivalent to the trivial connection and that there are two degrees of freedom for such connections. In electromagnetism, we also interpret $H^1 \cong \mathbb{R}^2$ by saying that there are two *Bohm–Aharonov modes* [6, 41].

For the sake of characterizing the topology more fully, we should calculate the other cohomology groups as well. We find

$$H^0 = \frac{\operatorname{span}\{v^1 - v^2\}}{\{0\}} \cong \frac{\mathbb{R}}{\{0\}} \cong \mathbb{R}$$

and

$$H^2 = \frac{C^2}{\operatorname{span}\{p^1 - p^2\}} \cong \frac{\mathbb{R}^2}{\mathbb{R}} \cong \mathbb{R}.$$

For doing path integrals on this lattice, it is convenient to take the following orthonormal basis for C^1:

$$e^{1'} = \tfrac{1}{\sqrt{2}}(e^1 - e^3), \qquad e^{2'} = \tfrac{1}{\sqrt{2}}(e^1 + e^3),$$
$$e^{3'} = \tfrac{1}{\sqrt{2}}(e^2 + e^4), \qquad e^{4'} = \tfrac{1}{\sqrt{2}}(e^2 - e^4),$$

since then $\{e^{1'}\}$ is a basis for $(Z^1)^\perp$ and $\{e^{1'}, e^{2'}, e^{3'}\}$ is a basis for $(B^1)^\perp$. In terms of this new basis, we can write the connection as

$$A = A_{i'}e^{i'} = \frac{(A_1 - A_3)}{\sqrt{2}}e^{1'} + \frac{(A_1 + A_3)}{\sqrt{2}}e^{2'} + \frac{(A_2 + A_4)}{\sqrt{2}}e^{3'} + \frac{(A_2 - A_4)}{\sqrt{2}}e^{4'}.$$

For simplicity, let us assume that the area of each plaquette is 1, so the action becomes

$$S(A) = \frac{1}{2e^2}((A_1 - A_3)^2 + (A_3 - A_1)^2) = \frac{2}{e^2}A_{1'}^2.$$

Now just as in the previous example, integrating over connections mod gauge transformations, i.e. C^1/B^1, is the same as integrating over $(B^1)^\perp = \mathrm{span}\{e^{1'}, e^{2'}, e^{3'}\}$, relative to which basis A has components $(A_{1'}, A_{2'}, A_{3'})$. But $S(A)$ is obviously a degenerate quadratic form in the variables $A_{1'}, A_{2'}, A_{3'}$, so the path integral diverges:

$$\int_{C^1/B^1} \mathrm{e}^{-S(A)}\mathcal{D}A = \int_{\mathbb{R}^3} \mathrm{e}^{-2A_{1'}{}^2/e^2} dA_{1'}dA_{2'}dA_{3'} = \infty.$$

Similarly, integrating over connections mod flat connections, i.e. C^1/Z^1, is the same as integrating over $(Z^1)^\perp$. But $S(A)$ is obviously nondegenerate in the variable $A_{1'}$, which is the coordinate of A relative to the basis $\{e^{1'}\}$ of $(Z^1)^\perp$. In fact, in this case we get

$$\int_{C^1/Z^1} \mathrm{e}^{-S(A)}\mathcal{D}A = \int_{\mathbb{R}} \mathrm{e}^{-2A_{1'}{}^2/e^2} dA_{1'} = \sqrt{\frac{\pi}{2}}e.$$

So we see that to make the path integral for $\mathbb{R}$ electromagnetism on the torus converge, we must kill off both of the Bohm–Aharonov modes, in addition to factoring out gauge freedom.

One *can* give a manifestly convergent description of p-form electromagnetism on the torus without ignoring the Bohm–Aharonov effect, but only by switching the gauge group to $U(1)$. We return to the torus example in section 7.4 where we in fact compute the $U(1)$ partition function more generally for any surface of genus $g \geqslant 0$.

6.3. Path integrals and the p-form Bohm–Aharonov effect

We have seen that the criterion in $\mathbb{R}$ electromagnetism for path integrals over the physical configuration space $\mathcal{A}/\mathcal{G}$ to converge is that the first cohomology be trivial, or in other words that the spacetime has no Bohm–Aharonov modes:

$$\left(\begin{array}{c} Z = \int_{\mathcal{A}/\mathcal{G}} \mathrm{e}^{-S(A)}\mathcal{D}A \\ \text{converges for } \mathbb{R} \text{ electromagnetism} \end{array}\right) \iff H^1 = 0.$$

More generally, p-form electromagnetism has what we might call the p-form Bohm–Aharonov effect, but whereas the ordinary Bohm–Aharonov effect applies to regions of spacetime with nontrivial *first* cohomology, the p-form version depends on the pth cohomology. We have also seen that getting path integrals in $\mathbb{R}$ p-form electromagnetism to converge requires killing off all 'p-form Bohm–Aharonov modes':

$$\left(\begin{array}{c} Z = \int_{\mathcal{A}/\mathcal{G}} \mathrm{e}^{-S(A)}\mathcal{D}A \\ \text{converges for } \mathbb{R} \ p\text{-form electromagnetism} \end{array}\right) \iff H^p = 0.$$

In fact, when H^p is trivial, we can sometimes simplify the process of taking path integrals further. To see this, note that the homomorphism $d_p : C^p \to C^{p+1}$ induces an isomorphism

$$\frac{C^p}{Z^p} \cong B^{p+1}.$$

In the case where H^p is trivial, $C^p/Z^p = C^p/B^p$, which is the physical configuration space. If in addition d_p is *onto*, so that $B^{p+1} = C^{p+1}$, we then have an isomorphism

$$\frac{\mathcal{A}}{\mathcal{G}} = \frac{C^p}{Z^p} \cong C^{p+1}.$$

In other words, when there are no p-form Bohm–Aharonov modes and d_p is surjective, we can just as well do path integrals over the space of curvatures F, rather than over the space of p-connections, at least in the absence of matter. In our first example in the previous subsection, H^1 is trivial and every possible curvature is d of some connection, so we could calculate the partition function more easily as

$$Z = \frac{1}{2\pi e^2\sqrt{V_1V_2}} \int_{C^2} \mathrm{e}^{-\frac{1}{2e^2}({F_1}^2/V_1+{F_2}^2/V_2)} dF_1 dF_2 = 1.$$

The value of the partition function is different in this case, but this does not matter since Z is just a normalization factor—what matters is that expectation values of observables are the same. Using this method, it is also a simple matter to calculate expectation values, for example

$$\langle {F_i}^2 \rangle = e^2 V_i, \qquad i = 1, 2.$$

For p-form electromagnetism with gauge group $\mathbb{R}$, we thus have several choices for what space to integrate over when we do path integrals.

(i) $\mathcal{A}$, the space of p-connections. This is the naive approach of section 2 and almost never gives convergent partition functions: only when there are no nontrivial flat p-connections.

(ii) $\mathcal{A}/\mathcal{G}$, the space of p-connections mod gauge transformations. This works beautifully when there are no p-form Bohm–Aharonov modes, since all path integrals over $\mathcal{A}/\mathcal{G}$ are then finite.

(iii) $\mathcal{A}/\mathcal{A}_0$, the space of p-connections mod flat p-connections. This always gives convergent path integrals and is the same as $\mathcal{A}/\mathcal{G}$ when the pth cohomology is trivial. However, in the case of nontrivial cohomology, it means ignoring the p-form Bohm–Aharonov modes.

(iv) $\mathcal{F} := C^{p+1}$, the space of curvatures of p-connections. In the case where the pth cohomology is trivial and $d{:}C^p \to C^{p+1}$ is onto, this is equivalent to the previous two options and generally far easier to calculate. This option is however less viable when we add matter, since matter fields typically couple to the A field, not F.

(v) $\mathcal{F}_0 := B^{p+1}$, the space of curvatures F, subject to the constraint $dF = 0$. This is like the previous option, but works even when $d{:}C^p \to C^{p+1}$ is not onto. The practical trade-off is that the constraint makes calculating integrals less straightforward. Like the previous option, this works best in the free field setting.

The major difficulty in the case of gauge group $\mathbb{R}$ is that none of these options allows us to take any p-form Bohm–Aharonov modes into account and get convergent path integrals. Of course, it is not convergent partition functions that matter in quantum field theory, but convergent *expectation values*. The fact that the infinities are not removable in nontrivial cohomology without factoring out p-form Bohm–Aharonov modes indicates that these infinities are actually an intrinsic physical aspect of the theory. While one could make a deeper analysis of the $\mathbb{R}$-theory, we instead resolve the issue in section 7 by switching to the *compact* gauge group $U(1)$, where all divergences disappear.

7. $U(1)$ p-form electromagnetism

7.1. Gauge groups for p-form electromagnetism: $\mathbb{R}$ versus $U(1)$

We have seen that eliminating all of the divergences from path integrals in $\mathbb{R}$ p-form electromagnetism requires that we factor out not only all of the gauge freedom, but also

the Bohm–Aharonov modes. This is undesirable: the Bohm–Aharonov effect is an empirical fact in ordinary 1-form electromagnetism. Neither should we rule out the *p-form* Bohm–Aharonov effect in our calculations. This is one advantage of doing p-form electromagnetism with gauge group $U(1)$: the $U(1)$ theory is free of divergences, even in cases where the pth cohomology is nontrivial! Better yet, $U(1)$ path integrals converge even before we factor out the gauge freedom. The reason for this is simple. The space of $U(1)$ p-connections is

$$\mathcal{A} = U(1)^{X_p},$$

where X_p is the set of p-cells in the lattice. But this is just a product of circles—a torus! Since a torus is compact,

$$\int_{U(1)^{X_p}} f(A)\, \mathrm{e}^{-S(A)} \mathcal{D}A$$

converges for any continuous function f of A. For f a physical observable, gauge invariance implies that the result is the same whether we integrate over $U(1)$ p-connections mod gauge transformations or over all $U(1)$ p-connections.

Though a compact gauge group clearly has certain advantages, as it turns out, the $U(1)$ theory can be obtained most easily by appealing to the $\mathbb{R}$ theory. In particular, without having developed the $\mathbb{R}$ theory, it would be difficult to guess the best action to use in evaluating $U(1)$ path integrals. The reason is that since the curvature lies not in the vector space $\mathbb{R}^{X_{p+1}}$ but in the mere group $U(1)^{X_{p+1}}$, we do not have the same analogy between cochains and differential forms as we had in the $\mathbb{R}$ case. In particular, we cannot rely directly on an inner product of cochains for the action as we did in the $\mathbb{R}$ case in section 4. What we will show is that we can turn p-form electromagnetism with gauge group $\mathbb{R}$ into p-form electromagnetism with gauge group $U(1)$, essentially by 'wrapping the real line around the circle'.

