IOP Publishing Classical and Quantum Gravity

Class. Quantum Grav. 31 (2014) 055008 (20pp) doi:10.1088/0264-9381/31/5/055008

Quantizing models of (2+1)-dimensional gravity on non-orientable manifolds

Si Chen, Donald M Witt and Steven S Plotkin

Department of Physics and Astronomy, University of British Columbia, Agricultural Road, Vancouver, V6T 1Z1, Canada

E-mail: si.chen@alumni.ubc.ca and steve@phas.ubc.ca

Received 18 October 2013, revised 15 January 2014
Accepted for publication 16 January 2014
Published 14 February 2014

Abstract

Motivated by the relation between the Chern–Simons gauge theory and (2+1)-dimensional gravity, we find a formulation of gauge theories which applies to both orientable and non-orientable manifolds, using orientation bundles and density-valued forms. We show that on a non-orientable manifold, (2+1)D gravity is equivalent to BF theory, which is still topological and can be mapped in turn to Chern–Simons theory on the orientable double cover. By quantizing U(1) BF theory on a non-orientable manifold, we find that non-orientability introduces additional constraints on the quantized BF theory beyond those present for an orientable manifold, such that the coupling constant can only adopt a small number of discrete values. Specifically, for both the Klein bottle of demigenus 2 (N_2) and the compact surface of non-orientable genus 3 (Dyck's surface or N_3), we find explicit representations for the holonomy, large gauge, and mapping class groups, as well as the Hilbert space; here the above symmetries along with the non-orientability of the surface constrain the coupling constant k to only take values $1/2$, 1, or 2.

Keywords: non-orientable manifolds, canonical quantization, mapping class group, quantum gravity, finite-dimensional Hilbert space
PACS numbers: 04.20.Cv, 04.60.Ds, 04.60.Kz

(Some figures may appear in colour only in the online journal)

1. Introduction

Models of gravity in (2+1)-dimensions are a prototype for (3+1)-dimensional gravity, wherein many of the conceptual foundations regarding the nature of time, observables, topology, and quantization may be addressed [1]. The Einstein–Hilbert action of (2+1)D gravity takes the same form as its higher dimensional counterparts,

$$I_{\mathrm{EH}}[g] = \frac{1}{16\pi G}\int_M \mathrm{d}^3x\sqrt{-\det g}(R-2\Lambda), \tag{1}$$

where M is the spacetime manifold, g is the metric tensor, G is the gravitational constant, R is the Ricci scalar, and Λ is the cosmological constant. It is well-known that the action of (2+1)D gravity can be reformulated as a gauge theory action [1–3], which is typically taken as the sum of two Chern–Simons actions, each of the form

$$I_{\mathrm{CS}}[A] = \frac{k_{\mathrm{CS}}}{4\pi}\int_M \mathrm{Tr}\left\{A\wedge \mathrm{d}A + \frac{2}{3}A\wedge A\wedge A\right\}, \tag{2}$$

where M is an orientable, three-dimensional, spacetime manifold with no spatial boundary, A is a 1-form field taking values in a Lie algebra $\mathfrak{g}$ [4], and k_{CS} is a coupling constant[1]. The relation between (2+1)D gravity and Chern–Simons theory has been exploited to calculate the entropy of Bañados–Teitelboim–Zanelli (BTZ) black holes by counting microscopic states [5].

Interest in (2+1)D Chern–Simons theory has also been driven by applications to quantum computing [6]. Quasiparticle excitations in low-dimensional systems, such as fractional quantum Hall states in 2D electron systems, are not constrained by Bose or Fermi statistics but instead obey anyon statistics corresponding to symmetry under the non-Abelian operations of the braid group [7]. Such excitations constitute a topological phase, in that at low energies, observable properties such as correlation functions are invariant under small deformations or diffeomorphisms of the spacetime manifold in which the system lives. The energy gap that accompanies a topological phase generally results in protection against decoherence [8] due to local perturbations such as electron-phonon or hyperfine interactions. Chern–Simons theory is the archetypal topological quantum field theory describing non-Abelian anyons.

The quantized Chern–Simons fields appearing in the gauge theory corresponding to (2+1)D gravity are defined on a given manifold of fixed topology. On the other hand, the calculation of observables in (2+1)D gravity involves a trace over all possible manifolds, including manifolds having different topologies, and both orientable and non-orientable manifolds. A quantizable gauge theory equivalent to quantum gravity that accounts for topology-changing processes has not been fully explored; here we take a step in this direction by exploring Chern–Simons theory on a non-orientable manifold.

Although Chern–Simons theory may provide a means of quantizing (2+1)D gravity, subtleties exist in relating the two approaches. One inequivalence is that some gauge theory configurations are mapped to degenerate metrics [9–11], which is putatively forbidden in gravity theories. As a result, the gauge theory phase space splits into several sectors, and the boundary between these sectors are composed of the states with degenerate metrics. If one insists to include only states with non-degenerate metric, the resulting 3D gravity phase space is only one of these sectors. Quantizations of these two different phase spaces are generally different [12].

Here we focus on resolving another apparent inequivalence between Chern–Simons theory and quantum gravity. The Chern–Simons action (2) is an integral of a 3-form, and thus is defined only when the manifold is orientable. On the other hand, the path integral calculation involved in (2+1)D gravity includes both orientable and non-orientable spacetime manifolds. But if the spacetime manifold is non-orientable, the conventional definition of an integral over a differential form fails. Fortunately, there is a generalization of such an integral [13, 14]: on an orientable or non-orientable n-manifold, n-form *densities* can be integrated. An n-form density is different from an n-form only in that, under a coordinate transformation with a negative

[1] More generally, A may be taken to be a connection of a G-bundle on M, where G is a Lie group corresponding to the Lie algebra $\mathfrak{g}$. Here we consider only single-valued (trivial) bundles.

Jacobian determinant, an n-form density obtains an extra minus sign. The mathematical definition of densities will be constructed in detail below. In terms of such integral, we find below that the action in (2+1)D gravity can generally be written in terms of a gauge theory known as *BF theory* [15]. We show that BF theory is closely related with Chern–Simons theory defined on the *orientable double cover* with additional parity conditions on the solution.

In reference [16], Louko has made progress in this direction by formulating the (2+1)D gravity action with vanishing cosmological constant as an integral of a 3-form density, and developing the classical solution space for the spacetime topology $\mathbb{R} \times$ (Klein bottle) in detail. The formalism we develop below generalizes these results to arbitrary cosmological constant, and arbitrary topology of the spacetime. We do not explore the space of classical solutions however, and leave this as a topic for future work.

The outline of this paper is as follows. In section 2, the general definition and properties of p-form densities are reviewed; in section 3, the relations between (2+1)D gravity, BF theory, and Chern–Simons theory are established using p-form densities, for the case of a non-orientable spacetime manifold; in section 4, the gauge theory formulation on a non-orientable manifold is applied by quantizing the BF theory with U(1) gauge group on the compact space manifolds of non-orientable genera 2 and 3, i.e. on the manifolds $\mathbb{R} \times N_g$ with $g = 2, 3$.

2. *p*-form densities and integration

In this section we review the mathematical formulation involved in the integration of densities, which can be used to deal with fields on non-orientable manifolds. A diffeomorphism of open subsets of $\mathbb{R}^n$ is *orientation-preserving* if the Jacobian determinant of the diffeomorphism is everywhere positive. Let M be covered by the atlas $\{U_\alpha, \phi_\alpha\}_{\alpha\in A}$, where U_α's are open sets that cover M, and for each α, the chart $\phi_\alpha : U_\alpha \to \mathbb{R}^n$ is a homeomorphism. The atlas is *oriented* if all the transition functions $t_{\alpha\beta} = \phi_\alpha \circ \phi_\beta^{-1}$ are orientation-preserving, and M is *orientable* if it has an oriented atlas. The set of all charts having orientation-preserving transition functions with each other is called an *orientation*. There are two orientations on an orientable manifold; either one is denoted by $[M]$. It can be shown that an n-dimensional manifold M is orientable if and only if there exist a nowhere vanishing n-form on M.

On an orientable n-dimensional manifold M, after choosing an orientation $[M]$, the (conventional) integration of a n-form is defined as follows. Given an oriented atlas $\{U_\alpha, \phi_\alpha\}_{\alpha\in A}$ within $[M]$, the integration of n-form τ is

$$\int_{[M]} \tau = \sum_{\alpha\in A} \int_{\mathbb{R}^n} \left(\phi_\alpha^{-1}\right)^* (\rho_\alpha \tau). \tag{3}$$

Here $\{\rho_\alpha\}_{\alpha\in A}$ is a partition of unity [4] of the atlas A, $\left(\phi_\alpha^{-1}\right)^*$ is the pullback of ϕ_α^{-1}, and the integrals on the right-hand side are Riemann integrals [17], or in principle Lebesgue integrals. Usually a fixed orientation $[M]$ is understood, and the integration is simply written as $\int_M \tau$. This definition of the integral $\int_M \tau$ has the property that it is independent of the atlas and the partition of unity.

With the above definitions regarding conventional integrals of forms in mind, we define several concepts to deal with integration on non-orientable manifolds. All of the following concepts and properties apply to both orientable and non-orientable cases: for the orientable cases they reduce to rather trivial counterparts of the conventional concepts.

As mentioned above, an n-form indeed cannot be integrated on a non-orientable n-dimensional manifold; the value of the integral would not be invariant of the partition of unity. Another set of objects however, the n-form densities, can be integrated. Let the n-dimensional manifold M be covered by the atlas $\{U_\alpha, \phi_\alpha\}_{\alpha\in A}$. The *orientation bundle*, denoted by (O, M, π)

[13, 14], is a $\mathbb{Z}_2$-fiber bundle over M, specified by transition functions $T_{\alpha\beta} = \mathrm{sgn}[\det(J_{\alpha\beta})]$, where $J_{\alpha\beta}$ is the Jacobian of the transition function $t_{\alpha\beta}$ between two charts. Let $\Omega^p(M)$ denote the bundle of smooth p-forms on M. A p-form density is a smooth section of the bundle $\Omega^p(M) \otimes_{\mathrm{p}} O$, where the notation $\otimes_{\mathrm{p}}$ means the tensor product is taken for the fibers at each point.

