The Quantum Electromagnetic Field on a Spatial Lattice

Randy S

Abstract This article introduces one of the simplest examples of a quantum model with a gauge field, treating D-dimensional space as a lattice so that the math is straightforward. The model is a special case of compact quantum electrodynamics (**compact QED**), namely the case with no electrically charged matter, so the quantum electromagnetic field is the only physical entity.

The adjective *compact* in the name refers to the fact that the model uses the compact group $U(1)$ as its gauged group, in contrast to traditional electrodynamics in which the gauged group is the noncompact group $\mathbb{R}$. The choice $U(1)$ is motivated by the fact that the electric charges of all known elementary particles appear to be precisely integer multiples of a single elementary unit of charge. The model constructed here does not include charged matter, but it uses $U(1)$ as the gauged group to prepare for models that do.

Contents

1 Introduction

Article 26542 sketches a model of the quantum electromagnetic field. That sketch isn't quite well-defined mathematically, because it tries to associate operators on a Hilbert space with individual points in continuous space. That sketch can be promoted to a mathematically well-defined construction by using only operators that are smeared in space,[1] but learning that consruction has a low value-to-cost ratio because it doesn't allow generalizing the model to include charged matter. This article uses a different approach to making the model well-defined, treating space as a large but finite number of closely spaced points.[2] This approach works just as well when charged matter is included.

As in article 26542, this article uses the hamiltonian formulation, in which time is continuous. The model also has a well-defined path-integral formulation in which time and space are both discrete. The path-integral formulation has several advantages, but the hamiltonian formulation makes the relationship to the general principles of quantum theory[3] more clear, so the hamiltonian formulation will be used here.

The equations of motion in this model are nonlinear,[4,5] even though interactions with charged matter are absent. The nonlinearity makes calculations challenging, but this article focuses on the easy part: constructing the model without any mathematical ambiguity, so that calculations have a solid place to start.

---

[1]Smearing in time works more generally than smearing only in space (Witten (2023)), but the smearing approach (in time and/or space) has not yet led to nonperturbative constructions of many models that are believed to have nontrivial continuum limits, much less of models that are not believed to have nontrivial continuum limits (like quantum electrodynamics with charged matter).

[2]Article 52890 uses this approach for scalar fields, for the same reason. The **holographic principle** (reviewed in Bousso (2002)) gives us a good reason to think that this lattice-like picture of space and the conventional picture of continuous space are *both* ultimately incorrect. With that in mind, the fact that we don't know how to define most quantum field models in continuous space is less disappointing, and the fact that we do know how to define so many quantum field models using a lattice-like approach is a welcome concession.

[3]Article 03431

[4]Section 25

[5]Section 26 will relate these nonlinear equations to Maxwell's equations, which are linear.

2 Outline

- Section 3 previews some notation.
- Section 4 introduces the gauged group.
- Sections 5-6 define the spatial lattice – actually two different versions of the spatial lattice, one periodic and one not, because they both have advantages when studying gauge theories in general.
- Sections 7-21 define a Hilbert space and the model's observables at time $t = 0$, represented as operators on that Hilbert space, and explain how this representation reproduces some of Maxwell's equations (the ones that don't involve time derivatives).
- Sections 22-23 introduce the hamiltonian and use it to define the model's observables at all times t in terms of those at $t = 0$.
- Sections 24-27 give some insight about the model's continuum limit. First, some simplistic calculations are used to help relate the lattice equations of motion to Maxwell's equations (the ones that involve time derivatives), and then some insights from more careful studies will be summarized.
- Section 28-30 clarify how to relate the definitions of magnetic flux in sections 15 and 19 to what experiments actually measure.

3 Preview of notation

For reference, here's a summary of some notation that will be introduced later in this article:

- $\mathbf{x}$ = a point (site) in the spatial lattice (section 5)
- $\mathcal{G}$ = the group of interior gauge transformations (section 6)
- ℓ = link (a pair of neighboring lattice sites)
- $\square$ = plaquette (section 9)
- $u(\ell) = U(1)$-valued link variable (section 7)
- $\theta(\ell)$ = angle-valued link variable (section 7)
- $E(\ell)$ = electric field operator associated with link ℓ (section 11)
- $W(\square)$ = plaquette operator (section 13)
- $W(C)$ = Wilson loop operator or Wilson line operator (section 13)
- $B(S)$ = magnetic flux through a surface S (sections 15 and 19)

This article uses the units conventions described in article 26542. That system of units uses a minimum electric charge that will be denoted q.

4 The gauged group

Article 70621 introduced the concept of a **principal *G*-bundle**, which is the mathematical foundation for the concept of a gauge field. In the physics literature, the group G is often called the **gauge group**, but that means something different in the math literature.[6] For clarity, this article calls G the **gauged group**. This name is not standard, but it is consistent with the important idea of *gauging* a symmetry group (using the word *gauge* as a verb).

In classical electrodynamics, the gauged group is usually taken to be $\mathbb{R}$, the additive group of real numbers. In quantum electrodynamics (QED), we have a good reason to take the gauged group to be the compact group $U(1)$ instead. This is called **compact QED**. Here's the reason: the magnitudes of the electric charges of all known elementary particles are precisely integer multiples of a single quantity, with no evidence of any deviations despite careful searches for exceptions.[7] This **charge quantization**[8] would be unexplained in models that use $G = \mathbb{R}$, but it is automatic in quantum models that use $G = U(1)$.[9] This is related to the fact that in a model with a charged entity, the term in the hamiltonian that implements its interaction with the electromagnetic field involves a link variable $e^{i\theta}$ (introduced in section 7) raised to the nth power, where n is the entity's electric charge expressed as a multiple of an elementary unit q of charge. When the gauged group is $U(1)$, the quantity θ is defined only modulo 2π, so n must be an integer for the nth power of $e^{i\theta}$ to make sense. If the gauged group were $\mathbb{R}$ instead, then n could be any real number, so the empirical quantization of charge would be unexplained.

The model constructed in this article doesn't include electrically charged objects, but it uses $G = U(1)$ anyway as practice for models that do.

---

[6] In the math literature, the name *gauge group* is used for the group of gauge transformations, which is much larger than G (article 76708). G is often called the **structure group**, but that can also be ambiguous (article 70621).

[7] Dylla and King (1973), Marinelli and Morpurgo (1984), Baumann *et al* (1988)

[8] Here, *quantization* means limited to a discrete (not continuous) set of values.

[9] Harlow and Ooguri (2021), section 3.4, page 76: "...it would be crazy to ignore the observational fact that the charges of the electron and proton are exact opposites to within one part in 10^{21} [ref]. By far the most plausible explanation of this remarkable agreement is that the gauge group of electrodynamics is indeed $U(1)$..."

5 The short-distance cutoff

Each element of the Hilbert space will be represented by a function of an enormous number of variables, nominally D variables for each point in D-dimensional space. To keep the number of variables finite, so that the model's construction is straightforward, we will need both a **short-distance (UV) cutoff** and a **long-distance (IR) cutoff**. This section describes the short-distance cutoff, and section 6 will modify this picture to implement a long-distance cutoff.

Start with D-dimensional euclidean space, and choose a set of D mutually orthogonal basis vectors $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_D$, all with the same magnitude ϵ. Choose any one point p in the D-dimensional space. The lattice consists of all points that may reached from p by adding integer multiples of the basis vectors $\mathbf{e}_k$. Each point in this infinite lattice has coordinates $(n_1, n_2, ..., n_D)$ in the chosen basis, where each coordinate n_k is an integer. The point p has coordinates $(0, 0, ..., 0)$.

Two points $\mathbf{x}$ and $\mathbf{y}$ in the lattice are called **nearest neighbors** if they have the same coordinates except for one coordinate in which they differ by ± 1, so the distance between them is ϵ. An ordered pair $(\mathbf{x}, \mathbf{y})$ of nearest neighbors will be called a **directed link**, and an unordered pair $\{\mathbf{x}, \mathbf{y}\}$ of nearest neighbors will be called an **undirected link**. The two directed links $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{y}, \mathbf{x})$ will be called **oppositely directed** compared to each other. The two points $\mathbf{x}$ and $\mathbf{y}$ will be called the **endpoints** of the link.

Instead of associating variables with each point in continuous space, variables will be associated only with the (directed) links in this lattice.[10] These variables will be called **link variables**. This is a type of short-distance cutoff, because the number of variables per unit volume is finite.

[10]Section 7

6 The long-distance cutoff

The Hilbert space inner product will be defined by integrating over all of the link variables. To ensure that this makes sense, a long-distance cutoff will be used so that the total number of integration variables is finite. Two different long-distance cutoffs will be considered. Both of them start with the lattice that was defined in section 5 and modify it to limit the total number of link variables.

