

Phase structure of self-dual lattice gauge theories in 4d

Mariia Anosova,^a Christof Gattringer,^{b,1} Nabil Iqbal^c and Tin Sulejmanpasic^c

^aUniversität Graz, Institut für Physik,²
 Universitätsplatz 5, 8010 Graz, Austria

^bFWF Austrian Science Fund,
 Sensengasse 1, 1090 Vienna, Austria

^cDepartment of Mathematical Sciences, Durham University,
 Stockton Road, Upper Mountjoy, Durham DH1 3LE, United Kingdom

E-mail: mariia.anosova@uni-graz.at, christof.gattringer@fwf.ac.at,
nabil.iqbal@durham.ac.uk, tin2019@gmail.com

ABSTRACT: We discuss U(1) lattice gauge theory models based on a modified Villain formulation of the gauge action, which allows coupling to bosonic electric and magnetic matter. The formulation enjoys a duality which maps electric and magnetic sectors into each other. We propose several generalizations of the model and discuss their 't Hooft anomalies. A particularly interesting class of theories is the one where electric and magnetic matter fields are coupled with identical actions, such that for a particular value of the gauge coupling the theory has a self-dual symmetry. The self-dual symmetry turns out to be a generator of a group which is a central extension of \mathbb{Z}_4 by the lattice translation symmetry group. The simplest case amenable to numerical simulations is the case when there is exactly one electrically and one magnetically charged boson. We discuss the phase structure of this theory and the nature of the self-dual symmetry in detail. Using a suitable worldline representation of the system we present the results of numerical simulations that support the conjectured phase diagram.

KEYWORDS: Lattice Quantum Field Theory, Phase Transitions, Anomalies in Field and String Theories, Spontaneous Symmetry Breaking

ARXIV EPRINT: [2203.14774](https://arxiv.org/abs/2203.14774)

¹On leave from: Universität Graz, Institut für Physik, Universitätsplatz 5, 8010 Graz, Austria.

²Member of NAWI Graz.

Contents

1	Introduction	1
2	Self-dual modified Villain models	4
2.1	The gauge field partition sum	4
2.2	Duality transformation	6
2.3	Self-duality and self-dual symmetry	8
2.4	Coupling electric and magnetic matter	10
2.5	Computable self-dual limits	11
2.6	Field-theoretical description away from self-duality	13
2.7	Generalization of the lattice model, symmetries and anomalies	15
2.7.1	General charge theories and 1-form 't Hooft anomalies	15
2.7.2	General numbers of electric and magnetic flavors	17
2.7.3	The self-dual non-abelian gauge theory	19
2.8	The phase diagram of self-dual scalar QED	19
3	Numerical simulation	21
3.1	Switching to a dual worldline formulation	21
3.2	The self-dual point revisited	23
3.3	Setup of the computation and general results	24
3.4	Analysis of the endpoints	27
4	Conclusion and future prospects	29
A	The RG equations	31
A.1	Setup	31
A.2	The RG analysis of scalar QED	34
A.2.1	Solving the equations (A.8)	36

1 Introduction

U(1) gauge theories are widely thought to be effective theories, not only for electrodynamics, but also for many condensed matter systems. In four-dimensional spacetime it is the common lore that the U(1) theory of electrodynamics is an effective theory, possibly arising from a non-abelian gauge theory in the UV. Such UV completions always require magnetic monopoles, whose masses and effects are fixed by the UV theory. Furthermore, it is well known that monopoles also emerge in Wilson's lattice regularization of U(1) gauge theories. In both of these UV completions, the simplest U(1) gauge theory necessarily has

minimally charged magnetic monopoles, even when no electric field is present. It is tempting to conclude that any UV completion of a U(1) gauge theory must necessarily contain unit charge magnetic monopoles.

The first hint that this cannot be true in general comes from lower dimensional anti-ferromagnetic spin-systems. Such systems are known to be well described by an emergent U(1) gauge theory. The most familiar example is that of an anti-ferromagnetic spin chain that is effectively described by a 1+1d O(3) model [1], which in turn flows to an effective 1+1d abelian gauge theory. In two spatial dimensions, anti-ferromagnets have a similar emergent U(1) gauge theory, but now in 2+1 space-time dimensions. Such an effective theory naively has a U(1) conserved current¹ $j_T^\mu = \frac{1}{2\pi} \epsilon^{\mu\nu\rho} F_{\nu\rho}$,² where $F_{\mu\nu}$ is the U(1) field strength tensor. But no such U(1) symmetry exists in the UV spin system, and so the effective theory must have operators that break the U(1) symmetry explicitly. Such operators are monopoles,³ and they are almost always relevant [2].⁴ This again agrees with the lore that UV completions of the U(1) gauge theory always requires monopoles.

However, there is something different in this setup. Haldane has shown that unit-charge monopoles couple to geometric phases of the underlying spin systems [1], and, depending on the underlying lattice symmetries, may obliterate the unit charge monopoles, preserving a discrete \mathbb{Z}_N subgroup of the U(1)-topological magnetic symmetry. Therefore such systems, while not preserving the full U(1)-topological symmetry, do preserve a subgroup of it. So at least in such systems monopoles of unit charge, albeit 2+1d monopole-instantons, are not part of an effective theory.

The discussion also makes clear that the absence or presence of dynamical monopoles is a question of symmetry. A theory without dynamical monopoles has more symmetries. Indeed in continuous 4d spacetime, the 2-index current $J_{\mu\nu}^m = \frac{1}{4\pi} \epsilon_{\mu\nu\rho\sigma} \partial^\nu F^{\rho\sigma}$ is identically conserved by the Bianchi identity in the absence of monopoles.⁵ So an abelian gauge theory a priori allows for more symmetries than its non-abelian counterparts.

Furthermore, free abelian gauge theories have an electric-magnetic self-duality,⁶ but such a duality is always explicitly broken in continuum descriptions of interacting U(1) gauge theory because of the inability to couple electric and magnetic matter simultaneously. In non-abelian gauge theories the situation is different. In the Coulomb regime of such theories, monopoles can appear as solitons, and ever since Montonen and Olive [8], the

¹The current is U(1) because the charges are quantized in magnetic flux units.

²The current is conserved off-shell, so it is commonly called a topological symmetry. Note, however, that this designation depends on the description of the theory. Indeed a free U(1) gauge theory is equivalent to a compact boson, where the topological symmetry becomes a normal Noether-like shift-symmetry of the boson.

³Note that monopoles are space-time localized in 2+1d, i.e., they are instantons.

⁴The monopole operators are known/conjectured to be irrelevant at certain 3d fixed points (see, e.g., [3–7]).

⁵The free 4d U(1) gauge theory has two 1-form symmetries in the continuum. These follow from the two sets of Maxwell equations $\partial_\mu F^{\mu\nu} = 0$ and $\partial_\mu \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} = 0$.

⁶Here we refer to self-duality in a broader sense, i.e., that the model is mapped to itself, but perhaps with some coupling changed. However, in this work we mostly will study the self-duality in a more restrictive sense, i.e., that the theory maps exactly to itself under the duality transformation. Self-duality then should be viewed as a symmetry of the theory.

pursuit of theories with electric and magnetic duality has been of interest, with all known cases being supersymmetric theories.⁷ The non-abelian structure of the theory therefore has two effects. 1: it furnishes a UV completion via asymptotic freedom, and 2: it allows for a local Lagrangian description of a theory with both electric and magnetic matter.

However, a non-abelian UV completion of an abelian theory, inevitably determines the matter content of the full theory to a large extent. For example a common feature of the abelianized non-abelian theory is the presence of magnetic monopoles with unit magnetic charge, and the charged W-bosons with unit fixed electric charge.

Still there should be no basic obstacle to formulating U(1) gauge theories with magnetic and electric matter on equal footing. Indeed such objects are mutually local, but the problem is that the continuum Lagrangian cannot be formulated with a local magnetic and electric gauge potential. One may therefore refer to such theories (if they exist as continuum theories) as *non-Lagrangian theories*.

However, in [9] two of us showed that such theories are possible on the lattice, where a Lagrangian can be formulated in such a way that it treats electric and magnetic matter on equal footing.⁸ In this formulation monopoles are not artifacts of a lattice theory, but can have an action associated with them. Such theories are not only interesting in their own right, but they also may lead to new quantum field theories which are beyond the one captured by non-abelian gauge theories. In this paper we continue the line of work initiated in [9] (compare also [14, 15]) and study the self-dual U(1) lattice system with a single electrically and a single magnetically charged boson. We discuss in detail the possible phase structure of the system and, using a suitable worldline representation, present results of a Monte Carlo simulation that allow for checking the conjectured phase diagram.

The results of the paper are presented in two main sections. Section 2 discusses generalities of self-dual modified Villain models and is organized as follows: in subsection 2.1 we review the pure-gauge construction of [9], and in subsection 2.2 we discuss the electric-magnetic duality. In subsection 2.3 we present the self-dual symmetry and discuss its mixing with lattice translation symmetries. In subsection 2.4 we introduce matter fields, focusing on bosonic matter primarily. In subsection 2.5 we discuss some computable limits of the scalar self-dual model, while in subsection 2.6 we present the perturbative analysis of scalar models far away from the self-dual point. In subsection 2.7 we discuss larger gauge charge and flavor generalizations, and 't Hooft anomalies of the flavor symmetries and 1-form symmetries. We also discuss a non-abelian model with exact electric- magnetic self-dual symmetry. In subsection 2.8 we condense all of the previous discussion into a self-dual scalar QED phase diagram.

Section 3 is devoted to the numerical simulation of one-flavor self-dual scalar QED. As a first step we introduce the necessary worldline representation in subsection 3.1 and then, in subsection 3.2, collect self-dual relations and introduce suitable order parameters which we need for the numerical analysis. In subsection 3.3 we discuss the numerical setup and

⁷The original conjecture of Montonen and Olive was for a non-supersymmetric Georgi-Glashow model, which, as is now known, does not enjoy a self-dual symmetry.

⁸Similar reasoning was also useful in 2d U(1) gauge theories with the θ -term [10–13].

general results of the condensation transition. We show that the theory has two transitions, one 1st order and one 2nd order. We analyze these transitions in subsection 3.4.

Finally in section 4 we conclude and discuss some future prospects, specifically addressing the search for the novel interacting conformal fixed points.

2 Self-dual modified Villain models

2.1 The gauge field partition sum

In this subsection we summarize the construction of modified U(1) Villain lattice models which was discussed in detail in [9].⁹ We consider a 4-d hypercubic lattice Λ with lattice extents $N_\mu, \mu = 1, 2, 3, 4$ and a total number of sites $V \equiv N_1 N_2 N_3 N_4$. We may think of Λ as the union of $\Lambda^{(0)}, \Lambda^{(1)}, \Lambda^{(2)}, \Lambda^{(3)}$ and $\Lambda^{(4)}$, which are the sets of all sites, links, faces, cubes and hypercubes.¹⁰ The lattice discretization in [9] relies on the Villain-like gauge action given by¹¹

$$S_g[A^e, n] = \sum_p \frac{1}{2} ((dA^e)_p + 2\pi n_p)^2, \tag{2.1}$$

where the sum runs over the plaquettes p . $(dA^e)_p$ is the discretized gauge field flux (defined below) around the links l of the plaquette p , built out of the gauge fields A_l^e living on the links l , where the superfix “ e ” stands for *electric*, emphasizing that this gauge fields naturally couples to *electric matter*. The n_p are integers living on plaquettes — the so-called *Villain variables*.

Explicitly, by labeling the square plaquettes on a hypercubic lattice¹² with the lattice site x of its root point and two indices $\mu, \nu = 1, 2, 3, 4$, we define the exterior lattice derivative operator d by

$$(dA^e)_{x,\mu\nu} = A_{x+\hat{\mu},\nu}^e - A_{x,\nu}^e - A_{x+\hat{\nu},\mu}^e + A_{x,\mu}^e, \tag{2.2}$$

where $\hat{\mu}$ indicates a vector in the direction μ . Using this we define the field strength living on the plaquettes p as

$$F_p \equiv (dA^e)_p + 2\pi n_p \quad (\Leftrightarrow F_{x,\mu\nu} \equiv (dA^e)_{x,\mu\nu} + 2\pi n_{x,\mu\nu}). \tag{2.3}$$

⁹This construction was also applied for describing fracton models in [16] and for formulating models with non-invertible symmetries [17].

¹⁰In general $\Lambda^{(r)}$ is the set of all r -cells, where a 0-cell is a vertex or site, a 1-cell is a link, edge or bond, a 2-cell is a face, etc.