The orientation bundle merely stores information about relative orientation between local charts—if two overlapping charts have the same orientation, their transition function is 1, otherwise it is -1. A given term in the p-form density is expressed as $(a_{i_1\cdots i_p}(x^{i_1}, \cdots, x^{i_p})\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p}, z)$, where z is in $\mathbb{Z}_2$. The combination of coefficients $z \cdot a_{i_1\cdots i_p}$ of a p-form density transforms between two coordinate charts (unprimed to primed) as [17]

$$z' \cdot a'_{j_1\cdots j_p} = \mathrm{sgn}(\det J) \sum_{i_1,\ldots,i_p} \frac{\partial x^{i_1}}{\partial x'^{j_1}} \cdots \frac{\partial x^{i_p}}{\partial x'^{j_p}} z \cdot a_{i_1\cdots i_p}. \tag{4}$$

That is, they transform in the same way as coefficients of regular p-forms, except that there is an extra minus sign when the coordinate orientation is reversed. Examples in physics include the magnetic field and angular momentum, and are generally called an axial scalar, vector, or tensor. p-form densities can thus be thought of as an axial generalization of p-forms.

The total space of the orientation bundle of M is called the *orientable double cover* of M, and is denoted as $\tilde{M}$. Explicitly, assuming that M is covered by the atlas $\{U_\alpha, \phi_\alpha\}_{\alpha\in A}$, $\tilde{M}$ is the set $(\bigcup_{\alpha\in A} U_\alpha \times \{\pm 1\})/\sim$, where $(x, z) \sim (x', z')$ if and only if $x = x' \in U_\alpha \cap U_\beta$ and $z = T_{\alpha\beta}(x)z'$. In this way $\tilde{M}$ is a two-fold cover of M, with the projection map $\pi : \tilde{M} \to M$ given by $\pi(x, z) = x$. $\tilde{M}$ as a manifold is described by the atlas $\{\tilde{U}_{\alpha,z}, \tilde{\phi}_{\alpha,z}\}_{\alpha\in A, z\in\mathbb{Z}_2}$, where the new atlas is labeled by two indices α and z, and $\tilde{\phi}_{\alpha,z} = (\phi_\alpha, z)$. If $U_\alpha \cap U_\beta \neq \emptyset$, then $\tilde{U}_{\alpha,z} \cap \tilde{U}_{\beta,t_{\alpha\beta}z} \neq \emptyset$, and the transition function between $\tilde{U}_{\alpha,z}$ and $\tilde{U}_{\beta,t_{\alpha\beta}z}$ is $\phi_\beta\phi_\alpha^{-1}$.

To show that the orientable double cover $\tilde{M}$ is orientable, one can construct a refinement $\{V_\beta\}_{\beta\in B}$ of $\{\tilde{U}_{\alpha,z}\}$, that is, for any $\beta \in B$, there exists α and z such that $V_\beta \subseteq \tilde{U}_{\alpha,z}$, and $\{V_\beta\}_{\beta\in B}$ still covers $\tilde{M}$. Define the functions $\chi_\beta(\tilde{x})$ such that $\chi_\beta(\tilde{x}) = 1$ for $\tilde{x} \in V_\beta$, $\chi_\beta(\tilde{x}) = 0$ for $\tilde{x} \notin \tilde{U}_{\alpha(\beta),z(\beta)}$, and $\chi_\beta(\tilde{x}) \geqslant 0$ everywhere. Consider the n-form

$$\sum_{\beta\in B} z\chi_\beta(\tilde{x})\,\mathrm{d}\tilde{x}^1_\beta \wedge \cdots \wedge \mathrm{d}\tilde{x}^n_\beta,$$

where for each β, z is the value in $\mathbb{Z}_2$ such that $V_\beta \subseteq \tilde{U}_{\alpha,z}$. It can be seen that for each point in $\tilde{M}$, z takes the same value. Also, for each point in $\tilde{M}$, at least one term in the summation over β is nonzero, since $\{V_\beta\}_{\beta\in B}$ covers $\tilde{M}$. Thus we have shown that this n-form is nowhere vanishing, and so $\tilde{M}$ is orientable.

Example 1. As a first example, let M be orientable. By the nowhere vanishing n-form on M, all coordinate charts fall into two families: charts on which the coefficient of the n-form is positive/negative. The orientation bundle is a trivial $\mathbb{Z}_2$-bundle—it is composed of two disconnected copies of M, i.e., $O = \tilde{M} = M \times \{\pm 1\}$. ■

Example 2. A simple non-orientable example is the Möbius strip. It can be covered by three coordinate charts (see figure 1) $(x, y) \in \phi_1(U_1) = (0, 1/2) \times (0, 1)$, $(x', y') \in \phi_2(U_2) = (1/3, 5/6) \times (0, 1)$, $(x'', y'') \in \phi_3(U_3) = (2/3, 7/6) \times (0, 1)$, and the transition functions $t_{\alpha\beta}$ are

$$\begin{aligned}
(x', y') &= t_{2,1}(x, y) = (x, y)\\
(x'', y'') &= t_{3,2}(x', y') = (x', y')\\
(x, y) &= t_{1,3}(x'', y'') = (x'' - 1, 1 - y'').
\end{aligned} \tag{5}$$

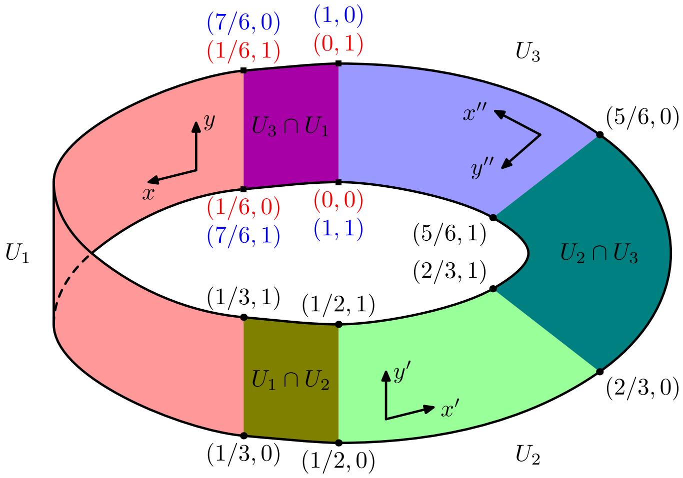

Figure 1. The Möbius strip can be covered by three charts—see example 2 in the text. The three charts in the example are denoted by red (U_1), green (U_2), and blue (U_3) regions. The regions overlapping two charts are colored olive ($U_1 \cap U_2$), teal ($U_2 \cap U_3$), and purple ($U_3 \cap U_1$). Coordinates delineating the charts are indicated- note the difference in coordinates and coordinate axes between U_3 and U_1.

Consider a constant 0-form α and a constant 0-form density β. Assume that on U_1, $\alpha = a$, where a is a constant real number. By the coordinate transformation law, it is easy to see that $\alpha = a$ on all charts. Assume that on U_1, $\beta = b$, where b is a constant real number. By the coordinate transformation law of densities, by $t_{2,1}$, $\beta = b$ on U_2; by $t_{3,2}$, $\beta = b$ on U_3; by $t_{1,3}$, $\beta = -b$ on U_1. Thus this constant 0-form density is inconsistent unless $b = 0$.

Now consider a general 2-form. It can be expressed on the above three charts as

$$\begin{aligned} U_1 &: \quad f(x, y)\, \mathrm{d}x \wedge \mathrm{d}y \\ U_2 &: \quad f'(x', y')\, \mathrm{d}x' \wedge \mathrm{d}y' \\ U_3 &: \quad f''(x'', y'')\, \mathrm{d}x'' \wedge \mathrm{d}y''. \end{aligned}$$

The transition functions (5) dictates that the coefficients satisfy

$$\begin{aligned} &f(x, y) = f'(x, y), \quad \forall (x, y) \in (1/3, 1/2) \times (0, 1), \\ &f'(x', y') = f''(x', y'), \quad \forall (x', y') \in (2/3, 5/6) \times (0, 1), \\ &f''(x'', y'') = -f(x'' - 1, 1 - y''), \quad \forall (x'', y'') \in (1, 7/6) \times (0, 1). \end{aligned}$$

As a consequence, one can see that at least one of f, f' and f'' must have a zero point, which is consistent with the general statement that on an n-dimensional non-orientable manifold, an n-form must have a zero point.

The orientable double cover of the Möbius strip by construction is covered by six coordinate charts, $\phi_{i,\pm 1}$, and the transition functions are

$$\begin{aligned} (x'_+, y'_+) &= t_{2,+1,1,+1}(x_+, y_+) = (x_+, y_+) \\ (x''_+, y''_+) &= t_{3,+1,2,+1}(x'_+, y'_+) = (x'_+, y'_+) \\ (x_-, y_-) &= t_{1,-1,3,+1}(x''_+, y''_+) = (x''_+ - 1, 1 - y''_+) \\ (x'_-, y'_-) &= t_{2,-1,1,-1}(x_-, y_-) = (x_-, y_-) \\ (x''_-, y''_-) &= t_{3,-1,2,-1}(x'_-, y'_-) = (x'_-, y'_-) \\ (x_+, y_+) &= t_{1,+1,3,-1}(x''_-, y''_-) = (x''_- - 1, 1 - y''_-). \end{aligned}$$

These charts and transitions simply describe a cylinder, where the above transition functions connect pairs of coordinate charts, with the last chart connected back to the first, and two of the transition functions flip the orientation. It can be shown that if M is non-orientable, its orientable double cover $\tilde{M}$ is always connected, and if M is orientable, $\tilde{M}$ is always disconnected. ■

With respect to the involution $\sigma : \tilde{M} \to \tilde{M}$ given by $\sigma(x, z) = (x, -z)$, p-forms on $\tilde{M}$ split into *even forms* and *odd forms*

$$\Omega^p(\tilde{M}) = \Omega^p_+(\tilde{M}) \oplus \Omega^p_-(\tilde{M}) \tag{6}$$

according to $(\sigma^*\tilde{\omega}_+)(\tilde{x}) = \tilde{\omega}_+(\tilde{x})$, and $(\sigma^*\tilde{\omega}_-)(\tilde{x}) = -\tilde{\omega}_-(\tilde{x})$. The pullback $\pi^* : \Omega^p(M) \to \Omega^p_+(\tilde{M})$ is a bijection, so regular forms on M are equivalent with even forms on $\tilde{M}$. Given a p-form density ξ on M and $v_1, \ldots, v_p \in T_{(x,z)}(\tilde{M})$, we can define $\tilde{\xi} \in \Omega^p_-(\tilde{M})$ by

$$\xi(\pi_*(v_1), \ldots, \pi_*(v_p)) = \tilde{\xi}(v_1, \ldots, v_p) \otimes (x, z), \tag{7}$$

which gives an identification of p-form densities on M and odd forms on $\tilde{M}$. In plain words, on M there exists p-forms and p-form densities, and equivalently we can work on the orientable double cover $\tilde{M}$, on which we only consider p-forms with a definite parity. It is obvious that the wedge product of two odd forms is an even form, the wedge product of two even forms is an even form, and the wedge product of an odd form and an even form is an odd form. These relations hold for both p-forms and p-form densities correspondingly; see the diagram below.

$$\begin{array}{lcc} \text{on } M: & \text{regular forms} & \text{densities} \\ & \updownarrow & \updownarrow \\ \text{on } \tilde{M}: & \text{even forms} & \text{odd forms.} \end{array}$$

We can also define Lie algebra valued p-form densities. They are sections of the tensor product space $\mathfrak{g} \otimes (\Omega^p(M) \otimes_{\mathrm{p}} O)$.

