Unitarity of Maxwell theory on curved spacetimes in the covariant formalism

William Donnelly and Aron C. Wall

Unitarity of Maxwell theory on curved spacetimes in the covariant formalism

William Donnelly

Department of Applied Mathematics,
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada*

Aron C. Wall

Department of Physics,
University of California, Santa Barbara
Santa Barbara, California 93106, USA [†]

Quantum field theory in curved spacetime may be defined either through a manifestly unitary canonical approach or via the manifestly covariant path integral formalism. For gauge theories, these two approaches have produced conflicting results, leading to the question of whether the canonical approach is covariant, and whether the path integral approach is unitary. We show the unitarity of the covariant U(1) Maxwell theory, defined via the Wick rotation of a Euclidean path integral. We begin by gauge-fixing the path integral, taking care with zero modes, large gauge transformations, and nontrivial bundles. We find an extra geometric factor in the partition function that has been overlooked in previous work, coming from the zero mode of the gauge symmetry, which affects the entropy and stress-energy tensor. With this extra factor, the covariant calculation agrees with the canonical result for ultrastatic manifolds, and in $D = 2$. Finally, we argue that if there exists a unitary (but not necessarily covariant) canonical formulation, then the covariant formulation must also be unitary, even if the two approaches disagree.

I. INTRODUCTION

The purpose of this article is to resolve some serious confusions about how to quantize gauge fields in curved spacetimes. There are claims in the literature that the canonical and covariant methods of quantization lead to different results from each other [1–4]. Analogous issues appear for gauge fields in the presence of dielectric media [5, 6] and for linearized gravity in curved spacetime [7]. This is disturbing because it makes it unclear how to construct a theory that is simultaneously unitary and covariant.

Since the Maxwell field is free, it is easy enough to promote the equations of motion to operator equations. But problems arise when one tries to calculate the dependence of the partition function Z on the background geometry. The geometry dependence is needed for e.g. calculations of the stress-energy tensor, or the geometric entropy [8]. To calculate the partition function one needs to gauge-fix the action, which can be done covariantly by introducing Faddeev-Popov ghosts. Although these ghosts do not interact with the gauge potential A_μ, they do contribute to the geometry-dependent part of Z. It is therefore important to check that they do not introduce any spurious unphysical effects.

We will focus on free Maxwell fields with gauge group U(1), on D-dimensional, spatially compact, connected[1]

and orientable spacetimes. In section II we will discuss the properties of the unfixed Maxwell partition function on compact Euclidean spacetimes without boundary. In section III we will derive the correct form of the covariant gauge-fixed theory in the continuum, including the correct normalization of the partition function Z, which is nontrivial. This establishes the equivalence of the fixed and unfixed actions on compact manifolds.

For the most part, our analysis will recover the traditional "vector minus two scalars" Faddeev-Popov action. But there are subtleties involving zero modes of the gauge symmetry and the gauge-fixing terms. These subtleties lead to extra terms in the partition function which have not been properly taken into account in previous work.

In section IV we will consider two cases in particular where the partition function can be calculated by other methods: 1) the thermal partition function of pure electrodynamics on an ultrastatic manifold, and 2) the partition function on two-dimensional compact oriented manifolds. We find that the above-mentioned extra terms must be included in the partition function in order to recover known results.

Finally, in section V we will discuss Maxwell fields on nonstatic spacetimes with boundaries. We will show that the unfixed covariant theory is unitary after Wick rotation to globally hyperbolic Lorentzian manifolds. Our analysis shows that the covariant formulation of Maxwell

---

* wdonnelly@uwaterloo.ca
[†] aroncwall@gmail.com
[1] The assumption of connectedness simplifies notation, but is

inessential since the partition function of a disconnected manifold is just the product of partition functions for each of its connected components.

fields is local and unitary, even if it turns out not to be equivalent to the canonical formulation.[2]

II.   A LOCAL COVARIANT ACTION

We start with the Euclidean Maxwell partition function, prior to gauge-fixing:

$$Z = \sum_{\text{bundles}} \int \frac{DA}{\text{Vol}(G)} e^{-S[F]}, \tag{1}$$

where S is the standard Maxwell action given by

$$S = \int d^D x \sqrt{g} \, F_{\mu\nu} F^{\mu\nu}/4, \tag{2}$$

$\text{Vol}(G)$ is the volume of the gauge group G consisting of all local gauge transformations defined by $\delta A = \nabla \alpha$ where α is a scalar parameter which is periodic in the range $[0, 2\pi/q)$, q being the fundamental unit of charge associated with the U(1) gauge symmetry. When the value of α increases by a nonzero multiple of $2\pi/q$ going around a closed curve it represents a large gauge transformation; otherwise it is a small gauge transformation.

Because the gauge group is U(1), the integral of A around around a closed curve is defined only modulo $2\pi/q$. As a result there can exist nontrivial bundles corresponding to magnetic instantons. To sum over these nontrivial bundles, one simply sums up e^{-S} for all possible harmonic 2-forms satisfying the Dirac quantization conditions.

The path integral measure DA is defined as follows: If one writes out modes of A in an orthonormal basis:

$$A_\mu = \sum_n A_n v_\mu^n, \qquad \int d^D x \sqrt{g} v_\mu^m v_\nu^n g^{\mu\nu} = \delta_{mn}, \tag{3}$$

then the path integral measure can be written out explicitly as

$$DA = \prod_n \mu_A \, dA_n, \tag{4}$$

where μ_A is a factor with dimensions of mass required to keep Z dimensionless. Similarly, in order to define $\text{Vol}(G)$, one needs a measure on the space of gauge transformations:

$$D\alpha = \prod_n \mu_\alpha^2 \, d\alpha_n, \tag{5}$$

where μ_α^2 is a factor with mass-dimension two.[3]

The partition function (1) has the advantage of being manifestly covariant. It is also manifestly local, in the sense that e^{-S} and the measure are products of local functions of the field variables. Unfortunately, it is also slightly ill-defined, because the pure-gauge modes of A have infinite fluctuations, while $\text{Vol}(G)$ is correspondingly infinite. Formally, these two infinities cancel each other out to get a finite answer: the precise finite value of ∞/∞ can be fixed by cutting off each pure gauge mode of A at the same value as the corresponding gauge symmetry mode $\nabla \alpha$. The zero mode α_0 corresponding to constant gauge transformations needs no cutoff, because it is periodic due to the U(1) nature of the gauge symmetry.