The key to understanding the relationship between the Abelian gauge groups $\mathbb{R}$ and $U(1)$ is the short exact sequence:

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{R} \longrightarrow U(1) \longrightarrow 0,$$

where the second map is the usual inclusion and the third sends $x \in \mathbb{R}$ to $\mathrm{e}^{2\pi \mathrm{i} x} \in U(1)$. This says that we map $\mathbb{R}$ to $U(1)$ by winding the line around the circle once every 2π and that the kernel of this map is precisely $\mathbb{Z}$.

Now given any $\mathbb{R}$ p-connection

$$A{:}C_p \to \mathbb{R},$$

we get a $U(1)$ p-connection

$$\hat{A} := \mathrm{e}^{2\pi \mathrm{i} A}{:}C_p \to U(1)$$

by composition with the homomorphism $\mathbb{R} \to U(1)$. We thus get a map

$$\begin{aligned} \hom(C_p, \mathbb{R}) &\to \hom(C_p, U(1)), \\ A &\mapsto \hat{A} \end{aligned}$$

and the fact that the above sequence of Abelian groups is exact implies that the kernel of this map is $\hom(C_p, \mathbb{Z})$. Moreover, since C_p is free, every $U(1)$ p-connection comes from an $\mathbb{R}$ p-connection in precisely this way. Given a p-connection $A{:}C_p \to U(1)$, map each generator c in the basis X_p of C_p to some $\tilde{A}(c) \in \mathbb{R}$ with $\exp(2\pi\, \mathrm{i}\tilde{A}(c)) = A(c)$. This defines an $\mathbb{R}$ p-connection $\tilde{A}$ which 'lifts' A:

$$\begin{array}{ccccc} & & C_p & & \\ & \overset{\tilde{A}}{\swarrow} & \downarrow{\scriptstyle A} & & \\ \mathbb{R} & \longrightarrow & U(1) & \longrightarrow & 0 \end{array}$$

That is, we have the identity

$$\hat{\hat{A}} = A.$$

In fact, each of the sequences

$$0 \longrightarrow \hom(C_k, \mathbb{Z}) \longrightarrow \hom(C_k, \mathbb{R}) \longrightarrow \hom(C_k, U(1)) \longrightarrow 0$$

is exact, thus giving a short exact sequence of chain maps:

$$\begin{array}{ccccccc}
0 & \longrightarrow & \hom(C_n, \mathbb{Z}) & \longrightarrow & \hom(C_n, \mathbb{R}) & \longrightarrow & \hom(C_n, U(1)) \longrightarrow 0 \\
 & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} \\
 & & \vdots & & \vdots & & \vdots \\
 & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} \\
0 & \longrightarrow & \hom(C_1, \mathbb{Z}) & \longrightarrow & \hom(C_1, \mathbb{R}) & \longrightarrow & \hom(C_1, U(1)) \longrightarrow 0 \\
 & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} & & \uparrow{\scriptstyle d} \\
0 & \longrightarrow & \hom(C_0, \mathbb{Z}) & \longrightarrow & \hom(C_0, \mathbb{R}) & \longrightarrow & \hom(C_0, U(1)) \longrightarrow 0
\end{array}$$

which we may write more succinctly as

$$0 \longrightarrow \hom(C_\bullet, \mathbb{Z}) \longrightarrow \hom(C_\bullet, \mathbb{R}) \longrightarrow \hom(C_\bullet, U(1)) \longrightarrow 0\,.$$

Lattice p-form electromagnetism with Abelian gauge group G is all about the cochain complex $\hom(C_\bullet, G)$. We may thus express the physical content of the above exact sequence of chain maps by saying that there is a projection from the theory with gauge group $\mathbb{R}$ to the theory with gauge group $U(1)$, as we wanted, and that the kernel of this projection is a theory with gauge group $\mathbb{Z}$. Metaphorically,

$$0 \longrightarrow \begin{pmatrix}\text{lattice } p\text{-form}\\ \text{electromagnetism}\\ \text{with gauge group}\\ \mathbb{Z};\end{pmatrix} \longrightarrow \begin{pmatrix}\text{lattice } p\text{-form}\\ \text{electromagnetism}\\ \text{with gauge group}\\ \mathbb{R};\end{pmatrix} \longrightarrow \begin{pmatrix}\text{lattice } p\text{-form}\\ \text{electromagnetism}\\ \text{with gauge group}\\ U(1)\end{pmatrix} \longrightarrow 0$$

is exact.

One critical implication of the exact sequence of cochain complexes is that two $\mathbb{R}$ p-connections A, $A' \in \hom(C_p, \mathbb{R})$ which are gauge equivalent, say

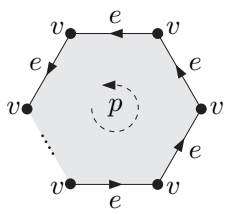
$$A - A' = d\phi, \qquad \phi \in \hom(C_{p-1}, \mathbb{R}),$$

project down to $U(1)$ p-connections $\hat{A}$, $\hat{A}' \in \hom(C_p, U(1))$ which are gauge equivalent:

$$\hat{A} - \hat{A}' = d\hat{\phi}, \qquad \hat{\phi} \in \hom(C_{p-1}, U(1)).$$

Briefly, $\mathbb{R}$ gauge equivalence implies $U(1)$ gauge equivalence.

But for some topologies, there is a genuine difference between the two choices of gauge group: $\mathbb{R}$-valued cohomology is oblivious to 'torsion'. A simple example is a space which is sometimes called an 'm-fold dunce cap', DC_m. This can be constructed from one vertex v, one edge e and one plaquette p which is sewn along its boundary around the looped edge in such a way that it wraps around m times:



The dunce cap has cohomology groups

$$H^0(\mathrm{DC}_m;\mathbb{R}) \cong \mathbb{R}, \qquad H^1(\mathrm{DC}_m;\mathbb{R}) = 0, \qquad H^2(\mathrm{DC}_m;\mathbb{R}) \cong 0$$

for $G = \mathbb{R}$, but

$$H^0(\mathrm{DC}_m;U(1)) \cong U(1), \qquad H^1(\mathrm{DC}_m;U(1)) \cong \mathbb{Z}/m, \qquad H^2(\mathrm{DC}_m;U(1)) = 0$$

for $G = U(1)$. The significant difference is in the first cohomology. The fact that H^1 is trivial in the $\mathbb{R}$ case shows that every $\mathbb{R}$-valued 1-cocycle is a 1-coboundary. Said another way, all flat $\mathbb{R}$ connections on DC_m are gauge equivalent:

$$dA = dA' = 0 \quad \Longrightarrow \quad d(A - A') = 0 \quad \Longrightarrow \quad A - A' = d\phi.$$

In fact, this result is not surprising given our particular construction of DC_m—there is only *one* flat $\mathbb{R}$ connection, the zero connection. But in the $U(1)$ case, H^1 is a cyclic group of order m. It is not hard to see why. There are m flat $U(1)$ connections on DC_m, given by assigning to the edge e any one of the m distinct mth roots of unity in $U(1)$.

7.2. Gaussian integrals on a torus

In the $U(1)$ version of lattice electromagnetism, the holonomy of the connection takes values on the circle $U(1)$. To do path integrals, we need the analogue of a Gaussian on a circle, which we obtain from the ordinary Gaussian by wrapping the real line around the circle using the exact sequence of groups discussed in the previous section. In particular, since

$$\frac{\mathrm{e}^{-x^2/2\sigma^2}dx}{\sqrt{2\pi}\sigma}$$

is a probability measure on $\mathbb{R}$,

$$\sum_{n\in\mathbb{Z}} \frac{\mathrm{e}^{-(\theta+2n\pi)^2/2\sigma^2}d\theta}{\sqrt{2\pi}\sigma} \tag{5}$$

is a probability measure on $U(1) = \mathbb{R}/2\pi\mathbb{Z}$. Since the normalized Haar measure on $U(1)$ is $d\theta/2\pi$, it is really the function

$$\frac{\sqrt{2\pi}}{\sigma}\sum_{n\in\mathbb{Z}} \mathrm{e}^{-(\theta+2n\pi)^2/2\sigma^2} \tag{6}$$

which is the appropriate analogue of the Gaussian function on the circle. Because of this analogy, we refer to both the measure (5) and the corresponding function (6) as 'circular Gaussians'.

We can use circular Gaussian measure to calculate the expected value of a function $f{:}U(1) \to \mathbb{R}$ of the random variable θ as follows:

$$\begin{aligned}\langle f\rangle &= \int_0^{2\pi} f(\theta)\sum_{n\in\mathbb{Z}} \frac{\mathrm{e}^{-(\theta+2n\pi)^2/2\sigma^2}d\theta}{\sqrt{2\pi}\sigma} \\ &= \frac{1}{\sqrt{2\pi}\sigma}\sum_{n\in\mathbb{Z}}\int_{2n\pi}^{2(n+1)\pi} f(\theta)\,\mathrm{e}^{-\theta^2/2\sigma^2}d\theta \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} f(\theta)\,\mathrm{e}^{-\theta^2/2\sigma^2}d\theta.\end{aligned}$$

What this calculation shows is that integrating a function $f{:}U(1) \to \mathbb{R}$ against the measure in (5) is the same as extending f to a periodic function $f{:}\mathbb{R} \to \mathbb{R}$ and then integrating this against the usual Gaussian measure.

More generally, a Gaussian on $\mathbb{R}^N$

$$\mathrm{e}^{-\frac{1}{2}x\cdot Qx} = \mathrm{e}^{-\frac{1}{2}Q^{ij}x_i x_j}$$

can be wrapped around a *torus*, with each of the N coordinates wrapping around a circle as in the one-dimensional case, to give

$$\sum_{n\in\mathbb{Z}^N} \mathrm{e}^{-\frac{1}{2}Q^{ij}(\theta_i+2n_i\pi)(\theta_j+2n_j\pi)},$$

which we might call a 'toroidal Gaussian'.