¹¹Whenever we write a sum \sum_c or a product \prod_c over the cells c of the lattice, we always take into account only one orientation of the cell. Moreover, since an r cell c of the hypercubic lattice is an r -dimensional hypercube uniquely defined by a vertex x and r orthonormal vectors $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_r$, we can always write instead of c , an ordered multiplet $(x, \mu_1 \mu_2, \dots, \mu_r)$ which uniquely determines c , such that the sum \sum_c is defined as $\sum_c \equiv \sum_x \sum_{\mu_1 < \mu_2 < \dots < \mu_r}$, and similarly for the product \prod_c .

¹²While these models can be formulated on arbitrary lattices, they are most natural on the hypercubic lattice. The reason is that the hypercubic lattice and its dual lattice are isomorphic, and since, as we will review, the self-dual symmetry maps the two into each other, such that the self-dual symmetry is only a genuine symmetry on the hypercubic lattice.

We can always restrict A_l^e ¹³ such that $A_l^e \in [-\pi, \pi)$. The partition function is defined as follows,

$$Z = \sum_{\{n\}} \int D[A^e] e^{-\beta S_g[A^e, n]}, \quad (2.4)$$

where we introduced $\beta = \frac{1}{e^2}$, which plays the role of the inverse electric coupling squared. We also defined some short-hand notation,

$$\int D[A^e] \equiv \prod_l \int_{-\pi}^{\pi} \frac{dA_l^e}{2\pi} = \prod_x \prod_{\mu} \int_{-\pi}^{\pi} \frac{dA_{x,\mu}^e}{2\pi}, \quad (2.5a)$$

$$\sum_{\{n\}} \equiv \prod_p \sum_{n_p \in \mathbb{Z}} = \prod_x \prod_{\mu < \nu} \sum_{n_{x,\mu\nu} \in \mathbb{Z}}. \quad (2.5b)$$

The lattice discretization summarized in (2.1) – (2.5) is the well known Villain formulation [18] of U(1) lattice gauge theory.

However, while the theory (2.4) has an electric 1-form center symmetry,¹⁴ it does not display the magnetic 1-form symmetry expected in four space-time dimensions. The reason is that if the sum over the n_p is unconstrained, then the model has dynamical monopoles, which are identified with the configurations for which $(dn)_c \neq 0$, where c is a 3-cube, labeled by one vertex x and three directions μ, ν, ρ (see, e.g., [9, 19–21]). Here d is again the exterior lattice derivative, with the action on 2-forms explicitly defined as

$$(dn)_{x,\mu\nu\rho} = n_{x+\hat{\rho},\mu\nu} - n_{x,\mu\nu} + n_{x+\hat{\mu},\nu\rho} - n_{x,\nu\rho} + n_{x+\hat{\nu},\rho\mu} - n_{x,\rho\mu}, \quad (2.6)$$

or, in short hand

$$(dn)_c = \sum_{p \in \partial c} n_p, \quad (2.7)$$

where ∂c is the boundary of c , i.e., the set of all (outward) oriented faces of the cube c .

To properly define a free gauge theory, we need to eliminate these monopoles by imposing the constraint

$$(dn)_c = 0 \quad \forall c \in \Lambda^{(3)}. \quad (2.8)$$

One way for implementing the constraint is to introduce Lagrange multipliers A_c^m assigned to the cubes c of the lattice Λ . Here the superscript m stands for *magnetic* since, as we will see, the Lagrange multipliers will turn out to be a natural definition of a magnetic gauge field. We thus may write the set of constraints (2.8) as

$$\prod_c \int_{-\pi}^{\pi} \frac{dA_c^m}{2\pi} e^{i(dn)_c A_c^m} = \int D[A^m] e^{i \sum_{c \in \Lambda^{(3)}} (dn)_c A_c^m}, \quad (2.9)$$

¹³One may be tempted to define a field strength to be just $(dA)_p$. However, this is a total derivative, and all fluxes of it over a closed 2-cycle on the lattice will vanish. Moreover, we can redundantly allow A_l to take values on the whole real line \mathbb{R} . Then the system enjoys a discrete 1-form gauge symmetry, and only the combination $dA_p + 2\pi n_p$ is gauge invariant [9].

¹⁴The 1-form symmetry can be seen by noting that shifts $A_l^e \rightarrow A_l^e + V_l$, with V_l such that $(dV)_p = 0$, are a symmetry. Naively, the symmetry group then is $H^1(M, \mathbb{R})$, where M is the underlying space-time manifold. However, shifts of A_l^e by $2\pi\mathbb{Z}$ are a gauge symmetry, and hence should not be considered a global symmetry. So the group is $H^1(M, U(1))$. For a 4-torus this is just $U(1) \times U(1) \times U(1) \times U(1)$, where each U(1) group corresponds to a different cycle of the 4-torus. The operators charged under the various U(1) parts are the (electric) Willson loops wrapping around these cycles.

where we defined,

$$\int D[A^m] \equiv \prod_c \int_{-\pi}^{\pi} \frac{dA_c^m}{2\pi}. \quad (2.10)$$

2.2 Duality transformation

Having defined the Villain form of the gauge field Boltzmann factor augmented with the constraint (2.9) we are now ready to discuss the duality transformation. We apply the partial integration formula (see the appendices of [9] and [15]) to the exponent of the constraint (2.9) and find,

$$\begin{aligned} \sum_c (dn)_c A_c^m &= \sum_x \sum_{\mu < \nu < \rho} (n_{x+\hat{\rho},\mu\nu} - n_{x,\mu\nu} + n_{x+\hat{\mu},\nu\rho} - n_{x,\nu\rho} + n_{x+\hat{\nu},\rho\mu} - n_{x,\rho\mu}) A_{x,\mu\nu\rho}^m = \\ &= \sum_x \sum_{\mu < \nu} n_{x,\mu\nu} (A_{x-\hat{\rho},\mu\nu\rho}^m - A_{x,\mu\nu\rho}^m) = - \sum_p n_p (\partial A)_p, \end{aligned} \quad (2.11)$$

where we defined the *divergence operator* ∂ on the lattice,

$$(\partial A^m)_{x,\mu\nu} = \sum_{\rho} (A_{x,\mu\nu\rho}^m - A_{x-\hat{\rho},\mu\nu\rho}^m), \quad (2.12)$$

which is the discrete version of the continuum divergence $\partial^\rho A_{\mu\nu\rho}^m$. The lattice divergence operator can also be written as

$$(\partial A^m)_p = \sum_{c \in \hat{\partial}p} A_c^m, \quad (2.13)$$

where $\hat{\partial}p$ is the coboundary of p , i.e., the set of all cubes c whose boundary contains p (see again the appendices of [9, 15] for details and generalizations).

Since for the cubic lattice also the dual lattice is cubic, we can define the *Hodge star* (again see, e.g., [9]) which maps r -cells (below denoted as $c_{x,\mu_1\mu_2\cdots\mu_r}^{(r)}$) of the original lattice Λ to $(D-r)$ -cells¹⁵ of the dual lattice $\tilde{\Lambda}$, where D is the space-time dimension, which here of course is $D = 4$. Analogously one can define a \star -operator which takes r -cells (below denoted as $\tilde{c}_{\tilde{x},\mu_1\mu_2\cdots\mu_r}^{(r)}$) of $\tilde{\Lambda}$ to $(D-r)$ -cells of Λ . We will denote both of these with the same symbol “ \star ”. The star operators are defined as follows

$$\star c_{x,\mu_1\mu_2\cdots\mu_r}^{(r)} = \sum_{\mu'_{r+1} < \mu'_{r+2} < \cdots < \mu'_D} \epsilon_{\mu_1\mu_2\cdots\mu_r\mu'_{r+1}\cdots\mu'_D} \tilde{c}_{\tilde{x}-\hat{\mu}'_{r+1}-\mu'_{r+2}\cdots-\hat{\mu}'_D,\mu'_{r+1}\mu'_{r+2}\cdots\mu'_D}, \quad (2.14a)$$

$$\star \tilde{c}_{\tilde{x},\mu_1\mu_2\cdots\mu_r}^{(r)} = \sum_{\mu'_{r+1} < \mu'_{r+2} < \cdots < \mu'_D} \epsilon_{\mu_1\mu_2\cdots\mu_r\mu'_{r+1}\cdots\mu'_D} c_{x+\hat{\mu}_1+\cdots+\hat{\mu}_r,\mu'_{r+1}\mu'_{r+2}\cdots\mu'_D}, \quad (2.14b)$$

where

$$\tilde{x} = x + \frac{\hat{1} + \hat{2} + \cdots + \hat{D}}{2}, \quad (2.15)$$

is a site on the dual lattice obtained by a translation of x . Some examples for the action of the \star -operator are illustrated in figure 1.

¹⁵A 0-cell is a vertex, a 1-cell is a link, a 2-cell is a face, etc.

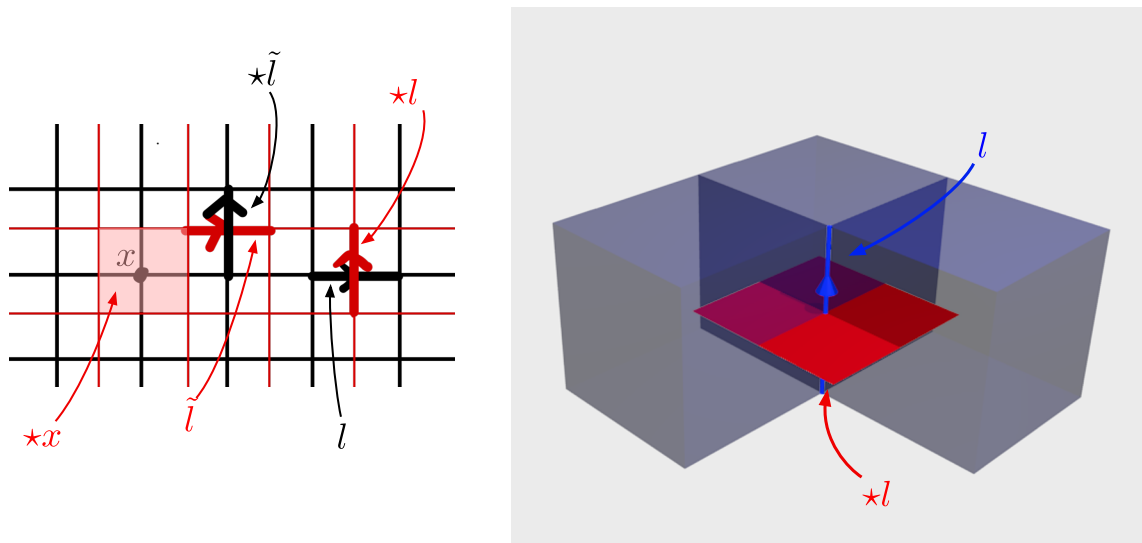


Figure 1. Examples for the action of the \star -operator in 2d (left) and 3d (right).

Furthermore, we can define the action of the \star -operator on an r -form $A_{c^{(r)}}$ living on the r -cells $c^{(r)} \in \Lambda^{(r)}$, as $(\star A)_{\tilde{c}^{(D-r)}} \equiv A_{\star \tilde{c}^{(D-r)}}$. In the case at hand, we have a 3-form $(A^e)_c$ living on cubes $c \in \Lambda^{(3)}$, such that we obtain for $(\star A^m)_{\tilde{l}}$ living on the dual links \tilde{l}

$$(\tilde{A}^m)_{\tilde{l}} \equiv (\star A^m)_{\tilde{l}} \equiv (A^m)_{\star \tilde{l}}. \quad (2.16)$$

We will need the important property that $d\star = \star\partial$, as well as the fact that \star^2 acting on a r -form $A_{c^{(r)}}$ is the identity up to a sign $(-1)^{r(D-r)}$, i.e., $\star^2 A_{c^{(r)}} = (-1)^{r(D-r)} A_{c^{(r)}}$.

Applying this machinery now to eq. (2.11) we find

$$\sum_c (dn)_c A_c^m = - \sum_p n_p (\partial A^m)_p = - \sum_p n_p (\star^2 \partial A^m)_p = - \sum_p n_p (d\tilde{A}^m)_{\star p}, \quad (2.17)$$

such that the partition function is given by

$$Z = \sum_{\{n\}} \int D[A^e] \int D[\tilde{A}^m] e^{-\frac{\beta}{2} \sum_p ((dA)_p + 2\pi n_p)^2 - i \sum_p n_p (d\tilde{A}^m)_{\star p}}. \quad (2.18)$$

Now we use the Poisson resummation formula

$$\sum_n e^{-\frac{\beta}{2} (\mathcal{F} + 2\pi n)^2 - i n \tilde{\mathcal{F}}} = \frac{1}{\sqrt{2\pi\beta}} \sum_m e^{-\frac{1}{2(2\pi)^2\beta} (\tilde{\mathcal{F}} + 2\pi m)^2 + i \mathcal{F} \tilde{\mathcal{F}} / (2\pi) + i m \mathcal{F}}, \quad (2.19)$$

to obtain yet another form of the partition function

$$Z = \frac{1}{(2\pi\beta)^{3V}} \sum_{\{m\}} \int D[A^e] \int D[A^m] e^{-\frac{1}{2(2\pi)^2\beta} \sum_p ((dA^m)_{\star p} + 2\pi m_p)^2 + i \sum_p m_p (dA^e)_p}, \quad (2.20)$$

where $V = N_1 N_2 N_3 N_4$ is again the total number of sites of the toroidal lattice.¹⁶ Note that the term $(dA^e)_p (dA^m)_{\star p}$ dropped out because it is zero¹⁷ when summed over p . By writing

¹⁶The overall factor $(2\pi\beta)^{-3V}$ is just $1/\sqrt{2\pi\beta}$ from (2.19) to the power of the number of plaquettes of the lattice $6V$.