One long-distance cutoff uses a **truncated lattice**.[11,12] To define this, choose a convex open set O of D-dimensional euclidean space that contains a very large but finite number of the original lattice points. Choose O so that its boundary doesn't pass through any lattice point. Points that are in O will be called **interior points**, any other point that is connected to an interior point by a single link will be called a **boundary point**, and those links will be called **boundary links**. This is illustrated in figure 1. Only links with at least one interior endpoint[13] will have associated link variables.[14]

The other long-distance cutoff will be called a **wrapped lattice**,[11] because space wraps back on itself like a torus. To define this, choose an integer $K \ggg 1$. Start with the same infinite lattice as before, but declare two points $\mathbf{x}$ and $\mathbf{y}$ to be equivalent (the same point) if each coordinate of $\mathbf{x}$ is equal to the corresponding coordinate of $\mathbf{y}$ modulo K. With this equivalence relation, each point still has $2D$ nearest neighbors, the lattice still has (discrete) translation symmetry, and it still doesn't have any boundary points or boundary links, but now the total number of points in the lattice is finite (equal to K^D). This is illustrated in figure 2.

The model's construction will be described in a way that works equally well with either of these two long-distance cutoffs. When using a wrapped lattice, statements that apply only to boundary points and boundary links may simply be ignored, because the wrapped lattice doesn't have any.

[11] This name is not standard.

[12] A truncated lattice is technically no longer a *lattice* in the usual mathematical sense of the word, but this article still calls it a lattice. This is common in the literature about "lattice" quantum field theory.

[13] An **interior endpoint** is an endpoint that is inside O.

[14] Section 7

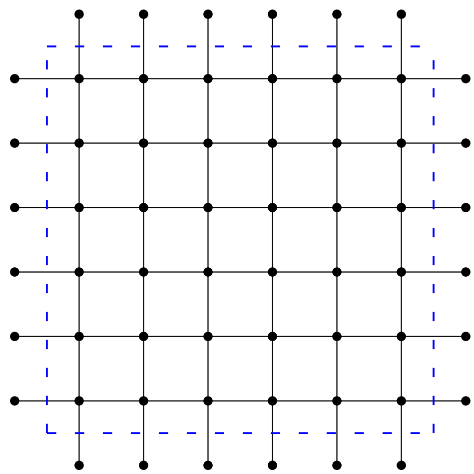


Figure 1 – Example of a two-dimensional truncated lattice ($D = 2$). Dots represent points in the lattice, and solid lines represent links. The dashed outline is the boundary of the region that was denoted O in the text. The points inside the dashed outline are interior points. The points outside the dashed outline are boundary points. Links that cross the dashed outline are boundary links. When $D = 3$, the one-dimensional dashed outline is replaced by a two-dimensional surface.

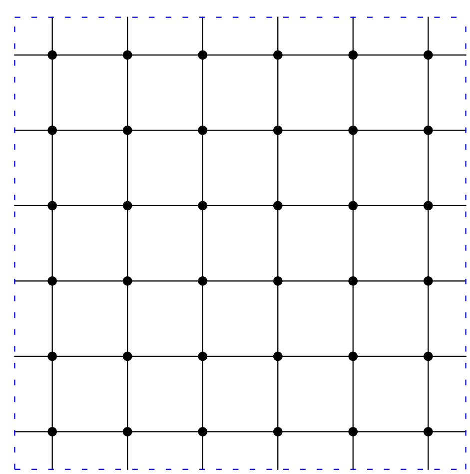


Figure 2 – Example of a two-dimensional wrapped lattice ($D = 2$). Opposite sides of the dashed outline are identified with each other, so space is topologically a torus. This lattice does not have any boundary points or links. Every point is an interior point with $2D$ nearest neighbors.

7 Link variables

Section 10 will construct the Hilbert space. Each element of the Hilbert space is a function of variables called *link variables.* This section introduces the link variables.

Gauge fields in smooth space are defined using a mathematical structure called a **principal *G*-bundle**.[15] The Lie group G is often called the **gauge group** in the physics literature, but the math literature uses that name for a much larger group (the group of all gauge transformations). To avoid confusion, this article will call G the **gauged group**, because the group of all gauge transformations is obtained by "gauging" the group G.[16]

In the standard hamiltonian formulation of lattice gauge theory, every directed link $(\mathbf{x},\mathbf{y})$ with at least one interior endpoint has an associated **link variable** $u(\mathbf{x},\mathbf{y})$, which takes values in the gauged group G. Link variables associated with oppositely-directed links are related to each other by the condition

$$u(\mathbf{x},\mathbf{y})u(\mathbf{y},\mathbf{x}) = 1. \tag{1}$$

If a link $(\mathbf{x},\mathbf{y})$ doesn't have an associated link variable,[17] then $u(\mathbf{x},\mathbf{y}) \equiv 1$.

In this article, the gauged group G is $U(1)$, the multiplicative group of complex numbers with magnitude 1, so each link variable may be written in terms of a real-valued **angle variable** $\theta(\mathbf{x},\mathbf{y})$ like this:

$$u(\mathbf{x},\mathbf{y}) = e^{i\theta(\mathbf{x},\mathbf{y})}. \tag{2}$$

The angle variable $\theta(\mathbf{x},\mathbf{y})$ is only defined modulo 2π. Equation (1) implies that $\theta(\mathbf{x},\mathbf{y})$ is equal to $-\theta(\mathbf{y},\mathbf{x})$ modulo 2π.

The collection of link variables represents the **gauge field**, and any assignment of specific values (specific elements of G) to all of the link variables will be called a **configuration** of the gauge field. The name *compact QED* refers to the fact that the gauged group $U(1)$ is compact as a smooth manifold.

[15] Article 76708

[16] Harlow and Ooguri (2021), section 3.1

[17] Remember that only links with at least one interior endpoint have associated link variables.

8 Gauge transformations

Section 10 will define the Hilbert space using functions that are invariant under a group of *gauge transformations*. This section explains what that means.

Let h be a map from the set of points in the lattice to the gauged group $U(1)$, so $h(\mathbf{x}) \in U(1)$ for each point $\mathbf{x}$. A transformation that replaces the original value of every link variable with the new value

$$u^h(\mathbf{x},\mathbf{y}) \equiv h(\mathbf{x})u(\mathbf{x},\mathbf{y})h^{-1}(\mathbf{y}) \tag{3}$$

will be called a **gauge transformation**. It will be called an **interior gauge transformation** if $h(\mathbf{x}) = 1$ whenever $\mathbf{x}$ is not an interior point.[18] The group of all interior gauge transformations will be denoted $\mathcal{G}$. A function $\Psi[u]$ of the link variables will be called **$\mathcal{G}$-invariant** if it is invariant under all interior gauge transformations:[19]

$$\Psi[u^h] = \Psi[u] \qquad\qquad \text{for all } h \in \mathcal{G}. \tag{4}$$

This is the only part of the construction that treats boundary points differently than interior points, so let's consider what would happen if we didn't require $h(\mathbf{x}) = 1$ for boundary points. The Hilbert space that will be defined in section 10 consists of $\mathcal{G}$-invariant functions of the link variables. If we required invariance under all gauge transformations, not just interior ones, then the space of invariant functions would be smaller, so the set of linear operators that can act on the Hilbert space would also be smaller. Requiring only $\mathcal{G}$-invariance accommodates a slightly larger set of operators on the Hilbert space.[20,21]

---

[18] This constraint is empty on a wrapped lattice (section 6).

[19] This article uses the **temporal gauge**, in which $A_0 = 0$. Only time-independent gauge transformations are considered here.

[20] The text between equations (3.21) and (3.22) in Harlow and Ooguri (2021) mentions a context in which this can be important.

[21] The set would be even larger if we didn't require gauge invariance at all, but then the model wouldn't be consistent with electrodynamics.