Since a periodic function on $\mathbb{R}^N$ (or, equivalently, a function on the N-torus) can be expanded in a Fourier series, we should work out the expected values for each element of the Fourier basis $\{\mathrm{e}^{\mathrm{i}m^k\theta_k}\}$. For any $m \in \mathbb{Z}^N$, we have

$$\begin{aligned}
\langle \mathrm{e}^{\mathrm{i}m^k\theta_k}\rangle &= \sqrt{\frac{\det(Q^{ij})}{(2\pi)^N}} \sum_{n\in\mathbb{Z}^N} \int_{U(1)^N} \mathrm{e}^{\mathrm{i}m^k\theta_k}\, \mathrm{e}^{-\frac{1}{2}Q^{ij}(\theta_i+2n_i\pi)(\theta_j+2n_j\pi)} d\theta \\
&= \sqrt{\frac{\det(Q^{ij})}{(2\pi)^N}} \int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}m^k\theta_k}\, \mathrm{e}^{-\frac{1}{2}Q^{ij}\theta_i\theta_j} d\theta \\
&= \sqrt{\frac{\det(Q^{ij})}{(2\pi)^N}} \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}Q^{ij}(\theta_i\theta_j - 2\mathrm{i}Q^{-1}_{jk}m^k\theta_i)} d\theta \\
&= \sqrt{\frac{\det(Q^{ij})}{(2\pi)^N}}\, \mathrm{e}^{-\frac{1}{2}Q^{-1}_{kl}m^k m^l} \int_{\mathbb{R}^N} \mathrm{e}^{-\frac{1}{2}Q^{ij}(\theta_i - \mathrm{i}Q^{-1}_{ik}m^k)(\theta_j - \mathrm{i}Q^{-1}_{jl}m^l)} d\theta \\
&= \mathrm{e}^{-\frac{1}{2}Q^{-1}_{kl}m^k m^l}.
\end{aligned} \tag{7}$$

In fact, what we are calling 'toroidal Gaussians' are actually famous special functions in their own right, though they are usually presented in a slightly different form. Using the Fourier coefficient $\langle \mathrm{e}^{\mathrm{i}m^k\theta_k}\rangle$ above, one can expand the toroidal Gaussian itself in the basis $\{\mathrm{e}^{\mathrm{i}m\theta}\}$, resulting in the identity

$$\sqrt{(2\pi)^N \det Q} \sum_{n\in\mathbb{Z}^N} \mathrm{e}^{-\frac{1}{2}Q^{ij}(\theta_i+2n_i\pi)(\theta_j+2n_j\pi)} = \sum_{n\in\mathbb{Z}^N} \mathrm{e}^{-\frac{1}{2}Q^{-1}_{ij}n^i n^j}\, \mathrm{e}^{\mathrm{i}n^k\theta_k}. \tag{8}$$

The form on the right-hand side of this equality makes it easier to recognize that our circular Gaussian is really

$$\vartheta\left(\frac{\theta}{2\pi}, \mathrm{i}\frac{Q^{-1}}{2\pi}\right),$$

where ϑ denotes the *theta function*. This is defined by

$$\vartheta(z,\Omega) := \sum_{n\in\mathbb{Z}^N} \mathrm{e}^{\pi\,\mathrm{i}n\cdot\Omega n + 2\pi\,\mathrm{i}n\cdot z},$$

where $z \in \mathbb{C}^N$ and Ω is a symmetric $N \times N$ complex matrix with positive definite imaginary part [33]. When convenient, we continue to write this theta function as $\mathrm{e}^{-\frac{1}{2}\theta\cdot Q\theta}$ and think of it as a 'toroidal Gaussian'. We engage in this gross abuse of notation because it is evocative of the analogy to Gaussians on $\mathbb{R}^N$. In particular, it is suggestive of how one actually calculates *integrals* using the corresponding 'theta measure'—by unrolling the torus to $\mathbb{R}^N$ and integrating an ordinary Gaussian.

What we have seen is that theta functions are to tori as Gaussians are to Euclidean spaces. This may be seen in other ways as well: the theta function is the fundamental solution of the

heat equation on a torus, just as the Gaussian is the fundamental solution of the heat equation in Euclidean space. For our purposes, the upshot is that theta functions are the natural kernels to use in path integrals in $U(1)$ lattice p-form electromagnetism. We turn now to the task of calculating such path integrals.

7.3. $U(1)$ path integrals

Recall that for gauge group $\mathbb{R}$, our action for discrete p-form electromagnetism is given by

$$S(A) = \frac{1}{2e^2}\langle F, F\rangle = \frac{1}{2e^2}h^{ij}F_iF_j,$$

where h^{ij} is the matrix of the inner product relative to the basis of $(p+1)$-cochains consisting of dual $(p+1)$-cells. As discussed in the previous subsection, making the transition from $\mathbb{R}$ to $U(1)$ involves replacing the ordinary Gaussian $\exp(-S)$ in the real variables F_i by a 'toroidal Gaussian' in the $U(1)$-valued variables F_i:

$$\begin{aligned} \mathrm{e}^{-S(A)} &= \sum_{n\in\mathbb{Z}^N} \mathrm{e}^{-\frac{1}{2e^2}h^{ij}(F_i+2n_i\pi)(F_j+2n_j\pi)} \\ &= \sum_{n\in\mathbb{Z}^N} \mathrm{e}^{-\frac{1}{2e^2}h^{ij}(A(\partial x_i)+2n_i\pi)(A(\partial x_j)+2n_j\pi)}, \end{aligned} \tag{9}$$

where $N = |X_{p+1}|$ is the number of $(p+1)$-cells in the spacetime, which we now assume to be finite. Note that we do not define the *action* S but only the analogue of its exponential e^{-S}, which is all we need for doing path integrals.

In defining the path integral in the $\mathbb{R}$ theory, we normalized the corresponding Gaussian function of the curvature with respect to Lebesgue measure. Accordingly, we here use a theta function normalized with respect to Haar measure on $U(1)^{X^{p+1}}$. The path integral is thus

$$Z = \sqrt{\det(2\pi h^{ij})}\int_{\mathcal{A}}\sum_{n\in\mathbb{Z}^N}\mathrm{e}^{-\frac{1}{2e^2}h^{ij}(A(\partial x_i)+2n_i\pi)(A(\partial x_j)+2n_j\pi)}\mathcal{D}A, \tag{10}$$

where, because of the compactness of the group, there is no need to mod out by gauge transformations and we simply integrate over all p-connections, with respect to Haar measure

$$\mathcal{D}A = \prod_{c_i\in X_p}\frac{dA_i}{2\pi}$$

on $U(1)^{X_p}$.

7.4. The $U(1)$ theory in $p+1$ dimensions and relation to 2D Yang–Mills

Just as in the case of gauge group $\mathbb{R}$, the $U(1)$ theory in $p+1$ dimensions is greatly simplified by the fact that the inner product on $\mathbb{R}$-valued $(p+1)$-cochains is diagonal in the canonical basis of dual $(p+1)$-cells, with volumes of $(p+1)$-cells as weights. The resulting partition function obtained from (10) is

$$\begin{aligned} Z &= \int_{U(1)^{X_p}}\prod_{c\in X_{p+1}}\sqrt{\frac{2\pi}{e^2\,\mathrm{vol}(c)}}\,\mathrm{e}^{-S(A|c)}\mathcal{D}A \\ &= \int_{U(1)^{X_p}}\prod_{c\in X_{p+1}}\sqrt{\frac{2\pi}{e^2\,\mathrm{vol}(c)}}\sum_{n\in\mathbb{Z}}\mathrm{e}^{-(A(\partial c)+2n\pi)^2/2e^2\,\mathrm{vol}(c)}\mathcal{D}A. \end{aligned} \tag{11}$$

It is well known that 2D Yang–Mills theory has a particularly nice lattice formulation, originally introduced by Migdal [30] and later worked out in greater generality by Witten [39]. We wish to show now that in the case $p = 1, n = 2$, our theory for discrete electromagnetism agrees with the $U(1)$ version of Witten's lattice Yang–Mills theory in two dimensions. In general, Witten defines the lattice Yang–Mills partition function to be

$$Z(V) = \int_{\mathcal{A}} \prod_{c \in X_2} \sum_{\rho \in \operatorname{Irrep}(G)} \dim(\rho) \chi_\rho(F(c))\, \mathrm{e}^{-e^2 \operatorname{vol}(c) \cdot C(\rho)/2} \mathcal{D}A, \tag{12}$$

where $\operatorname{Irrep}(G)$ is the set of isomorphism classes of irreducible representations of the gauge group G, χ_ρ is the character—the trace in the representation ρ—of the plaquette holonomy F and C is the quadratic Casimir operator corresponding to the Killing form on G. Following Witten's convention, we write $Z = Z(V)$ where $V = \sum \operatorname{vol}(c)$ is the total volume of spacetime.

Let us specialize to the case $G = U(1)$. $U(1) = \mathbb{R}/2\pi\mathbb{Z}$ has only one-dimensional irreducible representations ρ_n, labelled by integers, where $\rho_n(\theta){:}\mathbb{C} \to \mathbb{C}$ is simply multiplication by $\mathrm{e}^{\mathrm{i}n\theta}$. The characters are thus

$$\chi_{\rho_n}(F(c)) := \mathrm{e}^{\mathrm{i}nF(c)}.$$

Next, since $\mathfrak{u}(1)$ is an Abelian Lie algebra, the quadratic Casimir is not determined by the Killing form as it is for semisimple Lie algebras. However, the circle $U(1)$ has an obvious Riemannian metric, and the quadratic Casimir determined by this metric is simply $C(\rho_n) = n^2$. Thus, for 2D electromagnetism, the Yang–Mills partition function (12) becomes

$$Z(V) = \int_{U(1)^{X_2}} \prod_{c \in X_2} \sum_{n \in \mathbb{Z}} \mathrm{e}^{\mathrm{i}nA(\partial c)}\, \mathrm{e}^{-e^2 \operatorname{vol}(c) n^2/2} \mathcal{D}A.$$

At first, this seems rather different from (11). However, specializing the theta function identity (8) to one dimension, we obtain

$$\frac{\sqrt{2\pi}}{\sigma} \sum_{n \in \mathbb{Z}} \mathrm{e}^{-(\theta + 2n\pi)^2/2\sigma^2} = \sum_{n \in \mathbb{Z}} \mathrm{e}^{-n^2\sigma^2/2}\, \mathrm{e}^{\mathrm{i}n\theta}. \tag{13}$$

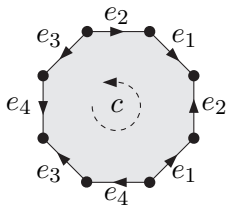
Substitution of this identity produces an alternative expression for the partition function:

$$Z(V) = \int_{U(1)^{X_2}} \prod_{c \in X_2} \sqrt{\frac{2\pi}{e^2 \operatorname{vol}(c)}} \sum_{n \in \mathbb{Z}} \mathrm{e}^{-(A(\partial c) + 2n\pi)^2/2e^2 \operatorname{vol}(c)} \mathcal{D}A,$$

which is the same as (11) with $p = 1$. The results obtained by Witten [39] are thus applicable to the theory we have developed here as well. In particular, 2D lattice electromagnetism as we have developed has the correct continuum limit. Better yet, just as in 2D Yang–Mills, the lattice theory is independent of the discretization, so that in path integral calculations one need not use a lattice finer than the observables one is interested in. Witten's results for Yang–Mills theory on a surface carry over directly. It is perhaps worth working this out explicitly in the $U(1)$ theory, since we have now developed the machinery to easily do so. For simplicity, we confine our attention to compact orientable surfaces.