The exterior derivative d commutes with σ^*, because in general, the exterior derivative commutes with pullbacks. The de Rham cohomology group splits accordingly,

$$H^p(\tilde{M}, \mathbb{R}) = H^p_+(\tilde{M}, \mathbb{R}) \oplus H^p_-(\tilde{M}, \mathbb{R}), \tag{8}$$

so the concept of harmonic p-forms generalizes to harmonic p-form densities on M.

The most important property of densities is that a p-form density can be consistently integrated on a p-dimensional manifold. From the above discussion, one can easily see that the following integration of p-form densities on a p-dimensional manifold M' is well defined,

$$\int_{M'} \xi = \frac{1}{2} \int_{\tilde{M}'} \tilde{\xi}, \tag{9}$$

where on the right side it is the regular integral of the corresponding odd p-form on the orientable double cover.

Consider a p-form density ξ supported on one coordinate chart. By equations (3) and (9), the integration of ξ is given in terms of the coordinates by

$$\int_{M'} \xi = \int z a_{i_1 \cdots i_p} \mathrm{d}x^{i_1} \cdots \mathrm{d}x^{i_p}.$$

Under a diffeomorphism of M', the integration transforms into

$$\int \bar{z}\bar{a}_{j_1 \cdots j_p} \mathrm{d}\bar{x}^{j_1} \cdots \mathrm{d}\bar{x}^{j_p} = \int \left[\mathrm{sgn}(\det J) \sum_{i_1, \ldots, i_p} \frac{\partial x^{i_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial x^{i_p}}{\partial \bar{x}^{j_p}} z a_{i_1 \cdots i_p} \right]$$
$$\times \left[\mathrm{sgn}(\det J) \sum_{i_1, \ldots, i_p} \frac{\partial \bar{x}^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial \bar{x}^{j_p}}{\partial x^{i_p}} \mathrm{d}x^{i_1} \cdots \mathrm{d}x^{i_p} \right].$$

So the transformations of the two parts cancel exactly, and the integration is invariant under arbitrary diffeomorphisms, as opposed to regular integrations, which are only invariant under orientation-preserving diffeomorphisms. This conclusion holds for integration of a general p-form density, because it can be decomposed into a sum of p-form densities supported on one coordinate chart, using a partition of unity for M'.

Example 3. 2-form density on the Möbius strip, with integral given in equation (9). Consider a general 2-form density ω defined on the Möbius strip, which is covered by the atlas in example 2. Assume that on the three charts, ω is expressed as

$$\begin{aligned} U_1 :&\quad (g(x, y)\mathrm{d}x \wedge \mathrm{d}y, 1)\\ U_2 :&\quad (g'(x', y')\mathrm{d}x' \wedge \mathrm{d}y', 1)\\ U_3 :&\quad (g''(x'', y'')\mathrm{d}x'' \wedge \mathrm{d}y'', 1).\end{aligned}$$

According to the transition functions (5), the coefficients must satisfy

$$\begin{aligned} &g(x, y) = g'(x, y), \quad \forall (x, y) \in (1/3, 1/2) \times (0, 1),\\ &g(x', y') = g''(x', y'), \quad \forall (x', y') \in (2/3, 5/6) \times (0, 1),\\ &g(x'', y'') = g'(x'' - 1, 1 - y''), \quad \forall (x, y) \in (1, 7/6) \times (0, 1).\end{aligned}$$

By the definition (9), ω can be integrated as

$$\int_M \omega = \int_0^1 \mathrm{d}y \left(\int_0^{1/3} g(x, y)\, \mathrm{d}x + \int_{1/3}^{2/3} g'(x, y)\, \mathrm{d}x + \int_{2/3}^{1} g''(x, y)\, \mathrm{d}x \right).$$

■

Example 4. The Klein bottle, which is a compact non-orientable surface, can be represented as a square with its sides identified as in figure 2(*a*). The schematic figures of its orientation bundle and the oriented double cover are shown in figures 2(*b*) and (*c*), respectively. One can see that the oriented double cover is a torus, and from this relation we can derive the harmonic forms on the Klein bottle. In section 4 we consider the quantization of $U(1)$ BF theory on a non-orientable surface, for which the space of harmonic 1-forms and 1-form densities must be identified.

On the torus, the space of harmonic 1-forms is two-dimensional, for which the basis can be taken as the two uniform 1-forms shown on the right side of figures 2(*d*) and (*e*). The 1-form on the right side of figure 2(*d*) is even with respect to the involution σ, so it corresponds to a harmonic 1-form on the Klein bottle, shown on the left side of figure 2(*d*). Similarly, the 1-form on the right side of figure 2(*e*) is odd with respect to σ, so it corresponds to a harmonic 1-form density on the Klein bottle, shown on the left side of figure 2(*e*). Conversely, because any harmonic form or harmonic form density on the Klein bottle corresponds to a harmonic form on the torus, the harmonic 1-form/1-form density that we have found on the Klein bottle is the complete basis for the space of harmonic 1-forms/1-form densities. ■

3. The relation between (2+1)D gravity, BF theory, and Chern–Simons theory on non-orientable manifolds

In the previous section the integral of n-form densities on non-orientable n-manifolds was defined. Using this construction we can define field theories on manifolds with any orientability by constructing an n-form density as the Lagrangian. In particular, this formalism can be used to express the (2+1)D gravity action as a gauge theory action, even if the spacetime manifold is non-orientable.

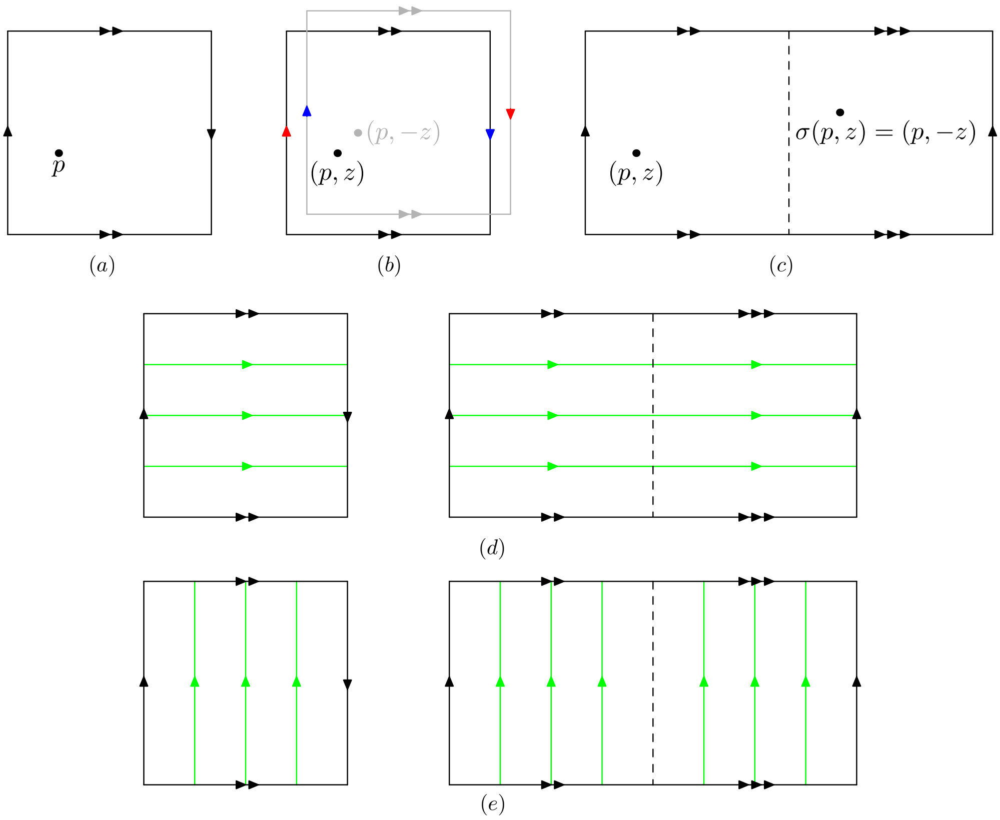

Figure 2. Schematic representations of (*a*) the Klein bottle; (*b*) the orientation bundle of the Klein bottle; (*c*) the orientable double cover of the Klein bottle; (*d*) the basis of harmonic 1-forms on the Klein bottle and corresponding even harmonic 1-forms on the orientable double cover; (*e*) the basis of harmonic 1-form densities on the Klein bottle and corresponding odd harmonic 1-forms on the orientable double cover.

The BF theory [15] is a gauge theory whose action is closely related to that in (2+1)D gravity. The BF action can be straightforwardly interpreted as the integral of a density, which in n dimensions is

$$I_{\mathrm{BF0}}[A,B]=\frac{k_{\mathrm{BF0}}}{2\pi}\int_M \mathrm{Tr}\{B\wedge F\}. \tag{10}$$

Here, B is a $\mathfrak{g}$-valued $(n-2)$-form density, $F=\mathrm{d}A+A\wedge A$, A is a Lie algebra $\mathfrak{g}$-valued 1-form field, and k_{BF0} is the coupling constant. In contrast to Chern–Simons theory, this theory is well-defined regardless of the orientability of M.

To map the BF theory to (2+1)D gravity with arbitrary cosmological constant (see (1)), a generalized BF theory is needed: To construct an n-form density, each term of the Lagrangian must include an odd number of densities. In $n=2+1$ dimensions, another term composed of A and B may be added to equation (10),

$$I_{\mathrm{BF}}[A,B]_M=\frac{k_{\mathrm{BF}}}{2\pi}\int_M \mathrm{Tr}\left\{B\wedge F-\frac{\lambda}{3}B\wedge B\wedge B\right\} \tag{11}$$

where k_{BF} and λ are parameters. When we relate this theory with 3D gravity below, λ will become the cosmological constant Λ. In the rest of this section, we will analyze this three-dimensional, generalized BF theory, and will refer to it simply as BF theory.