9 Examples of $\mathcal{G}$-invariant functions

Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ be any sequence of points for which each pair $(\mathbf{x}_n, \mathbf{x}_{n+1})$ of consecutive points is a link. Let C denote this set of directed links, and define

$$u(C) \equiv \prod_{\ell \in C} u(\ell) = u(\mathbf{x}_1, \mathbf{x}_2) u(\mathbf{x}_2, \mathbf{x}_3) \cdots u(\mathbf{x}_{N-1}, \mathbf{x}_N). \tag{5}$$

This is the product of link variables along a path C made of links.[22] The effect of a gauge transformation on this product is

$$u(C) \to h(\mathbf{x}_1) u(C) h^{-1}(\mathbf{x}_N).$$

The product is $\mathcal{G}$-invariant for either of these two types of path (figures 3-6):

- If the path's endpoints are both boundary points, then the product is $\mathcal{G}$-invariant because $h(\mathbf{x}) = 1$ for boundary points.[23]

- If the path is closed ($\mathbf{x}_N = \mathbf{x}_1$), then the product is $\mathcal{G}$-invariant. This works because the gauged group is abelian (commutative), so

$$h(\mathbf{x}_1) u(C) h^{-1}(\mathbf{x}_1) = u(C) h(\mathbf{x}_1) h^{-1}(\mathbf{x}_1) = u(C).$$

An important example of a closed path is the sequence of directed links that traces out the perimeter of a **plaquette**, the smallest possible loop in the lattice:

$$u(\square) \equiv \prod_{\ell \in \square} u(\ell) = u(\mathbf{x}_1, \mathbf{x}_2) u(\mathbf{x}_2, \mathbf{x}_3) u(\mathbf{x}_3, \mathbf{x}_4) u(\mathbf{x}_4, \mathbf{x}_1). \tag{6}$$

The product $u(\square)$ is called a **plaquette variable**. The plaquette $\square$ can have either of two possible **orientations**, corresponding to the two possible directions in which we can trace around the perimeter.

Any function of these $\mathcal{G}$-invariant products is still $\mathcal{G}$-invariant. This provides a rich supply of $\mathcal{G}$-invariant functions.

[22] Recall that $u(\mathbf{x}, \mathbf{y}) \equiv 1$ when $(\mathbf{x}, \mathbf{y})$ doesn't have an associated link variable (section 7).

[23] Section 8

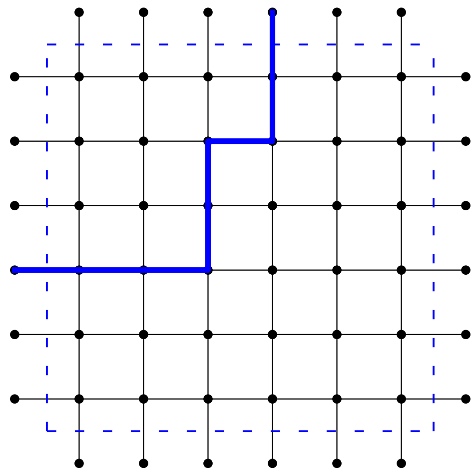


Figure 3 – Example of a path whose endpoints are both boundary points.

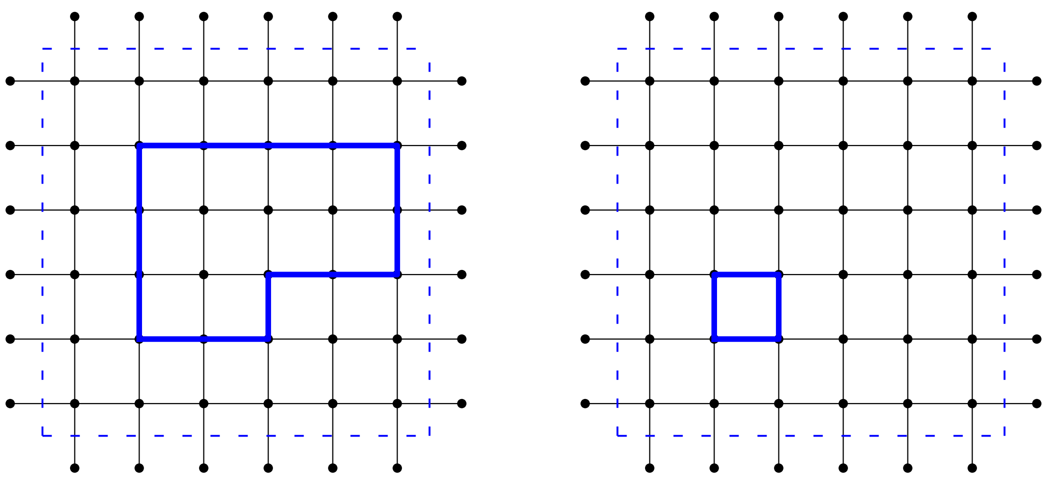


Figure 4 – Examples of closed paths. The example on the right is the boundary of a plaquette. The lattice in these pictures is two-dimensional ($D = 2$). On a D-dimensional lattice with $D \geq 3$, most paths do not lie in a single plane, but (the boundary of) a plaquette necessarily lies in a single plane no matter how many dimensions the lattice has.

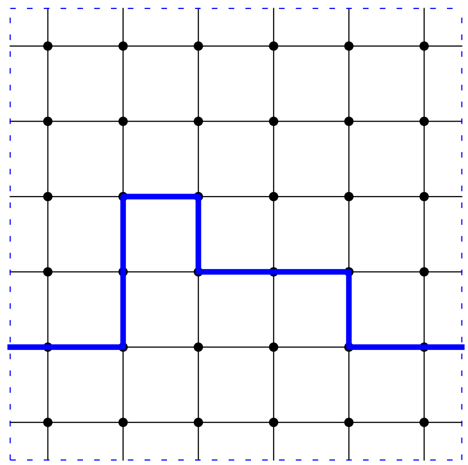


Figure 5 – Example of a closed path that is not the (whole) boundary of any surface made from plaquettes. Such paths exist on a wrapped lattice. If the space were continuous (but still topologically a torus), this loop would still not be contractible (article 61813).

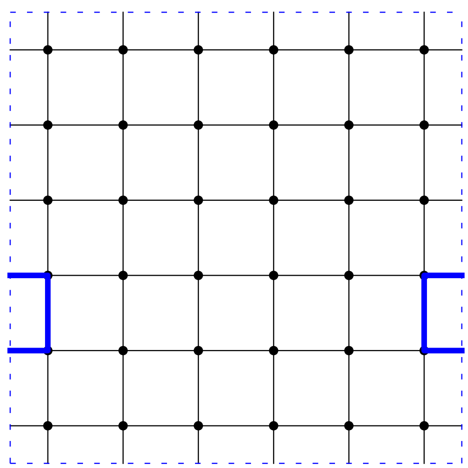


Figure 6 – Example of a closed path forming the boundary of a single plaquette, on a wrapped lattice. If space were continuous, this loop would be contractible.

10 The Hilbert space

Observables are represented by linear operators on a Hilbert space. This section constructs the Hilbert space that will be used for the rest of this article.

An element of the Hilbert space will be called a **state-vector**.[24] Each state-vector is represented by a $\mathcal{G}$-invariant[25] complex-valued function $\Psi[u]$ of the link variables. Given two states $\Psi_1[u]$ and $\Psi_2[u]$, their inner product is

$$\begin{aligned} \langle\Psi_1|\Psi_2\rangle &\equiv \int [du]\ \Psi_1^*[u]\Psi_2[u] \\ &\equiv \int \left(\prod_\ell du(\ell)\right)\ \Psi_1^*[u]\Psi_2[u] \end{aligned} \tag{7}$$

where the product is over all links that have associated link variables and where $du(\ell)$ is defined by

$$\int du(\ell)\cdots \equiv \int_0^{2\pi} d\theta(\ell)\cdots$$

with $u(\ell) = e^{i\theta(\ell)}$. The inner product is well-defined because the number of integration variables is finite (sections 5-7) and because the domain of each integration variable is finite.[26] This Hilbert space will be denoted $\mathcal{H}$.[27]

[24] Article 03431

[25] Section 8

[26] This is a technical advantage of working with the compact group $U(1)$ instead of the noncompact group $\mathbb{R}$. If the gauged group were noncompact, then defining the inner product would require *gauge fixing*.

[27] $\mathcal{H}$ is pronounced "curly H."

11 The electric field operators

This section defines some of the model's basic observables, namely the components of the electric field at a given locaion in space.

For each $U(1)$-valued link variable $u(\ell)$, define the angle-valued link variable $\theta(\ell)$ by writing $u(\ell)$ in the form (2), and define an operator $E(\ell)$ on $\mathcal{H}$ by

$$E(\ell)\Psi[u] \equiv i\kappa\frac{\partial}{\partial\theta(\ell)}\Psi[u], \tag{8}$$

where κ is a constant with units of mass/length that will be specified below.[28] The factor of i makes this operator self-adjoint. When $\ell = (\mathbf{x}, \mathbf{x}+\mathbf{e}_j)$, the operator $E(\ell)$ will also be denoted $E_j(\mathbf{x})$:

$$E_j(\mathbf{x}) \equiv E(\ell) \qquad\qquad \text{when } \ell = (\mathbf{x}, \mathbf{x}+\mathbf{e}_j). \tag{9}$$

These are the **electric field operators** representing the components of the electric field. The list of components $E_1(\mathbf{x}), E_2(\mathbf{x}), ..., E_D(\mathbf{x})$ will be abbreviated $\mathbf{E}(\mathbf{x})$.