¹⁷This follows from a ‘‘partial integration’’ formula, similar to what we used in (2.11), and the fact that $d^2 = 0$.

$m_p = (\star m)_{\star p} = (\tilde{m})_{\star p}$, $(A^e)_p = (A^e)_{\star(\star p)}$ and noting that we can replace the sum over p with the sum over $\tilde{p} = \star p$. Therefore the dual form of the partition function is given by

$$Z = \frac{1}{(2\pi\beta)^{3V}} \sum_{\{m\}} \int D[A^e] \int D[\tilde{A}^m] e^{-\frac{1}{2(2\pi)^2\beta} \sum_{\tilde{p}} ((dA^m)_{\tilde{p}} + 2\pi\tilde{m}_{\tilde{p}})^2 + i \sum_p \tilde{m}_{\tilde{p}} (dA^e)_{\star\tilde{p}}} . \quad (2.21)$$

This is our final expression for the duality transformation.¹⁸ Notice that it is almost identical (up to details we will discuss soon) to (2.18), with $\beta \rightarrow \tilde{\beta} = \frac{1}{(2\pi)^2\beta}$. Moreover when $\beta = \beta^* = \frac{1}{2\pi}$ the theory is self-dual, and hence has an enhanced symmetry. The self-duality in fact does not square to unity, but to charge-conjugation C . The reason for this is that in (2.21) the phase term has a different sign than in (2.18), which means that the self-duality requires not just exchanging the gauge field A^e with \tilde{A}^m , but also flipping the sign of one of them, hence squaring to a pure charge conjugation. However, even this is too naive in our lattice theory, as the exchange of the two gauge fields also requires shifting the lattice to the dual lattice, and we will see that the self-dual transformation squares to a charge conjugation C and an overall lattice shift. We discuss this in detail below.

2.3 Self-duality and self-dual symmetry

Comparing the dual form (2.21) with the original partition sum (2.18) we see that the change $(A^e, \tilde{A}^m) \rightarrow (-\tilde{A}^m, A^e)$, $2\pi\beta \rightarrow 1/(2\pi\beta)$ switches the original partition function to the structure of the dual one. However, we cannot simply replace (A^e, \tilde{A}^m) by $(-\tilde{A}^m, A^e)$, because A^e lives on the original lattice Λ , while \tilde{A}^m lives on the dual lattice $\tilde{\Lambda}$. So we must define the dual transformation such that it also incorporates the map from Λ to $\tilde{\Lambda}$. We accomplish this by defining the translation operator T , that shifts the lattice by the vector $\hat{s} = \frac{\hat{1} + \hat{2} + \dots + \hat{D}}{2}$, i.e., it shifts from Λ to $\tilde{\Lambda}$. Thus the operator T translates the r -cell of the lattice Λ to the r -cell of the dual lattice $\tilde{\Lambda}$ as follows

$$T c_{x, \mu_1 \mu_2 \dots \mu_r}^{(r)} = \tilde{c}_{x + \hat{s}, \mu_1 \mu_2 \dots \mu_r}^{(r)} . \quad (2.22)$$

Consequently, if we perform the replacement

$$\begin{aligned} A_l^e &\rightarrow -\tilde{A}_{T(l)}^m , \\ \tilde{A}_{\tilde{l}}^m &\rightarrow A_{T(\tilde{l})}^e , \\ \tilde{m}_{\tilde{p}} &\rightarrow n_{T(\tilde{p})} , \end{aligned} \quad (2.23)$$

in (2.21), the same form as in (2.18) is obtained,¹⁹ up to an overall constant and of course the change of coupling $\beta \rightarrow \tilde{\beta} = \frac{1}{(2\pi)^2\beta}$. Thus we find

$$Z(\beta) = (2\pi\beta)^{-3V} Z\left(\frac{1}{\beta(2\pi)^2}\right) , \quad (2.24)$$

¹⁸This duality was referred to as Kramers-Wannier duality in [9], which is a bit imprecise because the original Kramers-Wannier duality does not in fact map a theory back to itself, but maps it back to the same theory coupled to a TQFT [22]. In contrast the electric-magnetic duality we discuss here is an exact duality. It can be converted to a Kramers-Wannier type of duality by gauging one of the 1-form symmetries. This was discussed in [17] (see also [23]).

¹⁹Note that $T\star = \star T$ as can be checked from the expressions (2.14).

which we can also write in a more symmetric way as follows,

$$(2\pi\beta)^{\frac{3V}{2}} Z(\beta) = (2\pi\tilde{\beta})^{\frac{3V}{2}} Z(\tilde{\beta}) \quad \text{with} \quad \tilde{\beta} = \frac{1}{4\pi^2\beta}. \quad (2.25)$$

This relation for the partition sum generates relations for observables: taking the logarithm of both sides and differentiating with respect to β , we obtain a relation between the expectations values of the field strength squared as

$$\begin{aligned} \langle F^2 \rangle_\beta &\equiv -\frac{1}{3V} \frac{\partial}{\partial \beta} \ln Z(\beta) = -\frac{1}{3V} \frac{\partial}{\partial \beta} \ln \left((2\pi\beta)^{-3V} Z(\tilde{\beta}) \right) \\ &= \frac{1}{\beta} - \frac{1}{3V} \left(\frac{\partial}{\partial \tilde{\beta}} \ln Z(\tilde{\beta}) \right) \frac{d\tilde{\beta}}{d\beta} = \frac{1}{\beta} - \langle F^2 \rangle_{\tilde{\beta}} \frac{1}{4\pi^2\beta^2}, \end{aligned} \quad (2.26)$$

where we defined $F^2 \equiv \sum_{x,\mu<\nu} (F_{x,\mu\nu})^2 / (6V) = S_g[A^e, n] / (3V)$. The subscripts β and $\tilde{\beta}$ attached to $\langle F^2 \rangle$ indicate the coupling the respective vevs are computed with. The relation can be rearranged in a more symmetric form as [15]

$$\beta \langle F^2 \rangle_\beta + \tilde{\beta} \langle F^2 \rangle_{\tilde{\beta}} = 1. \quad (2.27)$$

Another derivative with respect to β generates a duality relation for the corresponding susceptibilities,

$$\beta \langle F^2 \rangle_\beta - \beta^2 \chi_\beta = \tilde{\beta} \langle F^2 \rangle_{\tilde{\beta}} - \tilde{\beta}^2 \chi_{\tilde{\beta}}, \quad (2.28)$$

with

$$\chi_\beta \equiv -\frac{\partial}{\partial \beta} \langle F^2 \rangle_\beta = 3V \left\langle (F^2 - \langle F^2 \rangle_\beta)^2 \right\rangle_\beta. \quad (2.29)$$

We conclude this subsection by identifying the *self-dual symmetry*. When $\beta = \tilde{\beta} = \frac{1}{4\pi^2\beta}$, the duality maps our theory to itself, and hence becomes a genuine symmetry. By virtue of (2.23), we define the self-duality operator \mathcal{S} to act on gauge fields as follows (and of course $\beta \rightarrow \tilde{\beta}$)

$$\mathcal{S} = \begin{cases} A_l^e \rightarrow -\tilde{A}_{T(l)}^m, \\ \tilde{A}_l^m \rightarrow A_{T(\tilde{l})}^e. \end{cases} \quad (2.30)$$

The square of the transformation is given by (note that $\tilde{\tilde{\beta}} = \beta$)

$$\mathcal{S}^2 : \begin{cases} A_l^e \rightarrow -A_{T^2(l)}^e, \\ \tilde{A}_l^m \rightarrow -\tilde{A}_{T^2(\tilde{l})}^m, \end{cases} \quad (2.31)$$

i.e., it acts as charge conjugation and a diagonal translation of the lattice in all directions by one lattice unit, which we will call \mathcal{D} . The translation \mathcal{D} forms a group \mathbb{Z} on an infinite lattice, while on our finite periodic lattice we find $\mathcal{D}^N = \mathbb{I}$ for some integer N , so that \mathcal{D} generates a \mathbb{Z}_N subgroup, which we will call Z . The duality transformation \mathcal{S} , the charge conjugation \mathcal{C} and diagonal shifts \mathcal{D} furnish a group G , which satisfies the following identities

$$\mathcal{S}^2 = \mathcal{C}\mathcal{D}, \quad (2.32)$$

$$\mathcal{C}^2 = \mathbb{I}. \quad (2.33)$$

The generator \mathcal{D} commutes with \mathcal{S} and \mathcal{C} and so Z is the center of G . Furthermore, $G/Z \cong \mathbb{Z}_4$. Indeed $\mathcal{S}^4 = \mathcal{D}^2$ is a pure center element, and so is equivalent to the identity element in G/Z . One way to phrase this is to say that G is a central extension of \mathbb{Z}_4 by \mathbb{Z} .

The fact that this algebra involves the lattice shift \mathcal{D} makes clear that the self-dual symmetry does not act in an *on-site* manner; its action necessarily involves lattice translations. This is consistent with the no-go theorem of [24], which argues that any theory with electric-magnetic self-duality cannot be regularized in a manner which realizes the symmetry in an on-site fashion.

If we assume, however, that the lattice symmetries are unbroken by the vacuum of the theory, then the vacuum can at most be acted on by $G/Z \cong \mathbb{Z}_4$. This \mathbb{Z}_4 group is generated by a self-duality generator \mathcal{S} for which $\mathcal{S}^2 = \mathcal{C}$ is the charge-conjugation. If we further assume that charge conjugation is unlikely to be broken in the vacuum (below we justify this with numerical results for the system we studied), we then conclude that the vacuum will transform at most under $\mathbb{Z}_2 \cong \mathbb{Z}_4/\mathbb{Z}_2$. If this happens we will say that self-dual symmetry is spontaneously broken.

2.4 Coupling electric and magnetic matter

We now couple *electrically charged matter* ϕ^e to the electric gauge field A^e and *magnetically charged matter* $\tilde{\phi}^m$ to the magnetic gauge field \tilde{A}^m . We write the partition function with gauge fields coupling to electric and magnetic matter fields in the form

$$Z(\beta, J_e, J_m) \equiv \int D[A^e] \int D[A^m] B_\beta[A^e, \tilde{A}^m] Z[A^e, J_e] \tilde{Z}[\tilde{A}^m, J_m], \quad (2.34)$$

where $B_\beta[A^e, \tilde{A}^m]$ is the weight of the free gauge theory derived in (2.18), i.e.,

$$B_\beta[A^e, \tilde{A}^m] = e^{-\frac{\beta}{2} \sum_p ((dA)_p + 2\pi n_p)^2 - i \sum_p n_p (d\tilde{A}^m)_{*p}}. \quad (2.35)$$

In (2.34) we introduced the partition sums Z and \tilde{Z} for electric and magnetic matter (defined below), which also depend on two coupling parameters J_e and J_m for the electric and magnetic matter fields. The electrically charged matter fields $\phi_x^e \in \text{U}(1)$ we parameterize as $\phi_x^e = e^{i\varphi_x^e}$ with $\varphi_x^e \in [-\pi, \pi)$. They couple to the electric background gauge field A^e via the partition sum

$$Z[A^e, J_e] \equiv \int D[\phi^e] e^{J_e S_e[\phi^e, A^e]}, \quad \int D[\phi^e] \equiv \prod_x \int_{-\pi}^{\pi} \frac{d\varphi_x^e}{2\pi}, \quad (2.36)$$

$$S_e[\phi^e, A^e] \equiv \frac{1}{2} \sum_{x,\mu} \left[\phi_x^{e*} e^{iA_{x,\mu}^e} \phi_{x+\hat{\mu}}^e + c.c. \right] = \sum_{x,\mu} \cos(\varphi_{x+\hat{\mu}}^e - \varphi_x^e + A_{x,\mu}^e). \quad (2.37)$$

The magnetically charged scalar field $\tilde{\phi}_{\tilde{x}}^m \in \text{U}(1)$, which we parameterize as $\tilde{\phi}_{\tilde{x}}^m = e^{i\tilde{\varphi}_{\tilde{x}}^m}$ with $\tilde{\varphi}_{\tilde{x}}^m \in [-\pi, \pi)$, lives on the sites \tilde{x} of the dual lattice and couples to the magnetic gauge field $\tilde{A}_{\tilde{x},\mu}^m$ on the links of the dual lattice. The corresponding partition sum has the same form as the partition sum (2.36) for the electric matter but for the magnetic matter

is defined entirely on the dual lattice:

$$\tilde{Z}[\tilde{A}^m, J_m] \equiv \int D[\tilde{\phi}^m] e^{J_m \tilde{S}_m[\tilde{\phi}^m, \tilde{A}^m]}, \quad \int D[\tilde{\phi}^m] \equiv \prod_{\tilde{x}} \int_{-\pi}^{\pi} \frac{d\tilde{\varphi}_{\tilde{x}}^m}{2\pi}, \quad (2.38)$$

$$\tilde{S}_m[\tilde{\phi}^m, \tilde{A}^m] \equiv \frac{1}{2} \sum_{\tilde{x}, \mu} \left[\tilde{\phi}_{\tilde{x}}^{m*} e^{i\tilde{A}_{\tilde{x}, \mu}^m} \tilde{\phi}_{\tilde{x}+\hat{\mu}}^m + c.c. \right] = \sum_{\tilde{x}, \mu} \cos(\tilde{\varphi}_{\tilde{x}+\hat{\mu}}^m - \tilde{\varphi}_{\tilde{x}}^m + \tilde{A}_{\tilde{x}, \mu}^m). \quad (2.39)$$

Here we have coupled electric and magnetic matter using U(1)-valued matter fields, but it is straightforward to generalize this construction to complex-valued bosonic matter or also to fermionic fields [9].