The reader may wonder whether it is really correct to include the zero mode α_0, since this transformation leaves the A field unchanged. The answer is yes. In its current form, e^{-S} and the measure are local functions of the fields, being a product of contributions associated with each point of space. But if α_0 were dropped, this would no longer be the case. $\text{Vol}(G)$ would be a nonlocal function of the geometry. In section V we will see that this locality property is needed to derive the unitarity of the theory. Hence, just as Feynman diagrams which are invariant under a permutation group Q have to have their amplitude multiplied by a symmetry factor $1/|Q|$, so also a spacetime history whose fields are invariant under a global gauge symmetry must have its amplitude divided by the volume of the α_0 moduli space. Integrating using the measure (5), one finds that the path integral has an overall symmetry factor

$$\frac{1}{\text{Vol}(\alpha_0)} = \frac{q}{2\pi\mu_\alpha^2 \sqrt{V}}, \tag{6}$$

where V is the volume of spacetime. This factor affects the trace of the stress-energy tensor

$$T = \frac{\delta \ln Z}{\delta \sqrt{g}} \tag{7}$$

as well as the entropy of a thermal manifold with inverse temperature β:

$$S = (1 - \beta \frac{\partial}{\partial \beta}) \ln Z \tag{8}$$

In section IV we will show that this factor is necessary in order to obtain correct results for thermodynamic calculations.

---

[2] We originally attempted to use BRST to prove unitarity, but it turns out that BRST is problematic for Maxwell fields on spatially compact manifolds. The representation of BRST is reducible [9], because there exists a canonically conjugate pair of spatially constant ghosts which are unpaired with any unphysical vector modes [10]. One can try to deal with this by adding ghosts-for-ghosts, but we were unable to deal with the resulting IR divergences in a satisfactory way.

[3] These factors of μ are similar to a UV cutoff Λ, insofar as they are dimensionful parameters needed to make the path integral well-defined. Indeed, in some regulator schemes (such as the heat kernel), the same dimensionful parameter plays the role of both Λ and μ. Nevertheless, it is important to recognize that this μ does not result from UV divergences, and can therefore play a role even for $D = 2$ Maxwell fields, where there are no local degrees of freedom.

III. GAUGE-FIXING

We now perform the Feynman-Faddeev-Popov-t'Hooft trick in order to gauge-fix the small gauge transformations. We would like to insert into the path integral a factor of unity which includes the delta function

$$\delta(\nabla_\mu A^\mu - \omega), \qquad (9)$$

with ω being an auxilliary scalar parameter. However this is overconstraining, since on a compact surface one of the constraints is redundant (if the integral of ω is zero) or impossible (if the integral of ω is nonzero).

To deal with this in a covariant way, we use a complete basis of orthonormal scalar modes ϕ_n of the Laplacian:

$$-\nabla^2 \phi_n = \lambda_n \phi_n, \qquad \int d^D x \sqrt{g} \phi_n^2 = 1. \qquad (10)$$

This allows us to expand out ω and $\nabla_\mu A^\mu$ in orthonormal modes of the Laplacian:

$$\omega = \sum_n \omega_n \phi_n, \qquad \nabla_\mu A^\mu = \sum_n (\nabla_\mu A^\mu)_n \phi_n. \qquad (11)$$

We can then insert into the path integral a factor of unity which includes every component of the delta function *except* the zero mode:

$$1 = \int D'\alpha \prod_{n \neq 0} \delta((\nabla_\mu A^\mu_{(\alpha)})_n - \omega_n) \det'(-\mu_\alpha^{-2}\nabla^2), \quad (12)$$

where $A^\mu_{(\alpha)} = A^\mu + \nabla^\mu \alpha$, $\det'(-\mu_\alpha^{-2}\nabla^2)$ is the usual Faddeev-Popov determinant, and the primes are reminders to omit zero modes. In the case of $D'\alpha$ we also

omit the large gauge transformations, which will be handled separately.

The determinant can be viewed as the quantum field theory of two scalar Grassmannian fields known as Faddeev-Popov ghosts, whose partition function is given by

$$\det'(-\mu_\alpha^{-2}\nabla^2) = \int D'c\, D'\bar{c}\, e^{-\int d^D x\, \bar{c}\nabla^2 c}, \qquad (13)$$

where the Grassmannian integrals are defined by

$$\int D'c\, D'\bar{c} = \prod_{n \neq 0} \mu_\alpha^{-2} \frac{\partial}{\partial c_n} \frac{\partial}{\partial \bar{c}_n} \qquad (14)$$

with the zero modes of c and $\bar{c}$ omitted. The factor of μ in Eq. (14) is necessary to make the partition function dimensionless; the fact that it is the same μ_α that appears in Eq. (5) is required so that the Faddeev-Popov trick is equivalent to inserting an identity operator.

One can then perform the $D\alpha$ integrals by changing variables from A to $A_{(\alpha)}$, making the integrand independent of α, leading to a factor of $\text{Vol}(G)$ that cancels (formally) with the one in Eq. (1). This eliminates all gauge symmetry except for the zero mode and large gauge transformations, giving the following gauge-fixed partition function:

$$Z = \frac{q}{2\pi\mu_\alpha^2 \sqrt{V}} \sum_{\text{bundles}} \int \frac{DA}{G_{\text{large}}} D'c\, D'\bar{c} \prod_{n \neq 0} \delta((\nabla_\mu A^\mu)_n - \omega_n)\, e^{-S[F,c,\bar{c}]}. \qquad (15)$$

The measure DA/G_{large} is an integration over equivalence classes of A under large gauge transformations; formally it can be thought of as dividing by the number of Gribov copies, i.e. configurations of A which satisfy $\nabla_\mu A^\mu = 0$ and are equivalent to $A = 0$ under a large U(1) gauge-transformation.