The value of the coefficient κ in equation (8) is[29]

$$\kappa \equiv \frac{q^2}{\epsilon^{D-1}}, \tag{10}$$

where q is the magnitude of the smallest electric charge that we would want the model to include when the model is extended to include charged matter. In the context of the full standard model of particle physics, the appropriate value would be 1/3 the charge of a proton.[30]

[28]This article uses units in which the speed of light is 1.

[29]The variable denoted θ here is related to the variable that was denoted a in article 26542 by equation (18) in section 16 (which writes A instead of a). That's why (10) doesn't include a factor of $\hbar$. To relate the factor ϵ^{D-1} in the denominator to article 26542, note that article 26542 implicitly uses $\partial a_j(\mathbf{x})/\partial a_k(\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y})$, whose lattice version has ϵ^D in the denominator, and the integral in (18) cancels one of those factors of ϵ. The sign is consistent with article 26542, because $E^j = -E_j$ when the mostly-minus convention is used for the Minkowski metric.

[30]Section 28

12 $\mathcal{G}$-invariance and the electric field

Gauss's law (one of Maxwell's equations) is implicit in the fact that the Hilbert space uses only $\mathcal{G}$-invariant functions. Given a gauge transformation (3), we can define a set of angle variables $\phi(\mathbf{x})$, one for each point $\mathbf{x}$, by $h(\mathbf{x}) = \exp(i\phi(\mathbf{x}))$. If $\Psi[u]$ is any smooth complex-valued function of the link variables, not necessarily $\mathcal{G}$-invariant, then equation (3) implies

$$\frac{\partial}{\partial\phi(\mathbf{x})}\Psi[u^h] \propto \nabla\cdot\mathbf{E}(\mathbf{x})\,\Psi[u^h] \tag{11}$$

with $\nabla\cdot\mathbf{E} \equiv \sum_j \nabla_j E_j$, where ∇ is this lattice version of the gradient:

$$\nabla_j f(\mathbf{x}) \equiv \frac{f(\mathbf{x}) - f(\mathbf{x}-\mathbf{e}_j)}{\epsilon}.$$

If the function $\Psi[u]$ is $\mathcal{G}$-invariant, then (11) implies

$$\nabla\cdot\mathbf{E}(\mathbf{x})\Psi[u] = 0 \qquad\qquad \text{for all } h\in\mathcal{G}. \tag{12}$$

This is the quantum version of Gauss's law in a model where the quantum electromagnetic field is the only physical entity (no charged matter). Equation (12) is another way to write equation (4).

13 Wilson loop and Wilson line operators

This section defines more of the model's basic observables. Section 16 will explain how these observables relate to the magnetic field.

Every $\mathcal{G}$-invariant function $\Psi[u]$ represents an element of the Hilbert space $\mathcal{H}$ that was constructed in section 10. Any $\mathcal{G}$-invariant function $\omega[u]$ may also be used to define a linear operator W on $\mathcal{H}$, like this:[31]

$$W\Psi[u] \equiv \omega[u]\Psi[u] \qquad\qquad \text{for all } \Psi \in \mathcal{H}. \tag{13}$$

Section 9 described examples of $\mathcal{G}$-invariant functions. One example is the product $u(C)$ of link variables around a closed path C. The corresponding operator $W(C)$, defined by

$$W(C)\Psi[u] \equiv u(C)\Psi[u], \tag{14}$$

is called a **Wilson loop operator** or just **Wilson loop**.[32] An important special case of a Wilson loop is the **plaquette operator**

$$W(\square)\Psi[u] = u(\square)\Psi[u] \tag{15}$$

with $u(\square)$ defined by (6). Another example is the product $u(C)$ of link variables along a path C whose endpoints are boundary points. In this case, the operator defined by (14) is called a **Wilson line**.[33]

If $\omega[u]$ is not a $\mathcal{G}$-invariant function, then (13) does not define an operator on the Hilbert space, because the product $\omega[u]\Psi[u]$ is not $\mathcal{G}$-invariant and so does not belong to the Hilbert space. In particular, multiplication by a single link variable does not define an operator on this Hilbert space.

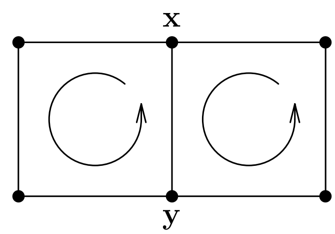
[31] All such operators clearly commute with each other. In particular, W commutes with its adjoint (defined by replacing $\omega[u]$ with its complex conjugate), so W is a **normal operator** (article 74088). This is important because operators representing observables should be normal. This is implicit in article 03431, using a relationship between normal operators and projection operators highlighted in article 74088.

[32] Some authors use the name *Wilson loop* for the expectation value of this operator, as in Montvay and Münster (1997), section 3.2.4. The way I'm using the name here is consistent with Peskin and Schroeder (1995), section 15.3.

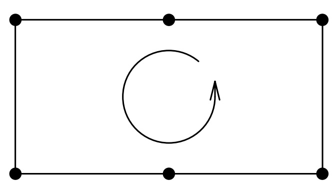
[33] Wilson loop operators and Wilson line operators are both called **line operators** to emphasize that they are localized along a one-dimensional curve, as opposed to being localized at a point (Aharony *et al* (2013), Gaiotto *et al* (2015)).

14 An identity

Consider two adjacent plaquettes, directed so that their shared link occurs with opposite directions, as shown here:



Call these two plaquettes $\square_1$ and $\square_2$, and define $u(\square_1)$ and $u(\square_2)$ as in equation (6). One plaquette includes the link $(\mathbf{y},\mathbf{x})$, and the other includes the link $(\mathbf{x},\mathbf{y})$. Equation (1) says that those two link variables cancel each other in the product $u(\square_1)u(\square_2)$, leaving the product of the six link variables around the perimeter of the pair, as illustrated here:



This is still true if the two adjacent plaquettes are not coplanar. This generalizes to any number of plaquettes whose shared links all occur in oppositely-directed pairs. Such a collection of plaquettes will be called **consistently directed**.

Consider a surface S formed by consistently directed plaquettes, and suppose for simplicity that its boundary ∂S is a single closed loop. Then the product of all of those plaquette variables satisfies

$$\prod_{\square\in S} u(\square) = \prod_{\ell\in\partial S} u(\ell) \tag{16}$$

because equation (1) says that the contributions of the other link variables (the ones that are not in ∂S) cancel each other in pairs, as illustrated above.

15 One definition of magnetic flux

Let S be any connected surface made from consistently oriented plaquettes, as in section 14, and suppose that its boundary ∂S is a single closed loop. Equation 16 shows that the Wilson loop operator $W(\partial S)$ that was defined in section 13 may be written

$$W(\partial S) = \prod_{\square \in S} W(\square),$$

so we can define an observable $B(S)$ corresponding to the **magnetic flux** through the surface S by

$$W(\partial S) = e^{iB(S)/\hbar}. \tag{17}$$

Sections 16-17 will help explain why $B(S)$ deserves to be called *magnetic flux*. The quantity (17) is unaffected when the $B(S)$ in the exponent is replaced by $B(S)+2\pi\hbar n$ for any integer n, so $B(S)$ is only defined modulo $2\pi\hbar$. Section 19 will introduce an alternate definition of magnetic flux that doesn't have this ambiguity. Sections 19-20 and 28-30 will explain how these two definitions relate to each other and to real-world experience.

16 Motivation for the definition (17)

Article 76708 introduces the concept of a **connection** on a principal G-bundle, which is the mathematical foundation for the general concept of a classical gauge field. A principal G-bundle associates a copy of the **fiber**, a smooth manifold that is almost the Lie group G but without the full structure of a group, to each point of the **base space**. In this article, the base space is ordinary three-dimensional physical space, generalized to D dimensions for more insight, and the group G is $U(1)$. A *connection* defines a way of lifting each path in the base space to a path through the collection of fibers. When the path in the base space is a closed loop C, the lifted path starts and ends in the same (copy of the) fiber, but not necessarily at the same point in the fiber. The transformation from one point to the other is reproduced by an element of G called the **holonomy** associated with the loop C.[34]

The quantity $u(C) \in G$ defined in equation (5) is a lattice version of the holonomy associated with C. To relate this to magnetic flux, consider classical electromagnetism. Article 76708 explains that when G is abelian, the holonomy has the form $\exp\left(i \int_C A\right)$, where A is the gauge field one-form and the integral is around the loop C. This is related to (5) through

$$\theta(\mathbf{y}, \mathbf{y}') = \frac{1}{\hbar} \int_{\mathbf{y}}^{\mathbf{y}'} A_k(\mathbf{x})\, dx^k, \tag{18}$$

where θ is the angle-valued link variable defined in (2). In classical electrodynamics with gauged group $\mathbb{R}$, if S is a two-dimensional surface whose boundary is C, then Stokes's theorem[35] gives

$$\int_C A = \int_S B, \tag{19}$$

where B is the magnetic field two-form. The integral $\int_S B$ defines the magnetic flux through the surface S in classical electromagnetism. This motivates the definition (17) of the magnetic flux operator in the quantum $U(1)$ model.