Self-duality of the full theory essentially follows from the self-duality of the pure gauge theory we already discussed above, combined with the interchange of electric and magnetic matter (see [9, 15] for a more detailed discussion). The corresponding self-duality relation for the partition function is given by

$$Z(\beta, J_e, J_m) = \left(\frac{1}{2\pi\beta} \right)^{3V} Z(\tilde{\beta}, \tilde{J}_e, \tilde{J}_m), \quad (2.40)$$

with $\tilde{\beta} = \frac{1}{4\pi^2\beta}, \quad \tilde{J}_e = J_m, \quad \tilde{J}_m = J_e.$

Again we can generate self-duality relations for observables by evaluating derivatives of $\ln Z$ with respect to the couplings. The pure gauge theory sum rules (2.27) and (2.28) generalize to,

$$\beta \langle F^2 \rangle_{\beta, J_e, J_m} + \tilde{\beta} \langle F^2 \rangle_{\tilde{\beta}, \tilde{J}_e, \tilde{J}_m} = 1, \quad (2.41)$$

and

$$\beta \langle F^2 \rangle_{\beta, J_e, J_m} - \beta^2 \chi_{\beta, J_e, J_m} = \tilde{\beta} \langle F^2 \rangle_{\tilde{\beta}, \tilde{J}_e, \tilde{J}_m} - \tilde{\beta}^2 \chi_{\tilde{\beta}, \tilde{J}_e, \tilde{J}_m}. \quad (2.42)$$

Derivatives with respect to J_e and J_m generate field expectation values for the electric and magnetic matter fields. Exploring the duality relation (2.40) one finds the following self-duality relation for the electric and the magnetic action densities $s_e \equiv S_e/4V$ and $\tilde{s}_m \equiv \tilde{S}_m/4V$:

$$\langle s_e \rangle_{\beta, J_e, J_m} \equiv \frac{1}{4V} \frac{\partial}{\partial J_e} \ln Z(\beta, J_e, J_m) = \langle \tilde{s}_m \rangle_{\tilde{\beta}, \tilde{J}_e, \tilde{J}_m}. \quad (2.43)$$

According to the self-duality relation (2.43), the electric and magnetic field expectation values are converted into each other when changing from weak to strong coupling and simultaneously interchanging the electric and magnetic coupling parameters.

2.5 Computable self-dual limits

Let us now discuss various limits of the model. First we consider the limit when $J_e = J_m = J \rightarrow 0$. In this case the model is a pure gauge theory model. Moreover the model has so-called 1-form magnetic and electric symmetries, which can be seen from (2.35) by taking $A^e \rightarrow A_l^e + \delta_l$ with δ_l^e such that $(d\delta^e)_l = 0$, and similarly for $\tilde{A}_l^m \rightarrow \tilde{A}_l^m + \tilde{\delta}_l^m$ with $(d\tilde{\delta}^e)_l = 0$. These 1-form symmetries are continuous U(1) symmetries in the case of a free theory. They act non-trivially on Wilson loops $W^e(C)$ and 't Hooft loops $W^m(\tilde{C})$ defined as

$$W^e(C) = e^{i \sum_{l \in C} A_l^e}, \quad W^m(\tilde{C}) = e^{i \sum_{l \in \tilde{C}} \tilde{A}_l^m}, \quad (2.44)$$

where C and \tilde{C} are closed paths on the lattice and the dual lattice, respectively.

Moreover the model has a 't Hooft anomaly between these two symmetries, which in a way guarantees that the photon is exactly massless in a free theory, as the anomaly is generically saturated by breaking either the electric or the magnetic 1-form symmetry, leaving a goldstone boson — the photon [25]. We discuss this anomaly, or rather the anomaly between the discrete subgroups of the electric and magnetic 1-form symmetries, in section 2.7. We can see the presence of this massless phase explicitly by integrating out the field \tilde{A}^m . We solve the constraint $(dn)_c = 0$ by setting

$$n_p = (dk)_p, \tag{2.45}$$

where k_l is an integer living on links. This ansatz is correct only locally and not fully general on our compact lattice, such that we now formally take the lattice to be infinite to illustrate our point.²⁰ Now we can remove k_l by a shift of gauge fields²¹ $A_l^e \rightarrow A_l^e - k_l$ and obtain the action

$$\beta S_g = \sum_p \frac{\beta}{2} (dA^e)_p^2, \tag{2.46}$$

which describes free photons and has no phase transition as a function of coupling because now the coupling can be absorbed in the field redefinition²² $A_l^e \rightarrow A_l^e / \sqrt{\beta}$.

Now let us discuss the opposite limit and take $J^m = J^e = J \rightarrow \infty$ in (2.36) and (2.38). In this case the matter contribution becomes dominant and will pin the gauge fields to values such that the phase difference between sites carries no action cost. This effectively imposes the constraint that

$$\begin{aligned} A_l^e &= 2\pi s_l^e, \\ A_l^m &= 2\pi s_l^m, \end{aligned} \quad s_l^e, s_l^m \in \mathbb{Z}. \tag{2.47}$$

Notice that the phase factor in the action (2.35) drops out and our total action is just

$$S_{\text{tot}} = \frac{\beta}{2} (2\pi)^2 \sum_p ((ds^e)_p + n_p)^2. \tag{2.48}$$

Moreover the Villain variables n_p -s are now unconstrained and we can shift them $n_p \rightarrow n_p - (ds^e)_p$ to eliminate s_l^e completely. Finally our partition function is given by a product of the plaquette contributions

$$Z_{J=\infty} \equiv \lim_{\substack{J^e \rightarrow \infty \\ J^m \rightarrow \infty}} Z(\beta, J^e, J^m) = \left(\sum_n e^{-\frac{\beta}{2\pi^2} n^2} \right)^{6V}, \tag{2.49}$$

²⁰Alternatively we can take the k_l to have general boundary conditions on our toroidal lattice.

²¹In order to do this we must a priori take $A_l \in \mathbb{R}$, which we can always do [9].

²²It is tempting to say that the action (2.46) describes a non-compact gauge field. However, this is not quite right, as the choice of gauge was not possible to perform on a compact manifold, nor is it compatible with the insertion of a 't Hooft line — two distinguishing features of the compact gauge theory. Rather this should be thought of as a compact gauge theory where there are no dynamical monopoles, i.e., they have been suppressed by the constraint on the Villain variables n_p . Such a theory is locally the same as the non-compact \mathbb{R} -gauge theory, but not globally.

where $6V$ is the total number of plaquettes. We can express the result via the elliptic function θ_3 and obtain

$$Z_{J=\infty} = \left(\theta_3(0, e^{-\frac{\beta}{2\pi^2}}) \right)^{6V} . \tag{2.50}$$

This theory is clearly trivially gapped, and has no phase transition as a function of β . The fact that the theory is now featureless and that there is a distinction between electric and magnetic condensation is the Shenker-Fradkin continuity [26].

What about intermediate values $J_e = J_m = J$? This regime is generally strongly coupled, and we will have to resort to lattice simulations to answer the question fully, but before we do so, let us perform a qualitative analysis to see what we can expect.

2.6 Field-theoretical description away from self-duality

Let us first consider moving away from self-duality by studying the regime where the inverse gauge coupling β is large. The electric matter is then weakly coupled, but the magnetic matter is strongly coupled.²³ So we expect the magnetic matter to get a large mass and decouple from the system, and the system is described by the condensation of electrically charged matter. Altering J is then related to the change of the mass squared parameter m^2 of the scalar. The limit where $J \rightarrow \infty$ corresponds to the deep Higgs phase where $m^2 \rightarrow -\infty$ causes the scalar to condense. On the other hand $J \rightarrow 0$ corresponds to the scalar decoupling limit $m^2 \rightarrow \infty$.

The transition between the two regimes is expected around $m^2 \sim 0$. If the mass is low enough, we expect a description in terms of continuum scalar QED. This is an extremely well-studied system, where the usual Coleman-Weinberg analysis [27] shows that the one-loop effective potential for the scalar generically has a minimum away from the origin, and thus the transition is expected to be first order.

This can also be understood from a renormalization group analysis, as we briefly review below. We present the argument for N_f^e scalar flavors, where the e in the superscript stands for “electric” as the matter is electrically coupled. In this section we will omit the superscript e (i.e., $N_f = N_f^e$) as we are considering only electrically coupled matter, but in later sections we will sometimes consider N_f^e electrically coupled flavors and N_f^m magnetically coupled flavors.

We now want to discuss the order of the transition, which changes for N_f sufficiently large. The continuum action describing the interaction of electrically charged matter with the photon takes the form

$$S[\phi, A] = \int d^4x \left(\frac{1}{4e^2} F^2 + \sum_{i=1}^{N_f} (|\partial\phi_i|^2 + m^2|\phi_i|^2) + \frac{\lambda}{4} \left(\sum_{i=1}^{N_f} |\phi_i|^2 \right)^2 \right) . \tag{2.51}$$

There are three couplings which are of interest: the electric coupling squared e^2 , the mass squared m^2 and the ϕ^4 coupling λ . When $m^2 > 0$, the relevance of this coupling drives

²³We note that by performing the duality transformation (2.40), the same analysis describes the large $\tilde{\beta}$ regime, and thus may be re-interpreted as the confinement transition associated with the condensation of magnetic charges. To avoid confusion we will use terminology appropriate to the Higgs transition.

the system to a free photon phase. When $m^2 < 0$ the system is likewise driven to a scalar condensed phase. We thus have to understand what happens exactly at $m^2 = 0$. At this point e^2 and λ are marginally irrelevant and an analysis of the beta functions shows that they flow logarithmically towards smaller values of the couplings e^2, λ .

The precise character of the transition now turns out to depend on N_f . The standard calculation gives the RG equations for $m^2 = 0$ ²⁴ (see e.g. [28–30]),

$$\frac{dg}{d\ell} = -\frac{N_f g^2}{3}, \quad (2.52a)$$

$$\frac{d\lambda}{d\ell} = -(N_f + 4)\lambda^2 + 6\lambda g - 6g^2, \quad (2.52b)$$

where $g = e^2$, and ℓ is the RG flow “time”.²⁵ As is commonly done, these RG equations only take into account the marginal couplings, and ignore the infinite set of irrelevant couplings that a-priori are not expected to be important.

The point $\lambda = e^2 = 0$ is a fixed point. The question is whether this fixed point is reached. If it is, then the transition at $m^2 = 0$ is 2nd order. Naively, since both g and λ are marginally irrelevant, it seems this is always true. However, note that it takes infinite flow time ℓ for g to reach zero, while it is not excluded that λ becomes zero in finite flow time. If that happens the last term $-6g^2$ in (2.52b) will push λ to negative values. When λ is negative, however, the would-be irrelevant couplings such as ϕ^6 become important as they stabilize the potential, i.e., they become dangerously irrelevant. If this happens the system flows to a Higgs phase at $m^2 = 0$, while for $m^2 > 0$ it is in a photon phase, rendering the transition discontinuous.

This indeed happens for N_f small enough as we explain in the appendix, and in particular for $N_f = 1$.²⁶ On the other hand, if one takes the large N_f limit in equations (2.52), implies that the fixed point $e^2 = \lambda = 0$ is generically reached for N_f sufficiently large.²⁷

Then there should exist a window $1 < N_f < N_f^*$ for which the couplings (e^2, λ) flow in the infrared towards the origin but then generically *miss* the mean-field point at zero coupling $(e^2, \lambda) = (0, 0)$, instead heading off towards large negative coupling λ , leading to a 1st order transition.