A surface of genus $g > 0$ can be described by the usual pairwise gluing of edges of a single polygon with $4g$ sides, with edges labelled $e_1, e_2, e_1^{-1}, e_2^{-1}, \ldots, e_{2g-1}, e_{2g}, e_{2g-1}^{-1}, e_{2g}^{-1}$.

Thanks to discretization independence, in lattice electromagnetism we are thus free to assume a lattice with just one plaquette c, $2g$ edges and one vertex. For example, for $g = 2$:



The plaquette has vanishing boundary, as do each of its edges since they each begin and end at the single vertex in the discretization. Hence, every discrete connection is flat and gauge transformations act trivially. The cochain complex is thus

$$\begin{array}{ccccc} C^0 & \xrightarrow{d} & C^1 & \xrightarrow{d} & C^2 \\ \| & & \| & & \| \\ U(1) & \xrightarrow{0} & U(1)^{2g} & \xrightarrow{0} & U(1) \end{array}$$

from which we can easily read off the cohomology: $H^1 \cong U(1)$, $H^2 \cong U(1)^{2g}$, $H^2 \cong U(1)$. For the path integral, we thus have

$$\begin{aligned} Z(V) &= \sqrt{\frac{2\pi}{e^2 V}} \int_0^{2\pi} \cdots \int_0^{2\pi} \sum_{n\in\mathbb{Z}} \mathrm{e}^{-(F(c)+2n\pi)^2/2e^2V} \prod_{i=1}^{2g} \frac{dA_i}{2\pi} \\ &= \sqrt{\frac{2\pi}{e^2 V}} \sum_{n\in\mathbb{Z}} \mathrm{e}^{-(2n\pi)^2/2e^2V} \\ &= \sum_{n\in\mathbb{Z}} \mathrm{e}^{-n^2e^2V/2} = \vartheta\left(0, \frac{\mathrm{i}e^2V}{2\pi}\right), \end{aligned}$$

where we have used the theta function identity (13). This formula is also valid for the case $g = 0$, a sphere, which can be constructed from a bigon by identifying its two edges. These results are of course just as we would obtain by specializing Witten's 2D Yang–Mills partition function to the case $G = U(1)$.

Yet our theory allows one to go further, in the Abelian case, describing the p-form theory on analogous decompositions of closed $(p + 1)$-manifolds. For example, we can describe a Kalb–Ramond field on the 3-torus, which has an obvious lattice decomposition which may be constructed by identifying opposite faces of a cube just as a torus is constructed by identifying sides of a square. The lattice suggested by this construction has a single 3-cell, three plaquettes, three edges and one vertex. Gauge transformations assign phases to edges, but these act trivially on the space of 2-connections, since each plaquette is itself a torus, hence has vanishing boundary. Similarly, all 2-connections on this lattice are flat, since the boundary of the single 3-cell also vanishes. Each of the differentials in the cochain complex thus vanishes:

$$\begin{array}{ccccccc} C^0 & \xrightarrow{d} & C^1 & \xrightarrow{d} & C^2 & \xrightarrow{d} & C^3 \\ \| & & \| & & \| & & \| \\ U(1) & \xrightarrow{0} & U(1)^3 & \xrightarrow{0} & U(1)^3 & \xrightarrow{0} & U(1) \end{array}$$

Calculation of the partition function is nearly identical to the 1-form case on the 2-torus. We get

$$Z(V) = \sum_{n\in\mathbb{Z}} \mathrm{e}^{-n^2e^2V/2}.$$

In fact, it is easy to see that this result generalizes to closed manifolds of any dimension.

Theorem 1. *For any connected closed orientable (smooth or piecewise-linear) $(p+1)$-manifold M of total volume V, the partition function for p-form electromagnetism on M is*

$$Z(V) = \sum_{n\in\mathbb{Z}} \mathrm{e}^{-n^2e^2V/2}.$$

Proof. Any connected closed oriented $(p+1)$-manifold can be constructed from a single $(p+1)$-cell by identifying portions of its p-dimensional boundary in an orientation-preserving way. There are various ways to see this. In the context of smooth manifolds, equip M with a Riemannian metric. Let U be a maximal connected open neighbourhood of $0 \in T_x(M)$ such that $\exp|_U : U \to M$ is injective. The closure $\bar{U}$ is a closed $(p+1)$-ball and the exponential map $\bar{U} \to M$ is a quotient map which glues together pieces of the boundary of $\bar{U}$. In the more general context of piecewise-linear manifolds, one can pick a triangulation of M and argue that the construction of M by gluing p-dimensional faces of the $(p+1)$-simplices may be done in two steps, first obtaining a single $(p+1)$-cell and then identifying the remaining faces. Either construction suggests a particular lattice decomposition with just one $(p+1)$-cell with vanishing boundary. Hence, every p-connection on this lattice is flat. The calculation of the partition function is thus formally identical to the analogous calculation for a closed surface above. Since $F = A \circ \partial = 0$, we have

$$Z(V) = \sqrt{\frac{2\pi}{e^2V}} \int_{\mathcal{A}} \sum_{n\in\mathbb{Z}} \mathrm{e}^{-(F+2n\pi)^2/2e^2V} \mathcal{D}A = \sum_{n\in\mathbb{Z}} \mathrm{e}^{-n^2e^2V/2}.$$

By discretization independence, this holds for any lattice decomposition M, hence also in the continuum limit. □

One can of course obtain similar results in nonorientable cases, but we shall not do so here.

We have said that p-form electromagnetism is 'volumetric' in $p+1$ dimensions. What is somewhat surprising, as the above result shows, is that it is in a sense *purely* volumetric: the Abelian nature of the theory makes it relatively insensitive to the topology. Indeed, considering Witten's non-Abelian 2D results together with the partition function for the Abelian p-form theory we have described, one might expect the partition function in 'p-form Yang–Mills' theory, if such a theory exists, to be of the form

$$Z(M) = \sum_{\rho} (\dim \rho)^{\chi(M)} \mathrm{e}^{-e^2V\cdot C(\rho)/2}$$

for a closed oriented $(p+1)$-manifold M of Euler characteristic $\chi(M)$ and volume V. Whether there actually is an exactly solvable non-Abelian 'p-Yang–Mills theory' in $p+1$ dimensions is an interesting question in higher gauge theory, which we leave open to future research.

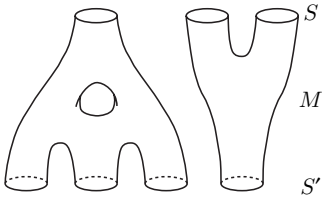
8. Chain field theory

We have seen in sections 4.3 and 7.4 that lattice p-form electromagnetism in $p+1$ dimensions has many similarities with topological quantum field theory—it is what we have called a 'volumetric field theory'. The fact that the volume of spacetime is the only nontopological input to the theory is special to $p+1$ dimensions and is intimately linked

to discretization independence, just as in 2D Yang–Mills. In more general dimension, discretization independence does not hold, and the calculation (7) of expectation values for theta integrals shows that the quantum theory depends in a local way on the chosen inner product on $(p+1)$-cochains.

But even in arbitrary spacetime dimension $n > p$, it is interesting to see to what extent p-form electromagnetism can be given a description which parallels Atiyah's functorial definition of TQFT [2]. In this section, we find that this goal can indeed be realized to a large extent. We warn the reader that this section requires a slight jump in mathematical sophistication. In particular, we need to assume an acquaintance with category theory, referring the reader to [28] for details not presented here.

In topological quantum field theory, 'space' at any given time is a compact $(n-1)$-dimensional manifold, whereas spacetime is an n-dimensional cobordism connecting two slices of space. This is easiest to visualize when $n = 2$, where space is necessarily just a union of circles and a cobordism between slices of space is a 2-manifold with boundary[4]:



A TQFT assigns to each $(n-1)$-manifold S representing space its Hilbert space of states $Z(S)$ and to each cobordism $M{:}S \to S'$ between $(n-1)$-manifolds a linear operator $Z(M){:}Z(S) \to Z(S')$, which we can think of as a 'time evolution' operator. More precisely, a *topological quantum field theory* is a symmetric monoidal functor [28]

$$Z{:}n\mathrm{Cob} \to \mathrm{Hilb}$$

from the category of smooth n-dimensional cobordisms to the category of Hilbert spaces.

Since we have described discrete spacetime in electromagnetism using chain complexes, we would like to define a *chain field theory* to be a symmetric monoidal functor

$$Z{:}n\mathrm{Chain} \to \mathrm{Hilb},$$

where nChain is a category patterned after the category nCob, but with manifolds replaced by chain complexes. However, as we have already noted, the Lagrangian for electromagnetism does not depend merely on topological data. In particular, it relies on having an inner product on $(p+1)$-forms. Consequently, the category nChain will have a richer structure than nCob.

Accordingly, an object S in the category nChain should be an $(n-1)$-complex (in the sense of definition 1)[5]:

$$0 \longleftarrow S_0 \longleftarrow S_1 \longleftarrow \cdots \longleftarrow S_{n-1}$$

equipped with an inner product on $\mathbb{R}$-valued $(p+1)$-cochains:

$$\langle\,,\,\rangle_S{:}\hom(S_{p+1},\mathbb{R}) \times \hom(S_{p+1},\mathbb{R}) \to \mathbb{R}.$$

To make our path integrals well defined, we demand that an object S has finitely many cells, so that, in particular, the dimension of S_{p+1} is finite.

[4] More properly, a cobordism is a diffeomorphism class of such 2-manifolds with boundary.

[5] Here we begin a notational practice which we continue throughout this section. Since we will be dealing with multiple complexes, representing both space and spacetime, we denote, for example, $C_k(S,\mathbb{R})$ or $C_k(S,U(1))$ by the shorthand S_k rather than C_k.

