

# The QCD $\theta$ -parameter in canonical quantization

Wen-Yuan Ai,<sup>a</sup> Björn Garbrecht<sup>b</sup> and Carlos Tamarit<sup>c</sup>

<sup>a</sup>*Theoretical Particle Physics and Cosmology, King's College London,  
Strand, London WC2R 2LS, UK*

<sup>b</sup>*Physik-Department T70, Technische Universität München,  
James-Frank-Straße, 85748 Garching, Germany*

<sup>c</sup>*PRISMA<sup>+</sup> Cluster of Excellence & Mainz Institute for Theoretical Physics,  
Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*

## Abstract

The role of the QCD  $\theta$ -parameter is investigated in pure Yang–Mills theory in the spacetime given by the four-dimensional Euclidean torus. While in this setting the introduction of possibly unphysical boundary conditions is avoided, it must be specified how the sum over the topological sectors is to be carried out. To connect with observables in real time, we perceive the partition function as the trace over the canonical density matrix. The system then corresponds to one of a finite temperature on a spatial three-torus. Carrying out the trace operation requires canonical quantization and gauge fixing. Fixing the gauge and demanding that the Hermiticity of the Hamiltonian is maintained leads to a restriction of the Hilbert space of physical wave functionals that generalizes the constraints derived from imposing Gauß' law. Consequently, we find that the states in the Hilbert space are properly normalizable under an inner product that integrates over each physical configuration represented by the gauge potential one time and one time only. The observables derived from the constrained Hilbert space do not violate charge-parity symmetry. We note that an exact hidden symmetry of the theory that is present for arbitrary values of  $\theta$  in the Hamiltonian is effectively promoted to parity conservation in this constrained space. These results, derived on a torus in order to avoid the introduction of boundary conditions, also carry over to Minkowski spacetime when taking account of all possible gauge transformations.

---

# Contents

1	Introduction	2
2	Topological quantization on the four-torus	5
3	Canonical quantization of gauge theory	9
4	Gauge redundancies and gauge fixing	11
5	Fixing the transition function and normalization of physical states	13
6	Constraints on the quantum states	16
7	Gauß' law constraint	21
8	Gauge covariance with respect to general changes of the transition function	22
9	Generalization to Minkowski spacetime	25
10	The hidden parity symmetry of QCD in the presence of a $\theta$ -term	26
11	Conclusions	31
A	Path integral over the gauge-fixed field configuration space	33

---

## 1 Introduction

The present paper concerns the topological term in Yang–Mills theory and the possibility of its observable effects [1–3] when the theory is canonically quantized. As a pure gauge theory, it is specified by the following Euclidean Lagrangian

$$\mathcal{L} = \frac{1}{2g^2} \text{tr} F_{\mu\nu} F_{\mu\nu} - \frac{i}{16\pi^2} \theta \text{tr} F_{\mu\nu} \tilde{F}_{\mu\nu}, \quad (1)$$

where  $F_{\mu\nu} = F_{\mu\nu}^a T^a$  with  $F_{\mu\nu}^a \equiv \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + f^{abc} A_\mu^b A_\nu^c$  is the field strength tensor and  $\tilde{F}_{\mu\nu} \equiv \frac{1}{2} \varepsilon_{\mu\nu\alpha\beta} F_{\alpha\beta}$  ( $\varepsilon_{1234} = 1$ ) is the Hodge dual of  $F_{\mu\nu}$ . We use the convention  $\text{Tr}(T^a T^b) = \delta^{ab}/2$ ,  $[T^a, T^b] = i f^{abc} T^c$  for the Lie algebra generators  $T^a$  and the structure constants  $f^{abc}$ , and we consider SU(2) for simplicity so that  $T^a = \tau^a/2$  with  $\tau^a$  being the Pauli matrices. The last term in Eq. (1) is odd under charge-parity ( $CP$ ) conjugation.

Through the remainder of the present text, we colloquially refer to Yang–Mills pure gauge theory with the gauge group  $SU(2)$  as quantum chromodynamics (QCD).

Before outlining matters of canonical quantization, we recall for comparison some salient points about the theory (1) when evaluated as a path integral in an infinite spacetime volume  $\Omega$ . The partition function is then given by

$$Z = \int \mathcal{D}A e^{-\lim_{\Omega \rightarrow \infty} \int_{\Omega} d^4x \mathcal{L}}. \quad (2)$$

There are a priori no boundary conditions imposed on the path integral. However, configurations of finite action are classified into different topological sectors characterized by integer  $\Delta n$ —the winding number. We refer to this as topological quantization. The integer  $\Delta n$  comes as a consequence of the limit  $\Omega \rightarrow \infty$ . This limit must therefore be taken before the sum over the sectors  $\Delta n$  is carried out. One then finds that there is no  $CP$  violation in the correlation functions [4]. Furthermore, leaving the boundary conditions in infinite spacetime unspecified allows the contours of path integration for different  $\Delta n$  to be connected by configurations of infinite action so that there exists a full integration contour over all  $\Delta n$  in the sense of the Cauchy theorem. For the individual sectors  $\Delta n$ , the contours can be constructed using steepest descent methods [5].

These points stand in opposition to the main share of literature on this topic (see e.g. Refs. [3] and [6] for a seminal paper and a recent review) in which the limits are taken the other way around and boundary conditions corresponding to field operator eigenstates that are pure gauge configurations are imposed on finite surfaces. These surfaces are only taken to infinity after the sum over  $\Delta n$  is carried out. In that approach, topological quantization does not follow from the definition of the partition function (2) but from the boundary conditions that are imposed ad hoc. Moreover, the field configurations with fixed boundary conditions on finite surfaces but different  $\Delta n$  cannot be continuously connected through segments of a contour in the gauge field configuration space which do not contribute to the path integral and arise from a nonsingular deformation of the original integration contour. Therefore, for these boundary conditions, there is no integration contour that would allow the application of the Cauchy theorem to the path integral. Calculationally, one finds in that order of limits, i.e.  $\Omega \rightarrow \infty$  after summation over  $\Delta n$ , there is  $CP$  violation in the correlation functions. For other recent works that question the existence of  $CP$  violation in QCD, see for example Refs. [7–9].

Given this context, it is therefore of interest to understand whether the vanishing of  $CP$  violation persists in systems that can be quantized in finite volume without imposing unphysical boundary conditions. These can be avoided by choosing manifolds without boundary. The apparently simplest case is to take the three-torus  $T^3$  for the spatial hypersurfaces on which canonical quantization is to be carried out. For path integral quantization, we need to specify the range of the temporal coordinate  $x_4$  (in the notation used for Euclidean time). In order to project on the ground state, one should take the range of  $x_4$  to infinity. This way, we do not need to specify boundary conditions on the spatial hypersurfaces  $T^3$  at  $x_4 \rightarrow \pm\infty$ . However, then we are back to the discussion on the partition function (2) in infinite spacetime volumes [4, 10].

It is therefore of interest to compactify time as well. A way to achieve this without specifying ad hoc boundary conditions is to impose periodicity in  $x_4$  so that the spacetime becomes  $T^4$ . Such a setup is physical as the resulting partition function corresponds to the one of a canonical ensemble kept at a temperature  $T$  where the range of  $x_4$  corresponds to  $1/T$ . Now, for the theory (1) on  $T^4$  once again the field configurations decompose into equivalence classes of integer  $\Delta n$ . While this is a well-known fact, we review the reasons for this topological quantization in Section 2 in order to relate it to the topological aspects that come into play in the subsequent discussion of canonical quantization, a central subject of the present paper.

To compute observables at finite temperature, we need an expression for the partition function that corresponds to Eq. (2) but now for  $T^4$  instead of an infinite  $\Omega$ . The answer is not as straightforward as it may seem. On  $T^4$ , there are no continuous deformations that connect sectors of different  $\Delta n$  and consequently, these do not lie on a single Cauchy contour. Furthermore, fixing the winding number  $\Delta n$  on  $T^4$  leads to a well-defined Euclidean quantum field theory for each  $\Delta n$  so that there is a priori no necessity for summing over  $\Delta n$ . We therefore need a more detailed reasoning to determine whether we need to sum over the topological sectors and how we must weigh the contributions in such a sum, in particular in terms of phase factors.

A unique way of specifying the finite-temperature partition function is to refer to its definition as

$$Z = \text{tr} e^{-H/T}, \quad (3)$$

where  $H$  is the Hamiltonian and the trace is over the physical states in the Hilbert space. We thus need to gather enough facts about the Hilbert space in order to derive the partition function. In particular, to carry out the trace operation, we must establish in what sense the states are normalizable and that we can fix the gauge in order not to sum over redundant states. While we are focusing here on canonical quantization, we note that when the observables obtained in this approach are independent of the particular gauge fixing, one can derive an expression for the path integral in which the gauge symmetry along with Lorentz invariance becomes manifest again [11]. Upon evaluating path integrals, gauge fixing is then in practice reintroduced in a different way using the standard functional methods, i.e. the Faddeev—Popov procedure.

To identify the Hamiltonian and the Hilbert space, we review some of the basic facts on canonical quantization of QCD theory in Section 3. In Section 4, we specify in what sense states should be normalizable after gauge fixing and make provisions in our notation for excluding fluctuations and derivatives from the Hamiltonian in the direction of gauge redundancies.

A particular set of gauge transformations that we discuss in Section 5 are those that change the transition functions. The latter are required to construct the maps from  $T^3$  into the vector potential  $\mathbf{A}$  in the temporal gauge. When we fix these transition functions, the remaining gauge transformations must be periodic over  $T^3$  and hence decompose again into integer equivalence classes. In particular, those periodic transformations that are not continuously connected with the identity are referred to as large

gauge transformations. The set of large gauge transformations (corresponding to  $n \neq 0$  below) can be generally expressed as

$$\{G_n\} = Z_n \otimes \{G_0\}, \quad (4)$$

where  $Z_n$  is the additive group of integers and  $\{G_0\}$  is the set of so-called small gauge transformations that are connected with the identity. At this point, one can introduce wave functionals that are eigenstates of large gauge transformations  $G_n$  with  $\exp(in\theta^{(i)})$  as their eigenvalues. The superscript here allows to write expressions involving wave functionals with different eigenvalues  $\exp(in\theta^{(j)})$  with  $\theta^{(i)} \neq \theta^{(j)}$  and also indicates that these eigenvalues are in the first place distinct from  $\theta$  in the Lagrangian (1).

However, we note that the states for different  $\theta^{(i)}$  based on a gauge-fixed inner product are not orthogonal. In turn, when relaxing the gauge fixing in the inner product we encounter the problem noted in Ref. [12] that the states are not properly normalizable. To resolve these issues, we pursue the direction that the Hilbert space with different  $\theta^{(i)}$  is too large. The correct value of  $\theta^{(i)}$  is uniquely determined by demanding that (i) one can fix the gauge; (ii) the operators corresponding to observables remain Hermitian; and (iii) the observables are the same in different gauges. In Section 6, we derive the constraints on the wave functionals that follow from these requirements. As a familiar example on which to apply this reasoning, in Section 7 we review that imposing Gauß' law restricts the Hilbert space to states that are gauge singlets with respect to transformations that leave the transition functions invariant. In turn, we can use the requirement of gauge invariance to restrict the Hilbert space to states that observe Gauß' law.

In Section 8 we make the analogous argument for transformations changing the transition function. As it turns out, this indeed determines the value of  $\theta^{(i)}$  so that the gauge-fixed inner product can be used without problem on the remaining Hilbert space. The allowed value for  $\theta^{(i)}$  is the one that complements the  $CP$ -odd phase from the Hamiltonian so that there remain no  $CP$ -violating effects. While the results to that point have been derived on a torus in order to avoid dealing with boundary conditions, in Section 9, we show that the same reasoning applies to Minkowski spacetime as well when one does not restrict gauge transformations at spatial infinity, and we also show how the results for the wave functionals explain  $CP$  conservation in amplitudes evaluated over finite time intervals without periodic temporal boundary conditions. In Section 10, a new discrete symmetry of the Hamiltonian operator is identified. It can be used to show that for the gauge fixed Hamiltonian, energy eigenstates in the physical Hermitian subspace can always be constructed as eigenstates of the usual parity. Conclusions can be found in Section 11.

## 2 Topological quantization on the four-torus

While it is not necessary to consider  $T^4$  in order to show  $CP$  conservation in canonical quantization on the spatial  $T^3$ , the canonical density matrix in its path integral representation on  $T^4$  is an object of special interest. In that context, it is essential that the

winding number  $\Delta n$  is an integer on  $T^4$ , i.e.

$$\Delta n = \frac{1}{16\pi^2} \int_{T^4} d^4x F \tilde{F} \in \mathbb{Z}. \quad (5)$$

Proofs of the topological quantization (5) can be found e.g. in Ref. [13]. In order to see in what way topology on  $T^3$  and  $T^4$  are related, we show here explicitly how Eq. (5) holds.

To that end, we recall that the integrand can be written as a total divergence,

$$\frac{1}{16\pi^2} \text{tr} F_{\mu\nu} \tilde{F}_{\mu\nu} = \partial_\mu K_\mu, \quad (6)$$

where

$$K_\mu = \frac{1}{4\pi^2} \epsilon_{\mu\nu\alpha\beta} \text{tr} \left[ \frac{1}{2} A_\nu \partial_\alpha A_\beta - \frac{i}{3} A_\nu A_\alpha A_\beta \right]. \quad (7)$$

We recall that for  $SU(2)$ ,  $\text{tr}(T^a T^b T^c) = i f^{abc}/4$  so that the above expression can be written in terms of the components of the gauge fields as

$$K_\mu = \frac{1}{16\pi^2} \epsilon_{\mu\nu\alpha\beta} \left[ A_\nu^a \partial_\alpha A_\beta^a + \frac{1}{3} f^{abc} A_\nu^a A_\alpha^b A_\beta^c \right]. \quad (8)$$

Next, consider the hypercube with edges of length  $a_i$ ,  $i = 1, \dots, 4$  from which we can obtain  $T^4$  by gluing the opposite three-faces together. On the opposite three-faces of the hypercube, the vector potential must match modulo a gauge transformation. For example,

$$\begin{aligned} A_\mu(\varepsilon, x_2, x_3, x_4) &= U_1^{-1}(x_2, x_3, x_4) A_\mu(a_1 - \varepsilon, x_2, x_3, x_4) U_1(x_2, x_3, x_4) \\ &\quad + i U_1^{-1}(x_2, x_3, x_4) \partial_\mu U_1(x_2, x_3, x_4), \end{aligned} \quad (9)$$

and accordingly for the remaining pairs of three-faces. Here,  $\varepsilon$  is an infinitesimal length that we have introduced in order to formulate this matching condition.