However, when $N_f \geq N_f^*$, the dynamics along the RG flow and thus the topology of the solution space is different such that there exists an open set of initial data with positive (e^2, λ) which are attracted in the far IR towards $(e^2, \lambda) = (0, 0)$, rather than generically “missing” the origin. Thus one can now arrive at the gapless weak-coupling critical point

²⁴The overall coefficient of the 2nd order RG equations can always be chosen by a simple redefinition $g_i \rightarrow c g_i$ for constant c . We have decided to completely remove the factors of π which come from the integration over the volume of the 3-sphere.

²⁵The “flow time” is given by $\ell = -\log(\mu/\mu_0)$ where μ_0 is the initial mass-dimension cutoff scale, and $\mu < \mu_0$ is the final cutoff scale.

²⁶For $N_f = 1$, the 1st order transition also follows from the standard effective potential of Coleman and Weinberg [27].

²⁷While the large N_f limit seems to indicate that this happens for any initial values, this is in fact not the case. The reason is that the decay of λ toward zero is faster than the decay of e^2 , rendering subleading terms at large N_f potentially important. A careful analysis is presented in the appendix.

by tuning only m^2 , and the transition may be second order, described by a mean-field phase transition with vanishing couplings.²⁸ In appendix A.2 we show that $N_f^* = 183$.

The upshot of this analysis is that for a small number of electric flavors N_f^e (and if magnetic monopoles are heavy) the Higgsing transition separating the Coulomb from the Higgs phase is expected to be first-order. By electric-magnetic duality, this also implies that for a small number of magnetic flavors N_f^m (and if electric charges are heavy), the confinement transition separating the Coulomb phase from a confined phase is expected to be first-order. Thus both of the lines bounding the Coulomb phase in figure 2 are first-order lines.

We note that in [32, 33] it was argued that the introduction of an ad hoc monopole mass term leads to a region of the phase diagram where the confinement-deconfinement transition is 2nd order. Since this setup is in the same universality class as the Abelian-Higgs model studied here, we conclude that the transition observed in [32, 33] is never continuous but is weakly 1st order. In fact in [33] it was observed that the alleged 2nd order transition is not universal. The likely explanation is the presence of two marginally irrelevant couplings which could keep the system in the vicinity of the non-interacting fixed point until exponentially large volumes are reached, thereby obscuring the 1st order nature of the transition.

2.7 Generalization of the lattice model, symmetries and anomalies

Here we discuss several generalizations of our lattice model, the symmetries which arise and their 't Hooft anomalies. This section is quite independent from the rest of the paper, and can be skipped at first reading.

The lattice model with one flavor of electric and magnetic matter has very little symmetries, and hence is not very constrained. However, it can be generalized in several ways, which introduces more symmetries with 't Hooft anomalies. We will first consider the generalization to general charges of the dynamical electric and magnetic matter, and then to introducing flavor multiplets. We conclude by formulating a non-abelian gauge theory with a self-dual electric-magnetic symmetry.

2.7.1 General charge theories and 1-form 't Hooft anomalies

We consider coupling matter with charges q_e and q_m larger than 1, i.e., the electric matter fields couple to link phases $e^{iq_e A_l}$ and the magnetic matter to $e^{iq_m \tilde{A}_l}$. The model then has 1-form global symmetries $\mathbb{Z}_{q_e}^{[1]} \times \mathbb{Z}_{q_m}^{[1]}$, as is clear from the invariance of the action under the transformation

$$A_l^e \rightarrow A_l^e + \frac{2\pi k_l}{q_e}, \tag{2.53}$$

$$A_l^m \rightarrow \tilde{A}_l^m + \frac{2\pi \tilde{k}_l}{q_m}, \tag{2.54}$$

where k_l and \tilde{k}_l are such that

$$(dk)_p = 0 \quad \forall p \in \Lambda, \quad \text{and} \quad (d\tilde{k})_{\tilde{p}} = 0 \quad \forall \tilde{p} \in \tilde{\Lambda}. \tag{2.55}$$

²⁸It is interesting to note that the only possible conformally-invariant critical point for a parity-invariant theory with a U(1) 1-form symmetry — in this case magnetic flux conservation — is the free fixed point [31].

We can also consider introducing background gauge fields for the 1-form symmetries. As we now show, there is an obstruction to putting background gauge fields for both 1-form symmetries simultaneously, which is a manifestation of the 't Hooft anomaly. Indeed, if we want to put background fields for the electric center symmetry, we must promote the shifts (2.53) to be valid for any k_l , not necessarily only those obeying the constraint $dk = 0$. We do this by introducing a background 2-form gauge field (i.e., living on plaquettes) $B_p^e = \frac{2\pi b_p^e}{q_e}$, $b_e \in \mathbb{Z}$, and replacing in the gauge action²⁹

$$(dA^e)_p + 2\pi n_p \rightarrow (dA^e)_p + 2\pi n_p + \frac{2\pi b_p^e}{q_e}. \tag{2.56}$$

The 1-form electric center symmetry is now a gauge symmetry

$$A_l^e \rightarrow A_l^e + \frac{2\pi k_l}{q_e}, \tag{2.57}$$

$$b_p^e \rightarrow b_p^e - (dk)_p. \tag{2.58}$$

However, note that now B_p^e is supposed to be a \mathbb{Z}_{q_e} -valued gauge field, which means that setting b_p^e to be an integer multiple of q_e , i.e., $b_p^e = m_p q_e$, $m_p \in \mathbb{Z}$, should be the same as not putting a background at all. Indeed for a gauge-kinetic term, we can always shift $n_p \rightarrow n_p - m_p$ to absorb such a field, but must recall that n_p also appears in the Lagrange multiplier term involving the magnetic field \tilde{A}_l^m . Hence we also must replace

$$i \sum_p (d\tilde{A}^m)_{\star p} n_p \rightarrow i \sum_p (d\tilde{A}^m)_{\star p} \left(n_p + \frac{b_p^e}{q_e} \right). \tag{2.59}$$

So far so good. We have managed to put a consistent background gauge field for a 1-form electric center symmetry. But now we find that in order to promote the magnetic center symmetry to a gauge symmetry, we must replace

$$(d\tilde{A}^m)_{\tilde{p}} \rightarrow \left((d\tilde{A}^m)_{\tilde{p}} + \frac{2\pi \tilde{b}_{\tilde{p}}^m}{q_m} \right), \quad \tilde{b}_{\tilde{p}}^m \in \mathbb{Z}. \tag{2.60}$$

In particular the Lagrange multiplier term in the action becomes

$$i \sum_p (d\tilde{A}^m)_{\star p} \left(n_p + \frac{b_p^e}{q_e} \right) \rightarrow i \sum_p \left((d\tilde{A}^m)_{\star p} + \frac{2\pi \tilde{b}_{\star p}^m}{q_m} \right) \left(n_p + \frac{b_p^e}{q_e} \right). \tag{2.61}$$

But in the above, the shift $\tilde{b}_{\tilde{p}}^m$ by arbitrary integer multiples of q_m is not a symmetry, as such shifts will get a contribution from the cross term between \tilde{b}^m and b^e . Indeed we find that under $\tilde{b}_{\tilde{p}}^m \rightarrow \tilde{b}_{\tilde{p}}^m + r_{\tilde{p}} q_m$, $r_{\tilde{p}} \in \mathbb{Z}$, the action changes by

$$i 2\pi \frac{r_{\star p}}{q_e} b_p^e. \tag{2.62}$$

²⁹Notice that despite b_e taking all integer values, it should be thought of as a \mathbb{Z}_{q_e} gauge field, because shifts of $b_e \rightarrow b_e + q_e$ can be absorbed by a corresponding shift of the dynamical Villain variables $n_p \rightarrow n_p - 1$.

Can this non-invariance be fixed by a local counter-term? Indeed such a term should be linear in b^e and \tilde{b}^m to reproduce the transformation above. Moreover, it should be invariant under the shift of b_p^e by integer multiples of q_e . The only such counter-term is given by

$$S_{\text{counter}} = -i \frac{2\pi P}{q_e} \sum_p \tilde{b}_{\star p}^m b_p^e, \quad P = 0, 1, 2, \dots, q_e. \quad (2.63)$$

Now the shift $b_{\tilde{p}}^m \rightarrow \tilde{b}_{\tilde{p}}^m + r_{\tilde{p}} q_m$ changes the counter-term as follows

$$\Delta S_{\text{counter}} = -i \frac{2\pi P q_m}{q_e} \sum_p r_{\star p} b_p^e, \quad (2.64)$$

which restricts P to be such that

$$P q_m = 1 \pmod{q_e}. \quad (2.65)$$

If such a P can be found then there is no mixed anomaly between the two 1-form symmetries. In particular the above condition translates to $\text{GCD}(q_m, q_e) = 1$. So we found that there is a mixed 't Hooft anomaly if $\text{GCD}(q_m, q_e) \neq 1$.

2.7.2 General numbers of electric and magnetic flavors

We now consider $\text{SU}(N_f^e)$ and $\text{SU}(N_f^m)$ matter multiplets, which can be either fermionic or bosonic. We couple these to the electric and magnetic gauge fields respectively. This is done in the standard way by with a gauge invariant hopping term

$$\sum_{x,\mu} \sum_{I=1}^{N_f^e} \phi_x^{I*} e^{i A_{x,\mu}^e} \phi_{x+\hat{\mu}}^I + c.c., \quad (2.66)$$

where ϕ_x is the a matter field on the lattice, coupling to an electric gauge field A^e . The index I is an $\text{SU}(N_f^e)$ flavor index. A similar expression can be written for the magnetic multiplet,

$$\sum_{\tilde{x},\mu} \sum_{I=1}^{N_f^m} \tilde{\phi}_{\tilde{x}}^{I*} e^{i \tilde{A}_{\tilde{x},\mu}^m} \tilde{\phi}_{\tilde{x}+\hat{\mu}}^I + c.c.. \quad (2.67)$$

Note that the global symmetry is not $\text{SU}(N_f^e) \times \text{SU}(N_f^m)$ but $\text{PSU}(N_f^e) \times \text{PSU}(N_f^m)$,³⁰ because the transformation by the center element which takes $\phi_x^I \rightarrow \exp\left(\frac{2\pi i}{N_f^e}\right) \phi_x^I$ is part of the $\text{U}(1)$ electric gauge transformation, and similarly for the transformation with the center of $\text{SU}(N_f^m)$.

Now let us couple the symmetry to a background gauge field. We promote

$$\sum_{x,\mu} \sum_{I=1}^{N_f^e} \phi_x^{I*} e^{i A_{x,\mu}^e} \phi_{x+\hat{\mu}}^I \rightarrow \sum_{x,\mu} \sum_{I,J=1}^{N_f^e} \phi_x^{I*} U_{x,\mu}^{e,IJ} \phi_{x+\hat{\mu}}^J \quad (2.68)$$

where $U_{x,\mu}^e$ is an $\text{SU}(N_f^e)$ matrix. However, recall that the symmetry we wish to gauge is really $\text{PSU}(N_f^e)$, while our link matrices $U_{x,\mu}^e$ are $\text{SU}(N_f^e)$. To do this we must make sure

³⁰ $\text{PSU}(N) \equiv \text{SU}(N)/\mathbb{Z}_N$.

that there exists a gauge symmetry $U_{x,\mu}^e \rightarrow \exp\left(i\frac{2\pi s_{x,\mu}}{N_f^e}\right) U_{x,\mu}^e$, with $s_{x,\mu} = 1, \dots, N_f^e - 1$, which effectively removes the center of $SU(N_f)$ and renders the gauge field a $PSU(N_f^e)$ gauge field. This can be accomplished by shifting $A_{x,\mu}^e \rightarrow A_{x,\mu}^e - \frac{2\pi s_{x,\mu}}{N_f^e}$, which is not a symmetry of the kinetic term however. To turn it into a symmetry we introduce a background 2-form $\mathbb{Z}_{N_f^e}$ gauge field b_p^e like before and replace the kinetic term and the Lagrange multiplier term by

$$\left((dA^e)_p + 2\pi n_p\right)^2 \rightarrow \left((dA^e)_p + 2\pi n_p + \frac{2\pi b_p^e}{N_f^e}\right)^2, \quad (2.69)$$

$$i(d\tilde{A}^m)_{\star p} \left(n_p + \frac{b_p^e}{q_e}\right) \rightarrow i(d\tilde{A}^m)_{\star p} \left(n_p + \frac{b_p^e}{N_f^e}\right). \quad (2.70)$$

Note that the gauge invariant flux is now $(dA^e)_p + 2\pi n_p + \frac{2\pi b_p^e}{N_f^e}$, and can be fractional.³¹