The BF theory is related with 3D gravity theory. From the 3D gravity action in equation (1), one can change the fundamental variable from the metric g to the local frame e and spin connection ω, according to the definitions

$$\eta_{ab} e_\mu^{\ a} e_\nu^{\ b} = g_{\mu\nu}, \tag{12}$$

$$\nabla_\mu e_\nu^{\ a} + \omega_\mu^{\ a}{}_b e_\nu^{\ b} = 0 \tag{13}$$

where $\eta = \text{diag}\{-1, 1, 1\}$. We also define the spin connection field with one local index

$$\omega^a = \tfrac{1}{2}\epsilon^{abc}\omega_{\mu bc}\,\mathrm{d}x^\mu \tag{14}$$

where ϵ^{abc} is totally antisymmetric with $\epsilon^{012} = 1$, and its indices are raised and lowered by η^{ab} and η_{ab}. Note that in this change of variable, according to (12), the local frame e is determined by g up to a local Lorentz rotation. This local Lorentz symmetry is an additional gauge symmetry of the resulting model. On the other hand, ω is fully determined by e classically from equation (13):

$$\omega_\mu^{\ a} = \epsilon^{abc} e^\nu_{\ c}(\partial_\mu e_{\nu b} - \partial_\nu e_{\mu b}) - \tfrac{1}{2}\epsilon^{bcd}\left(e^\nu_{\ b} e^\rho_{\ c}\partial_\rho e_{\nu d}\right) e_\mu^{\ a},$$

so there is no additional gauge symmetry associated with ω.

Now we rewrite the action (1) in terms of the new variables e and ω. The determinant of the metric is related to the local frame e by

$$(-\det g) = \left(\epsilon^{\mu\nu\lambda} e_\mu^{\ 0} e_\nu^{\ 1} e_\lambda^{\ 2}\right)^2,$$

where $\epsilon^{\mu\nu\lambda}$ is totally antisymmetric with $\epsilon^{txy} = 1$. When taking square root on both sides, there may or may not be an extra minus sign, so the volume element in the Einstein–Hilbert action (1) is

$$\sqrt{-\det g}\,\mathrm{d}^3x = \tfrac{1}{6}\epsilon_{abc} e^a \wedge e^b \wedge e^c \cdot \text{sgn}(\det e). \tag{15}$$

The left-hand side of (15) is a volume form that can always be integrated; the right side is a 3-form, and it is integrable on a non-orientable manifold only if e^a is a 1-form density. Thus we identify e^a as a 1-form density field with a local frame index. According to the definitions (13) and (14), ω^a is still a 1-form field with a local frame index. The Einstein–Hilbert action can then be cast into the so-called Palatini form [18]:

$$I'_{\mathrm{P}} = \frac{2}{16\pi G}\int_M \left\{\left[e^a \wedge \left(\mathrm{d}\omega_a + \frac{1}{2}\epsilon_{abc}\omega^b \wedge \omega^c\right) - \frac{\Lambda}{6}\epsilon_{abc} e^a \wedge e^b \wedge e^c\right] \cdot \text{sgn}(\det e)\right\}. \tag{16}$$

If the topology of M is fixed, this action takes the form of a well-defined gauge theory action, except for the potentially awkward term $\text{sgn}(\det e)$. For a model of gravity however, the metric is traditionally required to be non-degenerate everywhere [18], so for any solution of (16) corresponding a gravitational solution, the factor $\text{sgn}(\det e)$ does not change sign, and thus has no effect at all. However, as soon as we write the metric in terms of the local frame in this way, we allow the factor $\text{sgn}(\det e)$ to change sign within the spacetime, which is equivalent to allowing the metric to take degenerate values. In other words, the gauge theory defined by (16) is not exactly a reformulation of 3D gravity, but an extension whose classical solutions include the gravity solutions as a subset [9, 10]. For the same reason, one can choose to neglect the factor $\text{sgn}(\det e)$ in (16), and the result is another extension of 3D gravity. It has been conjectured that such a gauge theory extension may give a quantization of 3D gravity [3], which circumvents the difficulties in quantizing Einstein gravity directly. We follow this strategy here, and choose to neglect the factor $\text{sgn}(\det e)$ in equation (16), so the action becomes

$$I_{\mathrm{P}} = \frac{2}{16\pi G}\int_M \left\{e^a \wedge \left(\mathrm{d}\omega_a + \frac{1}{2}\epsilon_{abc}\omega^b \wedge \omega^c\right) - \frac{\Lambda}{6}\epsilon_{abc} e^a \wedge e^b \wedge e^c\right\}. \tag{17}$$

This Palatini gravity action is nothing but the BF action (11) with a cosmological constant term, if we interpret the local indices as labeling the components of a gauge field taking value in the Lie algebra $\mathfrak{sl}(2,\mathbb{R})$. That is, if we let

$$B = e^a T_a\,,\quad A = \omega^a T_a \tag{18a}$$

$$T_0 = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T_1 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad T_2 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{18b}$$

where T_a in (18b) are a set of $\mathfrak{sl}(2,\mathbb{R})$ generators that have the properties $[T_a, T_b] = \epsilon_{ab}{}^{c}T_c$, $\mathrm{Tr}(T_aT_b) = \frac{1}{2}\eta_{ab}$. Then

$$I_{\mathrm{P}} = I_{\mathrm{BF}} = \frac{1}{4\pi G}\int_M \mathrm{Tr}\left\{B\wedge(\mathrm{d}A + A\wedge A) - \frac{\Lambda}{3}B\wedge B\wedge B\right\}. \tag{19}$$

Equation (19) gives the relation between BF theory and 3D gravity, at the level of the action.

We may further relate Gravity to Chern–Simons theory by showing that the BF theory action is the sum of two Chern–Simons actions. In this relation, the Lie algebra of the BF theory is not restrained to $\mathfrak{sl}(2,\mathbb{R})$. However, this relation does depend on the sign of the parameter Λ as follows.

3.1. Case I: $\Lambda < 0$

If $\Lambda < 0$, let $\ell = 1/\sqrt{-\Lambda}$, and define $D^{\pm} = \tilde{A} \pm \ell^{-1}\tilde{B}$, where $\tilde{A}$ is the even form field on the orientable double cover $\tilde{M}$ corresponding to A, and $\tilde{B}$ is the odd form field on $\tilde{M}$ corresponding to B. Then

$$\begin{aligned}
I_{\mathrm{BF}}[A,B]_M &= \frac{k_{\mathrm{BF}}}{4\pi}\int_{\tilde{M}}\mathrm{Tr}\bigg\{\frac{\ell}{4}(D^+ - D^-)\wedge(\mathrm{d}D^+ + \mathrm{d}D^-)\\
&\quad + \frac{\ell}{8}(D^+ - D^-)\wedge(D^+ + D^-)\wedge(D^+ + D^-)\\
&\quad + \frac{\ell}{24}(D^+ - D^-)\wedge(D^+ - D^-)\wedge(D^+ - D^-)\bigg\}\\
&= \frac{\ell k_{\mathrm{BF}}}{16\pi}\int_{\tilde{M}}\mathrm{Tr}\left\{D^+\wedge\mathrm{d}D^+ + \frac{2}{3}D^+\wedge D^+\wedge D^+ - D^-\wedge\mathrm{d}D^- - \frac{2}{3}D^-\wedge D^-\wedge D^-\right\}\\
&= \frac{1}{2}(I_{\mathrm{CS}}[D^+]_{\tilde{M}} - I_{\mathrm{CS}}[D^-]_{\tilde{M}})
\end{aligned} \tag{20}$$

where $I_{\mathrm{CS}}[D^{\pm}]_{\tilde{M}}$ is the Chern–Simons action $\frac{k_{\mathrm{CS}}}{4\pi}\int_{\tilde{M}}\mathrm{Tr}\{D^{\pm}\wedge\mathrm{d}D^{\pm} + \frac{2}{3}D^{\pm}\wedge D^{\pm}\wedge D^{\pm}\}$, and the coupling constants of the two theories are matched by $k_{\mathrm{CS}} = \ell k_{\mathrm{BF}}/2$. Relation (20) has been well-studied for orientable manifolds [1, 3].

It is well-known that the Chern–Simons theory is a Schwarz-type topological theory [19], so from this relation it is clear that the BF theory is topological too. Because $\tilde{A}$ is an even 1-form field and $\tilde{B}$ is an odd 1-form field, $D^{\pm}$ cannot take arbitrary values. Rather, they must satisfy

$$\sigma^* D^+ = D^-, \tag{21}$$

where σ is the involution defined in section 2, and σ^* is its pullback. By this parity condition, D^- is determined by D^+. In fact, because σ^* commutes with the wedge product and the exterior derivative, and its operation amounts to a minus sign after integration,

$$I_{\mathrm{CS}}[D^-]_{\tilde{M}} = -I_{\mathrm{CS}}[D^+]_{\tilde{M}}, \tag{22}$$

and the action of BF theory can then be re-expressed simply as

$$I_{\mathrm{BF}}[A,B]_M = I_{\mathrm{CS}}[D^+]_{\tilde{M}}. \tag{23}$$

This is an exact equivalence between the BF theory on M and the Chern–Simons theory on the orientable double cover $\tilde{M}$. Properties of the BF theory on the left-hand side of (23), such as its gauge symmetry, classical moduli space, and quantization schemes, can be read off directly from the Chern–Simons theory on the right-hand side of (23) [20].

3.2. Case II: $\Lambda > 0$

If $\Lambda > 0$, we let $\ell = 1/\sqrt{\Lambda}$, and define $C = \tilde{A} + \mathrm{i}\ell^{-1}\tilde{B}$. Then

$$\begin{aligned} I_{\mathrm{BF}}[A, B]_M = {} & \frac{k_{\mathrm{BF}}}{4\pi}\int_{\tilde{M}} \mathrm{Tr}\left\{\frac{\ell}{4\mathrm{i}}(C - C^*)\wedge(\mathrm{d}C + \mathrm{d}C^*) + \frac{\ell}{8\mathrm{i}}(C - C^*)\wedge(C + C^*)\wedge(C + C^*)\right. \\ & \left. + \frac{\ell}{24\mathrm{i}}(C - C^*)\wedge(C - C^*)\wedge(C - C^*)\right\} \\ = {} & \frac{\ell k_{\mathrm{BF}}}{16\mathrm{i}\pi}\int_{\tilde{M}} \mathrm{Tr}\left\{C\wedge\mathrm{d}C + \frac{2}{3}C\wedge C\wedge C - C^*\wedge\mathrm{d}C^* - \frac{2}{3}C^*\wedge C^*\wedge C^*\right\} \\ = {} & \frac{1}{2\mathrm{i}}(I_{\mathrm{CS}}[C]_{\tilde{M}} - I_{\mathrm{CS}}[C^*]_{\tilde{M}}) \end{aligned} \tag{24}$$

where the coupling constants of the two theories are again matched by $k_{\mathrm{CS}} = \ell k_{\mathrm{BF}}/2$. The gauge group of the right hand side theory is $G^{\mathbb{C}}$, which is a group with algebra $\mathfrak{g}^{\mathbb{C}}$, which in turn is the complexification of $\mathfrak{g}$.