We can then give dynamics to the unphysical modes of A by integrating over ω using the identity

$$\frac{1}{N(\xi)} \int D\omega\, e^{-\int d^D x \sqrt{g}\omega^2/2\xi} = 1, \qquad (16)$$

where N is an infinite normalization factor. Since this normalization factor is local, and depends only on the spacetime volume $\sqrt{g}$, it can be dropped by absorbing it into a redefinition of the cosmological constant. One can then perform the $D\omega$ integrals to obtain the following

partition function:

$$Z = \frac{q}{\mu_\alpha^2} \sqrt{\frac{\xi}{2\pi V}} \sum_{\text{bundles}} \int \frac{DA}{G_{\text{large}}} D'c\, D'\bar{c}\, e^{-S[A,c,\bar{c}]}, \quad (17)$$

where the action is the usual covariant action in t'Hooft R_ξ gauge

$$S = \int d^D x \sqrt{g} \frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \frac{1}{2\xi}(\nabla_\mu A^\mu)^2 - \bar{c}\nabla^2 c, \quad (18)$$

and there is an extra factor coming from the Gaussian

integral over the unmatched zero mode ω_0:

$$\int d\omega_0 \, e^{-\omega_0^2/2\xi} = \sqrt{2\pi\xi}. \tag{19}$$

In this gauge-fixed form of the partition function, covariance is manifest, but locality has been hidden because the zero modes of the field are treated on a different footing than the nonzero modes. Nevertheless, the theory is equivalent by construction to the original unfixed action (1).

To further evaluate the partition function of Eq. (17), we divide the fields A_μ into the zero modes and non-zero modes. The non-zero modes can be evaluated directly as Gaussian integrals,

$$Z_T = \prod_{\lambda \in \sigma_T} \mu_A \sqrt{\frac{2\pi}{\lambda}} \tag{20}$$

$$Z_L = \prod_{\lambda \in \sigma_S} \mu_A \sqrt{\frac{2\pi\xi}{\lambda}} \tag{21}$$

$$Z_G = \prod_{\lambda \in \sigma_S} \mu_\alpha^{-2} \lambda \tag{22}$$

Here Z_T is the path integral over transverse modes, which depends on σ_T, the nonzero spectrum of the Hodge Laplacian on 1-forms restricted to act on transverse modes. Z_L is the path integral over longitudinal modes, and Z_G is the path integral over ghosts; both of these depend on σ_S, the nonzero spectrum of the scalar Laplacian.

Now consider the zero modes of A. The path integral over zero modes does not have the form of a Gaussian integral, as was the case for the non-zero modes. The path integral over these modes would be infinite, were it not cut off by identifying configurations related by large gauge transformations. Large gauge transformations identify A and $A + (2\pi/q)\eta$ for $\eta \in H^1(M, \mathbb{Z})$, where the cohomology group $H^1(M, \mathbb{Z})$ is the set of harmonic 1-forms whose holonomies are all integer-valued. The space of flat connections modulo large gauge transformations is the moduli space $\mathscr{M}$, and the path integral over zero modes is equal to its volume $\mathrm{Vol}(\mathscr{M})$.

It will be convenient to adopt a non-orthonormal basis for $\mathscr{M}$ consisting of harmonic 1-form fields $w_1, \ldots, w_{b_1}$ that form a $\mathbb{Z}$-basis for $H^1(M, \mathbb{Z})$, i.e. such that every 1-form with integer-valued holonomies can be written as a linear combination of the w_I with integer coefficients. In this basis, the flat connections modulo gauge transformations is the region $[0, 2\pi/q)^{b_1}$. The price to pay for this choice of basis is that the path integral measure acquires a factor $\sqrt{\det(\Gamma)}$, where Γ is the metric on moduli space:

$$\Gamma_{IJ} = \int w_I \wedge \star w_J = \int d^D x \sqrt{g} \, g^{ab} w_{Ia} w_{Jb}. \tag{23}$$

The volume of the moduli space is therefore

$$\mathrm{Vol}(\mathscr{M}) = \left(\frac{2\pi\mu_A}{q}\right)^{b_1} \sqrt{\det(\Gamma)}. \tag{24}$$

Finally, we have the monopole configurations. These are indexed by harmonic two-form fields F satisfying the Dirac quantization condition, $F \in \frac{2\pi}{q} H^2(M, \mathbb{Z})$, where the cohomology group $H^2(M, \mathbb{Z})$ is the set of 2-forms having integer flux through every closed two-dimensional surface $S \subset \Sigma$. The contribution of these field configuration to the partition function is

$$Z_{\mathrm{bundles}} = \sum_{F \in \frac{2\pi}{q} H^2(M, \mathbb{Z})} e^{-S[F]} \tag{25}$$

Thus combining all these factors we obtain

$$Z = \frac{q}{\mu_\alpha^2} \sqrt{\frac{\xi}{2\pi V}} \mathrm{Vol}(\mathscr{M}) Z_T Z_L Z_G \sum_{F \in \frac{2\pi}{q} H^2(M, \mathbb{Z})} e^{-S[F]} \tag{26}$$

IV. EXAMPLES

A. Two-dimensional case

Let us consider the case where the manifold M is two-dimensional and orientable. In two dimensions, the Maxwell field has no local degrees of freedom. This is reflected in the fact that the path integral simplifies considerably, due to a cancellation between the gauge potential and ghosts. In particular, the nonzero spectrum σ_T of the transverse Laplacian is identical to the nonzero spectrum of the scalar Laplacian σ_S, so the contribution of the nonzero modes is

$$Z_T Z_L Z_G = \prod_{\lambda \in \sigma_S} \frac{\mu_A^2}{\mu_\alpha^2} 2\pi \sqrt{\xi} = \frac{\mu_\alpha^2}{\mu_A^2} \frac{1}{2\pi\sqrt{\xi}} \tag{27}$$

Here we used the fact that if the product had been a product over all scalar modes (not just non-zero modes) then it could be absorbed into a local term.[4] Note that this factor of $\sqrt{\xi}$ cancels the one in Eq. (17), as required by gauge-invariance of the partition function.

To compute the volume of the moduli space, we appeal to Poincaré duality to show that the metric Γ (23) on moduli space has unit determinant.[5] In two dimensions, Poincaré duality implies that under the bilinear pairing

$$\langle \eta, \zeta \rangle = \int \eta \wedge \zeta, \tag{28}$$

the lattice $H^1(M, \mathbb{Z})$ is self-dual:

$$\{\eta : \langle \eta, \zeta \rangle \in \mathbb{Z} \quad \forall \zeta \in H^1(M, \mathbb{Z})\} = H^1(M, \mathbb{Z}). \tag{29}$$

---

[4] In this step we are essentially just reintroducing the factor of $N(\xi)$ that was dropped from Eq. (16).