Equation (9) is routinely imposed when putting gauge fields on a torus [14] and does not appear to require more explanation than given above. When we insist on stating it in a bit more technical terms, we can take an open cover  $\{\mathcal{U}_\alpha\}$  of  $T^4$ . Gauge-covariant quantities on  $T^4$  can then be defined with the help of transition functions  $t_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \rightarrow SU(2)$ . In particular, for  $x \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta$ , there is the compatibility condition

$$A_\mu^\beta = t_{\alpha\beta}^{-1} A_\mu^\alpha t_{\alpha\beta} + i t_{\alpha\beta}^{-1} \partial_\mu t_{\alpha\beta}, \quad (10)$$

where  $A_\mu^{\alpha,\beta}$  are the gauge potentials defined in the respective  $\mathcal{U}_{\alpha,\beta}$ . For consistency, the transition functions have to satisfy the cocycle condition: For  $x \in \mathcal{U}_\alpha \cap \mathcal{U}_\beta \cap \mathcal{U}_\gamma$  it must hold that  $t_{\alpha\gamma} = t_{\alpha\beta} t_{\beta\gamma}$ .

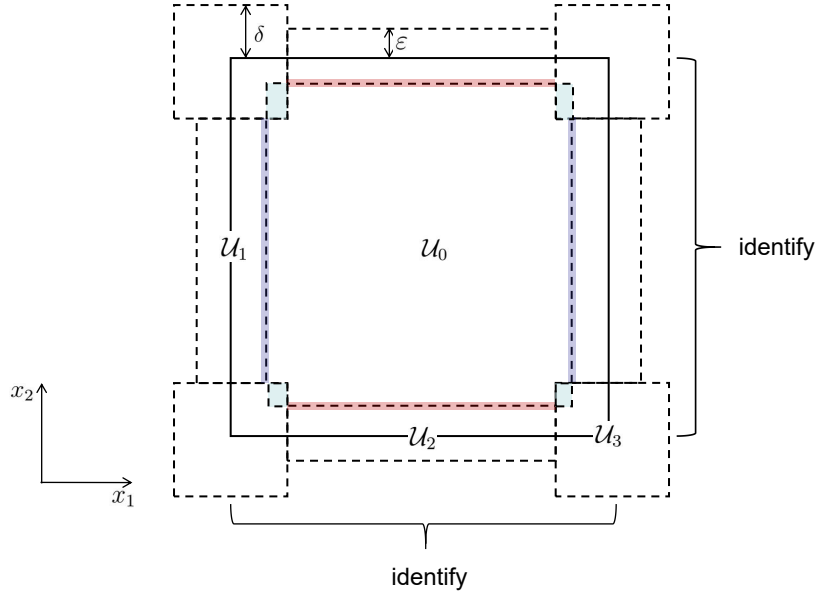


Figure 1: Illustration of a cover of  $T^2$  (the region enclosed by the solid line) using four open sets (regions enclosed by dashed lines and denoted as  $\mathcal{U}_\alpha$  with  $\alpha = 0, 1, 2, 3$ ) of which overlapping regions (shaded regions in color) have a minimal area when  $\delta, \epsilon \rightarrow 0$  (keeping  $\delta > \epsilon$  in taking the limit) [15].

To arrive at Eq. (9), it is advantageous to reduce the overlapping regions of the open sets  $\mathcal{U}_\alpha$  to a minimal “volume” [15]. For  $T^n$  the minimal number of the open sets for such a covering is  $2^n$ . One can illustrate this for  $T^2$  as an example, as shown in Figure 1. Note the identification for opposite edges and faces so that the four squares at the corners are identified as one single open set  $\mathcal{U}_3$ , the opposite rectangles are identified as  $\mathcal{U}_1$  and  $\mathcal{U}_2$  respectively. The respective overlaps between  $\mathcal{U}_0$  and  $\mathcal{U}_\alpha$ ,  $1 \leq \alpha \leq 3$  (e.g. the two lines in blue for  $\alpha = 1$ ) have disconnected parts. From the figure, one can see that  $\delta > \epsilon$  but eventually, one should take  $\delta, \epsilon \rightarrow 0$  to have a minimal area for the overlapping regions.

For  $T^4$ , we take  $\mathcal{U}_0$  as the open set that covers the hypercube just without its boundaries. Fifteen more sets  $\mathcal{U}_1, \dots, \mathcal{U}_{15}$  can be arranged along the cells (three-faces), faces (two-faces), edges (one-faces) and vertices (zero-faces) of the hypercube so that one obtains a complete cover. For definiteness, let the  $\mathcal{U}_i$  for  $i = 1, \dots, 4$  be the sets covering the cells with  $x_i = 0$  along with their opposites  $x_i = a_i$  up to the infinitesimal shell along the faces (analogues of  $\mathcal{U}_{1,2}$  in the  $T^2$  example in Figure 1). Then shrink  $\mathcal{U}_1, \dots, \mathcal{U}_{15}$  so that they cover just infinitesimal neighbourhoods of the cells, faces, edges and vertices. Analogous to the case of  $T^2$  the identification of  $x_i$  with  $x_i + a_i$  determines the number of open sets needed to cover the full  $T^4$ . Note that in this construction, the intersections  $\mathcal{U}_0 \cap \mathcal{U}_\alpha$ ,  $1 \leq \alpha \leq 15$  are not connected, as illustrated on the example of  $T^2$ . The transition functions over each pair of opposite cells of the hypercube can then be composed

from the transition functions on the disjoint subdomains according to e.g.

$$U_1(x_2, x_3, x_4) = t_{10}(\varepsilon, x_2, x_3, x_4)t_{01}(a_1 - \varepsilon, x_2, x_3, x_4) \quad (11)$$

so that Eq. (9) follows.

Then we choose a gauge so that the gauge potential is periodic in the three spatial directions [16], i.e.,  $A_\mu(x + a_i \hat{e}_i) = A_\mu(x)$  for  $i = 1, 2, 3$ . This can be accomplished because we do not impose boundary conditions on the allowed gauge transformations on the three-faces at  $x_4 = 0, a_4$ . As a consequence of this gauge choice, the net flux of  $K_\mu$  through the opposite three-faces  $x_{1,2,3} = 0$  and  $x_{1,2,3} = a_{1,2,3}$  vanishes. In addition, we can impose the temporal gauge so that

$$A_4 = 0 \quad (12)$$

everywhere. In Minkowski spacetime, this gauge condition reads  $A_0 = 0$ .

Now consider the opposite three-faces at  $x_4 = 0$  and  $x_4 = a_4$ . Given the gauge choices made thus far, all of the total flux of  $K_\mu$  through the surface of the hypercube must go through these two three-faces. Let

$$F_4 = [0, a_1] \times [0, a_2] \times [0, a_3] \quad (13)$$

be the set pertaining to these three-faces on which  $U_4$  is defined corresponding to Eq. (9). We can conclude that  $U_4(\mathbf{y}) = \mathbb{1}$  for  $\mathbf{y} \in \partial F_4$  because of the periodicity of the gauge potential in the spatial directions that is assumed here. Therefore, we can identify the points in  $\partial F_4$  so that in this sense, the three-face is homeomorphic with  $S^3$ . Now given the winding number

$$\nu(U; S) = \frac{1}{4\pi^2} \varepsilon_{ijk} \int_S d^3x \frac{1}{6} \text{tr} [U^{-1}(\partial_i U)U^{-1}(\partial_j U)U^{-1}(\partial_k U)] , \quad (14)$$

the transition functions  $U_4(\mathbf{x})$  therefore constitute homotopy classes with integer winding number, i.e.  $\nu(U_4; F_4) \in \mathbb{Z}$  for  $U_4 = \text{constant}$  on  $\partial F_4$ .

It remains to be shown that this implies Eq. (5), in particular when the field configurations on the three-face  $F_4$  are not a pure gauge. Consider the Chern–Simons functional

$$W[\mathbf{A}] = \frac{1}{4\pi^2} \varepsilon_{ijk} \int_S d^3x \text{tr} \left[ \frac{1}{2} A_i \partial_j A_k - \frac{i}{3} A_i A_j A_k \right] \quad (15)$$

over some three-dimensional hypersurface  $S$ . Now for a gauge transformation on  $S$

$$\mathbf{A}' = U \mathbf{A} U^{-1} + iU \nabla U^{-1} =: \mathbf{A}_U \quad (16)$$

the difference between the Chern–Simons functionals is [16]

$$W[\mathbf{A}'] - W[\mathbf{A}] = \nu(U; S) + \frac{i}{8\pi^2} \varepsilon_{ijk} \int_S d^3x \partial_j \text{tr} [A_k (\partial_i U^{-1}) U] . \quad (17)$$



In Eq. (16), we have introduced the shorthand notation  $\mathbf{A}_U$ , which we shall frequently use through the remainder of this work. Substituting  $U = U_4$  and  $S = F_4$  in Eq. (17), this is the difference between the Chern–Simons numbers that flows through the three-faces at  $x_4 = 0$  and  $x_4 = a_4$ . As we have made use of the gauge freedom so that  $U_4$  is constant on  $\partial F_4$  and that  $\mathbf{A}$  is periodic on  $F_4$ ,  $\text{tr}[A_k(\partial_i U_4)U_4^{-1}]$  is periodic on  $F_4$  and the integral vanishes by Gauß’ theorem. Furthermore, the Chern–Simons flux through the remaining faces vanishes and  $\nu(U_4, F_4) \in \mathbb{Z}$ . Hence, using Eq. (6), we recognize that  $\Delta n = W[\mathbf{A}'] - W[\mathbf{A}] \in \mathbb{Z}$ , i.e. Eq. (5) holds. Since Eq. (5) is manifestly gauge invariant, we can change the above construction to general gauges without changing this result.

### 3 Canonical quantization of gauge theory

Canonical quantization of non-Abelian gauge theory is reviewed in e.g. Ref. [17] which includes many details. This section contains a summary of the main points that are relevant in the present context.

To this end, we start by casting the Euclidean Lagrangian (1) to Minkowski spacetime

$$\mathcal{L} = -\frac{1}{2g^2} \text{tr} F_{\mu\nu} F^{\mu\nu} + \frac{1}{16\pi^2} \theta \text{tr} F_{\mu\nu} \tilde{F}^{\mu\nu}. \quad (18)$$

Note that we do not introduce different notations for the gauge field in the Euclidean and Minkowski spacetimes. To make our conventions explicit, we put a superscript “E” on the Euclidean gauge field  $A_\mu^E$  just for now. Then the Minkowskian gauge field  $A_\mu$  can be obtained from the Euclidean gauge field  $A_\mu^E$  via  $A_i^E = A_i$ ,  $A_4^E = -iA_0$  together with the conventional Wick rotation  $x_4 = it$ .

In the following, we make occasional use of the covariant derivative

$$D_\mu = \partial_\mu - iA_\mu^a \frac{\tau^a}{2}. \quad (19)$$

In terms of this operator, the field strength tensor

$$F_{\mu\nu} = F_{\mu\nu}^a \frac{\tau^a}{2} \quad (20)$$

can be expressed as

$$[D_\mu, D_\nu] = -iF_{\mu\nu}. \quad (21)$$

In canonical quantization, as long as we are not interested in the time evolution, we are concerned with  $T^3$  instead of  $T^4$ . However, the discussion in the previous section is related to the present context because we can, without loss of generality, take  $F_4$  as the spatial hypersurface  $T^3$  that the quantum field is defined on. And  $U_4$  can here be identified as a gauge transformation for the gauge field on the same spatial hypersurface  $T^3$ .

The most straightforward approach to canonical quantization of Yang–Mills theory appears to go through the temporal gauge (12), in which the non-Abelian electric and magnetic fields are given by

$$g\mathbf{E}^a = -\frac{\partial}{\partial t} \mathbf{A}^a, \quad (22)$$

$$g\mathbf{B}^a = \nabla \times \mathbf{A}^a - \frac{1}{2} f^{abc} \mathbf{A}^a \times \mathbf{A}^b. \quad (23)$$

In terms of these, the canonical momentum conjugate to  $\mathbf{A}^a$  is

$$g\boldsymbol{\Pi}^a \equiv g \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{A}}^a} = -\mathbf{E}^a + \frac{g^2}{8\pi^2} \theta \mathbf{B}^a. \quad (24)$$

The coordinates  $\mathbf{A}^a$ ,  $\boldsymbol{\Pi}^a$  are canonical and satisfy the corresponding Poisson brackets. This is the main advantage over imposing more restrictive gauge conditions, where one would need to generalize to Dirac brackets. While this is possible for the Abelian case, such procedure appears to face prohibitive complications for non-Abelian theories. On the other hand, using the above set of canonical coordinates requires extra steps when quantizing the theory. Since the temporal gauge leaves substantial gauge freedom in the form of the transformation (16), unphysical states and gauge redundancies must be eliminated subsequent to quantization. The present procedure may therefore colloquially be referred to as “first quantize then constrain” (modulo the constraint (12) that is imposed before quantization). We therefore yet need to specify how to narrow down the Hilbert space to physical states and eventually to fix the gauge completely.

The operators corresponding to the canonical coordinates in temporal gauge must observe the commutation relations:

$$[A^{a,i}(\mathbf{x}), \Pi^{b,j}(\mathbf{x}')] = i\delta^{ij}\delta^{ab}\delta^3(\mathbf{x} - \mathbf{x}'), \quad [\Pi^{a,i}(\mathbf{x}), \Pi^{b,j}(\mathbf{x}')] = 0. \quad (25)$$

These commutators hold for

$$\boldsymbol{\Pi}^a = \frac{\delta}{i\delta \mathbf{A}^a} + \theta_{\Pi} \frac{g}{8\pi^2} \mathbf{B}^a. \quad (26)$$

Here,  $\theta_{\Pi}$  is an arbitrary constant that is real for  $\boldsymbol{\Pi}^a$  to be Hermitian.

Do not confuse Eq. (24) with Eq. (26). The former is a general relation while the latter is just a representation of the operator  $\boldsymbol{\Pi}^a$  (in the “coordinate representation”, i.e. when acting on wave functionals  $\Psi[\mathbf{A}]$ ). Note that the only constraint imposed on the canonical formalism at this point is  $A_0 = 0$ , which can simply be satisfied without leading to complications in the commutation relations because this coordinate is not dynamical. In particular, we have not imposed periodicity of  $\mathbf{A}(\mathbf{x})$  and therefore should not assume this or any other constraint before making a consistent restriction to a gauge-fixed subspace.

Combining Eqs. (24) and (26) yields the operator  $\mathbf{E}^a$ ,

$$\mathbf{E}^a = -\left( g \frac{\delta}{i\delta \mathbf{A}^a} - \frac{g^2}{8\pi^2} (\theta - \theta_{\Pi}) \mathbf{B}^a \right). \quad (27)$$

Substituting the above into the Hamiltonian density, one obtains

$$\mathcal{H} = \frac{1}{2} [(\mathbf{E}^a)^2 + (\mathbf{B}^a)^2] = \frac{1}{2} \left[ \left( g \frac{\delta}{i\delta\mathbf{A}^a} - \frac{g^2}{8\pi^2} (\theta - \theta_\Pi) \mathbf{B}^a \right)^2 + (\mathbf{B}^a)^2 \right]. \quad (28)$$

The Hamiltonian

$$H = \int d^3x \mathcal{H} \quad (29)$$

acts on states represented through wave functionals  $\Psi[\mathbf{A}]$ . However, since we have not fixed the gauge prior to quantization, the Hilbert space of all functionals is too large, i.e. it contains unphysical states as well as gauge redundancies. In the following sections, we review how the unphysical states are eliminated.