Now the background gauge field $U_{x,\mu}^e \in SU(N_f^e)$ has its center promoted to a gauge symmetry, and hence represents a $PSU(N_f^e)$ gauge background. As part of this background we had to introduce 2-form gauge fields which are meaningful mod N_f^e , i.e., $b_p^e = 0, 1, \dots, N_f^e - 1 \pmod{N_f^e}$. The remaining steps closely follow the discussion of the 't Hooft anomaly with 1-form symmetries. We spell them out for convenience.³²

We first repeat the same for $PSU(N_f^m)$, and replace the expression (2.67) with

$$\sum_{\tilde{x},\mu} \sum_{I,J=1}^{N_f^m} \tilde{\phi}_{\tilde{x}}^{I*} U_{\tilde{x},\mu}^{m,IJ} \tilde{\phi}_{\tilde{x}+\hat{\mu}}^J + c.c. \quad (2.71)$$

Again we have to promote the shift of an $SU(N_f^m)$ by a center to a gauge symmetry, $U_{\tilde{x},\mu}^m \rightarrow \exp\left(\frac{2\pi i s_{\tilde{x},\mu}}{N_f^m}\right) U_{\tilde{x},\mu}^m$ must be a symmetry for any $s_{\tilde{x},\mu} = 1, 2, \dots, N_f^m - 1$ and for any link (\tilde{x}, μ) . Now we have a problem similar to the one we encountered when we spoke about center symmetries. To gauge the full $PSU(N_f^m)$ we must replace $(d\tilde{A}^m)_{\tilde{p}} \rightarrow (d\tilde{A}^m)_{\tilde{p}} + \frac{2\pi \tilde{b}_{\tilde{p}}}{N_f^m}$, where $\tilde{b}_{\tilde{p}} \in \mathbb{Z}$ serves as a $\mathbb{Z}_{N_f^m}$ 2-form gauge field, i.e., we replace

$$i \sum_p (d\tilde{A}^m)_{\star p} \left(n_p + \frac{b_p^e}{N_f^e}\right) \rightarrow i \sum_p \left((d\tilde{A}^m)_{\star p} + \frac{2\pi \tilde{b}_{\star p}^m}{N_f^m}\right) \left(n_p + \frac{b_p^e}{N_f^e}\right). \quad (2.72)$$

Here we have the same problem as in the case of higher charged matter, with the similar conclusion that an 't Hooft anomaly between $PSU(N_f^e)$ and $PSU(N_f^m)$ exists if $\text{GCD}(N_f^e, N_f^m) \neq 1$.

³¹We can additionally demand that $(db^e)_c = 0 \pmod{2\pi}$, i.e., that b_p^e is a representative of $H^2(M, \mathbb{Z}_{N_f^e})$, where M is the underlying space-time manifold. This will not change any arguments below. Then b_p^e is a representative of the well-known characteristic class $w_2 \in H^2(M, \mathbb{Z}_{N_f^e})$ of the $PSU(N_f^e)$ principal bundle.

³²A way to understand the similarity is to introduce $SU(N_f^e)$ gauge fields without introducing the b^e field. In the presence of $SU(N_f^e)$ gauge fields only, the setup has a global bonus 1-form $\mathbb{Z}_{N_f^e}^{[1]}$ symmetry, and the anomaly analysis largely reduces to the analysis of the global 1-form symmetries we discussed previously. This bonus 1-form symmetry is precisely there because by considering only $SU(N_f^e)$ background fields, we effectively ignore $PSU(N_f^e)$ gauge bundles which cannot be lifted to $SU(N_f^e)$ gauge bundles. Then we can think of placing background gauge fields b^e for this bonus 1-form symmetry, which is the same as introducing the $w_2 \in H^2(M, \mathbb{Z}_{N_f^e})$ classes we discussed previously.

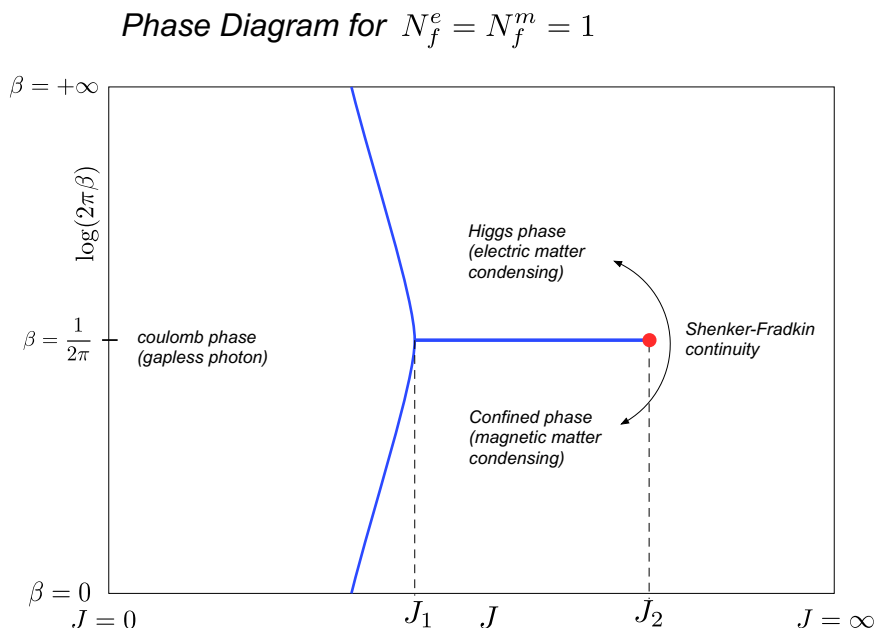


Figure 2. A cartoon of the phase diagram in the $J, \log(2\pi\beta)$ plane.

2.7.3 The self-dual non-abelian gauge theory

This will be our last generalization, and it is a small change on the multi-flavor theory we just discussed. The idea is to now promote the $SU(N_f^e)$ and $SU(N_f^m)$ link field U_l^e and U_l^m gauge background into a dynamical one. This leaves the theory as having two 1-form symmetries $\mathbb{Z}_{N_f^e}^{[1]} \times \mathbb{Z}_{N_f^m}^{[1]}$. If $\text{GCD}(N_f^e, N_f^m) \neq 1$, the two 1-form symmetries have mixed ‘t Hooft anomalies. This then is a model which is a fully fledged non-abelian gauge theory with gauge group $SU(N_f^e) \times SU(N_f^m)$, and a self-dual symmetry exchanging the two non-abelian gauge groups $SU(N_f^e)$ and $SU(N_f^m)$. This theory is extremely interesting, as it is a fully non-abelian theory, with electric-magnetic duality. However, all these theories have a complex-action problem, and cannot be simulated in the form we so far used, and so-called worldline representations of either electric or magnetic matter must be used. If the matter is charged under a dynamical gauge group, it is not clear that worldline representation will be useful in solving the complex-action problem.

2.8 The phase diagram of self-dual scalar QED

Here we collate all of the information we discussed so far to come up with a phase diagram for the self-dual lattice theories. We focus primarily on $N_f^e = N_f^m = 1$ with only bosonic matter, whose phase diagram is sketched in figure 2. We plot the diagram as a function of $J = J^e = J^m$ and $\log(2\pi\beta)$, because under self-duality $\log(2\pi\beta) \rightarrow -\log(2\pi\beta)$, such that the diagram has to have a symmetry around the line $\log(2\pi\beta) = 0$, which is the self-dual line.

On the one hand we said that as $J^e = J^m = J \rightarrow \infty$ we have a gapped phase, with no phase transition as the coupling β is varied. This is illustrated on the far right of the diagram. The phase $J = 0$ is a free photon phase, which is the far left of the diagram. The

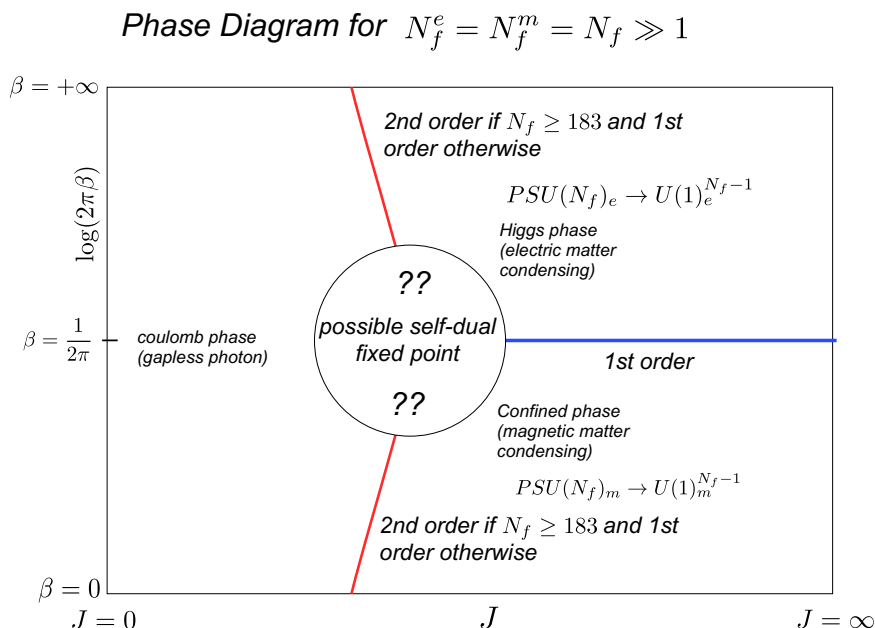


Figure 3. The conjectured phase diagram for sufficiently large $N_f^e = N_f^m = N_f$. The limits $\beta \sim 0$ or $\beta \sim +\infty$ can have a 2nd order phase transition if $N_f \geq 183$, as discussed in section 2.5.

limits $\beta \rightarrow \infty$ and $\beta \rightarrow 0$ are treatable perturbatively in the Abelian-Higgs model, and exhibit a 1st order transition for $N_f = 1$.

Now let us consider the self-dual line $\beta = 1/(2\pi)$ and change J from zero to infinity. Since the theory will be trivially gapped for J sufficiently large, we expect a phase transition. In fact dialing J to higher values is driving the system towards preferring a condensation of matter. But since condensing electric matter (the Higgs phase) confines magnetic matter, and condensing magnetic matter (the confined phase) confines electric matter, a tension is expected between the condensation of electric and magnetic matter so that a phase coexistence line should emerge at some value $J = J_1$. Since the electric and magnetic condensed phases coexist, this regime spontaneously breaks the self-dual symmetry. This is depicted by the horizontal blue line segment in figure 2.

On the other hand if we crank up J sufficiently high, we know that eventually the phase coexistence will disappear, as we discussed in section 2.5, and the coexistence phase should disappear at some value $J = J_2 > J_1$. The disappearance of the 1st order line is a critical point. Since the critical point restores the discrete self-dual symmetry, we expect this to be in the 4d Ising universality class, which is a Gaussian fixed point. Indeed our numerical results will agree with this.

A more interesting question is what happens at the transition between the free-photon phase and the self-dual broken phase, i.e., the leftmost point of the horizontal segment in figure 2. This is the point where the 1st order transitions which were computable in the $\beta = 0$ and $\beta = \infty$ limits, meet at the self-dual point. As we will see, numerical computations reveal this point to be a triple point, i.e., a coexistence point of three phases.

Let us quickly discuss the phase diagram for other generalizations. We first consider $N_f^e = N_f^m = 1$, theory with charges $q^m = q^e = q > 1$. As we discussed, this theory has a mixed anomaly between $\mathbb{Z}_q^{[1]}$ -electric and $\mathbb{Z}_q^{[1]}$ -magnetic 1-form symmetries. These anomalies are matched by either a photon phase or by spontaneous breaking of electric and/or magnetic 1-form symmetries. If the $\mathbb{Z}_q^{[1]}$ -electric symmetry is spontaneously broken, we call that the Higgs phase, while the confined phase spontaneously breaks $\mathbb{Z}_q^{[1]}$ -magnetic. However, now these two phases cannot be continuously connected, because they break different symmetries, so instead of the coexistence line ending, we expect that it continues all the way to $J = \infty$.

A similar picture is expected for $N_f^e = N_f^m = N_f > 1$ theories, except that now the Higgs and confined phases are respectively breaking $\text{PSU}(N_f)$ -electric and $\text{PSU}(N_f)$ -magnetic global symmetries, which will be gapless. Another point we should make is that for sufficiently large N_f , we saw that the regimes $\beta \rightarrow 0$ and $\beta \rightarrow \infty$ have 2nd order phase transitions, and so it is possible that this regime continues all the way to the self-dual point. This would indicate the existence of a new, self-dual fixed point. This is sketched in figure 3. A similar phenomenon is known to occur for \mathbb{Z}_2 gauge theory in 3d, where it is believed that two continuous Ising lines meet at a novel self-dual critical point [34].