The parity conditions on $\tilde{A}$ and $\tilde{B}$ are equivalent to the following condition on C,

$$\sigma^* C = C^*. \tag{25}$$

Note that this does not imply the C field is real—rather, this is a constrained Chern–Simons theory. A similar relation to equation (23) follows from the relation $I^*_{\mathrm{CS}}[C] = I_{\mathrm{CS}}[C^*]$, and so the BF action can be written as

$$I_{\mathrm{BF}}[A, B]_M = \mathrm{Im}\, I_{\mathrm{CS}}[C]_{\tilde{M}}. \tag{26}$$

The gauge transformations are parameterized by a $G^{\mathbb{C}}$-valued function $g^{\mathbb{C}}(\tilde{x})$,

$$C \to (g^{\mathbb{C}})^{-1} dg^{\mathbb{C}} + (g^{\mathbb{C}})^{-1} C g^{\mathbb{C}}. \tag{27}$$

Gauge transformations consistent with the parity conditions are of the form

$$\sigma^* g^{\mathbb{C}}(\tilde{x}) = Z(\tilde{x}) g^{\mathbb{C}*}(\tilde{x}), \tag{28}$$

where $Z(\tilde{x})$ belongs to the center of $G^{\mathbb{C}}$. The classical moduli space is $\mathrm{Hom}(\pi_1(\tilde{M}), G^{\mathbb{C}})/\mathrm{ad}\, G^{\mathbb{C}}$ with the constraint

$$\exp\left(\int_\alpha C\right) = \exp\left(\int_{\sigma\alpha} C^*\right). \tag{29}$$

Existing quantization methods of Chern–Simons theory [19, 21–26] can still be applied, but the constraint (25) or (29) restricts the phase space that will be so quantized.

4. Quantization of U(1) BF theory on non-orientable manifolds

In this section, we study the quantization of U(1) BF theory as an application of the formalism developed in the previous section. An exact solution for the quantized states of the BF action on a non-orientable manifold is tractable for the Abelian gauge group U(1). The choice of using the U(1) gauge group can also be motivated by considering the coupling of the gauge fields to an external electron current; then the U(1) gauge group arises naturally as a consequence of electron charge conservation. Bergeron and Semenoff [27] considered a Chern–Simons model including external U(1) current coupling, wherein they also found a finite-dimensional Hilbert space and quantization condition on the coupling constant k, by solving the clock algebra, large gauge transformation (LGT) group, and braid group for an orientable surface of any genus. We draw similar conclusions here for non-orientable surfaces; to do so we use the full mapping class group (MCG) rather than the braid group: the elements of the braid group can be viewed as equivalence classes of diffeomorphisms, thus the braid group is a subgroup of the MCG.

We follow the method in [28], where U(1) Chern–Simons theory was quantized on orientable manifolds. We look for explicit representations of the discrete symmetry groups of BF theory: (1) the LGT group; (2) the holonomy group; (3) the MCG.

4.1. General formalism

For the Abelian gauge group U(1), the BF theory action (11) simplifies to

$$I = \frac{k}{2\pi}\int_{\mathbb{R}\times N} B\wedge \mathrm{d}A \tag{30}$$

where $k = k_{\mathrm{BF}}$, B is a 1-form-density U(1) gauge field, and A is a regular 1-form U(1) gauge field. By decomposing the fields into time-like components and space-like components, the action becomes

$$I = \frac{k}{2\pi}\int_{N\times\mathbb{R}} \left(B_y\partial_t A_x - B_x\partial_t A_y + B_t F^A_{xy} + A_t F^B_{xy}\right) \tag{31}$$

where $F^A_{xy} = \left(\frac{\mathrm{d}}{\mathrm{d}x}A_y - \frac{\mathrm{d}}{\mathrm{d}y}A_x\right)\mathrm{d}x\wedge\mathrm{d}y$, $F^B_{xy} = \left(\frac{\mathrm{d}}{\mathrm{d}x}B_y - \frac{\mathrm{d}}{\mathrm{d}y}B_x\right)\mathrm{d}x\wedge\mathrm{d}y$. A_t, B_t are Lagrange multipliers and impose $F^A_{xy} = 0, F^B_{xy} = 0$. By choosing the gauge fixing conditions $A_t = 0, B_t = 0$, and regarding A and B as forms on a two-dimensional manifold, the constraint equation is $\mathrm{d}A = 0, \mathrm{d}B = 0$. Any classical solution thus can be Hodge-decomposed as

$$\begin{aligned} A &= \mathrm{d}U + \sum a_i\eta^i \\ B &= \mathrm{d}V + \sum b_i\xi^i \end{aligned} \tag{32}$$

where $\{\eta^i\}(\{\xi^i\})$ is a complete basis of harmonic 1-forms (1-form densities).

The topology of a non-orientable compact surface is specified by the non-orientable genus or demigenus g, and we denote the corresponding surface by N_g, so for example $g = 2$ corresponds to the Klein bottle N_2. The orientable double cover of N_{g+1} is the genus g orientable compact surface Σ_g [29], so for example the orientable double cover of the Klein bottle N_2 is the toroid Σ_1. Let $\#(\alpha,\beta)$ be the algebraic intersection number between two loops α and β. Generators of the fundamental group $\pi_1(\Sigma_g)$ can be taken as $\alpha_i, \beta_i, i = 1,\ldots,g$, where i labels the handles on the surface Σ_g. See figure 3 for the convention used for choosing the fundamental group generators. The generators α_i, β_i obey $\#(\alpha_i,\alpha_j) = 0, \#(\beta_i,\beta_j) = 0$ and $\#(\alpha_i,\beta_j) = \delta_{i-j}$. Each loop generator α,β on handle i corresponds to a harmonic form $\omega^{\alpha,i}, \omega^{\beta,i}$, such that

$$\int_{\alpha_i}\omega^{\alpha,j} = \int_{\beta_i}\omega^{\beta,j} = \delta_{i-j}, \quad \int_{\alpha_i}\omega^{\beta,j} = \int_{\beta_i}\omega^{\alpha,j} = 0, \tag{33}$$

which implies the following orthogonality conditions on the surface Σ_g:

$$\int_{\Sigma_g}\omega^{\alpha,i}\wedge\omega^{\beta,j} = \delta_{i-j}, \quad \int_{\Sigma_g}\omega^{\alpha,i}\wedge\omega^{\alpha,j} = \int_{\Sigma_g}\omega^{\beta,i}\wedge\omega^{\beta,j} = 0. \tag{34}$$

From the harmonic forms $\omega^{\alpha,i}, \omega^{\beta,i}$ on Σ_g, one can construct the corresponding harmonic forms and densities on N_{g+1}. For this purpose, let the surface Σ_g be embedded into $\mathbb{R}^3$, which

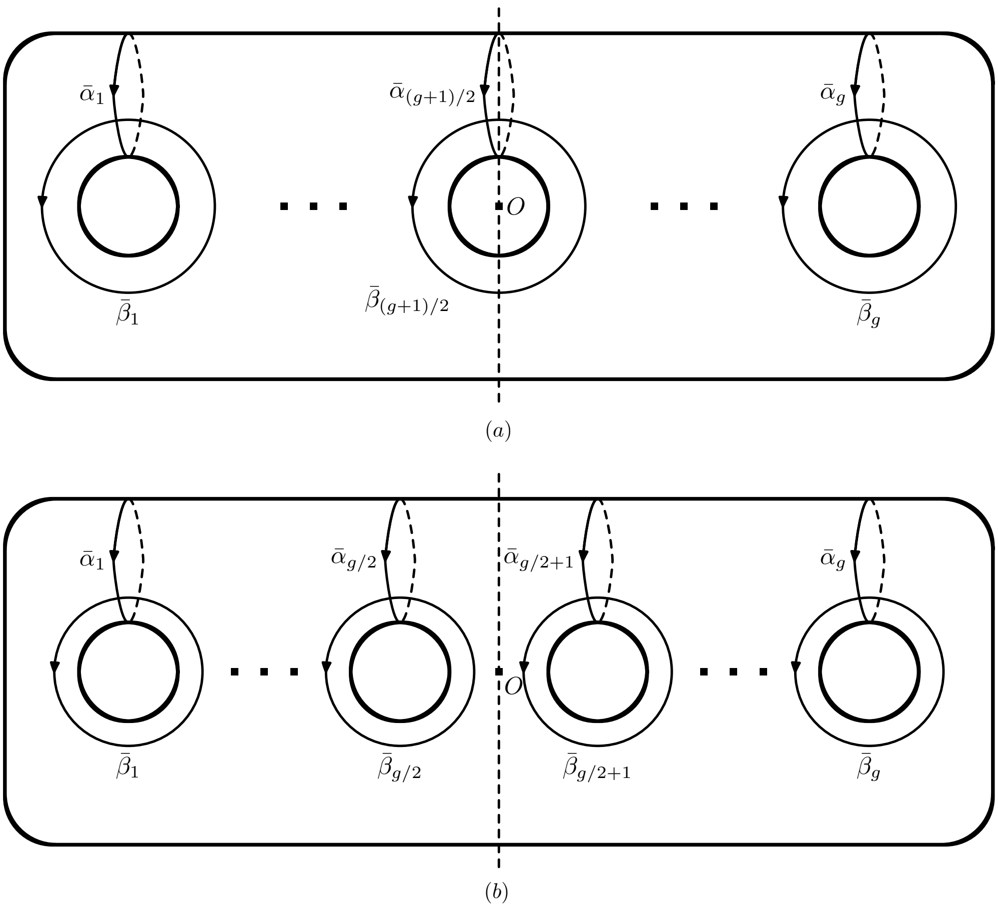

Figure 3. The orientable compact surface Σ_g embedded in $\mathbb{R}^3$, with point O at the origin. The case of odd g and even g are shown in (a) and (b), respectively. The generators of the fundamental group that we choose are shown as directed loops. If pairs of points on Σ_g are identified using the involution σ, then the resulting surface is N_g, composed of the part say to the left of the straight dashed line. For example, the orientable compact surface Σ_1 would consist of the top schematic with only the center hole present (a torus), and the Klein bottle N_2 would be the projected/involuted surface constituting say the left half. The surface Σ_2 would contain only the center two holes of the bottom schematic, and the non-orientable surface N_3 would be the projected/involuted surface consisting of say the left half.