## 4 Gauge redundancies and gauge fixing

As reviewed in Section 3, the quantized theory defined through the canonical commutation relations (25) may yet allow for unphysical states and entails gauge redundancies that must be removed by gauge fixing. In the present section, we therefore introduce some notations in order to deal with this matter.

We first get back to the point that there is still some residual gauge freedom. As different gauges are redundant descriptions of the same physical system, it is possible to fix the gauge in such a way that there is a one-to-one correspondence between the functions  $\mathbf{A}(\mathbf{x})$  and classical physical configurations, i.e. that there is a complete gauge fixing. This implies that we can express the inner product of two states as

$$\langle \Psi | \Phi \rangle = \int_{\mathcal{A}} \mathcal{D}\mathbf{A} f_{\mathcal{A}}[\mathbf{A}] \Psi^*[\mathbf{A}] \Phi[\mathbf{A}], \quad (30)$$

where the subscript on the integral indicates complete gauge fixing. That is, we integrate over a manifold  $\mathcal{A}$  on which each gauge inequivalent configuration is being represented one time and one time only. The factor  $f_{\mathcal{A}}[\mathbf{A}]$  is generally present in order to maintain the gauge covariance of the fixed integration measure, i.e. the measure  $\mathcal{D}\mathbf{A} f_{\mathcal{A}}[\mathbf{A}]$  is translation invariant when restricted to  $\mathcal{A}$ . For example, when fixing the gauge by inserting a gauge condition as an argument of a delta-function,  $f_{\mathcal{A}}(\mathbf{A})$  would be given by the absolute value of the inverse determinant of the gauge condition. That being said, we could also decide to integrate over redundant configurations if the corresponding integration volume is finite so that we can still properly normalize the states. Nonetheless, it should always be possible if not advantageous to use an inner product as in Eq. (30) that does not extend over redundant configurations. In turn, noting that Eq. (30) can represent a physical transition amplitude, it should be gauge invariant, so that extending the integration over an infinite measure of gauge redundant configurations leads to

unphysical infinite answers. In particular, were there no gauge fixing, the inner product of Eq. (30) would lead to states of infinite norm, in contradiction with the postulates of quantum mechanics.

The restriction of the Schrödinger equation to a certain subset of field configurations is naturally defined through bases of the field configuration space  $\{\mathbf{A}^a(\mathbf{x})\}$ . Let  $\mathbf{G}^a(\mathbf{x}; \sigma)$  be a basis for the configuration space, where  $\sigma$  is the multiindex that labels the particular basis modes. Then, a derivative in the direction of  $\mathbf{G}^a(\mathbf{x}; \sigma)$  can be defined as

$$\frac{\delta}{\delta A(\sigma)} \Psi[\mathbf{A}] = \int d^3y \sum_{i,a} G_i^a(\mathbf{y}; \sigma) \frac{\delta}{\delta A_i^a(\mathbf{y})} \Psi[\mathbf{A}]. \quad (31)$$

The operator  $\delta/\delta A(\sigma)$  may be understood as a derivative with respect to the corresponding coefficient in the expansion of  $\mathbf{A}^a(\mathbf{x})$  in the basis  $\mathbf{G}^a(\mathbf{x}; \sigma)$ . That is, writing

$$A_i^a(\mathbf{x}) = \int_{\sigma} c(\sigma) G_i^a(\mathbf{x}; \sigma), \quad (32)$$

one has

$$\frac{\delta}{\delta A(\sigma)} = \frac{\delta}{\delta c(\sigma)}. \quad (33)$$

The derivative  $\delta/\delta c(\sigma)$  is then an ordinary derivative in the case of discrete values of  $\sigma$  and a functional derivative in the case of continuous values.

Note that the basis  $G_i^a(\mathbf{y}; \sigma = \{j, b, \mathbf{x}\}) = \delta^3(\mathbf{y} - \mathbf{x}) \delta^{ij} \delta_{ab}$  reproduces the standard derivative  $\delta/\delta A_b^j(\mathbf{x})$ . If the basis is orthonormal, i.e. if the following identity is satisfied,

$$\int_{\sigma} G_i^a(\mathbf{x}; \sigma) G_j^b(\mathbf{y}; \sigma) = \delta^{(3)}(\mathbf{x} - \mathbf{y}) \delta^{ab} \delta^{ij}, \quad (34)$$

we can conversely write

$$\frac{\delta}{\delta A_i^a(\mathbf{x})} \Psi[\mathbf{A}] = \int_{\sigma} G_i^a(\mathbf{x}; \sigma) \frac{\delta}{\delta A(\sigma)} \Psi[\mathbf{A}]. \quad (35)$$

When we identify some basis modes  $\mathbf{G}^a(\mathbf{x}; \sigma_{\text{gauge}})$  with infinitesimal gauge transformations of a given configuration  $\mathbf{A}^a(\mathbf{x})$ , we can fix in the Schrödinger equation the gauge with respect to such transformations by omitting the derivatives  $\delta/\delta A(\sigma_{\text{gauge}})$ . In the following, we apply this to transformations generated by Gauß' law, which are periodic over the torus as well as to more general aperiodic transformations that change the transition function.

## 5 Fixing the transition function and normalization of physical states

To derive the Hamiltonian (28), no extra gauge fixing in addition to the temporal gauge has been applied. This is the assumption that we are going to work with in the subsequent sections. Before going ahead, in this section we consider what happens when we fix the transition function because this way one may arrive at constructions on the three-torus that correspond to the  $\theta$ -vacua usually discussed for infinite space. While the temporal gauge imposes no restriction on the transition function (9), let us consider here  $U_i = \mathbb{1}$  for  $i = 1, 2, 3$ . This way, the gauge potential is restricted to be periodic on the three torus and  $U_4 = \mathbb{1}$  on  $\partial F_4$ .

It is essential to note that other choices of the transition functions are also possible and generally lead to non-periodic gauge potentials  $\mathbf{A}$  on  $T^3$  even though the physical configuration remains periodic. The assumption made in the present section, that one can simply ignore field configurations that are not periodic in  $\mathbf{A}$  when restricting the form of the wave functionals, is something that should therefore be verified. Fixing the gauge and thereby the transition function before quantization corresponds to constraints that are difficult to implement in non-Abelian gauge theories. In Section 6, we therefore derive how to restrict the Hilbert space in order to allow for gauge fixing subsequent to quantization.

Considering again Eq. (17) with  $S = F_4$ ,  $U = U_4$  and assuming  $U_4 = \mathbb{1}$  on  $\partial F_4$ , we recognize that for a gauge transformation (16) under the imposed constraint of periodic vector potentials,  $W[\mathbf{A}'] - W[\mathbf{A}] \in \mathbb{Z}$ . We emphasize that this constraint that leads to topological quantization in this sense is a choice that one does not necessarily have to make. Going ahead notwithstanding this point, for a gauge transformation  $U_4^{(1)}$  so that  $W[\mathbf{A}'] - W[\mathbf{A}] = 1$ , using the fact that it commutes with the Hamiltonian  $H$ , one concludes that there are eigenstate wave functionals  $\Psi_{\theta^{(i)}}[\mathbf{A}]$  of  $H$  and  $U_4^{(1)}$  so that

$$\Psi_{\theta^{(i)}}[\mathbf{A}_{(U_4^{(1)})^n}] = e^{in\theta^{(i)}} \Psi_{\theta^{(i)}}[\mathbf{A}] \quad (36)$$

with some real  $\theta^{(i)}$  and where we use the notation introduced in Eq. (16). According to the Bloch theorem, these can be written as

$$\Psi_{\theta^{(i)}}[\mathbf{A}] = \psi_{\theta^{(i)}}[\mathbf{A}] e^{iW[\mathbf{A}]\theta^{(i)}}, \quad (37)$$

with Eq. (15) for  $W[\mathbf{A}]$ ,  $S = F_4$ , and  $\psi_{\theta^{(i)}}[\mathbf{A}] = \psi_{\theta^{(i)}}[\mathbf{A}']$ .

So far, there are three angles that appear in the canonical quantization of QCD:  $\theta$  in the Lagrangian,  $\theta_{\Pi}$  in the canonical field momentum operator and  $\theta^{(i)}$  in the wave functionals. (When introducing fermions, which we do not consider in the present work, there is an additional angular parameter from the Dirac mass.) The combination  $\theta - \theta_{\Pi}$  should appear in the Lagrangian of the path integral when one constructs the latter from the Hamiltonian (28). One might think that the freedom to choose  $\theta_{\Pi}$  could be used to remove the theta parameter. However, the path integral should be weighted by

wave functionals that arise from the projection from the initial and final physical states onto the gauge configurations (see Eq. (81) below). Therefore, the theta parameter usually discussed in QCD, where one often does not distinguish the additional angular parameters, corresponds to  $\theta - \theta_{\Pi} - \theta^{(i)}$  which is an invariant (because  $\theta_{\Pi} - \theta^{(i)}$  is invariant) [17]. However, as we shall show, a consistent canonical quantization of the QCD on  $T^3$  implies that the spectrum contains only states for which the invariant  $\theta - \theta_{\Pi} - \theta^{(i)}$  vanishes. As a consequence, the path integral formulation for the same theory, i.e. QCD in a finite volume, can only have a vanishing theta parameter. In Section 9, we also discuss how this result carries over to infinite volume, i.e.,  $\mathbb{R}^3$ , when we discard the hypothesis (79).

Now take two states  $\Psi_{\theta^{(i,j)}}^{(a,b)}$  of this type, where  $a, b$  are labels for states in orthonormal bases of the subspaces with fixed  $\theta^{(i)}, \theta^{(j)}$ , respectively. We may then note that

$$\begin{aligned} \int \mathcal{D}\mathbf{A} \Psi_{\theta^{(i)}}^{(a)*}[\mathbf{A}] \Psi_{\theta^{(j)}}^{(b)}[\mathbf{A}] &= \sum_{\nu=-\infty}^{\infty} \int_{0 \leq W[\mathbf{A}] < 1} \mathcal{D}\mathbf{A} e^{-i(\theta^{(i)} - \theta^{(j)})(W[\mathbf{A}] + \nu)} \psi_{\theta^{(i)}}^{(a)*}[\mathbf{A}] \psi_{\theta^{(j)}}^{(b)}[\mathbf{A}] \\ &= 2\pi \delta(\theta^{(i)} - \theta^{(j)}) \int_{0 \leq W[\mathbf{A}] < 1} \mathcal{D}\mathbf{A} e^{-i(\theta^{(i)} - \theta^{(j)})W[\mathbf{A}]} \psi_{\theta^{(i)}}^{(a)*}[\mathbf{A}] \psi_{\theta^{(j)}}^{(b)}[\mathbf{A}] \\ &= 2\pi \delta(\theta^{(i)} - \theta^{(j)}) \delta_{ab}, \end{aligned} \quad (38)$$

where we have assumed the normalization

$$\int_{0 \leq W[\mathbf{A}] < 1} \mathcal{D}\mathbf{A} \psi_{\theta^{(i)}}^{(a)*}[\mathbf{A}] \psi_{\theta^{(i)}}^{(a)}[\mathbf{A}] = 1. \quad (39)$$

As advertised above, the properties given by Eqs. (36), (38) are analogous to those of the so-called  $\theta$ -vacua that are commonly discussed in the literature,

$$|\theta\rangle = \sum_{n \in \mathbb{Z}} e^{in\theta} |n\rangle. \quad (40)$$

In the equation above, the  $|n\rangle$  are eigenstates of the operator corresponding to  $W[A]$  with integer eigenvalues, and hence in direct relation to classical states of minimal energy given by pure gauge configurations  $\mathbf{A} = iU^{-1}\nabla U$ , for which Eq. (17) implies that  $W[\mathbf{A}] = \nu(U; S) \in \mathbb{Z}$ . This means that the wave-functionals  $\Psi_{|\theta\rangle}[\mathbf{A}] = \langle \mathbf{A} | \theta \rangle$  corresponding to the states of Eq. (40) only have support in classical minima of the energy functional. In ordinary quantum mechanics, it is well known that in the case of multiple classical vacua the wave functions of the stationary states have support away from the classical minima, so that the states (40) are not expected to correspond to physical stationary states and therefore are in fact not “vacua”. Despite the analogies with the states (40), our present discussion is more general, as we never assume that the wave functionals  $\Psi_{\theta^{(i)}}(\mathbf{A})$  have support in classical field configurations of minimal energy only.

Going back to the general wave functionals  $\Psi_{\theta^{(i)}}(\mathbf{A})$ , a particular gauge fixing may be imposed such that the condition  $0 \leq W[\mathbf{A}] < 1$  is obeyed along with additional

constraints. But then with the corresponding gauge fixed inner product as in Eq. (30), there is no sum over  $\nu$  so that states with  $\theta^{(i)} \neq \theta^{(j)}$  are no longer orthogonal.

At this point, one may see two apparent ways of resolving this conflict.

- (a) We dismiss the requirement that there should be a gauge fixed inner product as in Eq. (30). Instead, we take (38) as an inner product even though we integrate over gauge-redundant configurations, which as mentioned before leads to states that are not properly normalizable. They are therefore not members of a Hilbert space, and taking these as physical states would be in conflict with the postulates of quantum mechanics.
- (b) We stick with the inner product (30). To avoid a contradiction, only states  $\Psi_{\theta^{(i)}}$  with a particular value of  $\theta^{(i)}$  (modulo  $2\pi$ ) can be in the Hilbert space. We will show that only one value of  $\theta^{(i)}$  (modulo  $2\pi$ ) is admissible, namely the one that leads to the prediction of no  $CP$  violation in the strong interactions.

The way we have put things here anticipates that in the present work, we take the view that Option (b) is the correct path to pursue. Nonetheless, we have to take due account of the fact that the vast majority of papers (see e.g. Ref. [18] for a review), save Refs. [12, 19], follow Option (a), propose the use of inner products over gauge redundant configurations and are not concerned with nonnormalizable states that are considered as unphysical in different contexts. (At best, in quantum mechanics, momentum eigenstates in a translation-invariant background, which are not normalizable, constitute a basis of the physical Hilbert space. They can be used to construct normalizable wave packets which are the physical states. When taking a trace over momentum eigenstates, the ensuing factors  $(2\pi)^3\delta^3(0)$  can be identified with the infinite volume  $V$ . It is not clear and may not be possible to interpret the improper normalization of the  $\theta$ -vacua in a corresponding way. In particular, wave-packet constructions, even if the wave packets are strongly peaked about a certain value of  $\theta^{(i)}$ , are in conflict with gauge invariance.) In addition to these objections, we therefore show explicitly that the restoration of gauge covariance after fixing the transition function requires that  $\theta^{(i)}$  takes values in accordance with  $CP$  conservation.