To define a morphism from S to $S' \in n\mathrm{Chain}$, let M be an n-complex, also with finitely many cells, whose chain complex $M_\bullet$ is equipped with injective chain maps

$$S_\bullet \xrightarrow{i} M_\bullet \xleftarrow{i'} S'_\bullet$$

sending basis elements to basis elements. That is, for each $k \in \{0, 1, \ldots, n-1\}$ we have subgroup inclusions $i{:}S_k \to M_k$ and $i'{:}S'_k \to M_k$ such that the following diagram is commutative.

$$\begin{array}{ccccccccccc}
0 & \longleftarrow & S_0 & \longleftarrow & S_1 & \longleftarrow & \cdots & \longleftarrow & S_{n-1} & & \\
 & & \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle i} & & & & \downarrow{\scriptstyle i} & & \\
0 & \longleftarrow & M_0 & \longleftarrow & M_1 & \longleftarrow & \cdots & \longleftarrow & M_{n-1} & \longleftarrow & M_n \\
 & & \uparrow{\scriptstyle i'} & & \uparrow{\scriptstyle i'} & & & & \uparrow{\scriptstyle i'} & & \\
0 & \longleftarrow & S'_0 & \longleftarrow & S'_1 & \longleftarrow & \cdots & \longleftarrow & S'_{n-1} & &
\end{array}$$

We also equip M with an inner product on $\mathbb{R}$-valued $(p+1)$-cochains:

$$\langle\,,\rangle_M{:}\hom(M_{p+1}, \mathbb{R}) \times \hom(M_{p+1}, \mathbb{R}) \to \mathbb{R}$$

and demand that M preserve the inner products from S and S' in the following sense. First, since M_{p+1} is free, we can think of $S^{p+1} = \hom(S_{p+1}, \mathbb{R})$ as a subgroup of $M^{p+1} = \hom(M_{p+1}, \mathbb{R})$ simply by extending any $\omega \in S^{p+1}$ trivially to M^{p+1}. That is, ω becomes a member of M^{p+1} by letting it be zero on $M^{p+1} \setminus S^{p+1}$. We then demand that the inner product $\langle\,,\rangle_M$ is precisely $\langle\,,\rangle_S$ when restricted to $S^{p+1} \times S^{p+1}$.

A morphism in $n\mathrm{Chain}$ will be defined as a certain equivalence class of the gadgets described in the previous paragraph. Suppose we have two such gadgets from S to S':

$$S_\bullet \xrightarrow{i} M_\bullet \xleftarrow{i'} S'_\bullet$$

and

$$S_\bullet \xrightarrow{i} M'_\bullet \xleftarrow{i'} S'_\bullet$$

We say that M and M' are equivalent, written as $M \sim M'$, if there exists an isomorphism of chain complexes

$$\phi{:}M_\bullet \to M'_\bullet$$

preserving both the chosen bases of the M_k and the inner product on M^{p+1}, such that the diagram

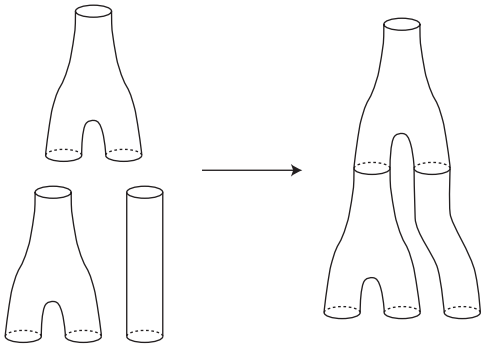
$$\begin{array}{ccccc}
S_\bullet & \xrightarrow{i} & M_\bullet & \xleftarrow{i'} & S'_\bullet \\
 & {\scriptstyle j}\searrow & \downarrow{\scriptstyle \phi} & \swarrow{\scriptstyle j'} & \\
 & & M'_\bullet & &
\end{array}$$

commutes. Finally, we define a morphism

$$[M]{:}S \to S'$$

in $n\mathrm{Chain}$, which we call a *chain cobordism*, to be the equivalence class under the relation $\sim$ of all gadgets $S_\bullet \longrightarrow M_\bullet \longleftarrow S'_\bullet$. Following standard practice, we will generally drop the brackets from $[M]$, writing simply M for our chain cobordism, blurring the distinction between the equivalence class and a representative.

Now if $M{:}S \to S'$ and $N{:}S' \to S''$ are chain cobordisms, we must describe a rule for composing them. Composing two cobordisms between *manifolds* amounts to stacking them on top of each other and gluing their boundaries together smoothly to get a new cobordism:



Similarly, when we compose two chain cobordisms $M{:}S \to S'$ and $N{:}S' \to S''$ represented by

$$S_\bullet \xrightarrow{i} M_\bullet \xleftarrow{i'} S'_\bullet, \qquad S'_\bullet \xrightarrow{j} N_\bullet \xleftarrow{j'} S''_\bullet,$$

we must get a new cobordism $NM{:}S \to S''$

$$S_\bullet \xrightarrow{i} (NM)_\bullet \xleftarrow{j'} S''_\bullet$$

by 'gluing M and N together along S''. Thinking of M and N as the physical lattices whence they derive, we want NM to have as n-cells all of the n-cells of both M and N. This leads us to define

$$(NM)_n := M_n \oplus N_n.$$

Likewise, NM should have all lower-dimensional faces of both M and N, except that we identify cells in M and N if they come from the same cell in S under the gluing maps

$$M_\bullet \xrightarrow{i'} S'_\bullet \xleftarrow{j} N_\bullet$$

The natural way to accomplish this is to let

$$(NM)_k := \frac{M_k \oplus N_k}{\langle i'(c) - j(c) | c \in S'_k \rangle}, \qquad k = 0, 1, \ldots, n-1.$$

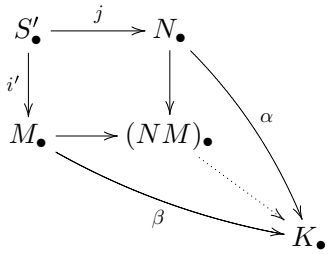
In other words, each NM_k is the direct sum $M_k \oplus N_k$ *modulo* relations declaring $c \in M_k \subset M_k \oplus N_k$ and $c' \in N_k \subset M_k \oplus N_k$ to be equivalent if they are equal as elements of S'_k. In particular, the diagram

$$\begin{array}{ccc} S'_\bullet & \xrightarrow{j} & N_\bullet \\ {\scriptstyle i'}\downarrow & & \downarrow \\ M_\bullet & \longrightarrow & (NM)_\bullet \end{array} \tag{14}$$

commutes, where the unlabelled arrows are the obvious injections. Moreover, this is the *best* way to make such a commutative square, in the sense that if $K_\bullet$ is any other chain complex of Abelian groups equipped with chain maps α and β such that

$$\begin{array}{ccc} S'_\bullet & \xrightarrow{j} & N_\bullet \\ {\scriptstyle i'}\downarrow & & \downarrow{\scriptstyle \alpha} \\ M_\bullet & \xrightarrow[\beta]{} & K_\bullet \end{array}$$

commutes, then there exists a unique chain map $(NM)_\bullet \to K_\bullet$ making



commute. Category theorists express this 'universal property' of NM by saying that diagram (14) is a *pushout square* [28]. We mention this here because the pushout property streamlines the proofs of theorems 2 and 3. Note that $(NM)_k$ are still *free*, since the direct sum of free Abelian groups is free and identifying two basis elements of a free Abelian group gives a free Abelian group.

We must define the inner product on $(p+1)$-forms on the composite NM. Since NM consists of all of the cells of N and M, there is a canonical way to do this. Given $F, G \in \hom((NM)_{p+1}, \mathbb{R})$, we define

$$\langle F, G\rangle_{NM} = \langle F|_N, G|_N\rangle_N + \langle F|_M, G|_M\rangle_M - \langle F|_{S'}, G|_{S'}\rangle_{S'}. \tag{15}$$

Here the first two terms are just the inner products on M and N, while the third term corrects for double counting on the common boundary S' of M and N. Specifically, if x_i and x_j are both $(p+1)$-cells in S', then since S' is a subcomplex of both M and N, each of the first terms in the inner product contains a term of the form $\langle x^i, x^j\rangle$. Subtracting $\langle F|_{S'}, G|_{S'}\rangle_{S'}$ thus eliminates the extra $\langle x^i, x^j\rangle$ term.

The reader already convinced, or else willing to accept on faith, that nChain as we have described does indeed form a category may skip ahead to theorem 3, where we prove the main result. For the more rigorously oriented, we collect some of the remaining technical details in the proof of the following theorem.

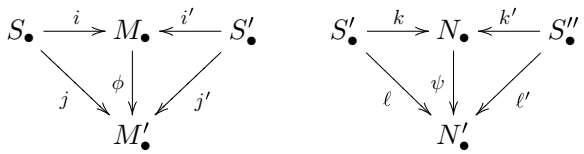
Theorem 2. nChain *is a symmetric monoidal category.*

Proof. To show nChain is a category, we must show that composition of morphisms is well defined (i.e. that it respects equivalence classes), that it is associative and that identity morphisms exist.

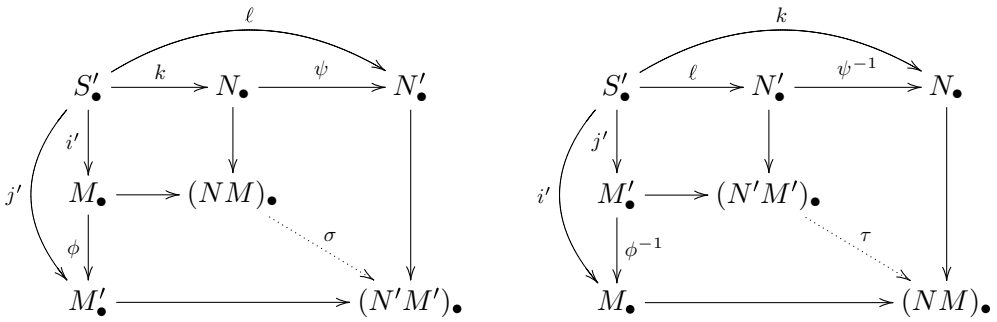
Suppose $M \sim M'$ are two representatives of a chain cobordism from S to S', and $N \sim N'$ representatives of a cobordism from S' to S''. We must show that $NM \sim N'M'$. By the equivalence relation $\sim$, there exist basis-preserving isomorphisms

$$\phi\colon M_\bullet \to M'_\bullet, \qquad \psi\colon N_\bullet \to N'_\bullet$$

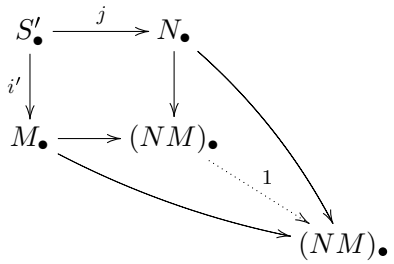
such that the following diagrams commute:



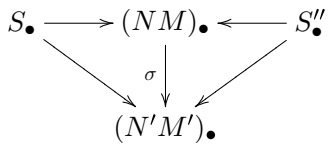
Here is where pushouts come in handy. By the fact that NM and $N'M'$ are both pushouts, there exist unique morphisms σ and τ making the following diagrams commute:



We now use a standard trick of category theory to see that these unique σ and τ are isomorphisms and in fact inverses of each other. Using the uniqueness part of the pushout property, we know that the identity chain map $(NM)_\bullet \xrightarrow{1} (NM)_\bullet$ is the only morphism for which the diagram



commutes. But the diagram also commutes if the dotted arrow is the map $\tau \circ \sigma$, and therefore $\tau \circ \sigma = 1$. An identical argument shows that $\sigma \circ \tau$ is the identity chain map $(N'M')_\bullet \xrightarrow{1} (N'M')_\bullet$. We thus have an isomorphism $\sigma{:}(NM)_\bullet \to (N'M')_\bullet$ such that the diagram



commutes. It is not hard to check that σ preserves the inner product on p-cochains, so that we have $NM \sim N'M'$.