[5] For a concise summary of the relevant aspects of Poincaré duality see Ref. [11]

Expressed in the basis w_I, this implies that the matrix

$$P_{IJ} = \int w_I \wedge w_J \qquad (30)$$

is invertible over the integers, and therefore $\det(P) = \pm 1$. The metric Γ_{IJ} is related to the bilinear form (28) by

$$\Gamma_{IJ} = \langle w_I, \star w_J \rangle \qquad (31)$$

where $\star$ denotes the Hodge star. In terms of matrices, $\Gamma = PS$ where S is the matrix $\star w_I = S_{IJ} w_J$. Since $\star^2 = -1$, the matrix S satisfies $\det(S) = \pm 1$. The determinant of Γ is then $\det \Gamma = \det P \det S = 1$ (the positive sign follows from the fact that Γ is positive-definite). Thus in two dimensions, the volume of the moduli space is

$$\mathrm{Vol}(\mathscr{M}) = \left(\frac{2\pi \mu_A}{q} \right)^{b_1} \sqrt{\det \Gamma} = \left(\frac{2\pi \mu_A}{q} \right)^{b_1}. \qquad (32)$$

On a two-dimensional connected manifold, the only harmonic two-forms are proportional to the volume form ϵ_{ab}, and the Dirac quantization condition gives

$$F_{ab} = \frac{2\pi}{q} \frac{n}{V} \epsilon_{ab} \qquad (33)$$

where $n \in \mathbb{Z}$. Their contribution to the partition function is

$$Z_{\mathrm{bundles}} = \sum_{n \in \mathbb{Z}} e^{-2\pi^2 n^2 / V q^2}. \qquad (34)$$

Combining these factors we obtain

$$Z = \sqrt{\frac{2\pi}{q^2 V}} \left(\frac{2\pi \mu_A}{q} \right)^{b_1 - 2} \sum_{n \in \mathbb{Z}} e^{-2\pi^2 n^2 / V q^2}. \qquad (35)$$

Since we assume M is connected and orientable, the exponent $b_1 - 2 = -\chi$, the Euler characteristic of M. Using the Gauss-Bonnet theorem $\chi = \frac{1}{4\pi} \int \sqrt{g} R$, this term can be absorbed into a local counterterm (the Einstein-Hilbert term) and we therefore discard it. By Poisson

summation the remaining part of the partition function can be put in the canonical form

$$Z = \sum_n e^{-\frac{1}{2} V q^2 n^2}. \qquad (36)$$

This agrees with the known result for the partition function of two-dimensional gauge theory, calculated using topological quantum field theory methods [12].

B. Ultrastatic manifold

We now consider the ultrastatic case in which M is of the form $S^1 \times \Sigma$, where the S^1 direction is the imaginary time coordinate τ with period β and Σ is compact and orientable, but otherwise arbitrary. We will show that the formula (26) reproduces the canonical form of the partition function $Z = \mathrm{tr}\, e^{-\beta H}$. We divide the field modes as follows:

1. Nonharmonic modes of Σ:

   (a) transverse spatial polarizations;

   (b) longitudinal spatial polarizations;

   (c) temporal polarizations;

   (d) ghosts;

2. Harmonic modes of Σ:

   (e) vector zero modes of Σ;

   (f) spatially constant temporal modes;

   (g) spatially constant ghosts;

3. Nontrivial U(1) bundles on $\Sigma \times S^1$.

In the first category, there is a factor of $(2 \sinh(\beta \sqrt{\lambda}/2))^{-1}$, the canonical partition function of a mode with frequency $\sqrt{\lambda}$, for each transverse vector mode with Laplacian eigenvalue λ. The contributions of the remaining modes to Z cancel each other exactly.

The second category includes, not only the spatially constant temporal mode and ghosts, but also $b_1 - 1$ vector zero modes of Σ. Here b_1 is the Betti number of M (which counts the number of harmonic vectors); we subtract 1 from this because the temporal mode is included separately. Excluding the integral over vectors that are constant in time (the moduli space), the contribution of all these modes to Z is

$$\prod_{n \neq 0} \mu_\alpha^{-2} \left(\frac{2\pi n}{\beta} \right)^2 \mu_A \sqrt{2\pi \xi} \left(\frac{2\pi n}{\beta} \right)^{-1} \left[\mu_A \sqrt{2\pi} \left(\frac{2\pi n}{\beta} \right)^{-1} \right]^{b_1 - 1} = \frac{\mu_\alpha^2 \beta^2}{\sqrt{\xi}} \left(\frac{1}{\sqrt{2\pi} \mu_A \beta} \right)^{b_1}, \qquad (37)$$

where we have used ζ-function regularization of the product, $\prod_{n \geq 1} a n = a^{\zeta(0)} e^{-\zeta'(0)} = \sqrt{2\pi/a}$. Note that the factor $\xi^{-1/2}$ cancels the explicit factor of $\xi^{1/2}$ in Eq. (26), as required by gauge invariance.

Next we calculate the volume of the moduli space. A basis for the cohomology group of M is given by a one-

form $w_1 = \beta \, d\tau$ wrapping the time direction together

with a basis $w_2, \ldots, w_{b_1}$ of $H^1(\Sigma, \mathbb{Z})$. Letting γ denote the metric on the moduli space of Σ, the metric Γ on the moduli space of M is given by

$$\Gamma = \begin{bmatrix} \frac{V_\Sigma}{\beta} & 0 \\ 0 & \beta \gamma_{IJ} \end{bmatrix} \quad (38)$$

and hence its volume is

$$\text{Vol}(\mathscr{M}) = \left(\frac{2\pi\mu_A}{q} \right)^{b_1} \beta^{(b_1-2)/2} \sqrt{V_\Sigma \det\gamma}. \quad (39)$$

The final kind of contribution comes from nontrivial U(1) bundles. These are proportional to suitably quantized harmonic 2-forms on M. They can be divided into two types: harmonic 2-forms on Σ constant in τ (magnetic fields), and those of the form $d\tau \wedge w_I$ where w_I is a harmonic 1-form on Σ (electric fields).

The magnetic fields are spatial 2-forms B satisfying the Dirac quantization condition $\int_S B \in \frac{2\pi}{q}\mathbb{Z}$ for every closed two-dimensional surface S. These contribute to the partition function as

$$Z_B = \sum_B e^{-\frac{\beta}{2} \int_\Sigma d^{D-1}x \sqrt{q}B^2} \quad (40)$$

which is precisely the contribution of the magnetic monopoles to the canonical partition function.