To make such an argument, we recall that, as stated above, any transition function  $U_4$ —also one that is not constant—on  $\partial F_4$  may be imposed as a partial gauge fixing before quantization. In order to restore gauge covariance, we may therefore consider the family of states that is obtained by having all possible transition functions  $U_4$  on  $\partial F_4$ , thus composing a Hilbert space with an extended basis. In this extended space, the Hamiltonian must act consistently, i.e. a simultaneous covariant gauge transformation of the Hamiltonian and its energy eigenstates should not change the eigenvalues. In other words, changing the transition function should not change the spectrum. This condition leads to a restriction of the possible values of  $\theta^{(i)}$ , which turns out to resolve the aforementioned issues with normalization and gauge fixing. A more direct way to obtain a gauge covariant construction with respect to the transition functions is of course to fix the latter only after quantization, which is the order in which we work in Sections 7 and 8.

Before proceeding, we comment on the case when space is  $\mathbb{R}^3$  and infinite as opposed to  $T^3$  and finite. For infinite space, one may have reasons to adopt the hypothesis that the gauge potentials must fall faster than  $1/r$  at large distances, which in turn limits the gauge transformations in Eq. (16) to those with  $U(|\mathbf{x}|) \rightarrow \text{constant}$  for  $|\mathbf{x}| \rightarrow \infty$  [17]. As a consequence, the points at infinity can then be identified so that space can be viewed as homeomorphic to the three-sphere  $S^3$ , for which  $\nu(U, S^3) \in \mathbb{Z}$ . In particular, there are transformations  $U^{(1)}$  with  $\nu(U^{(1)}, S^3) = 1$  so that Eq. (36) holds for  $U_4^{(1)} = U^{(1)}$ . When we assume that gauge transformations that are not constant for  $|\mathbf{x}| \rightarrow \infty$  are not permissible, we can also not discern which value of  $\theta^{(i)}$  gives the correct basis states by studying the gauge-covariant behaviour of the states and the Hamiltonian under the restricted gauge transformations. But as we recall, for infinite spacetime, correlation functions computed in the functional approach turn out to be independent of  $\theta$  and do not exhibit  $CP$  violation [4]. (The four-dimensional volume of space is to be taken to infinity before summing over the topological sectors in the four-dimensional spacetime.) The choice of  $\theta^{(i)}$  is then of no consequence for the predictions of observables. We can just fix it and therefore obtain a Hilbert space based on a gauge fixed inner product (30). We shall return to the case of infinite  $\mathbb{R}^3$  in Section 9, subsequent to presenting the main results.

Now, to further pursue Option (b) we need to work out the covariant behaviour of the states under gauge transformations. Considering gauge transformations that change the transition function, this will fix the value of  $\theta^{(i)}$  to be discussed in Section 8. Transformations that keep the transition function invariant are related with Gauß' law, as we recall next.

## 6 Constraints on the quantum states

Gauge fixing as stated in Section 5 corresponds to the choice of a constrained hypersurface  $\mathcal{A}$  for the field configurations  $\mathbf{A}(\mathbf{x})$  so that each gauge-inequivalent configuration appears one time and one time only. Under the induced inner product (30) without restricting the gauge configurations to a certain space  $\mathcal{A}$ , the space of square-integrable functionals  $\Psi$ , denoted here by  $\mathcal{C}$ , is generally no longer a Hilbert space. However, since in the present case the choice of  $\mathcal{A}$  corresponds to the removal of gauge redundancies, we can expect that for given  $\mathcal{A}$ , there is a subspace  $\subset \mathcal{C}$  that qualifies as a Hilbert space also for the inner product (30) and that is constituted by physical states.

To identify this subspace, we first notice that the independence of probabilities on the choice of  $\mathcal{A}$  requires the following form of the wave-functional for a state labeled with (a):

$$\Psi^{(a)}[\mathbf{A}] = \Psi_{\text{g.i.}}^{(a)}[\mathbf{A}] \exp(i\varphi^{(a)}[\mathbf{A}]), \quad (41)$$

where  $\Psi_{\text{g.i.}}^{(a)}[\mathbf{A}]$  is gauge-independent and  $\varphi^{(a)}[\mathbf{A}]$  is a real functional. That is, gauge transformations can induce at most a phase change in the wave functionals, as such



phases cancel when computing physical transition or probability amplitudes defined from inner products of wave functions.

Furthermore, the Hamiltonian as well as the operator  $-i\delta/\delta\mathbf{A}$  that it is built from should remain Hermitian operators. This leads to

$$\begin{aligned} & \int_{\mathcal{A}} \mathcal{D}\mathbf{A} f_{\mathcal{A}}[\mathbf{A}] \Psi_{\text{g.i.}}^{(a)*}[\mathbf{A}] \exp(-i\varphi^{(a)}[\mathbf{A}]) \frac{\delta}{i\delta A(\sigma)} \Psi_{\text{g.i.}}^{(b)}[\mathbf{A}] \exp(i\varphi^{(b)}[\mathbf{A}]) \\ &= \int_{\mathcal{A}} \mathcal{D}\mathbf{A} f_{\mathcal{A}}[\mathbf{A}] \left( \frac{\delta}{i\delta A(\sigma)} \Psi_{\text{g.i.}}^{(a)}[\mathbf{A}] \exp(i\varphi^{(a)}[\mathbf{A}]) \right)^* \Psi_{\text{g.i.}}^{(b)}[\mathbf{A}] \exp(i\varphi^{(b)}[\mathbf{A}]). \end{aligned} \quad (42)$$

Note that the above equality must hold for any pair of states such that  $\Psi^{(a)}$  and  $\Psi^{(b)}$  are not necessarily orthogonal here. Therefore, it must be that  $\varphi^{(a)}[\mathbf{A}] = \varphi^{(b)}[\mathbf{A}] = \varphi[\mathbf{A}]$ . For the terms with derivatives acting on  $\varphi[\mathbf{A}]$ , the above equation holds because  $\varphi[\mathbf{A}]$  is independent of  $(a)$ . This explains the Hermiticity of all derivatives in the directions  $\sigma = \sigma_{\text{gauge}}$  because the remaining factor  $\Psi_{\text{g.i.}}^{(a)}[\mathbf{A}]$  is gauge invariant. For all other directions  $\sigma$  the variational derivative is equivalent to one that is tangential to  $\mathcal{A}$ . By the translation-invariance of  $\mathcal{D}[\mathbf{A}]f_{\mathcal{A}}[\mathbf{A}]$  on  $\mathcal{A}$ , the above relation then follows from partial integration and vanishing boundary terms involving  $|\Psi_{\text{g.i.}}^{(a)*}[\mathbf{A}]|^2$ . In some directions in field space, there arise improper boundaries when the physical field strengths become unbounded and the wave functional goes to zero. In other particular directions boundaries occur for finite physical fields from e.g. the constraint  $0 \leq W[\mathbf{A}] < 1$ . The configurations on such gauge-equivalent locations on the hypersurfaces  $W[\mathbf{A}] = 0$  and  $W[\mathbf{A}] = 1$  are related by gauge transformations which leave  $|\Psi_{\text{g.i.}}^{(a)*}[\mathbf{A}]|^2$  invariant, so that the contributions from the hypersurfaces cancel. We conclude

$$\Psi^{(a)}[\mathbf{A}] = \Psi_{\text{g.i.}}^{(a)}[\mathbf{A}] \exp(i\varphi[\mathbf{A}]), \quad (43)$$

and note that  $\varphi[\mathbf{A}]$  is the only piece that still may depend on the choice of  $\mathcal{A}$ .

We proceed to show that there are indeed solutions for the wave functional that factorize into a piece  $\Psi_{\text{g.i.}}^{(a)}[\mathbf{A}]$  that is gauge independent and another piece  $p[\mathbf{A}]$  that may be gauge dependent but is universal for all states, i.e.

$$\Psi^{(a)}[\mathbf{A}] = \Psi_{\text{g.i.}}^{(a)}[\mathbf{A}] p[\mathbf{A}], \quad (44)$$

which is a more general form of Eq. (43) and hence a necessary condition. For this purpose, we use the fact that

$$\frac{\delta}{\delta\mathbf{A}(\mathbf{x})} W[\mathbf{A}] = \frac{g}{8\pi^2} \mathbf{B}(\mathbf{x}), \quad (45)$$

in order to carry out the following change of basis:

$$\Psi'[\mathbf{A}] = e^{-i(\theta - \theta_{\text{H}})W[\mathbf{A}]} \Psi[\mathbf{A}], \quad (46)$$

$$\begin{aligned} H' &= e^{-i(\theta - \theta_{\text{H}})W[\mathbf{A}]} H e^{i(\theta - \theta_{\text{H}})W[\mathbf{A}]} = \frac{1}{2} \int d^3x \operatorname{tr} \left[ -g^2 \frac{\delta^2}{\delta \mathbf{A}^2} + \mathbf{B}^2 \right] \\ &= -\frac{g^2}{2} \int_{\sigma} \frac{\delta^2}{\delta A^2(\sigma)} + \frac{1}{2} \int d^3x \operatorname{tr} \mathbf{B}^2. \end{aligned} \quad (47)$$

Now in the Hamiltonian  $H'$ , as opposed to Eq. (28), there are no cross terms  $\mathbf{B}(\mathbf{x}) \cdot (\delta/\delta \mathbf{A}(\mathbf{x}))$ , as we make explicit within the last expression in Eq. (47). In what follows, we thus refer to the basis of the Hilbert space given by  $\Psi'[\mathbf{A}]$  as the diagonal basis. To separate field directions related to gauge transformations, it is useful to write

$$\int_{\sigma} \frac{\delta^2}{\delta A^2(\sigma)} = \int_{\sigma_{\text{gauge}}} \frac{\delta^2}{\delta A^2(\sigma_{\text{gauge}})} + \int_{\sigma_{\parallel}} \frac{\delta^2}{\delta A^2(\sigma_{\parallel})}. \quad (48)$$

The directions parameterized by the  $\delta A(\sigma_{\parallel})$  correspond to field variations  $G_i^a(\mathbf{x}; \sigma_{\parallel})$  (cf. Eq. (31)) orthogonal to gauge transformations. Note that in general,  $\delta A(\sigma_{\parallel})$  is not tangent to the constrained surface  $\mathcal{A}$  as such tangents may be linear combinations of  $\delta A(\sigma_{\parallel})$  and  $\delta A(\sigma_{\text{gauge}})$ . However,  $\delta A(\sigma_{\parallel})$  can be projected onto  $\mathcal{A}$ , upon which a gauge transformation is added. As  $\Psi_{\text{g.i.}}^{(a)}[\mathbf{A}]$  is gauge independent by definition, integrating the Schrödinger equation in directions  $\delta A(\sigma_{\parallel})$  from a point  $\mathbf{A} \in \mathcal{A}$  also amounts to finding the solution at  $\mathbf{A} + \delta \mathbf{A} \in \mathcal{A}$  where  $\delta \mathbf{A}$  is  $\delta A(\sigma_{\parallel})$  projected on  $\mathcal{A}$ . It should be kept in mind here that the separation into  $\delta A(\sigma_{\text{gauge}})$  and  $\delta A(\sigma_{\parallel})$  depends on the point  $\mathbf{A} \in \mathcal{A}$ .

In the basis of Eq. (46), the functional Schrödinger equation is manifestly separable and can be solved by a product ansatz. In

$$\Psi'^{(a)}[\mathbf{A}] = \Psi_{\text{g.i.}}'^{(a)}[\mathbf{A}] p'[\mathbf{A}], \quad (49)$$

we take

$$\frac{\delta}{\delta A(\sigma_{\text{gauge}})} \Psi_{\text{g.i.}}^{(a)}[\mathbf{A}] = 0, \quad \frac{\delta}{\delta A(\sigma_{\parallel})} p'^{(a)}[\mathbf{A}] = 0 \quad (50)$$

and first solve the eigenvalue problem for the factor  $p'[\mathbf{A}]$

$$\int_{\sigma_{\text{gauge}}} \frac{\delta^2}{\delta A^2(\sigma_{\text{gauge}})} p'[\mathbf{A}] = \mu p'[\mathbf{A}], \quad (51)$$

where  $\mu$  is real due to Hermiticity. The first equality in Eq. (50) is in accordance with the general requirement (44) from gauge invariance. The Schrödinger equation corresponding to a state with energy  $E^{(a)}$  then gives

$$\left( -\frac{g^2}{2} \int_{\sigma_{\parallel}} \frac{\delta^2}{\delta A^2(\sigma_{\parallel})} + \frac{1}{2} \int d^3x \operatorname{tr} \mathbf{B}^2 \right) \Psi_{\text{g.i.}}'^{(a)} = \left( E^{(a)} + \frac{g^2 \mu}{2} \right) \Psi_{\text{g.i.}}'^{(a)}. \quad (52)$$

where we have used Eqs. (50) and (51).

The solutions to Eq. (51) however cannot be of the form  $p'[\mathbf{A}] = \exp(i\varphi[\mathbf{A}])$  assumed in Eq. (43) unless  $\varphi[\mathbf{A}] \equiv 0$ . This is because the solutions only depend on the gauge transformation and they can be written as product functions with respect to the particular directions  $\sigma_{\text{gauge}}$ . That is, they give rise to a one-dimensional representation (a ‘‘character’’) of the gauge group

$$\exp(i\varphi[\mathbf{A}_{U_3}]) = \exp(i\varphi[\mathbf{A}_{U_2}]) \exp(i\varphi[\mathbf{A}_{U_1}]), \quad \text{where } U_3 = U_2 U_1, \quad (53)$$

so that in particular  $\varphi[\mathbf{A}_{U_2 U_1} - \mathbf{A}_{U_1 U_2}] = 0$ . For a simple non-Abelian gauge group, this is only possible if  $\varphi[\mathbf{A}_U] \equiv 0$ , i.e. only trivial characters are allowed. Hence, the derivatives in the directions of particular generators of gauge transformations must be constant for solutions to Eq. (51).

Note that this argument can also be used to prove that the eigenfunctions of the Laplacian on a sphere of two or more dimensions, i.e. the spherical harmonics, cannot be pure phases over the whole sphere except for the constant solution with eigenvalue zero.

In summary, in the diagonal basis (47) for the Hamiltonian, the only permitted eigenvalue of

$$\int_{\sigma_{\text{gauge}}} \frac{\delta^2}{\delta A(\sigma_{\text{gauge}})^2} \quad (54)$$

under the condition that gauge transformations only induce pure phases on the wave functional is  $\mu = 0$ , and the phases must be constant. That is

$$\frac{\delta}{\delta A(\sigma_{\text{gauge}})} \Psi'[\mathbf{A}] = \int d^3 y \sum_{i,a} G_i^a(\mathbf{y}; \sigma_{\text{gauge}}) \frac{\delta}{\delta A_i^a(\mathbf{y})} \Psi'[\mathbf{A}] = 0. \quad (55)$$

In the diagonal basis of Eq. (46), this fixes the wave functional to be invariant under all gauge transformations that are continuously connected with the identity.