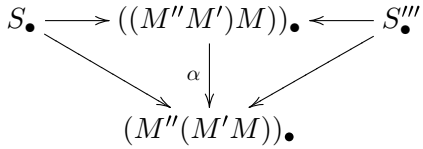
Now suppose we have three chain cobordisms

$$S_\bullet \xrightarrow{i} M_\bullet \xleftarrow{i'} S'_\bullet \qquad S_\bullet \xrightarrow{j} M'_\bullet \xleftarrow{j'} S''_\bullet \qquad S_\bullet \xrightarrow{k} M''_\bullet \xleftarrow{k'} S'''_\bullet$$

Showing that composition is associative involves constructing an isomorphism

$$((M''M')M))_\bullet \xrightarrow{\alpha} (M''(M'M))_\bullet$$

called the *associator* for (M'', M', M), such that the diagram



commutes, so that the two parenthesizations are equal as equivalence classes. In fact, α is just the obvious isomorphism. Checking the details is lengthy but not difficult. In fact, the equivalence relation in the definition of chain cobordism is motivated entirely by the need for associativity.

The identity morphism from S to S consists of two identity chain maps

$$S_\bullet \xrightarrow{1} S_\bullet \xleftarrow{1} S_\bullet$$

which in expanded notation is the diagram

$$\begin{array}{ccccccccccc}
0 & \longleftarrow & S_0 & \longleftarrow & S_1 & \longleftarrow & \cdots & \longleftarrow & S_{n-1} & & \\
 & & \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle 1} & & & & \downarrow{\scriptstyle 1} & & \\
0 & \longleftarrow & S_0 & \longleftarrow & S_1 & \longleftarrow & \cdots & \longleftarrow & S_{n-1} & \longleftarrow & 0 \\
 & & \uparrow{\scriptstyle 1} & & \uparrow{\scriptstyle 1} & & & & \uparrow{\scriptstyle 1} & & \\
0 & \longleftarrow & S_0 & \longleftarrow & S'_1 & \longleftarrow & \cdots & \longleftarrow & S_{n-1} & &
\end{array}$$

Here we explicitly write in the position $S_n = 0$ in the middle row just to emphasize that we are thinking of S both as an *object*, an $(n-1)$-complex, and as a *morphism*, an n-complex. We write this identity morphism as $1_S{:}S \to S$. To see that this does satisfy the left and right identity axioms, suppose $M{:}S \to S'$, $N{:}S'' \to S$. Then it is a routine exercise to check that the diagrams

$$\begin{array}{ccccc}
S_\bullet & \longrightarrow & (M1_S)_\bullet & \longleftarrow & S'_\bullet \\
 & \searrow & \downarrow & \swarrow & \\
 & & M_\bullet & &
\end{array}
\qquad \text{and} \qquad
\begin{array}{ccccc}
S''_\bullet & \longrightarrow & (1_S N)_\bullet & \longleftarrow & S_\bullet \\
 & \searrow & \downarrow & \swarrow & \\
 & & N_\bullet & &
\end{array}$$

commute, where the vertical arrows are the obvious isomorphisms induced by the inclusions of S into M and N.

This completes the proof that nChain is a category. A *monoidal* category is a category with a 'tensor product', so we must describe this product in nChain. In nCob the tensor product is just the disjoint union of spaces; in nChain the obvious analogue is the direct sum of chain complexes, at both object and morphism levels. Given objects S and C, define $S \oplus C$ to be the chain complex $S_\bullet \oplus C_\bullet$, with inner product on p-cochains given by

$$\langle\,,\rangle = \langle\,,\rangle_S + \langle\,,\rangle_C.$$

Similarly, for morphisms $M{:}S \to S'$ and $N{:}C \to C'$ define

$$M \oplus N{:} \quad S \oplus C \to S' \oplus C'$$

to be the class of basis-preserving and inner-product-preserving chain maps

$$S_\bullet \oplus C_\bullet \longrightarrow M_\bullet \oplus N_\bullet \longleftarrow S'_\bullet \oplus C'_\bullet.$$

The identity for $\oplus$ is the trivial chain complex $0_\bullet$:

$$0 \longleftarrow 0 \longleftarrow \cdots \longleftarrow 0$$

and the associator and unit laws are inherited from the obvious Abelian group isomorphisms $(G \oplus H) \oplus K \cong G \oplus (H \oplus K)$ and $G \oplus 0 \cong G \cong 0 \oplus G$.

Making nChain a *symmetric* monoidal category requires that we also specify a 'symmetry' or 'braiding' which lets us switch the order in the tensor product. This comes from the Abelian

group isomorphism $G \oplus H \cong H \oplus G$ in an obvious way. Checking that these constructions actually yield a symmetric monoidal category involves checking certain 'coherence laws' guaranteeing the associator, unit laws and braiding get along appropriately. This is a lengthy process, but not a difficult one. We refer the reader to [28] for the detailed definitions of monoidal and symmetric monoidal categories. □

Having established the category nChain as the analogue of nCob for chain complexes, we are now ready to formally state the definition of chain field theory.

Definition 2. *A chain field theory is a symmetric monoidal functor from* nChain *to* Hilb.

The main example is provided by the following.

Theorem 3. *Lattice p-form electromagnetism gives a chain field theory.*

Before proving this theorem, we remark that for cobordisms we must modify our definition of the *action* slightly. The idea of the modified action we shall use for cobordisms $M{:}S \to S'$ is that the action on p-cells in the 'boundary' $S + S'$ should only count *half* as much as the action on cells in the interior. This keeps us from 'overcounting' the action when we compose two cobordisms $S \xrightarrow{M} S' \xrightarrow{M'} S''$. If $M{:}S \to S'$ is a chain cobordism, let us rename our old naive action $\tilde{S}$, so that

$$\tilde{S}(A) = -\frac{1}{2e^2} h^{ij} F(x_i) F(x_j)$$

in the $\mathbb{R}$ case. If A is a p-connection on the spacetime cobordism $M{:}S \to S'$, we then define the action on M to be

$$S(A) = \tilde{S}(A) - \tfrac{1}{2}\tilde{S}(A|_S) - \tfrac{1}{2}\tilde{S}(A|_{S'}).$$

In the $U(1)$ case, we do not define the action, but only its 'exponential'—the theta function in equation (9) in section 7.3.

$$\mathrm{e}^{-\tilde{S}(A)} = \sum_{n \in \mathbb{Z}^N} \mathrm{e}^{-\frac{1}{2e^2} h^{ij} (A(\partial x_i) - 2n_i\pi)(A(\partial x_j) - 2n_j\pi)}.$$

This leads us to define for a p-connection A on $M{:}S \to S'$:

$$\mathrm{e}^{-S(A)} := \mathrm{e}^{-\tilde{S}(A)}\, \mathrm{e}^{\frac{1}{2}\tilde{S}(A|_S)}\, \mathrm{e}^{\frac{1}{2}\tilde{S}(A|_{S'})},$$

where $\mathrm{e}^{\frac{1}{2}\tilde{S}}$ just denotes the obvious thing—the reciprocal of the square root of the theta function $\mathrm{e}^{-\tilde{S}}$.

Proof of theorem 3. To define a functor

$$Z{:}n\mathrm{Chain} \to \mathrm{Hilb},$$

we must specify what Z does to objects and morphisms. When space is the $(n-1)$-complex S, the classical configuration space of p-form electromagnetism with gauge group $U(1)$ is the space of p-connections on S modulo gauge transformations:

$$\frac{\mathcal{A}(S)}{\mathcal{G}(S)} = \frac{C^p(S, U(1))}{B^p(S, U(1))}.$$

Quantization gives the Hilbert space of states, the space of all square-integrable functions on classical field configurations. We thus define

$$Z(S) := L^2\left(\frac{\mathcal{A}(S)}{\mathcal{G}(S)}\right).$$

Now Z must also assign to each morphism $M{:}S \to S'$ in nChain a linear operator $Z(M){:}Z(S) \to Z(S')$ between Hilbert spaces, corresponding to 'time evolution' of states on the slice S of space to the slice S'. To define $Z(M)$ it suffices to specify, for any $\psi \in Z(S) = L^2(\mathcal{A}(S)/\mathcal{G}(S))$ and $\phi \in Z(S') = L^2(\mathcal{A}(S')/\mathcal{G}(S'))$, the transition probability

$$\langle \phi, Z(M)\psi \rangle = \int_{\mathcal{A}(M)} \overline{\phi(A|_{S'})}\psi(A|_S)\, \mathrm{e}^{-S(A)}\mathcal{D}A, \tag{16}$$

where for convenience in this section we absorb the normalization factor into the definition of the measure $\mathcal{D}A$.