[34] If the gauged group were nonabelian, then this element of G would also depend on the starting point in the fiber.
[35] Article 91116

17 The magnetic flux through a closed surface

One of Maxwell's equations involves only the magnetic field. In continuous space, this equation may be written $dB = 0$, where B is the magnetic field two-form and d is the exterior derivative.[36] If S is a closed surface forming the boundary of a three-dimensional volume V, then[37,38]

$$\int_S B = \int_V dB = 0. \tag{20}$$

The magnetic flux defined in (17) also has this property, modulo $2\pi\hbar$. To show this, start with the identity (16). When ∂S is empty, that identity reduces to

$$\prod_{\square\in S} u(\square) = 1 \qquad\qquad \text{if } S \text{ is closed}$$

because all of the link variables cancel in pairs. Combine this with equations (15) and (17) to get

$$B(S) = 0 \text{ modulo } 2\pi\hbar \qquad\qquad \text{if } S \text{ is closed.} \tag{21}$$

This shows that the magnetic flux defined by equation (17) satisfies a lattice version of equation (20).

---

[36] Article 91116

[37] The first equality is another special case of Stokes's theorem $\int_{\partial M}\omega = \int_M d\omega$. This application uses $\omega = B$. Equation (19) uses $\omega = A$ and the relationship $B = dA$.

[38] Stokes's theorem assumes that the differential form ω in footnote 37 has compact support (article 91116). To appreciate why this is important, suppose that V is a 3-dimensional ball with one interior point deleted. Deleting that point makes V a non-compact manifold, and a two-form B on that manifold can satisfy both $dB = 0$ and $\int_{\partial V} B \neq 0$. This does not contradict (20) because (20) assumes that B has compact support as a two-form on V, but the condition $\int_{\partial V} B \neq 0$ requires B to have non-compact support as a two-form on V. (Specifically, it implies that B would have a singularity at the deleted point if that point were not deleted.) Having compact support on the boundary $S = \partial V$ is not sufficient.

18 A property of the low-energy limit

Space is being treated as a lattice to make the math clear, but we are ultimately only interested in taking a *continuum limit* in which lattice artifacts are negligible, so the energy range of interest is

$$\text{energy} < \xi\,\frac{\hbar}{\epsilon} \tag{22}$$

where ϵ is the lattice spacing and[39]

$$\xi \ll 1 \tag{23}$$

When space is three-dimensional ($D = 3$), the form of the hamiltonian that will be introduced in section 23 shows that the condition (22) requires using states in which the dimensionless quantity

$$\chi(\square) \equiv 2 - W(\square) - W^{-1}(\square)$$

effectively satisfies

$$\chi(\square) < \xi\,\frac{q^2}{\hbar} \qquad \text{when } D = 3. \tag{24}$$

The ratio $q^2/\hbar$ is $\lesssim 1$ in the real world,[40] so the conditions (23) and (24) imply

$$\chi(\square) \ll 1 \qquad \text{when } D = 3. \tag{25}$$

This will be important in section 19.

[39] This article uses units in which the speed of light is 1 (footnote 28 in section 11), so the quantity $\hbar/\epsilon$ has the same units as energy.

[40] In the system of units used here, $q^2/\hbar$ is the **fine structure constant** except for a factor of order 1. When $D = 3$, the model's properties are qualitatively different for large and small values of $q^2/\hbar$ (section 27). In QED with matter, *perturbation theory* relies on the smallness of $q^2/\hbar$.

19 Another definition of magnetic flux

If S is a surface made from consistently oriented plaquettes, then we can define a different observable $\overline{B}(\square)$ by

$$\overline{B}(S) \equiv \sum_{\square\in S} \overline{B}(\square) \qquad \text{with } \overline{B}(\square) \equiv \frac{W(\square) - W^{-1}(\square)}{2i}\,\hbar. \tag{26}$$

This observable is different than the observable $B(S)$ defined by equation (17), but this one also deserves to be called **magnetic flux**. To understand why, use (23) and (24) to infer that the flux $\overline{B}(\square)$ through a single plaquette is effectively restricted by the condition

$$(\overline{B}(\square)/\hbar)^2 \lesssim \xi\,\frac{q^2}{\hbar} \qquad \text{when } D = 3 \tag{27}$$

in the continuum limit. Use this with (23) to get the approximation

$$\exp\big(iB(\square)/\hbar\big) \approx \exp\big(i\overline{B}(\square)/\hbar\big) \qquad \text{when } D = 3.$$

The quantity ξ is arbitrarily small in the continuum limit, so if S is a macroscopic surface (made of an enormous number of plaquettes), then we can take ξ to be small enough so that the approximation

$$\exp\big(iB(S)/\hbar\big) \approx \exp\big(i\overline{B}(S)/\hbar\big) \qquad \text{when } D = 3$$

holds with negligible error in the continuum limit.

One appealing property of the earlier definition (17) is that $B(S)$ is exactly zero (modulo $2\pi\hbar$) whenever the surface S is closed (equation (21)). The observable $\overline{B}(S)$ defined by (26) only approaches this property in the low-energy limit, but sections 28-30 will explain why $\overline{B}(S)$ is a better representation of the magnetic flux that we actually measure in experiments. The key is to remember that measurements are physical processes. Any description of physical processes within the model uses its equations of motion (section 25), and those equations directly involve the quantities $W(\square)$.

20 The magnetic field

In continuous space, the components of the magnetic field may be expressed in terms of the magnetic flux like this:

$$B_{jk}(\mathbf{x}) = \lim_{\alpha(S)\to 0} \frac{B(S)}{\alpha(S)} \qquad \text{(in continuous space)}, \tag{28}$$

where S is a surface element[41] in the j-k plane with area $\alpha(S)$ containing the point $\mathbf{x}$. On a lattice, the minimum possible area is ϵ^2, where ϵ is the distance between neighboring lattice sites, so by analogy with (28), we can use either of these definitions:

$$B_{jk}(\mathbf{x}) \equiv B(\square)/\epsilon^2 \tag{29}$$

$$\overline{B}_{jk}(\mathbf{x}) \equiv \overline{B}(\square)/\epsilon^2 \tag{30}$$

where $\square$ is the plaquette whose links trace through this sequence of points:

$$\mathbf{x} \;\to\; \mathbf{x}+\mathbf{e}_j \;\to\; \mathbf{x}+\mathbf{e}_j+\mathbf{e}_k \;\to\; \mathbf{x}+\mathbf{e}_k \;\to\; \mathbf{x}. \tag{31}$$

This is consistent with the continuous-space relationship $B_{jk} = \nabla_j A_k - \nabla_k A_j$ when θ is related to A by equation (18).

In the definition (29), the flux is defined only modulo $2\pi\hbar$, so the field $B_{jk}(\mathbf{x})$ is defined only modulo $2\pi\hbar/\epsilon^2$. Section 29 will show that the period $2\pi\hbar/\epsilon^2$ is much larger than the magnetic field magnitudes encountered in real experiments, so in the context of states with low enough energy to be relevant to such experiments, we can refine the definition (29) by selecting the value with the smallest magnitude. That makes definitions (29) and (30) interchangeable in the low-energy limit.[42]

The definition (29) will be more convenient when deriving the formal continuum limit of the hamiltonian in section 24.

[41]To make this definition unambiguous, an orientation (one of the two directions around the boundary C) would need to be specified. The sequence (31) does that for the lattice version.

[42]Section 19

21 A commutation relation

Define $E(\ell)$ as in section 11, and let $W(C)$ be any Wilson loop or Wilson line as defined in section 13. Let ℓ^{rev} denote the link obtained by reversing the direction of ℓ, so if $\ell = (\mathbf{x}, \mathbf{y})$, then $\ell^{\text{rev}} = (\mathbf{y}, \mathbf{x})$. If the loop C does not intersect itself, then the definitions of $E(\ell)$ and $W(C)$ imply

$$\big[E(\ell), W(C)\big] = \begin{cases} -\kappa W(C) & \text{if } \ell \in C, \\ \;\;\kappa W(C) & \text{if } \ell^{\text{rev}} \in C, \\ \;\;0 & \text{otherwise,} \end{cases} \tag{32}$$

using the standard notation $[A, B] \equiv AB - BA$. In particular, the electric field operator $E(\ell)$ commutes with a Wilson loop or Wilson line $W(C)$ if the loop C does not include the link ℓ or its oppositely-directed version ℓ^{rev}. If it does, then $E(\ell)$ doesn't commute with $W(C)$.