Note that in the Hilbert space of functionals with the property (55) it is possible to evaluate the action of  $H'$  (i.e. in the diagonal basis) entirely in terms of derivatives  $\delta/\delta A(\sigma)$  pointing in directions that are tangential on any chosen  $\mathcal{A}$  without changing the spectrum of the Hamiltonian by this restriction. To isolate these directions in the individual terms that add up to the Hamiltonian, the absence of the cross term between the magnetic field and the functional derivative in Eq. (47) has been necessary.

As a consequence of the above, we can state that only derivatives

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\Psi'[\mathbf{A} + \varepsilon \Delta \mathbf{A}] - \Psi'[\mathbf{A}]) \quad (56)$$

for  $\Delta \mathbf{A}$  that are tangential on  $\mathcal{A}$  may contribute to observables such as the Hamiltonian. As follows from the line of reasoning in this section, this principle is equivalent to maintaining Hermiticity of the corresponding operators in the space of physical states. We

would therefore like to see whether the principle can also be phrased in terms of some notion of tangential derivatives in the case of  $H$  in the general basis. For this purpose, define

$$\mathbf{D}_{\mathbf{A}}\Psi[\mathbf{A}] = i \left( \frac{\delta}{i\delta\mathbf{A}} - (\theta - \theta_{\Pi}) \frac{g}{8\pi^2} \mathbf{B} \right) \Psi[\mathbf{A}]. \quad (57)$$

The operator  $\mathbf{D}_{\mathbf{A}}$  is of the same form as the momentum operator (26). It therefore satisfies the commutation relations (25) with the field operator  $\mathbf{A}$  which is the appropriate algebraic property to use it as a generator of translations in  $\mathbf{A}$ . When we denote the associated translation operator as  $T[\Delta\mathbf{A}]$ , it acts on  $\Psi[\mathbf{A}]$  as

$$T[\Delta\mathbf{A}]\Psi[\mathbf{A}] = \exp \{ -i(\theta - \theta_{\Pi}) (W[\mathbf{A} + \Delta\mathbf{A}] - W[\mathbf{A}]) \} \Psi[\mathbf{A} + \Delta\mathbf{A}] \quad (58)$$

as can be seen when using Eq. (45) and considering infinitesimal translations, for which

$$\begin{aligned} T[\varepsilon\Delta\mathbf{A}]\Psi[\mathbf{A}] &\approx \exp \left( -i(\theta - \theta_{\Pi}) \frac{g}{8\pi^2} \int d^3y \varepsilon \Delta\mathbf{A} \cdot \mathbf{B} \right) \Psi[\mathbf{A} + \varepsilon\Delta\mathbf{A}] \\ &\approx \Psi[\mathbf{A}] + \int d^3y \varepsilon \Delta\mathbf{A} \cdot \mathbf{D}_{\mathbf{A}}\Psi[\mathbf{A}], \end{aligned} \quad (59)$$

or, alternatively,

$$\int d^3y \Delta\mathbf{A} \cdot \mathbf{D}_{\mathbf{A}}\Psi[\mathbf{A}] = \lim_{\varepsilon \rightarrow 0} \frac{T[\varepsilon\Delta\mathbf{A}]\Psi[\mathbf{A}] - \Psi[\mathbf{A}]}{\varepsilon}, \quad (60)$$

i.e.  $\mathbf{D}_{\mathbf{A}}$  is indeed a generator of  $T$ . Now in terms of  $\mathbf{D}_{\mathbf{A}}$  and  $\Psi$ , the condition (55) can be written as

$$\frac{\delta}{\delta A(\sigma_{\text{gauge}})} \Psi'[\mathbf{A}] = \int d^3y e^{-i(\theta - \theta_{\Pi})W[\mathbf{A}]} \sum_{i,a} G_i^a(\sigma_{\text{gauge}}, \mathbf{y}) [D_{\mathbf{A}}]_i^a \Psi[\mathbf{A}] = 0, \quad (61)$$

while the Hamiltonian can be expressed as follows,

$$H = \frac{1}{2} \int d^3x \text{tr} [-g^2 \mathbf{D}_{\mathbf{A}}^2 + \mathbf{B}^2]. \quad (62)$$

The condition of Eq. (61) can be interpreted in terms of generalized derivatives  $D/DA(\sigma_{\text{gauge}})$  in the directions of gauge transformations, with

$$\frac{D}{DA(\sigma)} \equiv \int d^3y e^{-i(\theta - \theta_{\Pi})W[\mathbf{A}]} \sum_{i,a} G_i^a(\mathbf{y}; \sigma) [D_{\mathbf{A}}]_i^a. \quad (63)$$

In terms of the wave functionals  $\Psi$ , these may only have nonvanishing derivatives for  $D/DA(\sigma_{\parallel})$  associated with  $G_i^a(\mathbf{y}; \sigma_{\parallel})$  orthogonal to the infinitesimal gauge transformations  $G_i^a(\mathbf{y}; \sigma_{\text{gauge}})$ .

Hence, provided the condition (61) or, equivalently, Eq. (55), there are only contributions to the Hamiltonian  $H$  from changes in the wave functional  $\Psi[\mathbf{A}]$  under the translations  $T[\Delta\mathbf{A}]$  for  $\Delta\mathbf{A}$  that are tangent on any given  $\mathcal{A}$ . As long as this condition is satisfied for some translation operator  $T$  and the Hilbert space is restricted correspondingly, the observables do not depend on the choice of  $\mathcal{A}$  and gauge fixing following Eq. (30) is thus possible.

## 7 Gauß' law constraint

The presence of some of the unphysical states in the spectrum of the Hamiltonian is not a problem of quantization in the first place but can be traced to the fact that the canonical equations of motion for the Hamiltonian (28) do not contain Gauß' law. As it turns out, imposing Gauß' law, one restricts the Hilbert space in such a way that the unphysical states are removed and that it allows for gauge fixing. These matters are well known for gauge theory in Minkowski spacetime [17]. Here, we shall review these points in view of what happens on the three-torus  $T^3$ . We add to the discussion the point of view that one can also argue conversely that gauge invariance implies that the physical states when constrained according to the criteria derived in Section 6 satisfy Gauß' law.

In the temporal gauge, Hamilton's equations do not yield Gauß' law. It can be enforced at the level of the quantized theory by imposing

$$(\mathbf{D} \cdot \mathbf{E})^a \Psi[\mathbf{A}] = 0 \quad (64)$$

on the physical states represented by the wave functionals  $\Psi[\mathbf{A}]$ . (Explicitly,  $(\mathbf{D} \cdot \mathbf{E})^a = \nabla \cdot \mathbf{E}^a + f^{abc} \mathbf{A}^b \mathbf{E}^c$ .) In general, there will be states that do not observe this condition. Those will be removed from the Hilbert space of physical states.

For this construction to make sense, the condition (64) must define a subspace that is invariant under the time evolution. To show this is the case, one takes the gauge transformations (16) for symmetries with pertaining Noether charges. For an infinitesimal generator  $\Omega(\mathbf{x})$  of the transformation (defined as  $U(\mathbf{x}) = e^{i\alpha\Omega(\mathbf{x})}$  with  $\Omega^\dagger = \Omega$  such that, for  $\alpha \ll 1$ ,  $A_U^i(\mathbf{x}) - A^i(\mathbf{x}) = \alpha D^i \Omega(\mathbf{x})$ ), the Noether charge is given by

$$\begin{aligned} Q(\Omega) &= \int_V d^3x \operatorname{tr} [\Pi^i (D^i \Omega)] = \frac{1}{g} \int_V d^3x \operatorname{tr} \left[ \left( -E^i + \frac{g^2}{8\pi^2} \theta B^i \right) D^i \Omega \right] \\ &= \frac{1}{g} \int_V d^3x \operatorname{tr} \left[ \Omega D^i \left( E^i - \frac{g^2}{8\pi^2} \theta B^i \right) \right] + \frac{1}{g} \int_{\partial V} da^i \operatorname{tr} \left[ \Omega \left( -E^i + \frac{g^2}{8\pi^2} \theta B^i \right) \right]. \end{aligned} \quad (65)$$

Here, we have integrated by parts,  $V$  denotes the integration volume,  $\partial V$  its boundary and  $\mathbf{a}$  a normal surface element. In the present case, we take  $V = F_4$ , which can be identified with the cube that yields  $T^3$  when its opposite faces are identified.

We shall assume for now that the surface term in Eq. (65) vanishes. This is guaranteed if  $\Omega(\mathbf{x}) = 0$  for  $\mathbf{x} \in \partial V$ . Since  $D^i E^i \Psi = 0$  for the states that observe Eq. (64) and  $D^i B^i = 0$  per Eq. (23), it follows that  $Q(\Omega) \Psi = 0$ . Since  $Q(\Omega)$  is the operator corresponding to the Noether charge associated with gauge transformations, it is a generator of the latter in the Hilbert space of wave functionals. So under gauge transformations that are connected with the identity and that are generated by some  $\Omega$  that vanishes on  $\partial V$ , the  $\Psi$  satisfying Eq. (64) are invariant. Since the  $Q(\Omega)$  are Hermitian, states with a common eigenvalue of  $Q$  define a subspace of the Hilbert space of wave functionals  $\Psi[\mathbf{A}]$ . And Furthermore, since these are Noether charges,

$$[H, Q(\Omega)] = 0 \quad (66)$$

so that these subspaces have bases made up from eigenstates of the Hamiltonian and are thus invariant under time evolution. As  $\Omega(\mathbf{x})$  is arbitrary (up to the boundary condition  $\Omega(\mathbf{x}) = 0$  on  $\partial V$ ), the states that satisfy Gauß' law (64) indeed constitute a subspace that is invariant under the Hamiltonian evolution. The requirement  $\Omega(\mathbf{x}) = 0$  on  $\partial V$  is equivalent to that  $U_4(\mathbf{x})$  is constant there.

We can also turn the argument around, i.e. instead of imposing Gauß' law to derive how the states behave under gauge transformations we can obtain Gauß' law by imposing gauge invariant observables as discussed in Sections 4 and 6. To that end, we now consider gauge transformations in directions that we label with  $\sigma_{\text{Gauß}}$  (to be understood here as  $\sigma_{\text{Gauß}} \equiv \sigma_{\text{Gauß}}[\Omega]$ ), that have  $\Omega = 0$  on  $\partial V$  and therefore give no boundary term in Eq. (65). Note that we can relate this with  $Q(\Omega)$  from Eq. (65) as

$$\frac{\delta\Psi[\mathbf{A}]}{\delta A(\sigma_{\text{Gauß}})} = -i Q(\Omega)\Psi[\mathbf{A}], \quad (67)$$

for  $\Omega$  normalized as  $D^i\Omega(\mathbf{x}) = G^i(\mathbf{x}, \sigma_{\text{Gauß}})$ .

First, Eq. (55) for  $\sigma_{\text{gauge}} = \sigma_{\text{Gauß}}$  reads

$$\frac{\delta}{\delta A(\sigma_{\text{Gauß}})}\Psi'[\mathbf{A}] = 0. \quad (68)$$

Next note that  $W[\mathbf{A}]$  is invariant under gauge transformations  $U$  connected with the identity that are constant on the surface  $\partial F_4$ ,

$$\frac{\delta}{\delta A(\sigma_{\text{Gauß}})}W[\mathbf{A}] = 0. \quad (69)$$

To see this, we apply the Gauß theorem to the surface term in Eq. (17) as

$$-\frac{i}{8\pi^2}\varepsilon_{ijk}\int_S d^3x\partial_j\text{tr}[A_k(\partial_i U)U^{-1}] = -\frac{i}{8\pi^2}\int_{\partial S} da_j\varepsilon_{ijk}\text{tr}[A_k(\partial_i U)U^{-1}] = 0. \quad (70)$$

for  $S = F_4$ . As both  $\mathbf{a}$  and  $\nabla U$  are normal vectors on  $\partial S$ , the scalar triple product is zero.

Now combining Eq. (69) with Eq. (68) and recalling Eq. (46), we obtain

$$\frac{\delta}{\delta A(\sigma_{\text{Gauß}})}\Psi[\mathbf{A}] = 0. \quad (71)$$

By Eq. (67), this also means  $Q(\Omega)\Psi[\mathbf{A}] = 0$ . As  $\Omega$  is an arbitrary function vanishing on  $\partial V$ , by Eq. (65), Gauß' law (64) is implied. This means that Gauß' law is a result of gauge invariance in non-Abelian gauge theory and cannot be violated by any means.

## 8 Gauge covariance with respect to general changes of the transition function

Now, we turn to generators  $\Omega$  that do not vanish on  $\partial V$ . Taking for  $V = F_4$ , the pertaining transformations in general change the transition function on  $\partial F_4$ . This implies that

$U_4$  now can change in a general way, i.e. it is not fixed on  $\partial F_4$ . We label these transformations with  $\sigma_{\partial F_4}$ . The discussion is in large parts similar to the one for generators vanishing on  $\partial V$  from Section 7 but there are also crucial differences.

First, we notice that we can still associate with  $\Omega$  the Noether charges  $Q(\Omega)$  in Eq. (65). In the present case, in contrast to Section 7, we can no longer assume that the surface term vanishes.

Once again, we can perceive  $\Omega$  as a generator of a gauge transformation—there is no principle forcing us to have periodic gauge potentials even on the torus. In particular, no such restriction has been imposed in the canonical quantization carried out in Section 3. Therefore, we can determine what gauge invariance tells us about the form of the wave functionals  $\Psi[\mathbf{A}]$ . Again, this can be done by imposing that one can consistently fix the gauge by choosing a manifold  $\mathcal{A}$ , that the operators corresponding to observables remain Hermitian under the inner product (30) and that the observables do not depend on the particular choice of  $\mathcal{A}$ . As worked out in Section 6, this requires that we can omit derivatives in the direction of transformations generated by  $\Omega$  from the Hamiltonian, i.e. Eq. (55).

We first carry out the transformation given by Eqs. (46) and (47). Equation (55) with  $\sigma_{\text{gauge}} \rightarrow \sigma_{\partial F_4}$  reads

$$\frac{\delta}{\delta \mathbf{A}(\sigma_{\partial F_4})} \Psi'[\mathbf{A}] = 0. \quad (72)$$

One essential difference with the discussion in Section 7 is that this relation now entails information on the behaviour of the wave functional under translations of the field by the transition function  $U_4^{(1)}$  in Eq. (36). Since the transition functions are not required to be constant on  $\partial F_4$ , the same holds true for the allowed gauge transformations. Crucially, with gauge transformations unrestricted, and with general transformations in a simple gauge group being continuously connected with the identity, the usual large gauge transformations that are constant on  $\partial F_4$  and change  $W[\mathbf{A}]$  by an integer are therefore connected with the identity, which allows to relate these with the infinitesimal generators  $\Omega$ .