To show functoriality, we must show that this definition of Z respects identity morphisms and composition. First, suppose we apply Z to the identity cobordism $1_S{:}S \to S$, i.e.

$$S_\bullet \xrightarrow{1} S_\bullet \xleftarrow{1} S_\bullet$$

Then given two states $\psi, \phi \in Z(S) = L^2(\mathcal{A}(S)/\mathcal{G}(S))$, we have

$$\begin{aligned}\langle \phi, Z(1_S)\psi \rangle &= \int_{\mathcal{A}(S)} \overline{\phi(A|_S)}\psi(A|_S)\, \mathrm{e}^{-S(A)}\mathcal{D}A \\ &= \int_{\mathcal{A}(S)} \overline{\phi(A|_S)}\psi(A|_S)\mathcal{D}A = \langle \phi, \psi \rangle,\end{aligned}$$

where we have used the fact that for the identity cobordism 1_S, we have

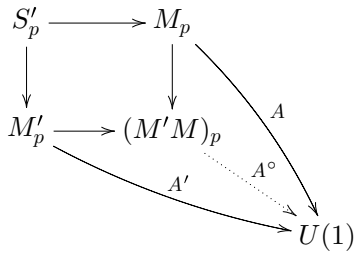
$$\mathrm{e}^{-S(A)} = \mathrm{e}^{-\tilde{S}(A)}\, \mathrm{e}^{\frac{1}{2}\tilde{S}(A)}\, \mathrm{e}^{\frac{1}{2}\tilde{S}(A)} = 1.$$

Hence, $Z(S)$ is the identity on $Z(S) = L^2(\mathcal{A}(S)/\mathcal{G}(S))$.

As usual, showing that Z respects composition is the nontrivial part of showing Z is a functor. Suppose we have two composable morphisms in nChain:

$$M{:}S \to S', \qquad M'{:}S' \to S''.$$

Since $M'M$ is a pushout, the diagram



shows that a p-connection A° on the composite $M'M$ is precisely a p-connection A on M together with a p-connection A' on M' such that A' and A agree when restricted to the common boundary S'. To enforce this agreement on S' in path integrals, we will use a delta function $\delta(A|_{S'} - A'|_{S'})$. If $\{\eta_\lambda : \lambda \in \Lambda\}$ is an orthonormal basis of $L^2(\mathcal{A}(S')/\mathcal{G}(S'))$, then expanding our delta function gives

$$\begin{aligned}\delta(A|_{S'} - A|_{S'}) &= \sum_{\lambda\in\Lambda} \langle \eta_\lambda(A'|_{S'}), \delta(A|_{S'} - A'|_{S'})\rangle \eta_\lambda(A'|_{S'}) \\ &= \sum_{\lambda\in\Lambda} \left(\int \overline{\eta_\lambda(A'|_{S'})}\delta(A|_{S'} - A'|_{S'})\mathcal{D}A' \right) \eta_\lambda(A'|_{S'}) \\ &= \sum_{\lambda\in\Lambda} \overline{\eta_\lambda(A|_{S'})}\eta_\lambda(A'|_{S'}).\end{aligned} \tag{17}$$

Now to show $Z(M'M) = Z(M')Z(M)$, it suffices to show

$$\langle \phi, Z(M'M)\psi\rangle = \langle \phi, Z(M')Z(M)\psi\rangle$$

for all $\psi \in Z(S)$ and $\phi \in Z(S'')$. We have

$$\begin{aligned}
\langle \phi, Z(M')Z(M)\psi\rangle &= \int_{\mathcal{A}(M')} \overline{\phi(A'|_{S''})}(Z(M)\psi)(A'|_{S'})\, \mathrm{e}^{-S(A')}\mathcal{D}A' \\
&= \int_{\mathcal{A}(M')} \overline{\phi(A'|_{S''})} \sum_{\lambda\in\Lambda} \langle \eta_\lambda, Z(M)\psi\rangle \eta_\lambda(A'|_{S'})\, \mathrm{e}^{-S(A')}\mathcal{D}A' \\
&= \int_{\mathcal{A}(M')}\int_{\mathcal{A}(M)} \overline{\phi(A'|_{S''})}\psi(A|_S) \sum_{\lambda\in\Lambda} \overline{\eta_\lambda(A|_{S'})}\eta_\lambda(A'|_{S'})\, \mathrm{e}^{-S(A)}\, \mathrm{e}^{-S(A')}\mathcal{D}A\mathcal{D}A' \\
&= \int_{\mathcal{A}(M'M)} \overline{\phi(A^\circ|_{S''})}\psi(A^\circ|_S)\, \mathrm{e}^{-S(A^\circ)}\mathcal{D}A^\circ \\
&= \langle \phi, Z(M'M)\psi\rangle,
\end{aligned}$$

as desired. Here we have used the expansion (17) and the fact that

$$\mathrm{e}^{-S(A)}\, \mathrm{e}^{-S(A')} = \mathrm{e}^{-S(A^\circ)}.$$

This comes from the definition of the action on a cobordism in the case where the gauge group is $\mathbb{R}$. Using this and definition (15) of the inner product for a composite of cobordisms, we obtain

$$\begin{aligned}
S(A) + S(A') &= \tilde{S}(A) + \tilde{S}(A') - \tfrac{1}{2}(\tilde{S}(A|_{S'}) + \tilde{S}(A'|_{S'})) - \tfrac{1}{2}\tilde{S}(A|_S) - \tfrac{1}{2}\tilde{S}(A'|_{S''}) \\
&= \tilde{S}(A^\circ|_M) + \tilde{S}(A^\circ|_{M'}) - \tilde{S}(A^\circ|_{S'}) - \tfrac{1}{2}\tilde{S}(A^\circ|_S) - \tfrac{1}{2}\tilde{S}(A^\circ|_{S''}) \\
&= \tilde{S}(A^\circ) - \tfrac{1}{2}\tilde{S}(A^\circ|_S) - \tfrac{1}{2}\tilde{S}(A^\circ|_{S''}) \\
&= S(A^\circ).
\end{aligned}$$

The result follows in the case where the gauge group is $U(1)$ from the relationship between Gaussians and theta functions. This completes the proof that Z is a functor.

The essential features of a *symmetric monoidal* functor are a natural isomorphism

$$\Phi: Z(-) \otimes Z(-) \xRightarrow{\sim} Z(- \oplus -)$$

between the functors

$$\begin{aligned}
&Z(-) \otimes Z(-): n\mathrm{Chain} \times n\mathrm{Chain} \to \mathrm{Hilb} \qquad \text{and} \\
&Z(- \oplus -): n\mathrm{Chain} \times n\mathrm{Chain} \to \mathrm{Hilb}
\end{aligned}$$

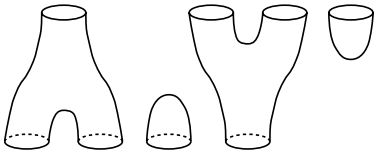
and an isomorphism

$$\phi: 1_{\mathrm{Hilb}} \xrightarrow{\sim} Z(0_{n\mathrm{Chain}})$$

between the identity in Hilb and Z applied to the identity in nChain. The existence of ϕ is obvious: $Z(0)$ is the space of L^2 functions on a one-point set, i.e. complex numbers, and $\mathbb{C}$ is the identity for the tensor product of Hilbert spaces. To complete the proof that the functor is symmetric monoidal, one must work out what Φ is and check that Φ and ϕ satisfy the required coherence laws. We leave these details to the reader, whom we again refer to Mac Lane's textbook [28] for the definitions. □

Let us turn briefly to examples. Naturally, the simplest examples are of the sort we have considered throughout the paper: p-form electromagnetism in $p + 1$ dimensions. To describe 2D electromagnetism, in particular, we can rely on our knowledge of 2D topological field theory. TQFTs in two dimensions are particularly simple to describe completely, since in the

category 2Cob every object is a finite disjoint union of circles and the morphisms can be given a finite description in terms of generators and relations. Specifically, morphisms in 2Cob are generated by a 3-punctured sphere (pair of trousers) and a disc, thought of, respectively, as cobordisms from one circle to two circles and from zero circles to one circle, together with inverted versions of both of these:

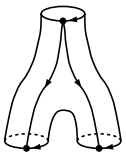


Likewise, in 2D lattice electromagnetism we can in some sense get away with a small list of chain cobordisms. We note, however, that as 2D electromagnetism is not merely topological, we must allow each chain cobordism to have arbitrarily chosen total area. By discretization independence, we may as well choose the simplest possible discretizations for both space and spacetime. Thus, we consider space to be a union of m circles, each consisting of a single vertex and edge. This corresponds to the chain complex:

$$\mathbb{Z}^m \xleftarrow{0} \mathbb{Z}^m.$$

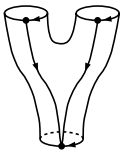
We build chain cobordisms between these by composing and tensoring those in the following list. In the list, we show both the chain cobordism M and the chain field theory time evolution operator $Z(M)$, computed using (16).

- *Trousers:*

$\Delta =$ 

$$\begin{array}{ccccc} \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & & \\ \downarrow & & \downarrow & & \\ \mathbb{Z}^3 & \longleftarrow & \mathbb{Z}^5 & \longleftarrow & \mathbb{Z} \\ \uparrow & & \uparrow & & \\ \mathbb{Z}^2 & \xleftarrow{0} & \mathbb{Z}^2 & & \end{array} \qquad \begin{array}{ccc} L^2(U(1)) & & e^{ikA} \\ {\scriptstyle Z(\Delta)}\downarrow & & \downarrow \\ L^2(U(1))^{\otimes 2} & & e^{-k^2e^2V/2}e^{ikA_1}e^{ikA_2} \end{array}$$

- *Upside-down trousers:*

$m =$ 

$$\begin{array}{ccccc} \mathbb{Z}^2 & \xleftarrow{0} & \mathbb{Z}^2 & & \\ \downarrow & & \downarrow & & \\ \mathbb{Z}^3 & \longleftarrow & \mathbb{Z}^5 & \longleftarrow & \mathbb{Z} \\ \uparrow & & \uparrow & & \\ \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & & \end{array} \qquad \begin{array}{ccc} L^2(U(1))^{\otimes 2} & & e^{ikA_1}e^{ikA_2} \\ {\scriptstyle Z(m)}\downarrow & & \downarrow \\ L^2(U(1)) & & \delta_{k_1k}\delta_{k_2k}e^{-k^2e^2V/2}e^{ikA} \end{array}$$

- *The cap:*

$\iota =$ 

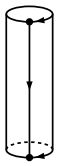
$$\begin{array}{ccccc} 0 & \xleftarrow{} & 0 & & \\ \downarrow & & \downarrow & & \\ \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & \xleftarrow{1} & \mathbb{Z} \\ \uparrow{\scriptstyle 1} & & \uparrow{\scriptstyle 1} & & \\ \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & & \end{array} \qquad \begin{array}{ccc} \mathbb{C} & & 1 \\ {\scriptstyle Z(\iota)}\downarrow & & \downarrow \\ L^2(U(1)) & & \displaystyle\sum_k e^{-k^2e^2V/2}e^{ikA} \end{array}$$