Now consider the electric fields: 2-forms proportional to $d\tau \wedge w_I$. If we integrate over a surface $S^1 \times \gamma_I$ the quantization condition implies that $F = \frac{2\pi}{q\beta} n_I d\tau \wedge w_I$ where $n_I \in \mathbb{Z}$. The contribution of these field configurations to the partition function is

$$Z_E = \sum_{n \in \mathbb{Z}^{b_1-1}} \exp\left[-\frac{1}{2\beta} \left(\frac{2\pi}{q} \right)^2 \sum_{I,J} \gamma_{IJ} n_I n_J \right]. \quad (41)$$

This is not of the canonical form, because the factor of β in the exponent of Eq. (41) appears in the denominator. However, we can remedy this using the Poisson summation formula,

$$\sum_{x \in \mathbb{Z}^n} e^{-x^T M x} = \det\left(\frac{M}{\pi} \right)^{-1/2} \sum_{y \in \mathbb{Z}^n} e^{-\pi^2 y^T M^{-1} y}, \quad (42)$$

where $M = \frac{1}{2\beta}\left(\frac{2\pi}{q} \right)^2 \gamma$. Thus we can rewrite Eq. (41) as

$$Z_E = \left(\frac{q^2\beta}{2\pi} \right)^{(b_1-1)/2} \frac{1}{\sqrt{\det(\gamma)}} \sum_{m \in \mathbb{Z}^{b_1-1}} \exp\left[-\frac{\beta\,q^2}{2} \gamma^{IJ} m_I m_J \right]. \quad (43)$$

where γ^{IJ} is the inverse of γ_{IJ} and Einstein summation is assumed.

The prefactors in Eqs. (26),(37),(39),(43) cancel, so the path integral has the form of a canonical partition function. All that remains to show is that the sum in Eq. (43) is the same as the canonical partition function of the harmonic E-fields satisfying the quantization condition $\int_S E \in q\mathbb{Z}$ for all closed $(D-2)$-surfaces S. In what follows we treat E as an $(D-2)$-form related to the usual vector field E by Hodge duality of vector fields on Σ. By Poincaré duality, the condition that $\int_S E \in q\mathbb{Z}$ for closed S is equivalent to $\int w_I \wedge E = qm_I$ where $m_I \in \mathbb{Z}$ for all I. Since E is a harmonic $(D-2)$-form, it can be expanded in the basis $\{\star w_I\}$, and takes the form

$$E = q\gamma^{IJ} m_I \star w_J. \quad (44)$$

The energy of such an E-field is

$$\frac{1}{2} \int E \wedge \star E = \frac{q^2}{2} \gamma^{IJ} m_I m_J. \quad (45)$$

Thus the sum in Eq (43) is precisely the canonical partition function of the harmonic E-fields satisfying the quantization condition.

The path integral expression (26) therefore gives the same result as the canonical partition function. Hence the covariant formulation of quantum electrodynamics is unitary on ultrastatic manifolds.

Note that this result depends crucially on using the correct normalization factor in the gauge-fixed partition function 26. Without this prefactor, we would not get the right result even in the trivial case of $D = 1$ electromagnetism, where (because Σ is 0-dimensional) there are no physical degrees of freedom at all. Yet the Faddeev-Popov action has two ghosts but only one vector mode, so it appears at first sight that there are -1 scalar degrees of freedom. This spurious contribution is cancelled when the zero modes are properly handled.

V. NON-STATIC UNITARITY

On non-static manifolds, it is not clear that the covariant and canonical approaches agree with one another [4]. Nevertheless, we will argue that the covariant partition function must be unitary once it is Wick rotated to Lorentzian signature (assuming without proof that this Wick rotation is possible). Our argument will rely on the *existence* of a unitary canonical approach to electromagnetism, but we will not assume that the two approaches agree in all respects.

In this section, we will go back to the original unfixed form (1) of the covariant Euclidean partition function Z.

In order to define unitarity, we need to be able to evaluate Z on a manifold M with boundary ∂M.

A. Manifolds with boundary

States. The partition function of a manifold M with boundary ∂M is a functional of the geometry of M and of a specified boundary value of the dynamical fields. Because the unfixed action does not propagate the temporal or longitudinal modes, in order to compute $Z(M)$ it is sufficient to specify the connection on ∂M. Let $A_{\partial M}$ refer to the pullback of the vector potential to ∂M (which includes only the $D-1$ components of A which point along ∂M). Then the connection is specified by $A_{\partial M}$ modulo gauge transformations $\alpha_{\partial M}$.

Since the value of $Z(M)$ depends on $A_{\partial M}$, the path integral can be regarded as outputting a wavefunction $\Psi[A_{\partial M}]$ on the space of boundary vector potentials. This wavefunction can be written as

$$\Psi[A_{\partial M}] = \sum_{\text{bundles}} \int_{A_{\partial M}} \frac{DA_M}{\text{Vol}(G_M)} e^{-S_M[F]}, \qquad (46)$$

where the integral is over connections with a specified pullback to ∂M and G_M refers to the group of gauge transformations α which vanish on ∂M. However, gauge symmetry implies that Ψ is constant along modes of A which are pure gauge on ∂M.

Inner Product. For any region $\Sigma \subset \partial M$, the inner product on the space of states is given by

$$\langle \Phi, \Psi \rangle_\Sigma = \sum_{\text{bundles}} \int \frac{DA_\Sigma}{\text{Vol}(G_\Sigma)} \Phi^*[A_\Sigma] \Psi[A_\Sigma], \qquad (47)$$

where G_Σ refers to gauge transformations on Σ that vanish on $\partial \Sigma$. Any two open manifolds M_1 and M_2 can be glued together along a shared boundary $\Sigma \subseteq \partial M_1, \partial M_2$ by using the symmetric bilinear $\langle \Phi^*, \Psi \rangle_\Sigma$ to evaluate the pair of states at Σ. In other words, one requires the fields at the boundary to match, and integrates over all possible fields while being careful to mod out by the newly introduced gauge symmetries. By doing this, one obtains Z for the combined region $M_1 \sqcup_\Sigma M_2$, where $\sqcup_X$ denotes the union of two disjoint submanifolds, glued together at a shared boundary X.[6]

Superselection Sectors. Suppose that a spatial slice Σ is decomposed into two open regions joined by a boundary B, so that $\Sigma = \Sigma_1 \sqcup_B \Sigma_2$. We would like to be able to say that the Hilbert space $\mathcal{H}_\Sigma$ decomposes into tensor factors $\mathcal{H}_{\Sigma_1} \otimes \mathcal{H}_{\Sigma_2}$. However, this is not quite true

---

[6] This subscript will sometimes be dropped when the identity of X is unimportant to the discussion, but the step where the manifolds are glued together at their shared boundary is still implied.

since the normal components of the electric and magnetic fields must match across the boundary B. Since these fields can be measured on both Σ_1 and Σ_2, by microcausality they must commute with all other operators localized in either Σ_1 or Σ_2, i.e. from the perspective of a single region they are superselection sectors.