Any transition function  $U_4^{(1)}$  that changes  $W[\mathbf{A}]$  by one unit can therefore be composed from a sequence of infinitesimal transformations  $\mathbb{1} + i\Omega_K$ ,  $K = 1, 2, \dots$ . The  $\Omega_K$  generally do not vanish on  $\partial V = \partial F_4$  even though  $U_4^{(1)}$  does. In fact, it is essential that some  $\Omega_K$  do not vanish on the boundary because otherwise they could not compose a transition function that changes  $W[\mathbf{A}]$ . But then the invariance of Eq. (72) of the wave functional under infinitesimal transitions implies that it is also invariant under large gauge transformations,

$$\Psi'[\mathbf{A}_{(U_4^{(1)})^n}] = \Psi'[\mathbf{A}] \quad \text{where } n \in \mathbb{Z}. \quad (73)$$

From the above, we conclude that in the primed basis the wave functional is gauge invariant,

$$\Psi'[\mathbf{A}] = \Psi'_{\text{g.i.}}[\mathbf{A}]. \quad (74)$$

Another essential difference with the case of transformations that vanish on  $\partial V = \partial F_4$  is now that there is no relation corresponding to Eq. (69) when taking  $\sigma_{\partial F_4}$  instead of  $\sigma_{\text{Gau\ss}}$ . Hence, the invariance (72) does not assure that  $\delta\Psi(\mathbf{A})/\delta\mathbf{A}(\sigma_{\partial F_4}) = 0$ . In fact, since  $U_4^{(1)}$  adds one unit to  $W[\mathbf{A}]$ , from Eq. (46) we can conclude that

$$\Psi[\mathbf{A}_{(U_4^{(1)})^n}] = e^{in(\theta - \theta_{\Pi})}\Psi[\mathbf{A}]. \quad (75)$$

Hence, in the original basis the phase acquired by the states under large gauge transformations is tied to the  $CP$ -odd parameters in the Hamiltonian. Using Eqs. (46) and (47), this phase can be simultaneously removed from the Hamiltonian and the wave functional so that there remains no  $CP$ -odd parameter.

We have thus seen how gauge invariant observables determine the eigenvalue in Eq. (36). This leads to a restriction of the Hilbert space that now consists of normalizable states under the inner product (30), in contrast to Eq. (38). To demonstrate this necessary restriction of the Hilbert space we have used the fact that the transition functions (and hence the gauge potentials) are not required to be constant on  $\partial F_4$ .

Equation (75) does not depend on whether the Euclidean time is compact or not. Therefore, it applies for the spectrum of QCD on a three-torus at both zero and finite temperatures. Of course, at zero temperature, one could deduce the  $CP$  conservation from the perspective given in Refs. [4, 10].

It has been noted in Ref. [20] that gauge invariance of the wave function corresponding to Eq. (74) can explain the absence of parity violation. However, the gauge transformations in that work are postulated and not derived.

Since the Hamiltonian  $H'$  as in Eq. (47) is equivalent to an ordinary Hamiltonian without a  $\theta$ -term, we can therefore conclude that the partition function corresponding to finite temperature field theory is given by

$$Z = \sum_{\Delta n=-\infty}^{\infty} \int \mathcal{D}A_{\Delta n} e^{-\frac{1}{2g^2} \int_{T^4} d^4x \text{tr} F_{\mu\nu} F_{\mu\nu} - S_{\text{E,g.f.}} - S_{\text{E,ghost}}}, \quad (76)$$

where  $S_{\text{E,g.f.}}$ ,  $S_{\text{E,ghost}}$  are the gauge-fixing and ghost contributions and an additional integral over the ghost fields should be understood. The derivation of the partition function from the canonically quantized theory is outlined in Appendix A. Here, the phase  $\theta$  from the Lagrangian (1) does not appear, having been cancelled by the correlated phases of the wave functionals of the states in the physical Hilbert space, leading to a partition function that is manifestly  $CP$  conserving. This is in contrast to the partition functions usually implemented in lattice studies, such as in Refs. [21–25], where  $CP$ -violation is sought after or found by calculating the path integral in finite volume and weighting the individual topological sectors with  $\exp(-i\theta)$ . According to the present analysis, such setups do not correspond to an evaluation of the trace of the canonical density matrix over the states in canonically quantized QCD.



## 9 Generalization to Minkowski spacetime

Throughout the previous sections of this article we have concentrated on the case of the torus in order to illustrate how the conclusion about  $CP$  conservation in QCD is not necessarily restricted to an infinite spacetime or zero temperature. In contrast to this, in Ref. [4], the lack of  $CP$  violation in QCD in Minkowski spacetime at zero temperature is derived from a path integral formalism, in which the infinite spacetime limit is tied to the necessity of guaranteeing that the partition function of QCD is evaluated at the correct ground state, which requires sending the time interval to infinity.

Estimating the dependence of the vacuum partition function  $\langle 0|e^{-iHT}|0\rangle$  on the  $CP$ -odd parameters without taking  $T \rightarrow \infty$  requires knowledge of the phase of the vacuum wave functional [5], which becomes possible with the results of this work, enabling an independent derivation of the results of Ref. [4].

The canonical quantization in Minkowski spacetime goes along the same lines as on the torus. As before, one can define a physical Hilbert space by restricting the inner product to a gauge-fixed surface  $\mathcal{A}$ . By demanding independence of the observables with respect to the choice of  $\mathcal{A}$ , one concludes that the wave functionals are as in Eq. (46),

$$\Psi[\mathbf{A}] = e^{i(\theta - \theta_{\pi})W[\mathbf{A}]} \Psi'[\mathbf{A}], \quad (77)$$

with  $\Psi'[\mathbf{A}]$  being invariant under gauge transformations that are continuously connected to the identity,

$$\frac{\delta}{\delta A(\sigma_{\text{gauge}})} \Psi'[\mathbf{A}] = 0. \quad (78)$$

The crucial observation is that, since gauge transformations are a redundancy, one should ensure invariance of physical observables with respect to arbitrary gauge transformations, without restrictions. This is in contrast to common assumptions [1, 17, 26] which restrict gauge transformations in the gauge  $A^0 = 0$  to those satisfying

$$U(|\mathbf{x}|) \rightarrow \text{constant} \quad \text{for} \quad |\mathbf{x}| \rightarrow \infty. \quad (79)$$

Under the latter restriction, the functional  $W[\mathbf{A}]$  in Minkowski spacetime changes by integer values under large gauge transformations  $(U^{(1)})^n$  that cannot be connected with the identity, which would then not fall under the constraint of Eq. (78), thus allowing for  $CP$ -odd phases associated with the transformation properties of  $\Psi'$  under the action of  $U^{(1)}$ . As mentioned before, Eq. (79) is only a hypothesis that is discussed e.g. in the classic review [17]. Furthermore, the constraint (78) has not been imposed on the derivation of the canonical momentum (26). Therefore, we cannot exclude transformations not satisfying Eq. (78) in the canonical formalism prior to eventually fixing the gauge by choosing the manifold  $\mathcal{A}$ . Motivated by the analysis for the torus case, we therefore lift the restriction of Eq. (79). Once this is done, all gauge transformations can be connected to the identity, and demanding Eq. (78) leads to the same conclusion as for the torus, namely

$$\Psi'[\mathbf{A}] = \Psi'_{\text{g.i.}}[\mathbf{A}] \Rightarrow \Psi[\mathbf{A}] = e^{i(\theta - \theta_{\pi})W[\mathbf{A}]} \Psi'_{\text{g.i.}}[\mathbf{A}]. \quad (80)$$

Again, there is no  $CP$  violation, as is apparent when working in the primed basis in which the states are invariant under gauge transformations and the Hamiltonian does not depend on  $\theta$ . When working in a general basis, the  $\theta$  dependence of the Hamiltonian is exactly correlated with the phases of the physical states, leading to cancellations in observables.

To make contact with the path integral results of Ref. [4], we can proceed as in the case of the torus and estimate the vacuum partition function at zero temperature as

$$\begin{aligned} \langle 0|e^{-iHT}|0\rangle &= \int \mathcal{D}\mathbf{A}_1(\mathbf{x}) \int \mathcal{D}\mathbf{A}_2(\mathbf{x}) \Psi^{(0)*}[\mathbf{A}_2] \langle \mathbf{A}_2(\mathbf{x})|e^{-iHT}|\mathbf{A}_1(\mathbf{x})\rangle \Psi^{(0)}[\mathbf{A}_1] \\ &= \int \mathcal{D}\mathbf{A}_1(\mathbf{x}) \int \mathcal{D}\mathbf{A}_2(\mathbf{x}) \Psi_{\text{g.i.}}^{\prime(0)*}[\mathbf{A}_2] \langle \mathbf{A}_2(\mathbf{x})|e^{-iH'T}|\mathbf{A}_1(\mathbf{x})\rangle \Psi_{\text{g.i.}}^{\prime(0)}[\mathbf{A}_1], \end{aligned} \tag{81}$$

where the superscript (0) indicates the ground state. The result is a path integral with  $\theta$  omitted from the Lagrangian, and with boundary conditions weighted by gauge-invariant vacuum wave functionals, which are  $\theta$ -independent as they solve Schrödinger equations with the Hamiltonian  $H'$ . In Ref. [4], the vacuum partition function was estimated without the need to know the wave functionals  $\Psi_0[\mathbf{A}]$  by leaving the boundary conditions free and setting  $T \rightarrow \infty$ , which eliminates the contamination of excited states for an amplitude of the form  $\langle \mathbf{A}_2|e^{-iHT}|\mathbf{A}_1\rangle$  with arbitrary  $\mathbf{A}_{1,2}$ . This procedure should give a result equivalent to Eq. (81) up to a normalization factor, and indeed the  $\theta$ -dependence of the resulting partition function is shown to be an unphysical global factor which drops out of observables. Hence, the results of this article show that the limiting procedure of Ref. [4], which is justified purely from the point of view of the path integral, is indeed projecting into the correct physical vacuum state, whose wave functional has a phase correlated with the  $\theta$ -term in the Hamiltonian, leading to a cancellation of the  $\theta$ -dependence in physical observables.

## 10 The hidden parity symmetry of QCD in the presence of a $\theta$ -term

Given the previous results about  $CP$  conservation in the strong interactions, it is natural to wonder whether there is an explanation in terms of a symmetry of the theory, which differs from the usual parity transformation

$$P : \quad \mathbf{x} \rightarrow -\mathbf{x}, \quad \mathbf{A}(\mathbf{x}) \rightarrow \mathbf{A}^P(\mathbf{x}) = -\mathbf{A}(-\mathbf{x}). \tag{82}$$

It turns out that such symmetry does exist. The Hamiltonian of Eq. (28) commutes with a modified parity operator  $\hat{P}$  with  $\hat{P}^2$  being equal to a phase. From this symmetry, we can deduce that energy eigenstates of the form (75), upon which the gauge-fixed Hamiltonian acts as a Hermitian operator, can be found to be also eigenstates of the usual parity  $P$ .

To construct  $\hat{P}$ , we start by considering large transformations  $U^{(1)}$  which change  $W[\mathbf{A}]$  by one unit, either on the three-torus or on a hypersurface of constant time in Minkowski spacetime,

$$W[U^{(1)}\mathbf{A}] = W[\mathbf{A}] + 1. \quad (83)$$

Note that in this section, we use a shorthand notation different from Eq. (16), i.e.

$$U\mathbf{A} := U\mathbf{A}U^{-1} + iU\nabla U^{-1}, \quad (84)$$

which we find here more suitable to display the order of the parity and gauge transformations. Under the parity transformations (82), one has

$$W[\mathbf{A}^P] = -W[\mathbf{A}], \quad (85)$$

as follows from Eq. (15). Then one has

$$\begin{aligned} W[U^{(1)}(U^{(1)}\mathbf{A}^P)^P] &= W[(U^{(1)}\mathbf{A}^P)^P] + 1 = -W[U^{(1)}\mathbf{A}^P] + 1 \\ &= -(W[\mathbf{A}^P] + 1) + 1 = W[\mathbf{A}]. \end{aligned} \quad (86)$$

The functional  $W[\mathbf{A}]$  is invariant under gauge transformations  $U^{(0)}$  that are continuously connected with the identity and that go to  $\mathbb{1}$  at the boundary. By ‘‘boundary’’ here we refer either to  $\partial F_4$  on the torus, or the region  $|\mathbf{x}| \rightarrow \infty$  on a spatial hypersurface in Minkowski spacetime. With  $W[\mathbf{A}] = W[U^{(1)}(U^{(1)}\mathbf{A}^P)^P]$  as in Eq. (86), it follows that for a generic  $U^{(1)}$ ,  $\mathbf{A}$  and  $U^{(1)}(U^{(1)}\mathbf{A}^P)^P$  are related to each other by a gauge transformation connected with the identity. As in the class of large gauge transformations  $U^{(1)}$  there is a freedom to multiply by transformations  $U^{(0)}$ , we can always find a representative  $\tilde{U}^{(1)}$  such that

$$\tilde{U}^{(1)}(\tilde{U}^{(1)}\mathbf{A}^P)^P = \mathbf{A}. \quad (87)$$

With this definition, one can consider the following operator on wave functionals,

$$\hat{P}\Psi[\mathbf{A}] = e^{i(\theta-\theta_\pi)(2W[\mathbf{A}]-1)}\Psi[\tilde{U}^{(1)}\mathbf{A}^P]. \quad (88)$$

The transformation  $\hat{P}$  can be seen to be physically equivalent to an ordinary parity transformation, where the parity transformed gauge field  $\mathbf{A}^P$  is gauge transformed by  $\tilde{U}^{(1)}$ . This additional gauge transformation should not affect physical observables. The relevance of  $\hat{P}$  is that it commutes with the Hamiltonian. Indeed, starting from Eqs. (28)

and (88) one can write

$$\begin{aligned}
& H(\hat{P}\Psi[\mathbf{A}]) \\
&= \int d^3x \frac{1}{2} \left[ \left( g \frac{\delta}{i\delta\mathbf{A}^a} - \frac{g^2}{8\pi^2} (\theta - \theta_\Pi) \mathbf{B}^a \right)^2 + (\mathbf{B}^a)^2 \right] e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} \Psi[\tilde{U}^{(1)} \mathbf{A}^P] \\
&= \int d^3x e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} \frac{1}{2} \left[ \left( g \frac{\delta}{i\delta\mathbf{A}^a} + \frac{g^2}{8\pi^2} (\theta - \theta_\Pi) \mathbf{B}^a \right)^2 + (\mathbf{B}^a)^2 \right] \Psi[\tilde{U}^{(1)} \mathbf{A}^P] \quad (89) \\
&= \int d^3x e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} \mathcal{H}[\mathbf{A}^P] \Psi[\tilde{U}^{(1)} \mathbf{A}^P] = e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} (H\Psi)[\tilde{U}^{(1)} \mathbf{A}^P] \\
&= \hat{P}(H\Psi[\mathbf{A}]).
\end{aligned}$$

In the equation above, when going from the second to the third line we made use of Eq. (45), while, going to the fourth line, we used parity invariance of the magnetic field and, in the fourth line, gauge invariance of the Hamiltonian density,  $\mathcal{H}[\mathbf{A}]\Psi[\mathbf{A}_{\tilde{U}^{(1)}}] = \mathcal{H}[\mathbf{A}_{\tilde{U}^{(1)}}]\Psi[\mathbf{A}_{\tilde{U}^{(1)}}]$ .