The commutation relation (32) is a lattice version of the commutation relation shown in article 26542. To infer this, use the fact that the representations of the electric and magnetic field operators in sections 11 and 15 are lattice versions of the representations of the electric and magnetic field operators in article 26542. The commutation relations are consequences of those representations, both in smooth space and on the lattice.

22 Time dependent observables

In quantum field theory, part of the task of defining a model is to associate observables with regions of spacetime.[43] Sections 11 and 13 defined the model's basic observables at time $t = 0$. These are the electric field operators $E(\ell)$ and the Wilson loop and Wilson line operators $W(C)$. This section defines observables at arbitrary times t in terms of those at $t = 0$, using a hamiltonian that will be specified in section 23.

If R is any region of space, then **observables** localized in R at time t are represented by operators of the form

$$\mathcal{O}(t) \equiv U(-t)\mathcal{O}U(t) \tag{33}$$

with

$$U(t) \equiv e^{-iHt/\hbar}, \tag{34}$$

where $\mathcal{O}$ is any linear operator on the Hilbert space – like an electric field operator $E(\ell)$ or a Wilson loop $W(C)$ – that can be expressed using only links in R, and the hamiltonian is defined by (36). The hamiltonian is (unbounded but) self-adjoint, so the operators (34) are unitary, and $\mathcal{O}(t)$ is self-adjoint for each t if $\mathcal{O}$ is self-adjoint.

The Hilbert space consists only of $\mathcal{G}$-invariant functions, so if $\mathcal{O}$ is a linear operator on the Hilbert space, then applying $\mathcal{O}$ to any $\mathcal{G}$-invariant function gives another $\mathcal{G}$-invariant function. In this sense, observables are $\mathcal{G}$-invariant.

Equations (32) and (33) imply

$$\big[E(\ell,t),\, W(\square,t)\big] = U^{-1}(t)\big[E(\ell),\, W(\square)\big]U(t) = \big[E(\ell),\, W(\square)\big] \tag{35}$$

for all t. This is the **equal-time commutation relation**. It is a lattice version of the commutation relation between the electric and magnetic field operators that was described in article 26542.

[43] Article 21916

23 The hamiltonian

This section introduces the hamiltonian, the operator that section 22 used to define the time dependence of the model's observables. The hamiltonian is[44,45]

$$
\begin{aligned}
H &= \frac{1}{q^2}\left(\epsilon^D\sum_\ell \frac{E^2(\ell)}{4} + \epsilon^D\sum_\square \frac{1-W(\square)}{2\epsilon^4}\hbar^2\right) + \text{constant}\\
&= \frac{1}{q^2}\left(\epsilon^D\sum_\ell \frac{E^2(\ell)}{4} + \epsilon^D\sum_\square \frac{2-W(\square)-W^\dagger(\square)}{4\epsilon^4}\hbar^2\right) + \text{constant}.
\end{aligned}
\tag{36}
$$

The operators $E(\ell)$ and $W(\square)$ are defined by equations (8) and (15). The sum over ℓ is over all directed links that have associated link variables.[46] The sum over $\square$ is over all oriented plaquettes,[47] including plaquettes that involve fewer than four link variables.[48,49]

For any t, equation (33) and the obvious identity $U^{-1}(t)HU(t) = H$ imply that the hamiltonian (36) may also be written[50]

$$
H = \frac{1}{q^2}\left(\epsilon^D\sum_\ell \frac{E^2(\ell,t)}{4} + \epsilon^D\sum_\square \frac{1-W(\square,t)}{2\epsilon^4}\hbar^2\right) + \text{constant}. \tag{37}
$$

Section 24 will show that this is a lattice version of the more familiar hamiltonian for electrodynamics in continuous space.

[44] This is the **Kogut-Susskind hamiltonian** specialized to the gauged group $G = U(1)$.

[45] The two expressions for H are equal because reversing the direction of a plaquette is the same as replacing $W(\square) \to W^\dagger(\square)$.

[46] Recall that only links with at least one interior endpoint have associated link variables (section 7).

[47] Equation (36) has an extra factor of 2 in the denominator compared to equation (3.66) in Montvay and Münster (1997), because the sum in their equation includes only one orientation of each unoriented plaquette.

[48] Harlow and Ooguri (2021), text below equation (3.26)

[49] A plaquette always involves four links, but they might not all have associated link variables (section 7). Including these plaquettes makes the hamiltonian depend on all of the link variables, including those with only one interior endpoint.

[50] Footnote 45 in section 23

24 Formal continuum limit of the hamiltonian

This section shows that the hamiltonian (36) is a lattice version of the more familiar hamiltonian for electrodynamics in continuous space.

Use the relationships (17) and (29) to get

$$\begin{aligned}\sum_{\square}\frac{2-W(\square)-W^\dagger(\square)}{4\epsilon^4}\hbar^2 &= \sum_{\square}\frac{2-e^{iB(\square)/\hbar}-e^{-iB(\square)/\hbar}}{4\epsilon^4}\hbar^2\\ &= \sum_{\mathbf{x},j,k}\frac{2-e^{i\epsilon^2 B_{jk}(\mathbf{x})/\hbar}-e^{-i\epsilon^2 B_{jk}(\mathbf{x})/\hbar}}{4\epsilon^4}\hbar^2\\ &= \sum_{\mathbf{x},j,k}\frac{\big(B_{jk}(\mathbf{x})\big)^2}{4}+O(\epsilon^2).\end{aligned}$$

Use this and (9) in (36) to get

$$\begin{aligned}H &\approx \frac{1}{q^2}\left(\epsilon^D\sum_{\ell}\frac{E^2(\ell)}{4}+\epsilon^D\sum_{\mathbf{x},j,k}\frac{\big(B_{jk}(\mathbf{x})\big)^2}{4}\right)+\text{constant}\\ &= \frac{1}{2q^2}\,\epsilon^D\sum_{\mathbf{x}}\left(\sum_j E_j^2(\mathbf{x})+\sum_{j<k}\big(B_{jk}(\mathbf{x})\big)^2\right)+\text{constant},\end{aligned}$$

where "≈" means up to terms that are negligible when the resolution is low compared to the lattice scale ϵ, ignoring the fact that the magnetic flux is only defined modulo $2\pi\hbar$.[51] This matches the form of the hamiltonian that was used in article 26542.

[51]Section 16 explained why it's only defined modulo $2\pi\hbar$.

25 Equations of motion

This section derives expressions for the time derivatives of $E(\ell,t)$ and $W(C,t)$, and section 26 will show that the resulting equations are a lattice version of Maxwell's equations.[52] This section ignores the long-distance cutoff (section 6).

Use equation (37) to get[53]

$$
\begin{aligned}
i\hbar\dot{E}(\ell,t) &= \big[E(\ell,t),H\big] = \frac{-\hbar^2}{2q^2}\epsilon^{D-4}\sum_{\square}\big[E(\ell,t),\,W(\square,t)\big] \\
i\hbar\dot{W}(\square,t) &= \big[W(\square,t),H\big] = \frac{1}{4q^2}\epsilon^{D}\sum_{\ell}\Big(E(\ell,t)\big[W(\square,t),\,E(\ell,t)\big] \\
&\qquad\qquad + \big[W(\square,t),\,E(\ell,t)\big]E(\ell,t)\Big),
\end{aligned}
$$

and then use the commutation relations (32) and (35) to get

$$
i\dot{E}(\ell,t) = \frac{\kappa\hbar}{2q^2}\epsilon^{D-4}\left(\sum_{\square\ni\ell}W(\square,t) - \sum_{\square\ni\ell^{\text{rev}}}W(\square,t)\right) \tag{38}
$$

$$
\begin{aligned}
i\hbar\dot{W}(\square,t) &= \frac{\kappa}{4q^2}\left(\epsilon^{D}\sum_{\ell\in\square}\Big(E(\ell,t)W(\square,t)+W(\square,t)E(\ell,t)\Big)\right. \\
&\qquad \left. - \,\epsilon^{D}\sum_{\ell^{\text{rev}}\in\square}\Big(E(\ell,t)W(\square,t)+W(\square,t)E(\ell,t)\Big)\right) \\
&= \frac{\kappa}{2q^2}\epsilon^{D}\sum_{\ell\in\square}\Big(E(\ell,t)W(\square,t)+W(\square,t)E(\ell,t)\Big).
\end{aligned} \tag{39}
$$

The sum $\sum_{\square\ni\ell}$ is over all directed plaquettes that include the directed link ℓ, and the sum $\sum_{\ell\in\square}$ is over all directed links that occur in the directed plaquette $\square$.