In order to implement the magnetic constraint, we require that 1) the choice of bundle on B must match on both sides, and 2) so must the connection; in other words the vector potential A_B must agree up to gauge transformations (including large ones). Note that if B has noncontractable curves, this constraint also requires the holonomy of A around those curves to agree.

The electric constraint has to do with G_B, the group of gauge transformations on B. The fields in a single region (e.g. Σ_1) do not have to be invariant under gauge transformations which affect the boundary. Instead they can transform under nontrivial unitary representations of G_B. Since G_B is abelian, the irreps are all 1-dimensional, and are labelled by the choice of electric flux across B. However, the state of Σ must be invariant under G_B, so the irrep of Σ_1 must be conjugate to the irrep of Σ_2. This enforces the constraint that the normal component of the electric field matches across the boundary.

Evolution Operator. Now suppose we divide the boundary of a Euclidean spacetime region into two parts joined by B, so that $\partial M = \Sigma_1 \sqcup_B \Sigma_2$. Then the partition function can also be used to calculate a time evolution operator $\mathcal{E}_{\text{cov}}(\Sigma_1 \to \Sigma_2)$ which evolves a state of the fields in Σ_1, to a state of the fields in the complement Σ_2, while preserving superselection data at B:

$$\mathcal{E}_{\text{cov}} = (gJ \otimes I)Z(M), \qquad (48)$$

where J is the antiunitary time-reversal operator (i.e. complex conjugation of the position-space wavefunction), and g is the inner product defined in Eq. (47). Here $Z(M)$ is regarded as a state using Eq. (46), gJ is an operator whose domain is the Hilbert space $\mathcal{H}_{\Sigma_1}$ but whose target is the dual vector space $\mathcal{H}_{\Sigma_1}^*$, and I is the identity acting on $\mathcal{H}_{\Sigma_2}$. Thus $\mathcal{E}_{\text{cov}}$ takes states from $\mathcal{H}_{\Sigma_1}$ to $\mathcal{H}_{\Sigma_2}$.

Wick Rotation. We now wish to define $\mathcal{E}_{\text{cov}}$ for a globally hyperbolic Lorentzian manifold M with boundary $\partial M = \Sigma_1 \sqcup_B \Sigma_2$, such that Σ_1 is an initial spacelike slice and Σ_2 is a final spacelike slice. To do this, we consider the class of all Euclidean metrics for which ∂M has the same geometry. $\mathcal{E}_{\text{cov}}$ is defined in terms of the partition function $Z(M)$, which in turn depends on the metric $g_{\mu\nu}$. We will assume without proof that, as usual in QFT, $Z(M)$ is an analytic function of $g_{\mu\nu}$, which can be analytically continued through complex values of $g_{\mu\nu}$ to the Lorentzian signature, without encountering obstructions or ambiguities. Because the Euclidean partition function is real, the Lorentzian partition function is invariant under time-reversal T.

Converting from $Z(M)$ to $\mathcal{E}_{\text{cov}}$ does not depend on the signature, and can therefore be done either before or after the Wick rotation.

Canonical Approach. Given a particular folation of a globally hyperbolic Lorentzian manifold (with boundary) into Cauchy surfaces $\Sigma(t)$, we will assume that there is some way to define a canonical theory of free electromagnetic fields, in terms of the on-shell degrees of freedom on each time $\Sigma(t)$, modulo gauge transformations. Rather than constructing this theory, we will take a more axiomatic approach and list the properties which we expect it to have.

In the canonical approach, the evolution operator $\mathcal{E}_{\text{can}}(\Sigma_1 \to \Sigma_2)$ must be a manifestly unitary evolution operator. (On a non-static background, the evolution operator $\mathcal{E}$ is actually a map between two *different* Hilbert spaces. When we call such an evolution operator "unitary", we really mean that it is an isomorphism between these two Hilbert spaces.) However, we do not assume that $\mathcal{E}_{\text{can}} = \mathcal{E}_{\text{cov}}$, or even that $\mathcal{E}_{\text{can}}$ is necessarily covariant under a change of foliation.

Instead, we assume that the canonical and covariant theories both satisfy the classical linear field equations, interpreted as operator equations. Using these equations of motion, we can fix the evolution of the wavefunction from Σ_1 to Σ_2 up to an overall numerical factor X which can depend only on the geometry of M:

$$Z_{\text{cov}}(M, A_{\partial M}) = Z_{\text{can}}(M, A_{\partial M}) X(M). \qquad (49)$$

B. Argument for unitarity

We will now argue that $\mathcal{E}_{\text{cov}}$ is unitary in Lorentz signature. The normal method for proving unitarity is to show that the covariant theory is equivalent to the canonical formulation using a reduced phase space. However, according to Ref. [4] these two formulations are not in general equivalent. Nevertheless, it is still extremely useful to know that a manifestly unitary canonical formulation of Maxwell exists, because we can use Eq. (49) to rewrite the covariant evolution operator as

$$\mathcal{E}_{\text{cov}} = X\mathcal{E}_{\text{can}} = |X|U, \qquad (50)$$

where X is a complex number, and U is some unitary evolution rule (i.e. an isometry). In the second equality we have performed a polar decomposition of X, and absorbed the phase into U. After properly renormalizing ultraviolet divergences, both evolution operators should be finite, so $X \neq 0$. Hence one can define a unitary evolution operator as:

$$\frac{\mathcal{E}_{\text{cov}}}{|X|} = U. \qquad (51)$$

Although we have used $\mathcal{E}_{\text{can}}$ to prove the existence of U, once we know it exists it can be defined using $\mathcal{E}_{\text{cov}}$ alone, by choosing the one and only positive real number $|X|$ that makes the right-hand-side unitary. So by Curie's principle (unique solutions to symmetric problems have symmetric solutions), U and $|X|$ are themselves covariant and T-invariant.