3 Numerical simulation

3.1 Switching to a dual worldline formulation

We now want to study the minimal interacting self-dual theory with $N_f^e = N_f^m = 1$ numerically and analyze the phase diagram sketched in figure 2. However, self-dual $U(1)$ gauge theory coupled to electric and magnetic matter as introduced in section 2.4 is not yet suitable for a numerical simulation, since the gauge field Boltzmann factor (2.34) has a complex phase and does not give rise to a real and positive weight that can be used in a Monte Carlo simulation. In this subsection we now show that this complex action problem can be overcome by switching to a worldline formulation for the magnetic matter.

In order to prepare the Boltzmann factor (2.35) for the worldline formulation we rewrite the second exponent in (2.35) by switching to the dual lattice using the identity³³

$$(dn)_{x,\mu\nu\rho} = - \sum_{\sigma} \epsilon_{\mu\nu\rho\sigma} (\partial \tilde{n})_{\tilde{x}-\hat{\sigma},\sigma} , \tag{3.1}$$

which is straightforward to check (see the appendix of [15]). Thus the gauge field Boltzmann factor assumes the form

$$B_{\beta}[A^e, A^m] = \sum_{\{n\}} e^{-\beta S_g[A^e, n]} \prod_{\tilde{x},\mu} e^{-i\tilde{A}_{\tilde{x},\mu}^m (\partial \tilde{n})_{\tilde{x},\mu}} , \tag{3.2}$$

where we have converted the sum in the second exponent of the Boltzmann factor into a product over all links of the dual lattice.

³³We remark that the step of switching to the dual Villain variables is not essential but simplifies the discussion here and also the actual computer code used in the simulations discussed in subsections 3.3 and 3.4.

The second step is to use the well known worldline representation for U(1) gauge field theories (see, e.g., [35, 36]). It is straightforward to convert this worldline representation to the dual lattice where the magnetic matter partition sum (2.38) is defined. The worldline representation then reads (compare the appendix of [9] for the notation used here)

$$\tilde{Z}[\tilde{A}^m, J_m] = \sum_{\{\tilde{k}\}} \left[\prod_{\tilde{x}, \mu} I_{\tilde{k}_{\tilde{x}, \mu}}(J_m) \right] \left[\prod_{\tilde{x}} \delta((\partial \tilde{k})_{\tilde{x}}) \right] \left[\prod_{\tilde{x}, \mu} e^{i \tilde{A}_{\tilde{x}, \mu}^m \tilde{k}_{\tilde{x}, \mu}} \right], \quad (3.3)$$

where $I_n(x)$ denotes the modified Bessel functions. The partition function is a sum over configurations of the dual flux variables $\tilde{k}_{\tilde{x}, \mu} \in \mathbb{Z}$ assigned to the links (\tilde{x}, μ) of the dual lattice, where

$$\sum_{\{\tilde{k}\}} \equiv \prod_{\tilde{x}, \mu} \sum_{\tilde{k}_{\tilde{x}, \mu} \in \mathbb{Z}}. \quad (3.4)$$

The flux variables are subject to vanishing divergence constraints

$$(\partial \tilde{k})_{\tilde{x}} \equiv \sum_{\mu=1}^d [\tilde{k}_{\tilde{x}, \mu} - \tilde{k}_{\tilde{x}-\hat{\mu}, \mu}] = 0 \quad \forall \tilde{x}, \quad (3.5)$$

which in (3.3) are implemented with the product of Kronecker deltas. These constraints enforce flux conservation at each site \tilde{x} of the dual lattice, such that the $\tilde{k}_{\tilde{x}, \mu}$ form closed loops of flux on the dual lattice. At every link (\tilde{x}, μ) of the dual lattice the dual magnetic gauge field $\tilde{A}_{\tilde{x}, \mu}^m$ couples in the form $e^{i \tilde{A}_{\tilde{x}, \mu}^m \tilde{k}_{\tilde{x}, \mu}}$ which gives rise to the second product in (3.3). The configurations of the dual flux variables $\tilde{k}_{\tilde{x}, \mu}$ come with real and positive weight factors given by the Bessel functions.

With the gauge field Boltzmann factor in the form (3.2) and the dependence of the partition sum $\tilde{Z}[\tilde{A}^m, J_m]$ on the dual magnetic gauge field $\tilde{A}_{\tilde{x}, \mu}^m$ given by the last factor in (3.3) we can now completely integrate out the dual magnetic gauge field. The corresponding integral reads (compare (2.34) and use $\int D[A^m] = \int D[\tilde{A}^m]$),

$$\begin{aligned} \int D[\tilde{A}^m] \left[\prod_{\tilde{x}, \mu} e^{-i \tilde{A}_{\tilde{x}, \mu}^m (\partial \tilde{n})_{\tilde{x}, \mu}} \right] \left[\prod_{\tilde{x}, \mu} e^{i \tilde{A}_{\tilde{x}, \mu}^m \tilde{k}_{\tilde{x}, \mu}} \right] &= \prod_{\tilde{x}, \mu} \int_{-\pi}^{\pi} \frac{d \tilde{A}_{\tilde{x}, \mu}^m}{2\pi} e^{i \tilde{A}_{\tilde{x}, \mu}^m [\tilde{k}_{\tilde{x}, \mu} - (\partial \tilde{n})_{\tilde{x}, \mu}]} \\ &= \prod_{\tilde{x}, \mu} \delta(\tilde{k}_{\tilde{x}, \mu} - (\partial \tilde{n})_{\tilde{x}, \mu}). \end{aligned} \quad (3.6)$$

Integrating out the dual magnetic gauge fields has generated link-based constraints that completely determine the flux variables as

$$\tilde{k}_{\tilde{x}, \mu} = (\partial \tilde{n})_{\tilde{x}, \mu} \quad \forall (\tilde{x}, \mu). \quad (3.7)$$

Note that the configurations (3.7) also obey the vanishing divergence constraints $\partial \tilde{k} = 0$ from (3.5), due to $\partial^2 = 0$ (see the appendix of [15]).

Thus we may summarize the final form of self-dual scalar lattice QED with a worldline representation for the magnetic matter:

$$Z(\beta, J_e, J_m) = \int D[A^e] \sum_{\{n\}} \int D[\phi^e] e^{-\beta S_g[A^e, n] + J_e S_e[\phi^e, A^e]} \left[\prod_{\tilde{x}, \mu} I_{(\partial \tilde{n})_{\tilde{x}, \mu}}(J_m) \right]. \quad (3.8)$$

Obviously all weight factors in (3.8) are real and positive, such that this form now is accessible to numerical Monte Carlo simulations. Note that here the Villain variables are not subject to any constraints, which in some aspects makes a numerical simulation of (3.8) simpler than the simulation of the pure gauge theory (2.18), where configurations of the Villain variables need to obey the closedness constraint (2.8).

It is interesting to consider the limit $J_m \rightarrow 0$. Using the fact that for the Bessel functions $\lim_{x \rightarrow 0} I_0(x) = 1$ and $\lim_{x \rightarrow 0} I_n(x) = 0 \quad \forall n \neq 0$, one finds that in (3.8) only those configurations of the Villain variables survive where the dual Villain variables obey

$$(\partial \tilde{n})_{\tilde{x}, \mu} = 0 \quad \forall (\tilde{x}, \mu) \quad \iff \quad (dn)_{x, \mu\nu\rho} = 0 \quad \forall (x, \mu\nu\rho), \quad (3.9)$$

where in the second form we used (3.1) to identify this constraint as the closedness condition for the Villain variables on the original lattice. Thus we find

$$\lim_{J_m \rightarrow 0} Z(\beta, J_e, J_m) = \int D[A^e] \sum_{\{n\}} e^{-\beta S_g[A^e, n]} Z[A^e, J_e] \left[\prod_x \prod_{\mu < \nu < \rho} \delta((dn)_{x, \mu\nu\rho}) \right]. \quad (3.10)$$

Although not self-dual, this is an interesting theory in its own right, as it describes U(1) lattice gauge fields coupled to electric matter without magnetic monopoles that appear in the usual lattice discretization of this system. Finally we remark that a second limit $J_e \rightarrow 0$ reduces the partition sum to our pure gauge theory partition sum (2.18) without monopoles.

We conclude the discussion of the worldline form by expressing the expectation value $\langle \tilde{s}_m \rangle_{\beta, J_e, J_m}$ that appears in the duality relation (2.43) in terms of the worldline variables. The expectation value is obtained from a derivative of $\ln Z$ with respect to J_m , and this derivative can of course also be applied to $Z(\beta, J_e, J_m)$ in the form (3.8). A few lines of algebra give

$$\langle \tilde{s}_m \rangle_{\beta, J_e, J_m} = \frac{1}{4V} \frac{\partial}{\partial J_m} \ln Z(\beta, J_e, J_m) = \frac{1}{4V} \left\langle \sum_{\tilde{x}, \mu} \frac{I'_{(\partial \tilde{n})_{\tilde{x}, \mu}}(J_m)}{I_{(\partial \tilde{n})_{\tilde{x}, \mu}}(J_m)} \right\rangle_{\beta, J_e, J_m}. \quad (3.11)$$

3.2 The self-dual point revisited

To prepare for the numerical simulations presented in the next two subsections we here discuss suitable observables at the self-dual point of the inverse gauge coupling, i.e., at

$$\beta = \tilde{\beta} = \beta^* \equiv \frac{1}{2\pi}. \quad (3.12)$$

Furthermore we may set J_e and J_m equal to the same value J such that

$$J_e = J_m = \tilde{J}_e = \tilde{J}_m = J. \quad (3.13)$$

In other words, the theory has only one remaining parameter, i.e., the coupling J .

With this choice for the couplings the self-duality relation (2.41) simplifies to,

$$\beta^* \langle F^2 \rangle_{\beta^*, J} + \beta^* \langle F^2 \rangle_{\beta^*, J} = 1, \quad (3.14)$$

which implies that $\langle F^2 \rangle_{\beta^*, J}$ is constant,

$$\langle F^2 \rangle_{\beta^*, J} = \pi \quad \forall J. \quad (3.15)$$

We remark, that the self-duality relation (2.42) for the second moment of F^2 does not constrain the susceptibility for the self-dual couplings (3.12), (3.13).

The self-duality relation (2.43) that links the electric and the magnetic action densities, for the self-dual couplings (3.12), (3.13) assumes the form

$$\langle s_e \rangle_{\beta^*, J} = \langle \tilde{s}_m \rangle_{\beta^*, J} \quad \forall J. \quad (3.16)$$

An interesting question, already touched upon in the previous subsections, is whether self-duality can be broken spontaneously as a function of J . Such a symmetry breaking should become manifest in a violation of the two relations (3.15) and (3.16).

For the further analysis we introduce the two order parameters

$$M_g \equiv |F^2 - \pi| \quad \text{and} \quad M_m \equiv |s_e - s_g|, \quad (3.17)$$

which are normalized such that

$$\langle M_g \rangle_{\beta^*, J} \neq 0 \quad \text{and} \quad \langle M_m \rangle_{\beta^*, J} \neq 0, \quad (3.18)$$

signal the breaking of self-duality. The absolute value in the definitions (3.17) was introduced to allow for a non-zero expectation value also on a finite lattice. We will also analyze the corresponding susceptibilities

$$\chi_g \equiv V \left\langle \left(M_g - \langle M_g \rangle_{\beta^*, J} \right)^2 \right\rangle_{\beta^*, J} \quad \text{and} \quad \chi_m \equiv V \left\langle \left(M_m - \langle M_m \rangle_{\beta^*, J} \right)^2 \right\rangle_{\beta^*, J}, \quad (3.19)$$

as well as the Binder cumulants

$$U_g \equiv 1 - \frac{\langle (M_g)^4 \rangle_{\beta^*, J}}{3 \langle (M_g)^2 \rangle_{\beta^*, J}^2} \quad \text{and} \quad U_m \equiv 1 - \frac{\langle (M_m)^4 \rangle_{\beta^*, J}}{3 \langle (M_m)^2 \rangle_{\beta^*, J}^2}. \quad (3.20)$$

3.3 Setup of the computation and general results

In this subsection we present our results for the simulation of the full theory at the self-dual point as discussed in subsection 3.2, i.e., at $\beta = \beta^* \equiv 1/2\pi$ with $J_e = J_m = J$. In a second study we keep J fixed and vary β in the vicinity of β^* for studying the nature of the transition when crossing the critical line of the phase diagram shown in figure 2.