- *The cup:*

$\varepsilon =$ 

$$\begin{array}{ccccc} \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & & \\ \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle 1} & & \\ \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & \xleftarrow{1} & \mathbb{Z} \\ \uparrow & & \uparrow & & \\ 0 & \xleftarrow{} & 0 & & \end{array} \qquad \begin{array}{ccc} L^2(U(1)) & & e^{ikA} \\ {\scriptstyle Z(\varepsilon)}\downarrow & & \downarrow \\ \mathbb{C} & & e^{-k^2e^2V/2} \end{array}$$

- *The cylinder:* if we were describing 2D topological field theory, this list would have ended with the previous item. Since 2D electromagnetism is 'volumetric', one must describe how states evolve not only at topology changes, but also as 'time' passes—actually, as spacetime volume is swept out—when no topology change takes place. This is described by the cylinder:

$C =$ 

$$\begin{array}{ccccc} \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & & \\ \downarrow & & \downarrow & & \\ \mathbb{Z}^2 & \xleftarrow{0} & \mathbb{Z}^3 & \xleftarrow{1} & \mathbb{Z} \\ \uparrow & & \uparrow & & \\ \mathbb{Z} & \xleftarrow{0} & \mathbb{Z} & & \end{array} \qquad \begin{array}{c} L^2(U(1)) \\ \Big\downarrow{\scriptstyle Z(C)} \\ L^2(U(1)) \end{array} \qquad \begin{array}{c} e^{ikA} \\ \vdots \\ e^{-k^2e^2V/2}e^{ikA} \end{array}$$

This example verifies the volumetric character of 2D electromagnetism. For example, a torus T^2 of volume V is the composite of a cap ι, trousers Δ, upside-down trousers m and a cup ε whose volumes $V_1, \ldots, V_4$ add up to V. This gives an evolution operator

$$Z(\varepsilon \circ \Delta \circ m \circ \iota) : \mathbb{C} \to \mathbb{C}$$

which is just a complex number. It is easy to calculate that this number is precisely the partition function

$$Z(T^2) = \sum_{k \in \mathbb{Z}} \mathrm{e}^{-k^2e^2V/2}.$$

A further interesting discussion related to this example appears in notes by Baez [8]. Much of what Baez discusses is of course equally applicable to the p-form theory in $p + 1$ dimensions.

The higher-dimensional cases of p-form electromagnetism in $p + 1$ dimensions provide additional chain field theory examples which are simple to work out in as many particular cases as one wishes. A complete characterization is harder in these cases, of course, since there are more compact orientable p-manifolds—more choices of topology for 'space'—when $p > 1$.

We wish to emphasize, however, that while our examples have been drawn from the $(p + 1)$-dimensional theory, the machinery of chain field theory is more broadly applicable. One can calculate time evolution operators on any abstract chain cobordism, given any inner product on $(p + 1)$-cochains. Of course, one imagines the chain cobordisms as coming from cellular decompositions of smooth cobordisms, and in this case one would ideally pick this inner product such that the classical theory has the correct continuum limit. Though we have skirted this issue for $n > p + 1$ in the present paper in favour of developing the chain field theory idea, an interesting continuation of this project would be to study, for example, chain field theories coming from 3D lattice electromagnetism or 4D lattice Kalb–Ramond theory. Doing these will require a more detailed analysis of discrete differential geometry than what we have needed here, but should be tractable, at least in the case of a regular lattice. In particular, one could consider objects in nChain arising from cubical discretizations of flat tori.

Acknowledgments

I thank John Baez for numerous discussions on this subject. This paper would not have been written without all I learned from John before and during its writing. I also thank James Dolan, Eric Forgy, Xiao-Song Lin, Jeffrey Morton, Hendryk Pfeiffer and Urs Schreiber for useful discussions or correspondence. Finally, I thank the referee, whose comments led to significant improvements in the final product.

References

[1] Álvarez M C 2004 Loop quantization versus Fock quantization of *p*-form electromagnetism on static spacetimes *PhD Thesis* University of California, Riverside, CA (*Preprint* math-ph/0412032)

[2] Atiyah M F 1990 *Geometry and Physics of Knots* (Cambridge: Cambridge University Press)

[3] Atiyah M F and Bott R 1982 The Yang–Mills equations over Riemann surfaces *Phil. Trans. R. Soc.* A **308** 523

[4] Baez J C 1998 Spin foam models *Class. Quantum Grav.* **15** 1827–58 (*Preprint* gr-qc/9709052)

[5] Baez J C 2000 An introduction to spin foam models of quantum gravity and BF theory *Geometry and Quantum Physics* ed H Gausterer and H Grosse (Berlin: Springer) (*Preprint* gr-qc/9905087)

[6] Baez J C and Muniain J P 1994 *Gauge Fields, Knots and Gravity (Series on Knots and Everything* vol 4*)* (Singapore: World Scientific)

[7] Baez J C and Schreiber U 2005 Higher gauge theory *Preprint* math.DG/0511710
Baez J C and Schreiber U 2004 Higher gauge theory: 2-connections on 2-bundles *Preprint* hep-th/0412325

[8] Baez J C *Quantum Gravity Seminar Notes* http://math.ucr.edu/home/baez/qg-winter2001 (weeks 14, 15 and 16 of Track 2)

[9] Bartels T 2004 Higher gauge theory: 2-bundles *Preprint* math.CT/0410328

[10] Bossavit A 1998 *Computational Electromagnetism* (New York: Academic)

[11] Brylinski J-L 1993 *Loop Spaces, Characteristic Classes and Geometric Quantization* (Basle: Birkhauser)

[12] Choquet-Bruhat Y, DeWitt-Morette C and Dillard-Bleick M 1982 *Analysis, Manifolds, and Physics* revised edn (Amsterdam: North-Holland)

[13] Ciufolini I and Wheeler J A 1995 *Gravitation and Inertia (Princeton Series in Physics)* (Princeton, NJ: Princeton University Press)

[14] Dimakis A and Müller-Hoissen F 1999 Discrete Riemannian geometry *J. Math. Phys.* **40** 1518–48 (*Preprint* gr-qc/9808023)

[15] Dodziuk J 1976 Finite difference approach to the Hodge theory of harmonic forms *Am. J. Math.* **98** 79–104

[16] Forgy E and Schreiber U 2004 Discrete differential geometry on causal graphs *Preprint* math-ph/0407005

[17] Forman R 1999 Combinatorial differential topology and geometry *New Perspectives in Algebraic Combinatorics* vol 38 ed L J Billera, A Björner, C Green, R Simion and R P Stanley (Cambridge: Cambridge University Press) http://www.msri.org/publications/books/Book38/contents.html

[18] Geroch R 1985 *Mathematical Physics* (Chicago, IL: University of Chicago Press)

[19] Grandis M and Mauri L 2003 Cubical sets and their site *Theory Appl. Categ.* **11** 185–211 http://www.tac.mta.ca/tac/volumes/11/8/11-08abs.html

[20] Green M B, Schwarz J H and Witten E 1987 *Superstring Theory* vol 1 (Cambridge: Cambridge University Press)

[21] Greenberg M J and Harper J R 1981 *Algebraic Topology: A First Course* revised edn (New York: Perseus)

[22] Hatcher A 2002 *Algebraic Topology* (Cambridge: Cambridge University Press) http://www.math.cornell.edu/~hatcher/AT/ATpage.html

[23] Hog-Angeloni C, Metzler W and Sieradski A 1993 *Two-Dimensional Homotopy and Combinatorial Group Theory (London Mathematical Society Lecture Note Series* vol 197*)* (Cambridge: Cambridge University Press)

[24] Hudson J 1969 *Piecewise Linear Topology* (Reading, MA: Benjamin)

[25] Kalb M and Ramond P 1974 Classical direct interstring action *Phys. Rev.* D **9** 2273–84

[26] Kogut J B and Stephanov M A 2004 *The Phases of Quantum Chromodynamics: From Confinement to Extreme Environments* (Cambridge: Cambridge University Press)

[27] Mackaay M and Picken R 2002 Holonomy and parallel transport for Abelian gerbes *Adv. Math.* **170** 287–339 (*Preprint* math.DG/0007053)

[28] Mac Lane S 1998 *Categories for the Working Mathematician* 2nd edn (Berlin: Springer)

[29] May J P 1992 *Simplicial Objects in Algebraic Topology* (Chicago, IL: University of Chicago Press)

[30] Migdal A 1975 *Zh. Eksp. Teor. Fiz.* **69** 810
Migdal A 1975 *Sov. Phys.—JETP* **42** 413 (Engl. Transl.)

[31] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (New York: Freeman)

[32] Montvay I and Münster G 1994 *Quantum Fields on a Lattice* (Cambridge: Cambridge University Press)

[33] Mumford D 1982 *Tata Lectures on Theta I (Progress in Mathematics* vol 28*)* (Basle: Birkhauser)

[34] Pfeiffer H 2003 Higher gauge theory and a non-Abelian generalization of *p*-form electromagnetism *Ann. Phys.* **308** 447–77 (*Preprint* hep-th/0304074)
Girelli F and Pfeiffer H 2004 Higher gauge theory—differential versus integral formulation *J. Math. Phys.* **45** 3949–71 (*Preprint* hep-th/0309173)

[35] Rourke C and Sanderson B 1972 *Introduction to Piecewise-Linear Topology* (Berlin: Springer)

[36] Tonti E 1998 Algebraic topology and computational electromagnetism *4th Int. Workshop on Electric and Magnetic Fields (Marseille, France)* pp 284–94
Tonti E 2001 Finite formulation of the electromagnetic field *Prog. Electromagn. Res.* **32** 1–44
[37] Wilson S 2005 On the algebra and geometry of a manifold's chains and cochains *PhD Thesis* Stony Brook University http://www.math.umn.edu/scottw/main.pdf
[38] Wise D 2005 Lattice p-form electromagnetism and chain field theory *Preprint* gr-qc/0510033
[39] Witten E 1991 On quantum gauge theories in two dimensions *Commun. Math. Phys.* **141** 153–209
Witten E 1992 Two-dimensional gauge theories revisited *J. Geom. Phys.* **9** 303–68
[40] Witten E 1988 2+1 gravity as an exactly soluble system *Nucl. Phys.* B **311** 46–78
[41] Zee A 2003 *Quantum Field Theory in a Nutshell* (Princeton, NJ: Princeton University Press)