is equipped with a coordinate system such that the embedded surface is symmetric with respect to $x \leftrightarrow -x$, $y \leftrightarrow -y$ and $z \leftrightarrow -z$, separately (see figure 3). The involution σ can be constructed as $\sigma(x, y, z) = (-x, -y, -z)$. With the generators of the fundamental group chosen as in figure 3, one can see, using relations such as $\sigma^*(\omega^{\alpha,i}) = -\omega^{\alpha,(g+1-i)}$, that the following harmonic forms are even

$$\begin{aligned}
&\tilde{\eta}^i = \omega^{\alpha,i} - \omega^{\alpha,(g+1-i)}, \quad i = 1, \cdots, \lfloor g/2 \rfloor, \\
&\tilde{\eta}^{g-i} = \omega^{\beta,i} + \omega^{\beta,(g+1-i)}, \quad i = 1, \cdots, \lfloor g/2 \rfloor, \\
&\tilde{\eta}^{(g+1)/2} = 2\omega^{\beta,(g+1)/2}, \quad \text{if } g \text{ is odd,}
\end{aligned} \tag{35}$$

and the following harmonic forms are odd

$$\begin{aligned} &\tilde{\xi}^{i} = \omega^{\beta,i} - \omega^{\beta,(g+1-i)}, \quad i = 1, \cdots, \lfloor g/2 \rfloor, \\ &\tilde{\xi}^{g-i} = -\omega^{\alpha,i} - \omega^{\alpha,(g+1-i)}, \quad i = 1, \cdots, \lfloor g/2 \rfloor, \\ &\tilde{\xi}^{(g+1)/2} = -\omega^{\alpha,(g+1)/2}, \quad \text{if } g \text{ is odd}. \end{aligned} \tag{36}$$

These harmonic forms are normalized such that

$$\int_{\Sigma_g} \tilde{\eta}^i \wedge \tilde{\xi}^j = 2\delta_{i-j}, \quad \int_{\Sigma_g} \tilde{\eta}^i \wedge \tilde{\eta}^j = \int_{\Sigma_g} \tilde{\xi}^i \wedge \tilde{\xi}^j = 0. \tag{37}$$

Applying the general definition (7) to $\tilde{\eta}$ and $\tilde{\xi}$, it follows that $\widetilde{\eta \wedge \xi} = \tilde{\eta} \wedge \tilde{\xi}$, so that from (9), the corresponding harmonic forms and densities on N_g satisfy

$$\int_{N_g} \eta^i \wedge \xi^j = \delta_{i-j}, \quad \int_{N_g} \eta^i \wedge \eta^j = \int_{N_g} \xi^i \wedge \xi^j = 0. \tag{38}$$

See figure 2 for the example of Klein bottle, which has $g = 1$ ($N_{g+1} = N_2$). The harmonic forms and densities we found above should be complete, because it can be shown any harmonic form or density on N_{g+1} induces a harmonic form on Σ_g, and there are $2g$ harmonic forms on Σ_g, while we already found g harmonic forms and g harmonic densities on N_{g+1}. Using these explicit harmonic forms on N_{g+1}, we obtain the symplectic structure of the phase space.

Using continuous (small) gauge transformations, in (32), the factors U and V can be eliminated. Then substituting (32) and (38) into (31), the BF action simplifies to

$$I_{\text{BF}} = \frac{k}{2\pi} \int \mathrm{d}t \sum_{i=1}^{g} a_i \partial_t b_i, \tag{39}$$

from which we can read off the canonical commutation relation $[a_i, b_j] = \frac{-\mathrm{i}2\pi}{k} \delta_{i-j}$. After the small gauge symmetries are fixed, the BF action is still invariant under the LGT, which effectively translates a_i and/or b_i by multiples of $\mathrm{i}2\pi$. Thus instead of quantizing the U(1)-valued coordinates a_i, b_i, we quantize the LGT-invariant U(1)-valued holonomies

$$\begin{aligned} \tau_i^A &= \begin{cases} \exp\left(\oint_{\alpha_i} A\right), & i < g/2 \\ \exp\left(\oint_{\beta_{g+1-i}} A\right), & i \geqslant g/2 \end{cases} \\ &= \exp\left(a_i\right), \\ \tau_i^B &= \begin{cases} \exp\left(\oint_{\beta_i} B\right), & i < g/2 \\ \exp\left(-\oint_{\alpha_{g+1-i}} B\right), & i \geqslant g/2 \end{cases} \\ &= \exp(b_i), \end{aligned} \tag{40}$$

where A and B are the fields appearing in (32), and the loops α_i, β_i are on N_{g+1}—they are projections of the loops α_i, β_i on Σ_g. Here the holonomies are defined in terms of loops and fields on N_{g+1}, because in this way it is easier to keep track of how they transform under MCG. We also quantize the U(1)-valued LGT group itself, which has the generators

$$\begin{aligned} \rho_i^A &= \begin{cases} \exp\left(k \oint_{\alpha_i} A\right), & i < g/2 \\ \exp\left(k \oint_{\beta_{g+1-i}} A\right), & i \geqslant g/2 \end{cases} \\ &= \exp\left(k a_i\right), \\ \rho_i^B &= \begin{cases} \exp\left(k \oint_{\beta_i} B\right), & i < g/2 \\ \exp\left(-k \oint_{\alpha_{g+1-i}} B\right), & i \geqslant g/2 \end{cases} \\ &= \exp(k b_i). \end{aligned} \tag{41}$$

Classically, the holonomy group and the LGT group formed from these generators are Abelian. After quantization however, these two groups are each deformed to be non-Abelian within themselves due to the canonical commutators, i.e.:

$$\tau_i^A \tau_j^B = \tau_j^B \tau_i^A \exp\left(\frac{-\mathrm{i}2\pi}{k}\delta_{i-j}\right),$$
$$\rho_i^A \rho_j^B = \rho_j^B \rho_i^A \exp(-\mathrm{i}2\pi k \delta_{i-j}). \tag{42}$$

Still, these two groups commute with each other, so we can treat them separately. Their representations are related by the duality transformation $k \leftrightarrow 1/k$, as pointed out by Polychronakos [30].

In addition to the holonomy and LGT groups, we consider the MCG, which is defined as

$$\mathrm{MCG}(M) = \mathrm{Diff}(M)/\mathrm{Diff}_0(M), \tag{43}$$

where $\mathrm{Diff}(M)$ is the diffeomorphism group on M, and $\mathrm{Diff}_0(M)$ is identity component of $\mathrm{Diff}(M)$. Thus states that are covariant with respect to MCG constitute 'toy' U(1) analogues of quantum states in general relativity; we will require here that the quantum states form a representation of the MCG.

The MCG of the Klein bottle $\mathrm{MCG}(N_2)$ was found to be $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ in [31]. It was shown in [29] that the MCG of a non-orientable surface can be derived from its oriented double cover, and an explicit presentation of $\mathrm{MCG}(N_3)$ was derived there. In [32] an algorithm was devised to provide an explicit presentation of the MCG for any non-orientable surface, and this was applied to N_4 in [33]. However, the resulting presentation of $\mathrm{MCG}(N_4)$ is quite complicated, and we shall not calculate the representation of this group in the quantization of BF theory. Beyond N_4, no explicit presentation is presently known.

4.2. Quantization on the surface N_2 (Klein bottle)

For the Klein bottle N_2 having non-orientable genus or demigenus 2, the fundamental group is generated by two loops α_1, β_1, with the relation

$$\alpha_1\beta_1\alpha_1\beta_1^{-1} = 1. \tag{44}$$

It has one even harmonic form $\eta^1 = \omega^{\beta,1}$ and one odd harmonic form $\xi^1 = \omega^{\alpha,1}$, so we define two holonomies

$$\tau^A \equiv \tau_1^A = \mathrm{e}^{\oint_{\beta_1} A} = e^a,\ \ \tau^B \equiv \tau_1^B = \mathrm{e}^{-\oint_{\alpha_1} B} = e^b. \tag{45}$$

From $[a, b] = \frac{-\mathrm{i}2\pi}{k}$, we obtain the clock algebra [27]

$$\tau^A\tau^B = \tau^B\tau^A\omega^{-1}, \tag{46}$$

where $\omega = \exp\left(\frac{\mathrm{i}2\pi}{k}\right)$.

The MCG of the Klein bottle is generated by two elements, a Dehn twist A and a cross-cap slide Y [31]. If one establishes a coordinate system (x, y) with $0 < x, y < 1$ on the surface of the Klein bottle (figure 2), then the action of the cross-cap slide is $y \to 1 - y$ and the action of the Dehn twist is $y \to \begin{cases} x+y & \text{if } x+y<1, \\ x+y-1, & \text{if } x+y>1 \end{cases}$. These operations act on the loops α, β as:

$$\mathsf{A}(\alpha_1, \beta_1) = (\alpha_1, \alpha_1\beta_1) \tag{47}$$

$$\mathsf{Y}(\alpha_1, \beta_1) = (\alpha_1^{-1}, \beta_1). \tag{48}$$

It can be shown using (44) that $\mathsf{A}^2 = 1, \mathsf{Y}^2 = 1, \mathsf{AY} = \mathsf{YA}$, which confirms the MCG is $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

The induced operations on the holonomies are

$$\mathsf{A}^{\dagger}(\tau^B, \tau^A)\mathsf{A} = (\tau^B, \tau^A), \tag{49}$$

$$\mathsf{Y}^{\dagger}(\tau^B, \tau^A)\mathsf{Y} = ((\tau^B)^{-1}, \tau^A). \tag{50}$$

Note that although the Dehn twist A maps the loop β to $\alpha\beta$, the holonomies are not affected by the operator A.