We will now show that $|X| = 1$, i.e. $\mathcal{E}_{\text{cov}}$ is unitary. First we observe that $|X|$ by construction depends only on the geometry, not the state. Next we exploit the gluing property of Eq. (1):

$$\mathcal{E}_{\text{cov}}(M_1 \sqcup M_2) = \mathcal{E}_{\text{cov}}(M_1)\mathcal{E}_{\text{cov}}(M_2). \qquad (52)$$

Because the product $U(M_1)U(M_2)$ of two unitary operators is unitary, and $U(M_1 \sqcup M_2)$ is the unique unitary operator equal to a real number times $E_{\text{cov}}(M_1 \sqcup M_2)$, it follows that

$$U(M_1 \sqcup M_2) = U(M_1)U(M_2). \qquad (53)$$

Hence $|X|$ is also local:

$$|X|(M_1 \sqcup M_2) = |X|(M_1)|X|(M_2). \qquad (54)$$

Using these relations, we can chop up the manifold into tiny pieces, in order to show that $|X|$ is given by a local action:

$$|X| = e^{iS[M]}. \qquad (55)$$

Here the action S is imaginary, and depends only on the metric g_{ab}. This action might include boundary terms at $\partial M = \Sigma_1 + \Sigma_2$. Temporarily ignoring these boundary terms, we can expand out the action in derivatives of the metric to obtain

$$S = i \int_M d^D x \sqrt{-g} [\, a + bR + (cR^2 + dR_{ab}^2 + eR_{abcd}^2) + \dots]. \qquad (56)$$

When this action is Wick rotated back to the Euclidean signature, it remains purely imaginary. This is because each power of the Riemann tensor has an even number of time-derivatives, while the i coming from rotating the volume form $\sqrt{-g} = i\sqrt{g}$ turns e^{iS} into e^{-S}. It follows that such terms cannot exist by T-invariance. (Although this fact could be deduced from the Lorentz signature theory, it is easiest to diagnose violations of T in Euclidean signature since there T consists of complex conjugation only.) Therefore the covariant partition function is unitary up to a local boundary action.

Terms in the boundary action could depend on the metric on the boundary, as well as its normal derivatives. However, it turns out that no boundary term can be found having the necessary properties. We start by mapping each of the initial and final Hilbert spaces into a specific Hilbert space $\mathcal{H}$ described by an orthonormal basis of states $(\psi_1, \psi_2, \psi_3...)$. This allows U to be regarded as a unitary operator in $\mathcal{H}$. In this basis, the inner product $\langle \Phi, \Psi \rangle$ can be written as the identity matrix. Since raising and lowering with respect to the identity matrix is trivial, the boundary terms must take the same form for the "initial" slice Σ_1 and the "final" slice Σ_2.

The absence of allowed boundary terms is easiest to see if $\ln|X|$ is rotated back to Euclidean signature. In order to satisfy the gluing property, the boundary term would have to contain an odd number of normal derivatives. (Terms such as $\text{Vol}(\partial M)$ containing an even number of derivatives would not cancel when gluing together

two regions.) T-reversal symmetry says the boundary term has to be real. Now an *odd, real* boundary term in Euclidean signature Wick rotates into a contribution to the Lorentzian $\ln|X|$ which is *odd* and *imaginary*. But $\ln|X| = iS$ is real in Lorentz signature by construction (and in any case, an imaginary contribution to iS would not lead to a violation of unitarity.)

Since no contributions to $|X|$ are possible, it follows that $\mathcal{E}_{\text{cov}}$ is unitary. This is true even if it disagrees with $\mathcal{E}_{\text{can}}$ on nonstatic manifolds.

VI. DISCUSSION

In section II, we defined the covariant electromagnetic partition function on spatially compact manifolds, taking into account issues involving zero modes of the field and gauge symmetry, as well as nontrivial U(1) bundles and large U(1) gauge transformations.

Section III showed how to implement the Faddeev-Popov trick on compact Euclidean manifolds. In Eq. (17) we found several kinds of departures from the usual "vector minus two scalars" partition function. In addition to summing over nontrivial bundles, modding out by large gauge transfromations (i.e. Gribov copies), and omitting ghost zero modes, there is also a nontrivial normalization factor $(q/\mu_\alpha^2)\sqrt{\xi/2\pi V}$, depending on the charge q, the t'Hooft gauge parameter ξ, the spacetime volume V, and the dimensionful parameter μ_α which defines the measure on the space of gauge-transformations.

Because the normalization factor depends on the geometry through V, it contributes to the vacuum stress-energy tensor, and to the entropy of e.g. de Sitter space. In electromagnetism with one spacetime dimension (which has no degrees of freedom at all), this factor is necessary to obtain a trivial partition function.

We also showed that the covariant approach is local and unitary, although the Faddeev-Popov trick obscures these properties. In our proof of unitarity (section V), we relied on the existence of a canonical formulation of electromagnetism, but we did not explicitly construct the canonical formulation, nor did we prove that the two approaches are generally equivalent.

We were able to explicitly compare to the canonical formulation for two special categories of Euclidean manifolds (section IV). One was $D = 2$ electromagnetism (where there are no local degrees of freedom), treated in [10, 12]. The second was ultrastatic manifolds of the form $S^1 \times \Sigma$, where the canonical partition function takes the form $\text{tr}\, e^{-\beta H}$ summed over physical modes. In both of these cases, we found an agreement with the covariant approach. Although we have not done the calculation, we expect that the two approaches will always agree for static, horizonless manifolds.

However, we expect that serious complications will arise for $D > 2$ time-dependent manifolds. In these cases, the two ghosts do not in general cancel out the temporal and longitudinal modes of A_μ [4, 10]. In other words, the unphysical modes give rise to physical effects, which are hard to explain in terms of the canonical degrees of freedom. Similarly, in the case of manifolds with horizons, there are effects arising from the fact that vectors and scalars have different boundary conditions at the horizon. This seems to be related to the mysterious "contact term" discovered by Kabat [13], who calculated that Maxwell fields contribute a negative divergence to the horizon entropy. This contact term cannot come from the entanglement entropy of the canonical modes, which is intrinsically positive. A similar discrepancy appears for gravitons [14, 15].