[52] Some of Maxwell's equations don't involve time derivatives. Sections 12 and 16 already showed how those are reproduced in this model.

[53] $\dot{X}$ denotes the derivative of X with respect to t.

26 Relationship to Maxwell's equations

To show that equation (38) is a lattice version of one of Maxwell's equations, use equation (17) and ignore the modulo-$2\pi\hbar$ ambiguity in the magnetic flux to get

$$\dot{E}(\ell,t) = \frac{\kappa}{q^2}\epsilon^{D-4}\sum_{\square\ni\ell} B(\square,t) + O(B^2).$$

Then use (9), (29), and (31) to get

$$\begin{aligned}\dot{E}_j(\mathbf{x},t) &= \frac{\kappa}{q^2}\epsilon^{D-2}\left(\sum_k \big(B_{jk}(\mathbf{x},t) - B_{jk}(\mathbf{x}-\mathbf{e}_k,t)\big) + O(\epsilon^2)\right)\\ &= \frac{\kappa}{q^2}\epsilon^{D-1}\left(\sum_k \nabla_k B_{jk}(\mathbf{x},t) + O(\epsilon^2)\right)\end{aligned} \tag{40}$$

where ∇ is a lattice version of the gradient. To show that equation (39) is a lattice version of one of Maxwell's equations, use equations (17) and (29) on the right-hand side to get

$$i\hbar\dot{W}(\square,t) = \frac{\kappa}{q^2}\epsilon^{D}\left(\sum_{\ell\in\square} E(\ell,t) + O(\epsilon^2)\right).$$

Then use (17) and (29)-(31) again on the left-hand side to get

$$\begin{aligned}-\dot{B}_{jk}(\mathbf{x},t) &= \frac{\kappa}{q^2}\epsilon^{D-2}\big(E_j(\mathbf{x}) + E_k(\mathbf{x}+\mathbf{e}_j) - E_j(\mathbf{x}+\mathbf{e}_k) - E_k(\mathbf{x})\big)\\ &= \frac{\kappa}{q^2}\epsilon^{D-1}\big(\nabla_j E_k(\mathbf{x},t) - \nabla_k E_j(\mathbf{x},t) + O(\epsilon^2)\big)\end{aligned} \tag{41}$$

where ∇ is another lattice version of the gradient. When κ is given by (10), equations (40) and (41) agree with the zero-current case of Maxwell's equations as presented in article 31738, up to terms that are negligible when the resolution is low compared to ϵ.

27 Notes about the continuum limit

Equations (38)-(39) look more complicated than Maxwell's equations in smooth space, for two reasons. One reason is that space is being treated here as a lattice instead of as a continuum, so we have sums $\epsilon^D \sum_{\mathbf{x}} \cdots$ in place of integrals $\int d^D x \cdots$. Another reason is that these equations are nonlinear in the basic observables E and W, both explicitly and implicitly. The products of E with W on the right-hand side make them explicitly nonlinear, and a further nonlinearity is implicit in the constraint $W^\dagger W = 1$.

The fact that the equations of motion are nonlinear makes extracting this model's predictions more difficult. Even something as basic as the existence of a Lorentz-symmetric continuum limit governed by Maxwell's equations is far from obvious. As a concession, section 26 showed that equations (38)-(39) are a lattice version of Maxwell's equations, in the sense that they reproduce Maxwell's equations when the parameter ϵ is formally sent to zero, ignoring the noncommutativity of the operators and the periodicity of the magnetic flux. The periodicity of the flux does need to be taken into account when studying the continuum limit of the model as a whole, though. The rest of this section summarizes some insights from such studies.

For $D \geq 3$, studies using the path-integral formulation have shown that compact QED has a **Coulomb phase** with massless photons when the overall coefficient of the action is large enough. For a given ϵ, the coefficient of the action is proportional to $1/q^2$, just like the coefficient of the hamiltonian, so the Coulomb phase with massless photons occurs when q^2 is small enough (for fixed ϵ). Masslessness implies infinite correlation length, so this is evidence that the expected continuum limit exists.[54] For larger values of q^2, compact QED is in a **confinement phase** with no massless particles. When $D = 3$, the phase transition between the Coulomb phase and the confinement phase appears to be weakly first order.[55] That means

[54]These statements are based on Frölich and Spencer (1982), section 2.11 (for $D = 3$) and remark 1 at the end of section 2.12 (for $D > 3$). Their analytic results agree with numerical studies for $D = 3$, recent examples of which include Lewis and Woloshyn (2018), Loveridge *et al* (2021), and Loveridge *et al* (2021b).

[55]Torres *et al* (2024), section 1

that the correlation length doesn't diverge in units of the lattice spacing ϵ when the transition is approached from the confinement phase, so a strict continuum limit probably doesn't exist for the confinement phase[56] even though it does for the Coulomb phase.

For $D = 2$, compact QED is in the confinement phase for all nonzero values of q^2 when the lattice spacing is finite, but what happens in the continuum limit depends on how the continuum limit is defined.[57] By definition of *continuum limit*, the correlation length in units of the lattice spacing must diverge, but the correlation length in unspecified units may either diverge or remain finite. We are free to choose which of these two options we use to define the limit.[58] If we choose the limit in which the correlation length diverges, then the result is a model of free massless photons.[59,60] If we choose the limit in which the correlation length remains finite (even though it diverges in units of the lattice spacing), then the result is a model whose only particle is a massive spinless particle with no interactions.[61] Both continuum limits are valid, and interactions are absent in both of them.[62]

---

[56]Majumdar *et al* (2004), Espriu and Tagliacozzo (2003)

[57]Athenodorou and Teper (2019), paragraph containing equations (3.1)-(3.3)

[58]Article 07611 explains this in more detail.

[59]This limit is mentioned in Banks *et al* (1977) (second-to-last paragraph in section 2, with additional insight in the text below equation (A.9) about what happens as the limit is approached) and in Harlow and Ooguri (2021) (footnote 47 on page 59).

[60]Article 26542 shows that the angular momentum of a photon is zero when $D = 2$, but they are still often called *photons*.

[61]This limit is used in Göpfert and Mack (1982) and in Athenodorou and Teper (2019).

[62]Confinement – defined as a linear potential between static external charges – is also absent in both continuum limits (Athenodorou and Teper (2019), section 3.4 and the text below equation (3.3)).

28 Quantifying the magnetic flux period

Equation (17) defines the magnetic flux $B(S)$ only modulo the period $2\pi\hbar$. Instead of using the units convention that was used in this article,[63] we can remove a factor of q from the definitions of the electric and magnetic fields to recover the engineering convention. With that convention, the period is $2\pi\hbar/q$. As explained in section 4, the constant q should be the magnitude of the smallest nonzero electric charge that the model includes after charged matter is included.[64] The smallest nonzero electric charge in the standard model of particle physics is 1/3 the charge of a proton. If we use this as the value of q, then the period is[65]

$$\frac{2\pi\hbar}{q} \approx 10^{-14} \text{ weber} \tag{42}$$

in standard international units. The assertion that magnetic flux is defined only modulo $2\pi\hbar/q$ means that for any given surface S, values of the magnetic flux $B(S)$ that differ from each other by an integer multiple of the quantity (42) are physically equivalent to each other.

For comparison, a typical value for the flux of the earth's magnetic field through one square meter is more than 10^{-5} weber,[66] which is enormous compared to (42), and other values we encounter routinely are even larger. The alternate definition that was given in section 19, which is not periodic, is evidently a better representation of the magnetic flux that we determine through real measurements. Section 30 will explain how this can be anticipated by thinking about how the model would describe the physical process of measurement.

[63] The units convention used in this article is described in detail in article 26542.

[64] If a nonzero electric charge could be arbitrarily small, then the flux period would be infinite – the gauged group would be $\mathbb{R}$ instead of $U(1)$.

[65] The **flux quantum** that is famous in the study of conventional (BCS) superconductivity has the form $2\pi\hbar/q$, but in that case q is two times the magnitude of an electron's charge. References are cited in Loder *et al* (2007), which points out that the flux quantum can have a different value in unconventional superconductors.

[66] https://climate.nasa.gov/news/3105/

29 Quantifying the magnetic field period

Section 20 used the magnetic flux through a plaquette $\square$ to define the magnetic field like this:

$$\text{magnetic field} = \frac{\text{magnetic flux through } \square}{\epsilon^2} \tag{43}$$

where ϵ is the distance between neighboring lattice sites. If we use the definition (17) for the magnetic flux, then the periodicity of the magnetic flux implies a periodicity for the magnetic field. This section quantifies that period.