We now seek a representation of τ^B, τ^A, A, Y that satisfies (46), (49), and (50). We solve the clock algebra equation (46) first. For the MCG to have a finite-dimensional representation, k must be rational: $k = p/q$, with p, q coprime [30]. Up to a unitary transformation, it can be shown that the holonomies α and β have the block-diagonal form, with one holonomy diagonal. Specifically, we let the matrices representing the holonomies be composed of r blocks of $p \times p$ sub-matrices, where r is an arbitrary integer:

$$\tau^A = \mathrm{diag}\{\tilde\beta(k, \theta_0^\beta), \ldots, \tilde\beta(k, \theta_{r-1}^\beta)\}, \quad \tau^B = \mathrm{diag}\{\tilde\alpha(k, \theta_0^\alpha), \ldots, \tilde\alpha(k, \theta_{r-1}^\alpha)\}, \tag{51}$$

where $\theta_i^{\alpha,\beta}$ is an arbitrary angle in $[0, \frac{2\pi}{p})$, and the $p \times p$ sub-matrices are given by

$$\begin{aligned}\tilde\beta(k, \theta^\beta)_{ij} &= \delta_{i-j}\omega^i \mathrm{e}^{\mathrm{i}\theta^\beta},\\ \tilde\alpha(k, \theta^\alpha)_{ij} &= \delta_{(i-j-1)\mathrm{mod}p}\mathrm{e}^{\mathrm{i}\theta^\alpha}.\end{aligned} \tag{52}$$

In (52), i, j start from zero, so e.g. for $k = 3$:

$$\tilde\beta(3, \theta^\beta) = \begin{pmatrix}1 & & \\ & \exp(\mathrm{i}2\pi/3) & \\ & & \exp(\mathrm{i}4\pi/3)\end{pmatrix}\mathrm{e}^{\mathrm{i}\theta^\beta} \quad \tilde\alpha(3, \theta^\alpha) = \begin{pmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\mathrm{e}^{\mathrm{i}\theta^\alpha}.$$

Solving (49) yields trivial representation for A. To solve the equations $\mathsf{Y}^\dagger\beta\mathsf{Y} = \beta^{-1}$, $\mathsf{Y}^\dagger\alpha\mathsf{Y} = \alpha$ in (50), we decompose Y into $r \times r$ blocks of $p \times p$ elements as we did for the holonomies. Substituting (51) into (50), we find that for each block

$$\begin{aligned}\tilde{\mathsf{Y}}_{ij} &= \tilde{\mathsf{Y}}_i\delta_{i+j\mathrm{mod}p}\\ \tilde{\mathsf{Y}}_{ij} &= \tilde{\mathsf{Y}}_{i+1,j+1},\end{aligned}$$

where $\tilde{\mathsf{Y}}_i$ are constants, for which a solution exists only if $p = 1$ or 2; the expression for each block of Y is simply

$$\mathsf{Y}_{mn} = u_{mn}^{\mathsf{Y}}\tilde{\mathsf{Y}}, \quad \tilde{\mathsf{Y}} = 1 \text{ or } \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

where u_{mn}^{Y} is a complex number. Because $\tilde{\mathsf{Y}}$ does not depend on the indices m, n, Y can be written as the tensor product form $\mathsf{Y} = U^{\mathsf{Y}} \otimes \tilde{\mathsf{Y}}$, where U^{Y} is a unitary matrix.

The conditions $\mathsf{A}^2 = 1$, $\mathsf{Y}^2 = 1$, $\mathsf{AY} = \mathsf{YA}$ are trivially satisfied. Note that in this case for the surface N_2, the angles $\theta_i^{\alpha,\beta}$ are not fixed by the MCG and remain arbitrary. We will see below that for the surface N_3, the corresponding angles are constrained to be zero.

The representations for the LGT group is just the dual of the representation of the holonomy group, with $p \leftrightarrow q$. Thus q can also only take value 1 or 2, i.e., k is quantized to be:

$$k = \tfrac{1}{2}, \; 1, \; \text{or } 2.$$

The full representation of these discrete groups is

$$\begin{aligned}&\tau^B = U^\alpha \otimes \tilde\alpha(k, 0) \otimes I_q, \quad \tau^A = U^\beta \otimes \tilde\beta(k, 0) \otimes I_q,\\ &\rho^B = U^\rho \otimes I_p \otimes \tilde\alpha(1/k, 0), \quad \rho^A = U^\sigma \otimes I_p \otimes \tilde\beta(1/k, 0),\\ &\mathsf{A} = U^{\mathsf{A}} \otimes I_p \otimes I_q, \quad \mathsf{Y} = U^{\mathsf{Y}} \otimes \tilde{\mathsf{Y}}(k) \otimes \tilde{\mathsf{Y}}(1/k),\end{aligned} \tag{53}$$

where U^{A} and U^{Y} form a unitary representation of the generators of the MCG $\mathbb{Z}_2 \oplus \mathbb{Z}_2$, U^α, U^β are diagonal unitary matrices such that

$$(U^{\mathsf{A}})^\dagger (U^\alpha, U^\beta) U^{\mathsf{A}} = (U^\alpha, U^\beta)$$
$$(U^{\mathsf{Y}})^\dagger (U^\alpha, U^\beta) U^{\mathsf{Y}} = ((U^\alpha)^{-1}, U^\beta),$$

and similarly for U^ρ, U^σ.

4.3. Quantization on the surface N_3 (Dyck's surface)

According to the general formalism in section 4.1, the canonical commutators are $[a_i, b_i] = \frac{-\mathrm{i}2\pi}{k}$, $i = 1, 2$. We define the holonomies as

$$\tau_1^A = \mathrm{e}^{\oint_{\alpha_1} A} = \mathrm{e}^{a_1},\ \ \tau_1^B = \mathrm{e}^{\oint_{\beta_1} B} = \mathrm{e}^{b_1},$$
$$\tau_2^A = \mathrm{e}^{\oint_{\beta_1} A} = \mathrm{e}^{a_2},\ \ \tau_2^B = \mathrm{e}^{-\oint_{\alpha_1} B} = \mathrm{e}^{b_2}. \tag{54}$$

Then again we find the clock algebra $\alpha_i\beta_i = \beta_i\alpha_i\omega^{-1}$, $i = 1, 2$. The clock algebra representation can be written as

$$\tau_1^A = \mathrm{diag}\{\tilde{\beta}(k, \theta_0^{\beta,1}) \otimes I_p, \ldots, \tilde{\beta}(k, \theta_{r-1}^{\beta,1}) \otimes I_p\},$$
$$\tau_1^B = \mathrm{diag}\{\tilde{\alpha}(k, \theta_0^{\alpha,1}) \otimes I_p, \ldots, \tilde{\alpha}(k, \theta_{r-1}^{\alpha,1}) \otimes I_p\},$$
$$\tau_2^A = \mathrm{diag}\{I_p \otimes \tilde{\alpha}(k, \theta_0^{\alpha,2}), \ldots, I_p \otimes \tilde{\alpha}(k, \theta_{r-1}^{\alpha,2})\},$$
$$\tau_2^B = \mathrm{diag}\{I_p \otimes \tilde{\beta}^{-1}(k, \theta_0^{\beta,2}), \ldots, I_p \otimes \tilde{\beta}^{-1}(k, \theta_{r-1}^{\beta,2})\},$$

where $\tilde{\beta}$ and $\tilde{\alpha}$ are given by (52).

The MCG generators act on the loops as [29]

$$\mathsf{A}(\alpha_1, \beta_1) = (\alpha_1, \beta_1\alpha_1),$$
$$\mathsf{B}(\alpha_1, \beta_1) = (\beta_1^{-1}\alpha_1, \beta_1),$$
$$\mathsf{Y}(\alpha_1, \beta_1) = (\alpha_1^{-1}, \beta_1),$$

and they have the relations $\mathsf{ABA} = \mathsf{BAB}$, $(\mathsf{BAB})^4 = 1$, $\mathsf{Y}^2 = 1$, $\mathsf{YAY} = \mathsf{A}^{-1}$, $\mathsf{YBY} = \mathsf{B}^{-1}$. Their induced operations on holonomies are

$$\mathsf{A}^\dagger(\tau_1^A, \tau_1^B, \tau_2^A, \tau_2^B)\mathsf{A} = (\tau_1^A, \tau_1^B(\tau_2^B)^{-1}, \tau_2^A\tau_1^A, \tau_2^B),$$
$$\mathsf{B}^\dagger(\tau_1^A, \tau_1^B, \tau_2^A, \tau_2^B)\mathsf{B} = ((\tau_2^A)^{-1}\tau_1^A, \tau_1^B, \tau_2^A, (\tau_1^B)^{-1}\tau_2^B),$$
$$\mathsf{Y}^\dagger(\tau_1^A, \tau_1^B, \tau_2^A, \tau_2^B)\mathsf{Y} = ((\tau_1^A)^{-1}, \tau_1^B, \tau_2^A, (\tau_2^B)^{-1}).$$

To find representations of the MCG, we again decompose A, B and Y into $r \times r$ blocks of $p \times p$ elements. Let us first focus on the (m, m)th block. A_{mm} and B_{mm} have the solution

$$\mathsf{B}_{mm} = u_{mm}^{\mathsf{B}}\tilde{\mathsf{B}}(k, \theta_m^{\beta 1}, \theta_m^{\beta 2}, \theta^{\mathsf{B}}),$$
$$\mathsf{A}_{mm} = u_{mm}^{\mathsf{A}}\tilde{\mathsf{A}}(k, \theta_m^{\alpha 1}, \theta_m^{\alpha 2}, \theta^{\mathsf{A}}),$$

where

$$\tilde{\mathsf{B}}(k, \theta_m^{\beta 1}, \theta_m^{\beta 2}, \theta^{\mathsf{B}})_{i_1 j_1, i_2 j_2} = \omega^{-i_1 i_2}\mathrm{e}^{-\mathrm{i}(i_1\theta_m^{\beta 2} + i_2\theta_m^{\beta 1})}\delta_{i_1 - j_1}\delta_{i_2 - j_2}\mathrm{e}^{\mathrm{i}\theta_{\mathsf{B}}}$$
$$\tilde{\mathsf{A}}(k, \theta_m^{\alpha 1}, \theta_m^{\alpha 2}, \theta^{\mathsf{A}})_{i_1 j_1, i_2 j_2} = \frac{1}{p}\omega^{(i_1 - j_1)(i_2 - j_2)}\mathrm{e}^{-\mathrm{i}[(i_1 - j_1)\theta_m^{\alpha 2} + (i_2 - j_2)\theta_m^{\alpha 1}]}\mathrm{e}^{\mathrm{i}\theta_A}.$$

The periodicity of $\tilde{\mathsf{A}}$ and $\tilde{\mathsf{B}}$ enforces that $\theta_m^{\beta 1}, \theta_m^{\beta 2}, \theta_m^{\alpha 1}, \theta_m^{\alpha 2}$ are all multiples of $2\pi/p$. However a unitary transformation on $\tilde{\alpha}$ and $\tilde{\beta}$ can shift any of these angles by a multiple of $2\pi/p$, hence we can take all of these angles to be 0. This restriction on the phase angles is stronger than the restriction on the surface N_2, where each of the phase angles could take any value on the interval $[0, 2\pi/p)$. After fixing these phase angles, the holonomies take the simplified form

$$\begin{aligned}\beta_1 &= I_r \otimes \tilde{\beta}(k, 0) \otimes \tilde{I}_p,\\ \alpha_1 &= I_r \otimes \tilde{\alpha}(k, 0) \otimes \tilde{I}_p,\\ \beta_2 &= I_r \otimes \tilde{I}_p \otimes \tilde{\beta}(k, 0),\\ \alpha_2 &= I_r \otimes \tilde{I}_p \otimes \tilde{\alpha}(k, 0).\end{aligned}$$

and it is straightforward to solve for all blocks of A, B,

$$\begin{aligned}\mathsf{B} &= U^{\mathsf{B}} \otimes \tilde{\mathsf{B}}(k, 0, 0, \theta^{\mathsf{B}}),\\ \mathsf{A} &= U^{\mathsf{A}} \otimes \tilde{\mathsf{A}}(k, 0, 0, \theta^{\mathsf{A}}),\end{aligned}$$

where U^{B} and U^{A} are arbitrary r-dimensional unitary matrices.