In fact, the original motivation of this work was to investigate the validity of this contact term on compact manifolds (where one need not worry about complications due to a boundary at infinity). In an earlier article [10], we showed that the contact term is absent in $D = 2$ for canonically quantized electromagnetism. This article confirms that the covariant $D = 2$ theory is equivalent, so the contact term is absent there as well. The case of higher dimensions will be treated in a future article.

Nevertheless, quite aside from the contact term, it would be disturbing if the covariant and canonical formulations gave different answers to physical questions. This article goes halfway towards resolving the issue, by showing how to properly treat the covariant formulation. Further progress on the discrepancy will require a better understanding of the canonical approach.

It is not surprising that little work has been done on canonical Maxwell fields on nonstatic spacetimes, because the required conceptual apparatus is technically difficult. For each time slice, one has to construct a Hilbert space based on physical modes only. In order to write down the dynamics in terms of a time-dependent Hamiltonian, one must first select an isomorphism between the Hilbert spaces at each moment of time (on a static manifold one can use time translation symmetry to do this, but on a nonstatic manifold one has to make some arbitrary choices). At the end, one has to prove covariance, i.e. that the dynamics would have been the same regardless of the choice of slicing.

In Vassilevich's approach to canonical QFT, one slices up the spacetime, identifies the "physical modes" which would contribute to the canonical analysis, and then quantizes these modes using covariant methods [4]. Vassilevich found a noncovariant result, which disagrees with the covariant approach for e.g. Euclidean de Sitter space S^4 or S^2. However, Ref. [4] did not include the above normalization factor in his canonical approach. It also used the wrong measure in the path integral. (In the $D = 2$ case, the path integral measure on an expanding spacetime is a function of time. The correct measure is easiest to calculate if one uses the integrated spacetime volume as the "time" parameter, as in Ref. [10].) It may be that a more sophisticated version of the canonical approach might give an answer which agrees with the covariant approach.

The analysis of section V shows that the canonical and

covariant approaches agree up to a geometry-dependent phase factor. Therefore, so long as one is willing to insert such phases as a "fudge factor", the canonical method can be made to work. The question remains whether these phases can be derived elegantly in a canonical framework.

In Wald's approach to canonical QFT in curved spacetime [16], one identifies the physical modes of the field in a slice-independent way, and quantizes this using the symplectic structure. This approach is manifestly covariant, but Wald was not able to directly calculate the stress-energy tensor. Instead he lists some axioms (including covariance) which the stress-energy tensor should satisfy, and used this to pin down their correct form.

Assuming this approach can be applied to gauge fields (which Ref. [16] claims, but does not work out in detail), it should be possible to derive the evolution operator $\mathcal{E}_{\text{can}}$, use this to prove that $\mathcal{E}_{\text{cov}}$ is unitary, and then show that the covariant stress-energy tensor obeys Wald's axioms. Unfortunately, this roundabout method obscures the physical origin of the effects. It would be good to have a more direct way to derive the stress-energy tensor from the canonical formulation. If our treatment of the covariant calculation is any indication, this will probably require being very careful when handling the gauge symmetries.

[1] D. Vassilevich, "On the Hamiltonian QED in curved space-time," *Nuovo Cim.* **A104** (1991) 743–754.

[2] D. Vassilevich, "Path integral measure for Hamiltonian QED in curved space," *Nuovo Cim.* **A105** (1992) 649–653.

[3] H. Fukutaka, "Anomalous Jacobian factor in the Polyakov measure for Abelian gauge field in curved space-times," `arXiv:hep-th/9201022 [hep-th]`.

[4] D. V. Vassilevich, "QED on curved background and on manifolds with boundaries: Unitarity versus covariance," *Phys.Rev.* **D52** (1995) 999–1010, `arXiv:gr-qc/9411036 [gr-qc]`.

[5] M. Bordag, K. Kirsten, and D. Vassilevich, "Path integral quantization of electrodynamics in dielectric media," *J.Phys.* **A31** (1998) 2381–2389, `arXiv:hep-th/9709084 [hep-th]`.

[6] X. Bei and Z. Liu, "Electromagnetic Field Quantization in Time-Dependent Dielectric Media," `arXiv:1104.2453 [quant-ph]`.

[7] H. Fukutaka, "Path integral measure of linearized gravity in curved space-time," `arXiv:hep-th/9203049 [hep-th]`.

[8] C. G. Callan and F. Wilczek, "On geometric entropy," *Phys.Lett.* **B333** (1994) 55–61, `arXiv:hep-th/9401072 [hep-th]`.

[9] G. Barnich, F. Brandt, and M. Henneaux, "Local BRST cohomology in gauge theories," *Phys.Rept.* **338** (2000) 439–569, `arXiv:hep-th/0002245 [hep-th]`.

[10] W. Donnelly and A. C. Wall, "Do gauge fields really contribute negatively to black hole entropy?," *Phys.Rev.* **D86** (2012) 064042, `arXiv:1206.5831 [hep-th]`.

[11] M. Headrick, "Compendium of useful formulas," Nov., 2012. `http://people.brandeis.edu/~headrick/compendium.pdf`.

[12] S. Cordes, G. W. Moore, and S. Ramgoolam, "Lectures on 2-d Yang-Mills theory, equivariant cohomology and topological field theories," *Nucl.Phys.Proc.Suppl.* **41** (1995) 184–244, `arXiv:hep-th/9411210 [hep-th]`.

[13] D. N. Kabat and M. Strassler, "A Comment on entropy and area," *Phys.Lett.* **B329** (1994) 46–52, `arXiv:hep-th/9401125 [hep-th]`.

[14] D. V. Fursaev and G. Miele, "Cones, spins and heat kernels," *Nucl.Phys.* **B484** (1997) 697–723, `arXiv:hep-th/9605153 [hep-th]`.

[15] S. N. Solodukhin, "Entanglement Entropy of Black Holes," *Living Reviews in Relativity* **14** no. 8, (2011) , `arXiv:1104.3712 [hep-th]`. `http://www.livingreviews.org/lrr-2011-8`.

[16] R. M. Wald, *Quantum field theory in curved spacetime and black hole thermodynamics*. University of Chicago Press, 1994.