The final property of  $\hat{P}$  to be stated here is that  $\hat{P}^2$  evaluates to a constant phase, again reflecting that under  $\hat{P}^2$  one should get a state which is physically equivalent to the original one. To see this, we use successive applications of the definition of Eq. (88), leading to

$$\begin{aligned}
\hat{P}^2\Psi[\mathbf{A}] &= \hat{P}(e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} \Psi[\tilde{U}^{(1)} \mathbf{A}^P]) \\
&= e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} e^{i(\theta - \theta_\Pi)(2W[\tilde{U}^{(1)} \mathbf{A}^P] - 1)} \Psi[\tilde{U}^{(1)}(\tilde{U}^{(1)} \mathbf{A}^P)^P]. \quad (90)
\end{aligned}$$

With  $W[\tilde{U}^{(1)} \mathbf{A}^P] = 1 + W[\mathbf{A}^P] = 1 - W[\mathbf{A}]$ , and using Eq. (87), one gets

$$\hat{P}^2\Psi[\mathbf{A}] = \Psi[\mathbf{A}]. \quad (91)$$

Given Eq. (91), the allowed eigenvalues of  $\hat{P}$  are  $\pm 1$ , similar to an ordinary parity transformation. A consequence of  $[\hat{P}, H] = 0$  is that the  $\hat{P}, H$  can be diagonalized simultaneously. Hence, an eigenstate of  $H$  can always be written as a sum of eigenstates of  $\hat{P}$ .

Let us now derive the representation of  $\hat{P}$  in the primed basis. From Eq. (77), one has

$$\begin{aligned}
\hat{P}\Psi[\mathbf{A}] &= \hat{P}e^{i(\theta - \theta_\Pi)W[\mathbf{A}]} \Psi'[\mathbf{A}] = e^{i(\theta - \theta_\Pi)(2W[\mathbf{A}] - 1)} e^{i(\theta - \theta_\Pi)W[\tilde{U}^{(1)} \mathbf{A}^P]} \Psi'[\tilde{U}^{(1)} \mathbf{A}^P] \\
&= e^{i(\theta - \theta_\Pi)W[\mathbf{A}]} \Psi'[\tilde{U}^{(1)} \mathbf{A}^P]. \quad (92)
\end{aligned}$$

Since in the primed basis,

$$\hat{P}' = e^{-i(\theta - \theta_\Pi)W[\mathbf{A}]} \hat{P} e^{i(\theta - \theta_\Pi)W[\mathbf{A}]} \quad (93)$$

one then obtains

$$\hat{P}'\Psi'[\mathbf{A}] = \Psi'[\tilde{U}^{(1)}\mathbf{A}^P]. \quad (94)$$

From Eqs. (88) and (94), it follows that ordinary parity transformations can be written as follows:

$$\begin{aligned} \Psi[\mathbf{A}^P] &= e^{-i(\theta-\theta_\pi)(1+2W[\mathbf{A}])}(\hat{P}\Psi)[(\tilde{U}^{(1)P})^{-1}\mathbf{A}], \\ \Psi'[\mathbf{A}^P] &= (\hat{P}'\Psi')[(\tilde{U}^{(1)P})^{-1}\mathbf{A}]. \end{aligned} \quad (95)$$

Above, we defined  $U^{(1)P}(\mathbf{x}) \equiv U^{(1)}(-\mathbf{x})$ , while in the derivation we made use of the fact that Eqs. (17) and (14) imply

$$W[(\tilde{U}^{(1)P})^{-1}\mathbf{A}] = W[\mathbf{A}] + 1. \quad (96)$$

Equation 95 has a profound implication: ordinary parity is equivalent to the  $\hat{P}$  symmetry of the Hamiltonian, up to non-dynamical phases and gauge transformations. The issue of whether parity is conserved or not is then tied to how gauge transformations act on the physical Hilbert space, as argued next. First, as gauge transformations commute with the Hamiltonian, while at the same time they are a redundancy that should not affect physical observables, one expects that physical states should be simultaneous eigenstates of the Hamiltonian and gauge transformations, with the eigenvalues under the latter being pure phases. In this physical basis, it is not guaranteed that  $\hat{P}$  is diagonal: while  $\hat{P}$  commutes with  $H$ , it does not necessarily commute with gauge transformations. Hence, a logical possibility is that physical states are not eigenvectors of  $\hat{P}$ , and in that case Eq. (95) would imply that the states would not have a well-defined quantum number under parity, leading to apparent parity violation. On the other hand, if  $\hat{P}$  were to commute with gauge transformations, physical states would be simultaneous eigenstates of  $H$ , gauge transformations and  $\hat{P}$ , and then from Eq. (95) it would follow that the states would be eigenstates under ordinary parity, with the latter being conserved.

Consider then a physical state in the primed basis, assumed to transform with a phase under gauge transformations,

$$\Psi'[U\mathbf{A}] = e^{i\theta_\Psi[U]}\Psi'[\mathbf{A}]. \quad (97)$$

To simplify the notation, we shall define ordinary parity and local gauge transformation operators acting on wave functions in the primed basis as follows,

$$P\Psi'[\mathbf{A}] \equiv \Psi'[\mathbf{A}^P], \quad U\Psi'[\mathbf{A}] \equiv \Psi'[U\mathbf{A}]. \quad (98)$$

By considering the parity transform of Eq. (97), one has

$$PU\Psi'[\mathbf{A}] = P\Psi'[U\mathbf{A}] = \Psi'[U^P\mathbf{A}^P] = e^{i\theta_\Psi[U^P]}\Psi'[\mathbf{A}^P] = U^P\Psi'[\mathbf{A}^P] = U^PP\Psi'[\mathbf{A}], \quad (99)$$

which implies

$$PU = U^P P. \quad (100)$$

With the notation of Eq. (98), we can write the operator  $\hat{P}'$  in Eq. (94) as

$$\hat{P}' = \tilde{U}^{(1)} P. \quad (101)$$

From the latter result, as well as from Eqs. (100), (98) and (97), it follows that

$$\begin{aligned} \hat{P}' U \Psi'[\mathbf{A}] &= \tilde{U}^{(1)} U^P P \Psi'[\mathbf{A}] = e^{i(\theta_\Psi[U^P] + \theta_\Psi[\tilde{U}^{(1)}])} \Psi'[\mathbf{A}^P], \\ U \hat{P}' \Psi'[\mathbf{A}] &= U \tilde{U}^{(1)} P \Psi'[\mathbf{A}] = e^{i(\theta_\Psi[U] + \theta_\Psi[\tilde{U}^{(1)}])} \Psi'[\mathbf{A}^P]. \end{aligned}$$

From the previous identities, one can conclude that for generic phases  $\theta_\Psi[U] \neq \theta_\Psi[U^P]$ ,  $[\hat{P}', U] \neq 0$ , and parity is not conserved. However, Eq. (97) implies that the physical states are in a one-dimensional representation of the gauge group. As mentioned in Section 6, if the gauge group is simple and non-Abelian as in QCD, only trivial representations are allowed, so that  $\theta_\Psi[U] = 0$ ,  $[\hat{P}', U] = 0$  and states have a well-defined, conserved parity. While the discussion above assumed on reasonable grounds that quantization allows for Eq. (97), the results of Section 6 provide explicit proof that this is the case. As we have seen, the constraints that remove the gauge redundancy in the Hilbert space amount to  $\Psi'$  being invariant under all gauge transformations, in particular under  $\tilde{U}^{(1)}$ . Then Eq. (95) leads to

$$\begin{aligned} \Psi'[\mathbf{A}^P] &= \pm \Psi'[\mathbf{A}] \quad \text{for physical states, diagonal basis,} \\ \Psi[\mathbf{A}^P] &= \pm e^{-2i(\theta - \theta_\pi)W[\mathbf{A}]} \Psi[\mathbf{A}] \quad \text{for physical states, general basis,} \end{aligned} \quad (102)$$

where we have used  $\hat{P}^2 = \hat{P}'^2 = 1$  and thus  $\hat{P}\Psi = \pm\Psi$ ,  $\hat{P}'\Psi' = \pm\Psi'$ , the invariance of  $\Psi'$  under gauge transformations and Eq. (96). That is, in both the diagonal and the general basis, up to a phase the states are eigenvectors of parity, which thus remains conserved under time evolution. In the primed diagonal basis one recovers the usual  $\pm 1$  eigenvalues of  $P$ , while in the general basis, the states acquire an unobservable,  $\theta$ -dependent phase.

To conclude this section, let us discuss how this parity symmetry can be seen from the path integral formulation once the projection to physical states has been correctly taken into account. To that end, we recall that a proper path integration should weigh the boundary conditions with the vacuum wave functionals, as in Eqs. (111) and (81). While a parity transformation of  $\mathbf{A} \rightarrow \mathbf{A}^P$  is not a symmetry of the Lagrangian, so that the integrand  $\exp(iS)$  acquires a phase, this is compensated by the phases acquired by the wave functionals under parity, as in Eq. (102). Consider for example the partition function of Eq. (81), where we replace the transition amplitude with an appropriate gauge-fixed path integration with boundary conditions fixed by  $\mathbf{A}_1$  and  $\mathbf{A}_2$ ,

$$\langle 0 | e^{-iHT} | 0 \rangle = \int \mathcal{D}\mathbf{A}_1(\mathbf{x}) \int \mathcal{D}\mathbf{A}_2(\mathbf{x}) \int_{\mathbf{A}_1}^{\mathbf{A}_2} \mathcal{D}\mathbf{A}(t, \mathbf{x}) \Psi^{(0)*}[\mathbf{A}_2] \Psi^{(0)}[\mathbf{A}_1] e^{i(S[\mathbf{A}] + S_{\text{g.f.}} + S_{\text{ghost}})}, \quad (103)$$

where we have suppressed the notation of the functional integral over the ghost fields. The  $\theta$ -term in the action  $S$  has to be correlated with the  $CP$ -odd quantity  $\theta - \theta_{\Pi}$  appearing in the Hamiltonian of Eq. (28). In the following, we choose  $\theta_{\Pi} = 0$ , which leads to the usual path integral with

$$S \supset S_{\theta} = \frac{\theta}{16\pi^2} \text{tr} \int d^4x F_{\mu\nu} \tilde{F}^{\mu\nu}. \quad (104)$$

Choosing a gauge  $A^0 = 0$  and with  $\mathbf{A} \rightarrow 0$  for  $|\mathbf{x}| \rightarrow \infty$ ,  $S_{\theta}$  becomes

$$S_{\theta}[\mathbf{A}] = \theta \Delta n = \theta(W[\mathbf{A}_2] - W[\mathbf{A}_1]). \quad (105)$$

Under parity,  $W[\mathbf{A}]$  changes sign, so that

$$e^{iS[\mathbf{A}^P]} = e^{-2i\theta(W[\mathbf{A}_2] - W[\mathbf{A}_1])} e^{iS[\mathbf{A}^P]}. \quad (106)$$

The extra phase in  $e^{iS[\mathbf{A}^P]}$  is exactly cancelled by the rephasings of the wave functionals in Eq. (103) under parity transformations, as follows from Eq. (102) with  $\theta_{\Pi} = 0$ . Hence the partition function is invariant under parity, regardless of  $\theta$ . When one estimates the partition function without knowledge of  $\Psi^{(0)}[\mathbf{A}]$  by sending  $T \rightarrow \infty$  and keeping boundary conditions free, as in Ref. [4], the partition function should be invariant under parity up to a change of the unphysical normalization factor, which is indeed the case for the results of Ref. [4]. Again, the limiting procedure of that reference can be justified by requiring consistency with the results of the canonical formalism.

## 11 Conclusions

On  $T^3$ , The existence of eigenstates of the Hamiltonian of the Bloch form (36) relies on fixing the gauge so that the transition function  $U_4 = \mathbb{1}$  on  $\partial F_4 = \partial W_3$ . This then implies that the vector potentials  $\mathbf{A}$  are periodic on  $T^3$ . As this is not a general gauge, we have to check that the conclusions drawn from fixing the gauge in this way are gauge invariant and correspond to expectation values of Hermitian operators. These considerations reveal that the phase  $\theta^{(i)}$  associated with the periodicity of the Bloch state is pinned in such a way that it cancels the phase  $\theta - \theta_{\Pi}$  in the Hamiltonian. Furthermore, we can use an inner product that does not extend over gauge redundant configurations. As a consequence, the finite temperature field theory for the Lagrangian (1) does not yield  $CP$ -odd correlation functions.

This constraint on  $\theta^{(i)}$  also determines the Hilbert space of physical wave functionals. Within this space, the absence of  $CP$  violation can be understood in terms of an exact symmetry of the QCD Hamiltonian which exists for arbitrary values of  $\theta$  and can be viewed as a deformation of the ordinary parity symmetry by additional unphysical phases and gauge transformations.

We have focused here on the three-torus, in order to avoid a discussion of boundary conditions that one has to carry out at spatial infinity in Minkowski spacetime. If one

does not adapt the common assumption that the relevant pure gauge configurations at infinity must satisfy  $U(|\mathbf{x}|) \rightarrow \text{constant}$  for  $|\mathbf{x}| \rightarrow \infty$ , the reasoning for the torus applies also to Minkowski spacetime, which is complementary to the findings for path integrals in infinite volume [4]. It would also be interesting to relate the present conclusions to spacetimes with a cosmological horizon, as discussed in Refs. [9, 27].

In either case, Minkowski spacetime or three-torus, and in contrast to the usual  $\theta$ -vacua, the Hilbert space of physical states is by construction restricted in such a way that these are properly normalizable. It thus turns out that the topological configurations in QCD pose no obstacle to a straightforward quantum-mechanical interpretation of squared amplitudes as probabilities.

## Acknowledgments

The work of WYA is supported by the UK Engineering and Physical Sciences Research Council (EPSRC), under Research Grant No. EP/V002821/1. C.T. acknowledges support by the Cluster of Excellence “Precision Physics, Fundamental Interactions, and Structure of Matter” (PRISMA<sup>+</sup> EXC 2118/1) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the German Excellence Strategy (Project No. 390831469).