The lattice is artificial, so when we treat space as a lattice for the purpose of defining the model, we should take ϵ to be much smaller than any practical measurement can resolve. As an example, suppose we choose $\epsilon \sim 10^{-20}$ meter, so that the lattice is indistinguishable from a continuum for most practical purposes. Then the period (42) for the magnetic flux implies this period for the magnetic field:

$$\frac{2\pi\hbar/q}{\epsilon^2} \sim 10^{25} \text{ tesla.} \tag{44}$$

This is much greater than any of the magnetic field strengths that been measured so far.[67] This confirms that we can refine the definition (29) by selecting the value with the smallest magnitude, as stated at the end of section 20.

If we could take the strict $\epsilon \to 0$ limit of compact QED, then the range of distinguishable values of the magnetic field strengths would be unbounded even with the definition (29). A nontrivial strict limit $\epsilon \to 0$ might be obstructed when charged matter is included[68] but that's not a problem for physical applications, because (44) is plenty big enough already.

[67] According to the text below equation (1) in Kong *et al* (2022), the magnetic field of a particular neutron star, with a strength of $\sim 10^{13}$ Gauss $= 10^9$ Tesla, became the new record-holder in 2022.

[68] Göckeler *et al* (1998a) and Göckeler *et al* (1998b)

30 Measuring magnetic flux

To study the physical process of measurement using a model that includes the microscopic details of the measurement equipment would be too difficult, and the model defined in this article doesn't include those details anyway – it doesn't include any matter at all. This model does share some features with more comprehensive models, though, and we can use those features to anticipate that the definition of *magnetic flux* in equation (26) is a more direct representation of what we actually measure than the definition given by equation (17).

Measurements are physical processes, and to describe physical processes within a model, we use its equations of motion. In the present model, the equations of motion (38)-(39) involve only single-valued functions of $W(\square)$. The corresponding equations of motion in compact QED with matter also have that property,[69] so the quantities that we measure in experiments should also expressible as single-valued functions of $W(\square)$. The definition of $\overline{B}(\square)$ in equation (26) satisfies this condition. The definition of $B(\square)$ in equation (17) does not.

This suggests that compared to the definition (17), the definition (26) of the magnetic flux and the corresponding definition (43) of the field is a better representation of what we actually measure in experiments,[70] after a suitable degree of smearing[71] to make the resolution much coarser than the lattice spacing ϵ.

[69]They also involve other gauge invariant combinations of link variables and matter fields, but that doesn't change the message here.

[70]The derivation of the formal continuum limit of the hamiltonian in section 24 and of the equations of motion in section 26 used the definition (17), but that's justified because those derivations implicitly use the refinement that was described at the end of section 20, and that refinement makes the two definitions of the magnetic field in section 20 consistent with each other at sufficiently low energies.

[71]**Smearing** a local observable means forming a weighted sum of translated versions of the observable over a region whose size corresponds to the resolution (article 22792).

31 References

(Open-access items include links.)

Aharony *et al*, 2013. "Reading between the lines of four-dimensional gauge theories" *JHEP* **2013(08)**: 115, https://doi.org/10.1007/JHEP08(2013)115

Athenodorou and Teper, 2019. "On the spectrum and string tension of U(1) lattice gauge theory in 2 + 1 dimensions" *JHEP* **2019(01)**: 063, https://doi.org/10.1007/JHEP01(2019)063

Banks *et al*, 1977. "Phase transitions in Abelian lattice gauge theories" *Nuclear Physics B* **129**: 493-510, https://lib-extopc.kek.jp/preprints/PDF/1977/7709/7709073.pdf

Baumann *et al*, 1988. "Experimental limit for the charge of the free neutron" *Phys. Rev. D* **37**: 3107-3112

Bousso, 2002. "The holographic principle" *Rev. Mod. Phys.* **74**: 825-874, https://arxiv.org/abs/hep-th/0203101

Dylla and King, 1973. "Neutrality of Molecules by a New Method" *Phys. Rev. A* **7**: 1224-1229

Espriu and Tagliacozzo, 2003. "Compact lattice U(1) and Seiberg-Witten duality" *Phys. Lett. B* **557**: 125-133, https://arxiv.org/abs/hep-th/0301086

Frölich and Spencer, 1982. "Massless Phases and Symmetry Restoration in Abelian Gauge Theories and Spin Systems" *Commun. Math. Phys.* **83**: 411-454, https://projecteuclid.org/euclid.cmp/1103920870

Gaiotto *et al*, 2015. "Generalized global symmetries" *JHEP* **2015(02)**: 172, https://doi.org/10.1007/JHEP02(2015)172

Göckeler *et al*, 1998a. "Is there a Landau Pole Problem in QED?" *Phys. Rev. Lett.* **80**: 4119-4122, https://arxiv.org/abs/hep-th/9712244

Göckeler *et al*, 1998b. "Resolution of the Landau pole problem in QED" *Nucl. Phys. Proc. Suppl.* **63**: 694-696, https://arxiv.org/abs/hep-lat/9801004

Göpfert and Mack, 1982. "Proof of Confinement of Static Quarks in 3-Dimensional U(1) Lattice Gauge Theory for all Values of the Coupling Constant" *Commun. Math. Phys.* **82**: 545-606, https://projecteuclid.org/euclid.cmp/1103920658

Harlow and Ooguri, 2021. "Symmetries in Quantum Field Theory and Quantum Gravity" *Communications in Mathematical Physics* **383**: 1669-1804, https://arxiv.org/abs/1810.05338

Kong *et al*, 2022. "Insight-HXMT discovery of the highest energy CRSF from the first Galactic ultra-luminous X-ray pulsar Swift J0243.6+6124" *The Astrophysical Journal Letters* **933**: L3, https://doi.org/10.3847/2041-8213/ac7711

Lewis and Woloshyn, 2018. "Photon operators for lattice gauge theory" *Phys. Rev. D* **98**: 034502, https://doi.org/10.1103/PhysRevD.98.034502

Loder *et al*, 2007. "Magnetic flux periodicity of h/e in superconducting loops" *Nature Physics* **4**: 112-115, https://doi.org/10.1038/nphys813

Loveridge *et al*, 2021. "The lattice Landau gauge photon propagator for 4D compact QED" *Phys. Rev. D* **103**: 094519, https://doi.org/10.1103/PhysRevD.103.094519

Loveridge *et al*, 2021b. "Lattice pure gauge compact QED in the Landau gauge: the photon propagator, the phase structure and the presence of Dirac strings" *Phys. Rev. D* **104**: 114511, https://doi.org/10.1103/PhysRevD.104.114511

Majumdar *et al*, 2004. "Glueball masses in 4d U(1) lattice gauge theory using the multi-level algorithm" *Nucl. Phys. B* **677**: 273-288, https://arxiv.org/abs/hep-lat/0309003

Marinelli and Morpurgo, 1984. "The electric neutrality of matter: A summary" *Physics Letters B* **137**: 439-442

Montvay and Münster, 1997. *Quantum Fields on a Lattice.* Cambridge University Press

Peskin and Schroeder, 1995. *An Introduction to Quantum Field Theory.* Addison-Wesley

Torres *et al*, 2024. "Tricriticality in 4D U(1) Lattice Gauge Theory" https://arxiv.org/abs/2404.15907

Witten, 2023. "Algebras, Regions, and Observers" https://arxiv.org/abs/2303.02837

32 References in this series

Article **03431** (https://cphysics.org/article/03431):
"The Core Principles of Quantum Theory and the Nature of Measurement"

Article **07611** (https://cphysics.org/article/07611):
"Asymptotic Freedom and the Continuum Limit of Yang-Mills Theory"

Article **21916** (https://cphysics.org/article/21916):
"Local Observables in Quantum Field Theory"

Article **22792** (https://cphysics.org/article/22792):
"The Free Scalar Quantum Field: Waves"

Article **22871** (https://cphysics.org/article/22871):
"Time Evolution in Quantum Theory"

Article **26542** (https://cphysics.org/article/26542):
"The Free Quantum Electromagnetic Field in Smooth Space"

Article **31738** (https://cphysics.org/article/31738):
"The Electromagnetic Field and Maxwell's Equations in Any Number of Dimensions"

Article **52890** (https://cphysics.org/article/52890):
"Defining Scalar Quantum Fields on a Spatial Lattice"

Article **61813** (https://cphysics.org/article/61813):
"Homotopy, Homotopy Groups, and Covering Spaces"

Article **70621** (https://cphysics.org/article/70621):
"Principal Bundles and Associated Vector Bundles"

Article **74088** (https://cphysics.org/article/74088):
"Linear Operators on a Hilbert Space"

Article **76708** (https://cphysics.org/article/76708):
"Connections, Local Potentials, and Classical Gauge Fields"

Article **91116** (https://cphysics.org/article/91116):
"A Quick Review of Differential Forms and Stokes's Theorem"