The relation $\mathsf{ABA} = \mathsf{BAB}$ gives $\theta^{\mathsf{A}} = \theta^{\mathsf{B}}$. Note that in this case no quadratic Gauss sum is involved in the equation. The relation $(\mathsf{BAB})^4 = 1$ gives $\mathrm{e}^{\mathrm{i}12\theta^{\mathsf{B}}} = 1$.

However, when solving for Y, it turns out that solution exists only if $p = 1$ or 2, and $\mathsf{Y} = \pm I_p \otimes I_p$, $\theta^{\mathsf{A}} = \theta^{\mathsf{B}} = 0$ or π. As the previous cases, these choices of phases can be absorbed into the Abelian representation of the MCG, which leaves the part of MCG representation that interacts with the holonomy group trivial when $p = 1$, and

$$\tilde{\mathsf{B}} = \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad \tilde{\mathsf{A}} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & -1\\ 1 & 1 & -1 & 1\\ 1 & -1 & 1 & 1\\ -1 & 1 & 1 & 1 \end{pmatrix}, \quad \tilde{\mathsf{Y}} = \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 \end{pmatrix}$$

when $p = 2$.

As for the representation of LGT and the part of MCG interacting with LGT, the same result can be found, with $p \leftrightarrow q$. Thus when $\Sigma = N_3$, the quantization condition for k is the same as for the Klein bottle N_2, namely that k can only take values:

$$k = \tfrac{1}{2},\ 1,\ \text{or } 2.$$

5. Discussion

In this paper, we have taken the formalism involving the integration of p-form densities on non-orientable manifolds in section 2, and established the relations between (2+1)D gravity, BF theory, and Chern–Simons theory. We found that the well-known relation between (2+1)D gravity and Chern–Simons theory can be extended to the case of non-orientable manifolds: then (2+1)D gravity defined on a non-orientable manifold is equivalent to Chern–Simons theory on the orientable double cover, and thus Chern–Simons theory continues to give a well-defined quantization scheme for (2+1)D gravity. This formalism was then applied in section 4 to quantize U(1) BF theory defined on the spacetime manifold $\mathbb{R} \times N_{g+1}$, where N_{g+1} is the non-orientable surface whose orientable double cover is Σ_g.

In quantizing U(1) BF theory, the holonomy group and the large gauge transformation (LGT) group are deformed according to the canonical commutation relations, while the mapping class group (MCG) is not deformed. For non-orientable surfaces having non-orientable genera 2 and 3 (the Klein bottle N_2 and Dyck's surface N_3), explicit and tractable presentations of MCG are known; for these cases we found explicit, finite-dimensional representations of the discrete groups. In order to consistently quantize the system, the values of the coupling constant k are strongly restricted to be either 1/2, 1 or 2. We suspect, though we have not proved it, that the allowable values of k are at least this restricted for higher genus surfaces as well.

For the non-orientable surface N_3, the phase angles associated the holonomy group generators and the LGT group generators are fully fixed by the value of k, due to the requirement of representing the MCG; the same situation occurs for the case of an orientable surface Σ_g [28]. For the Klein bottle N_2 on the other hand, the MCG contains only 4 elements, and does not fix the phase angles, which remain continuously adjustable on the interval $[0, 2\pi/p)$. Although no straightforwardly applicable presentation of the MCG is known for N_{g+1} with $g > 2$, promotion of the MCG to quantum operators is evidently important in characterizing the allowable dimension of the Hilbert space $\sim (pq)^g$ where $k = p/q$: before applying the MCG, arbitrarily large Hilbert spaces were allowed.

It is finally worth noting that the quantum states on the non-orientable manifold are not simply the quantum states on the oriented double cover with the correct parity. For example, on N_2 and N_3, when k does not take a value among $1/2$, 1 or 2, there is no quantum state consistent with a representation of the MCG, while on Σ_1 and Σ_2, k can take any rational value with the numerator or the denominator being even, and quantum states with any parity can be constructed. The extra restriction on k on non-orientable manifolds is due to non-trivial effects arising from the quantization of the MCG. Specifically, in the calculation of section 4.3, representing the Dehn twists A and B, which have analogues on the orientable surface Σ_2, does not introduce this extra quantization condition. It is only when we calculate the representation of the Y-homeomorphism Y, which is unique to the non-orientable surface, that the quantization condition appears.

Acknowledgments

We thank Gordon Semenoff for helpful discussions. SSP and DMW acknowledge funding from the Natural Sciences and Engineering Research Council. SC acknowledges support from the 4YF program at UBC Vancouver.

References

[1] Carlip S 2005 Quantum gravity in 2+1 dimensions: the case of a closed universe *Living Rev. Rel.* **8** 1
[2] Achucarro A and Townsend P K 1986 A Chern–Simons action for three-dimensional anti-de Sitter supergravity theories *Phys. Lett.* B **180** 89
[3] Witten E 1988 (2+1)-dimensional gravity as an exactly soluble system *Nucl. Phys.* B **311** 46
[4] Warner F W 1971 *Foundations of Differentiable Manifolds and Lie Groups* (*Graduate Texts in Mathematics*) (Berlin: Springer)
[5] Carlip S 1995 The statistical mechanics of the (2+1)-dimensional black hole *Phys. Rev.* D **51** 632–7
[6] Nayak C, Simon S H, Stern A, Freedman M and Das Sarma S 2008 Non-Abelian anyons and topological quantum computation *Rev. Mod. Phys.* **80** 1083–159
[7] Einarsson T 1990 Fractional statistics on a torus *Phys. Rev. Lett.* **64** 1995–8
[8] Kitaev A Yu 2003 Fault-tolerant quantum computation by anyons *Ann. Phys.* **303** 2–30
[9] Matschull H-J 1999 On the relation between (2+1) Einstein gravity and Chern–Simons theory *Class. Quantum Grav.* **16** 2599–609
[10] Buffenoir E and Noui K 2003 Unfashionable observations about three-dimensional gravity arXiv:gr-qc/0305079
[11] Witten E 2007 Three-dimensional gravity revisited arXiv:0706.3359 [hep-th]
[12] Woodhouse N M J 1997 *Geometric Quantization* (Clarendon: Oxford)
[13] Bott R and Tu L W 1982 *Differential Forms in Algebraic Topology* (*Graduate Texts in Mathematics* vol 82) (New York: Springer)

[14] Murdoch T A 1988 Twisted calibrations and the cone over the Veronese surface *PhD Thesis* Rice University

[15] Horowitz G T 1989 Exactly soluble diffeomorphism invariant theories *Commun. Math. Phys.* **125** 417

[16] Louko J 1995 Witten's 2+1 gravity on R x (Klein bottle) *Class. Quantum Grav.* **12** 2441–68

[17] Rham G 1984 *Differentiable Manifolds: Forms, Currents, Harmonic Forms* (*Grundlehren der mathematischen Wissenschaften*) (Berlin: Springer)

[18] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (San Francisco: Freeman)

[19] Witten E 1989 Quantum field theory and the Jones polynomial *Commun. Math. Phys.* **121** 351

[20] Dunne G V 1999 Aspects of Chern–Simons theory *Aspects Topologiques de la Physique en Basse Dimension. Topological Aspects of Low Dimensional Systems* (Berlin: Springer) pp 177–263

[21] Moore G W and Seiberg N 1989 Taming the conformal zoo *Phys. Lett.* B **220** 422

[22] Axelrod S, Pietra S D and Witten E 1991 Geometric quantization of Chern–Simons gauge theory *J. Diff. Geom.* **33** 787–902

[23] Bos M and Nair V P 1990 Coherent state quantization of Chern–Simons theory *Int. J. Mod. Phys.* A **5** 959

[24] Alekseev A Yu, Grosse H and Schomerus V 1995 Combinatorial quantization of the Hamiltonian Chern–Simons theory *Commun. Math. Phys.* **172** 317–58

[25] Alekseev A Yu, Grosse H and Schomerus V 1995 Combinatorial quantization of the Hamiltonian Chern–Simons theory: 2. *Commun. Math. Phys.* **174** 561–604

[26] Alekseev A Yu and Schomerus V 1996 Representation theory of Chern–Simons observables *Duke Math. J.* **85** 447–510

[27] Bergeron M and Semenoff G W 1996 Canonical Chern–Simons theory and the braid group on a Riemann surface *Ann. Phys.* **245** 1–22

[28] Chen Si 2012 Mapping class group and U(1) Chern–Simons theory on closed orientable surfaces *Mod. Phys. Lett.* A **27** 1250087

[29] Birman J S and Chillingworth D R J 1972 On the homeotopy group of a non-orientable surface *Proc. Cambridge Phil. Soc.* **71** 437–48

[30] Polychronakos A P 1990 On the quantization of the coefficient of the Abelian Chern–Simons term *Phys. Lett.* B **241** 37

[31] Lickorish W B R 1963 Homeomorphisms of non-orientable two-manifolds *Proc. Cambridge Phil. Soc.* **59** 307–17

[32] Szepietowski B 2008 A presentation for the mapping class group of a non-orientable surface from the action on the complex of curves *Osaka J. Math.* **45** 283–326

[33] Szepietowski B 2009 A presentation for the mapping class group of the closed non-orientable surface of genus 4 *J. Pure Appl. Algebra* **213** 2001–16