## A Path integral over the gauge-fixed field configuration space

In this appendix, we derive the path integral formulation for the partition function (3) on  $T^4$  in the constrained configuration space  $\mathcal{A}$ . The partition function is defined as a trace over the physical Hilbert space,

$$\begin{aligned} Z &= \sum_a \langle \Psi^{(a)} | e^{-\beta H} | \Psi^{(a)} \rangle \\ &= \sum_n \int \mathcal{D}\mathbf{A}_1(\mathbf{x}) \int \mathcal{D}\mathbf{A}_2(\mathbf{x}) \langle \Psi^{(a)} | \mathbf{A}_1(\mathbf{x}) \rangle \langle \mathbf{A}_1(\mathbf{x}) | e^{-\beta H} | \mathbf{A}_2(\mathbf{x}) \rangle \langle \mathbf{A}_2(\mathbf{x}) | \Psi^{(a)} \rangle. \end{aligned} \quad (107)$$

Above,  $|\Psi^{(a)}\rangle$  represent the physical states in the constrained Hilbert space, and we have inserted the spectral resolution of the identity in the full unconstrained space in terms of projectors into eigenvectors  $|\mathbf{A}(\mathbf{x})\rangle$  of the field operators in the Heisenberg picture,

$$\mathbb{1}_{\text{phys}} = \int \mathcal{D}\mathbf{A}(\mathbf{x}) |\mathbf{A}(\mathbf{x})\rangle \langle \mathbf{A}(\mathbf{x})| P_{\text{phys}}. \quad (108)$$

In the previous equation,  $P_{\text{phys}}$  is a projector into the space of physical states  $|\Psi^{(a)}\rangle$ . In particular,  $P_{\text{phys}}|\Psi^{(a)}\rangle = |\Psi^{(a)}\rangle$ , which allows us to drop the projectors in Eq. (107). The inner product  $\langle \bullet | \bullet \rangle$  is defined over the full unconstrained field space, and so it also integrates over all gauge-equivalent configurations. We can identify the inner products with our wave-functionals,

$$\langle \mathbf{A}(\mathbf{x}) | \Psi^{(a)} \rangle = \Psi^{(a)}[\mathbf{A}] = e^{i(\theta - \theta_{\Pi})W[\mathbf{A}]} \Psi'^{(a)}[\mathbf{A}] = e^{i(\theta - \theta_{\Pi})W[\mathbf{A}]} \Psi'_{\text{g.i.}}{}^{(a)}[\mathbf{A}], \quad (109)$$

where we have used Eqs. (46) and (74). From Eq. (109), and using

$$\begin{aligned} &e^{-i(\theta - \theta_{\Pi})W[\mathbf{A}_1]} \langle \mathbf{A}_1(\mathbf{x}) | e^{-\beta H} | \mathbf{A}_2(\mathbf{x}) \rangle e^{i(\theta - \theta_{\Pi})W[\mathbf{A}_2]} \\ &= \langle \mathbf{A}_1(\mathbf{x}) | e^{-i(\theta - \theta_{\Pi})W[\mathbf{A}]} e^{-\beta H} e^{i(\theta - \theta_{\Pi})W[\mathbf{A}]} | \mathbf{A}_2(\mathbf{x}) \rangle = \langle \mathbf{A}_1(\mathbf{x}) | e^{-\beta H'} | \mathbf{A}_2(\mathbf{x}) \rangle \end{aligned} \quad (110)$$

we obtain

$$Z = \sum_a \int \mathcal{D}\mathbf{A}_1(\mathbf{x}) \int \mathcal{D}\mathbf{A}_2(\mathbf{x}) \Psi'_{\text{g.i.}}{}^{(a)*}[\mathbf{A}_1] \Psi'_{\text{g.i.}}{}^{(a)}[\mathbf{A}_2] \langle \mathbf{A}_1(\mathbf{x}) | e^{-\beta H'} | \mathbf{A}_2(\mathbf{x}) \rangle. \quad (111)$$

One can recognize  $\sum_a \Psi'_{\text{g.i.}}{}^{(a)*}[\mathbf{A}_1] \Psi'_{\text{g.i.}}{}^{(a)}[\mathbf{A}_2]$  as a projector into the physical eigenstates in the diagonal basis,

$$\sum_a \Psi'_{\text{g.i.}}{}^{(a)*}[\mathbf{A}_1] \Psi'_{\text{g.i.}}{}^{(a)}[\mathbf{A}_2] = P'_{\text{phys}}(\mathbf{A}_1, \mathbf{A}_2). \quad (112)$$

As seen in Section 6, in the diagonal basis the physical eigenstates satisfy Eq. (55), i.e. they are gauge invariant. And as seen in Section 7, the invariance under gauge

transformations with parameters  $\Omega(\mathbf{x})$  that go to zero at the boundary can be enforced by imposing Gauß' law on the physical states. Therefore, one can take  $P'_{\text{phys}}$  to be proportional to a projector  $P_{\text{Gauß}}$  enforcing Gauß' law. Further, when considering an inner product over the unconstrained field space, one should reflect the fact that field configurations related by gauge transformations are physically equivalent. Hence, one may take

$$P'_{\text{phys}}(\mathbf{A}, \mathbf{A}') \propto \int \mathcal{D}U(\mathbf{x}) \delta(\mathbf{A}_U - \mathbf{A}') P_{\text{Gauß}}(\mathbf{A}, \mathbf{A}') \equiv P_U(\mathbf{A}, \mathbf{A}') P_{\text{Gauß}}(\mathbf{A}, \mathbf{A}'). \quad (113)$$

With this choice, the partition function in Eq. (111) becomes

$$Z \propto \int \mathcal{D}U(\mathbf{x}) \mathcal{D}\mathbf{A}(\mathbf{x}) P_{\text{Gauß}}(\mathbf{A}, \mathbf{A}) \langle \mathbf{A}(\mathbf{x}) | e^{-\beta H'} | \mathbf{A}_U(\mathbf{x}) \rangle. \quad (114)$$

The amplitude  $\langle \mathbf{A}(\mathbf{x}) | e^{-\beta H'} | \mathbf{A}_U(\mathbf{x}) \rangle$  corresponds to a path integration —with paths going from  $\mathbf{A}_U(\mathbf{x})$  at  $\tau = 0$  to  $\mathbf{A}(\mathbf{x})$  at  $\tau = \beta$ — by following the standard procedure of inserting at each point in time spectral resolutions of the identity in the unconstrained field space in terms of eigenstates of the field operators. To compensate for using the unconstrained field space, one must insert  $P'_{\text{phys}}$  in each amplitude. As seen explicitly in Ref. [11], the factors of  $P_{\text{Gauß}}$  can be expressed in terms of integrations over field components  $A_0$ , with the result that the action appearing in the path integration is promoted from the  $A_0 = 0$  gauge-fixed result to the standard action, which is invariant under space-time dependent gauge transformations. Due to this emerging gauge-invariance under arbitrary transformations, and due to the fact that given a space-dependent  $U(\mathbf{x})$  one can always find a space-time dependent  $\tilde{U}(t, \mathbf{x})$  that interpolates between  $U^{-1}(\mathbf{x})$  at an initial time and the identity at a final time, one has

$$P'_{\text{phys}} \langle \mathbf{A}(\mathbf{x}) | e^{-\Delta\tau H'} | \mathbf{A}'_U(\mathbf{x}) \rangle = P'_{\text{phys}} \langle \mathbf{A}(\mathbf{x}) | e^{-\Delta\tau H'} | \mathbf{A}'(\mathbf{x}) \rangle. \quad (115)$$

This renders the gauge group integrations trivial, so that up to gauge group volume factors the result is

$$\begin{aligned} Z &\propto \int \mathcal{D}\mathbf{A}(\mathbf{x}) P_{\text{Gauß}}(\mathbf{A}, \mathbf{A}) \int_{\mathbf{A}(\mathbf{x})} \mathcal{D}A(t, \mathbf{x}) e^{-S'_E} \\ &= \int_{\text{periodic, Gauß}} \mathcal{D}A(t, \mathbf{x}) e^{-S'_E}, \end{aligned} \quad (116)$$

where now one has to consider generic periodic field configurations with no temporal component at  $\tau = 0, \beta$ , which must satisfy Gauß' law. The subscript  $\mathbf{A}(\mathbf{x})$  indicates the boundary conditions at  $t = 0, \beta$ . As the Hamiltonian  $H'$  in the primed basis does not depend on  $\theta$ , the action appearing in the path integral has no  $\theta$  term. Note that the periodic boundary conditions do not imply that there is no sum over topological sectors: while the periodicity condition implies that the sum of the topological flux through the spatial volumes at  $\tau = 0$  and  $\tau = \beta$  is zero, topological flux can still permeate

through the timelike surfaces of the torus, and per the reasoning given in Section 2, the topological charge remains quantized on the four-torus. The  $\theta$ -independent path integral in Eq. (116) still has redundancies associated with the four-dimensional gauge invariance, which give an overall factor associated with the integration over the gauge group. The latter, although immaterial as it cancels in physical observables, can be removed with the usual procedure of gauge-fixing, leading to a partition function

$$Z = \int_{\text{periodic, Gau\ss}} \mathcal{D}A(t, \mathbf{x}) e^{-S'_E - S_{\text{g.f.}} - S_{\text{E,ghost}}} . \quad (117)$$

## References

- [1] R. Jackiw and C. Rebbi, “Vacuum Periodicity in a Yang-Mills Quantum Theory,” *Phys. Rev. Lett.* **37** (1976) 172–175.
- [2] C. G. Callan, Jr., R. F. Dashen, and D. J. Gross, “The Structure of the Gauge Theory Vacuum,” *Phys. Lett. B* **63** (1976) 334–340.
- [3] C. G. Callan, Jr., R. F. Dashen, and D. J. Gross, “Toward a Theory of the Strong Interactions,” *Phys. Rev. D* **17** (1978) 2717.
- [4] W.-Y. Ai, J. S. Cruz, B. Garbrecht, and C. Tamarit, “Consequences of the order of the limit of infinite spacetime volume and the sum over topological sectors for CP violation in the strong interactions,” *Phys. Lett. B* **822** (2021) 136616, [arXiv:2001.07152 \[hep-th\]](#).
- [5] W.-Y. Ai, J. S. Cruz, B. Garbrecht, and C. Tamarit, “The limits of the strong CP problem,” *PoS DISCRETE2020-2021* (2022) 084, [arXiv:2205.15093 \[hep-th\]](#).
- [6] F. Yu, “Primer on Axion Physics,” *Annalen Phys.* **2023** (8, 2023) 2300106, [arXiv:2308.08612 \[hep-ph\]](#).
- [7] Y. Nakamura and G. Schierholz, “The strong CP problem solved by itself due to long-distance vacuum effects,” *Nucl. Phys. B* **986** (2023) 116063, [arXiv:2106.11369 \[hep-ph\]](#).
- [8] N. Yamanaka, “Unobservability of the topological charge in nonabelian gauge theory: Ward-Takahashi identity and phenomenological aspects,” [arXiv:2212.11820 \[hep-ph\]](#).
- [9] G. Torrieri and H. D. Truran, “The strong CP problem, general covariance, and horizons,” *Class. Quant. Grav.* **38** no. 21, (2021) 215002, [arXiv:2007.13183 \[hep-th\]](#).
- [10] W.-Y. Ai, B. Garbrecht, and C. Tamarit, “CP conservation in the strong interactions,” [arXiv:2404.16026 \[hep-ph\]](#).

- [11] E. Fradkin, *Quantum Field Theory: An Integrated Approach*. Princeton University Press, 3, 2021.
- [12] S. Okubo and R. E. Marshak, “Argument for the nonexistence of the ‘strong CP problem’ in QCD,” *Prog. Theor. Phys.* **87** (1992) 1059–1062.
- [13] H. Leutwyler and A. V. Smilga, “Spectrum of Dirac operator and role of winding number in QCD,” *Phys. Rev. D* **46** (1992) 5607–5632.
- [14] G. ’t Hooft, “A Property of Electric and Magnetic Flux in Nonabelian Gauge Theories,” *Nucl. Phys. B* **153** (1979) 141–160.
- [15] P. van Baal, “Some Results for SU(N) Gauge Fields on the Hypertorus,” *Commun. Math. Phys.* **85** (1982) 529.
- [16] M. Luscher, “Some Analytic Results Concerning the Mass Spectrum of Yang-Mills Gauge Theories on a Torus,” *Nucl. Phys. B* **219** (1983) 233–261.
- [17] R. Jackiw, “Introduction to the Yang-Mills Quantum Theory,” *Rev. Mod. Phys.* **52** (1980) 661–673.
- [18] L. Di Luzio, M. Giannotti, E. Nardi, and L. Visinelli, “The landscape of QCD axion models,” *Phys. Rept.* **870** (2020) 1–117, [arXiv:2003.01100 \[hep-ph\]](#).
- [19] J. M. Gracia-Bondía and J. C. Várilly, “Ensuring locality in QFT via string-local fields,” 7, 2022. [arXiv:2207.06522 \[hep-th\]](#).
- [20] V. F. Tokarev, “Does the strong CP problem really exist?,” *Mod. Phys. Lett. A* **8** (1993) 531–541.
- [21] F. K. Guo, R. Horsley, U. G. Meissner, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, and J. M. Zanotti, “The electric dipole moment of the neutron from 2+1 flavor lattice QCD,” *Phys. Rev. Lett.* **115** no. 6, (2015) 062001, [arXiv:1502.02295 \[hep-lat\]](#).
- [22] A. Shindler, T. Luu, and J. de Vries, “Nucleon electric dipole moment with the gradient flow: The  $\theta$ -term contribution,” *Phys. Rev. D* **92** no. 9, (2015) 094518, [arXiv:1507.02343 \[hep-lat\]](#).
- [23] J. Dragos, T. Luu, A. Shindler, J. de Vries, and A. Yousif, “Confirming the Existence of the strong CP Problem in Lattice QCD with the Gradient Flow,” *Phys. Rev. C* **103** no. 1, (2021) 015202, [arXiv:1902.03254 \[hep-lat\]](#).
- [24] L. Giusti and M. Lüscher, “Topological susceptibility at  $T > T_c$  from master-field simulations of the SU(3) gauge theory,” *Eur. Phys. J. C* **79** no. 3, (2019) 207, [arXiv:1812.02062 \[hep-lat\]](#).

- [25] D. Albandea, G. Catumba, and A. Ramos, “The Strong CP Problem in the Quantum Rotor,” [arXiv:2402.17518 \[hep-lat\]](#).
- [26] M. Shifman, *Advanced Topics in Quantum Field Theory*. Cambridge University Press, 4, 2022.
- [27] A. D. Linde, “VACUUM STRUCTURE IN GAUGE THEORIES: THE PROBLEM OF STRONG CP VIOLATION AND COSMOLOGY,” *Phys. Lett. B* **93** (1980) 327–330.