These simulations are based on the partition sum in the worldline form (3.8), which uses the electric gauge fields $A_\mu^n(x) \in [-\pi, \pi)$ and the Villain variables $n_{x,\mu\nu} \in \mathbb{Z}$ for the gauge field degrees of freedom, and $\phi_x^e = e^{i\varphi_x}$, $\varphi_x \in [-\pi, \pi)$ for the electric matter. There are no constraints left for these variables and they can be updated efficiently using local Metropolis updates, which we organize in sweeps, i.e., one Metropolis update of all degrees of freedom. For equilibration we use 10^6 sweeps followed by 10^5 measurements of our observables separated by 20 sweeps for decorrelation. For the finite size scaling analysis of the critical

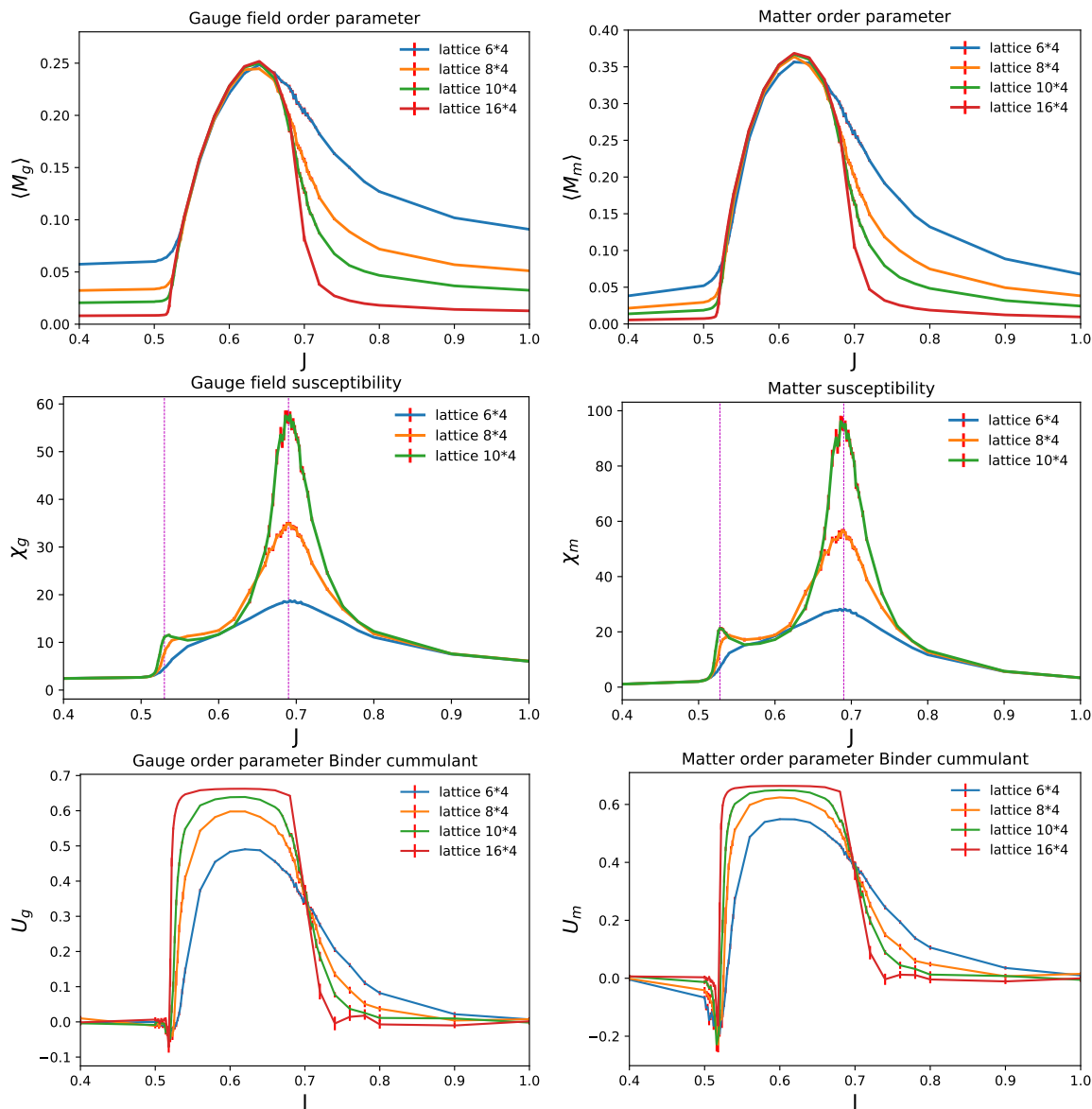


Figure 4. Vacuum expectation value (top row), susceptibility (middle row) and Binder cumulant (bottom row) for the gauge field order parameter M_g (lhs. column) and the matter order parameter M_m (r.h.s. column). The results are for the self-dual value $\beta = \beta^* \equiv 1/2\pi$ and are plotted as a function of J . We compare different volumes as indicated in the legends.

exponents the number of measurements is increased to 10^6 . The error bars we show are statistical errors determined with the jackknife method combined with a blocking analysis.

In figure 4 we show our results for the order parameters $\langle M_g \rangle$, $\langle M_m \rangle$ (top row), for the susceptibilities χ_g , χ_m (middle row) and for the Binder cumulants U_g and U_m (bottom row). The lhs. column shows the result for the respective gauge field quantities, while the rhs. column is for the matter fields. The observables are plotted as a function of J and were determined for volumes 4^4 , 6^4 , 8^4 and 10^4 at fixed gauge coupling $\beta = \beta^* = 1/2\pi$, i.e., the self-dual value.

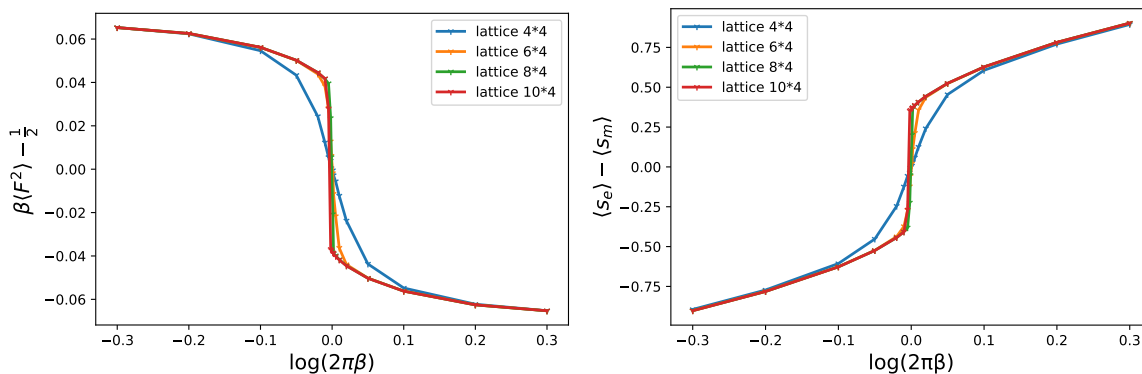


Figure 5. The gauge field order parameter (lhs.) and the matter order parameter (rhs.) as a function of the rescaled gauge coupling $\ln(2\pi\beta)$, for values across the self-dual value $\beta^* \equiv 1/2\pi$. The matter field coupling is set to $J = 0.6$ here.

All observables suggest that there is indeed spontaneous symmetry breaking as a function of J with endpoints located at $J_1 \sim 0.52$ and $J_2 \sim 0.7$. Below J_1 and above J_2 the order parameters $\langle M_g \rangle$ and $\langle M_m \rangle$ approach 0 in the infinite volume limit, while inside the interval (J_1, J_2) they remain finite. The corresponding susceptibilities develop peaks near J_1 and J_2 that scale with the volume. Finally the Binder cumulants allow for a first assessment of the nature of the endpoints: near J_1 they develop minima which hints at a first order endpoint at J_1 . At J_2 the Binder cumulants for the different volumes intersect in a common point which indicates second order behavior at the endpoint $J = J_2$. We will study the nature of the two endpoints in more detail in the next subsection.

As already announced, we also want to cross-check the nature of the transition when vertically crossing the critical line in the phase diagram figure 2. For this study we now keep the matter field coupling fixed at $J = 0.6$ and vary the gauge field coupling β in the vicinity of $\beta = \beta^*$. The corresponding plots for the gauge field order parameter (lhs. plot) and the matter order parameter (rhs.) are shown in figure 5. In order to fully display the symmetry of the first order transition we plot the gauge field coupling on the horizontal axis in the rescaled form $\ln(2\pi\beta)$, which is odd under duality transformations and gives 0 at the self-dual point. The gauge field and the matter order parameters on the vertical axes are plotted in the form

$$\beta\langle F^2 \rangle - \frac{1}{2} \quad \text{and} \quad \langle s_e \rangle - \langle s_m \rangle, \tag{3.21}$$

which is a form that is again odd under duality transformations, as can be seen from (2.41) and (2.43). Since both, the rescaled coupling and the order parameters are odd under duality transformations, the plots for the order parameters must be antisymmetric functions of $\ln(2\pi\beta)$. This is indeed what we observe in figure 5. When comparing the different volumes we find that the observables quickly develop the discontinuity at $\ln(2\pi\beta) = 0$, which is the expected first order signature when driving the symmetry breaking coupling β through the self-dual point. Thus we confirm that the vertical line in the phase diagram figure 2 is indeed of first order.

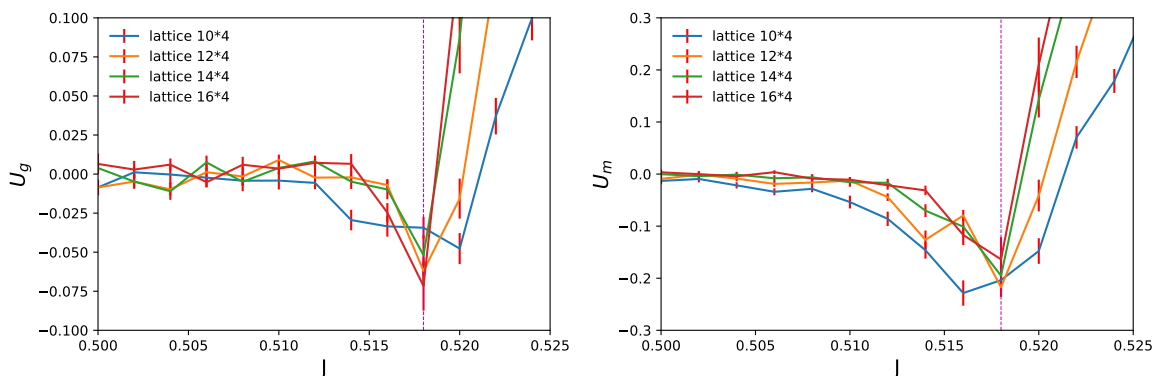


Figure 6. The Binder cumulant for the gauge field order parameter (lhs.) and the matter order parameter (rhs.). For different volumes we show the results at $\beta = \beta^*$ as a function of J and zoom into the region near the first order transition point at $J_1 \sim 0.52$.

3.4 Analysis of the endpoints

The first round of analysis in the previous subsection suggested that the endpoint at $J_1 \sim 0.52$ is first order, while the one at $J_2 \sim 0.7$ is of second order. In this subsection we now aim at determining more precisely the values of J_1 and J_2 and at characterizing the two endpoints.

For the transition at J_1 we observed the formation of minima in the two Binder cumulants. The positions of the minima converge towards the true value J_1 when increasing the volume. Since in the two bottom plots of figure 4 this is a little hard to see, in figure 6 we zoom into the region near J_1 . We find that both Binder cumulants form minima and that for all volumes except for the smallest volume 10^4 the positions of the minima agree. Thus we conclude that we find a first order transition at $J_1 = 0.518(2)$, where the error is given by the stepsize in J we use, i.e., $\Delta J = 0.002$.

For the transition at $J_2 \sim 0.7$ figure 4 suggests that the transition is of second order. In that case the Binder cumulants are expected to obey the finite size scaling formula

$$U \sim A + B L^{1/\nu} (J - J_2), \quad (3.22)$$

where A and B are constants and ν is the critical exponent for the scaling of the correlation length. Thus, when plotting the Binder cumulants as a function of $L^{1/\nu} (J - J_2)$ the results for different volumes $V = L^4$ should collapse to universal straight lines — given that ν and J_2 are chosen correctly. As discussed in subsection 2.8, we conjecture that the transition is in the 4d Ising universality class, i.e., a Gaussian fixed point, such that we expect $\nu = 1/2$. We test this hypothesis by setting $\nu = 1/2$ in the scaling formula (3.22) and in figure 7 plot the Binder cumulants as function of $L^2 (J - J_2)$, where J_2 is treated as a free parameter which we choose such that we find the best collapse of the data for different L . The lhs. plot in figure 7 shows the results for the gauge field Binder cumulant, while the rhs. is for the matter Binder cumulant. In both cases the collapse confirms the expected critical exponent $\nu = 1/2$ and the critical coupling is determined to be $J_2 = 0.700(1)$.

Finally we also try to confirm the universality class of the second order transition by analyzing the susceptibilities and the corresponding critical exponent γ , which for the

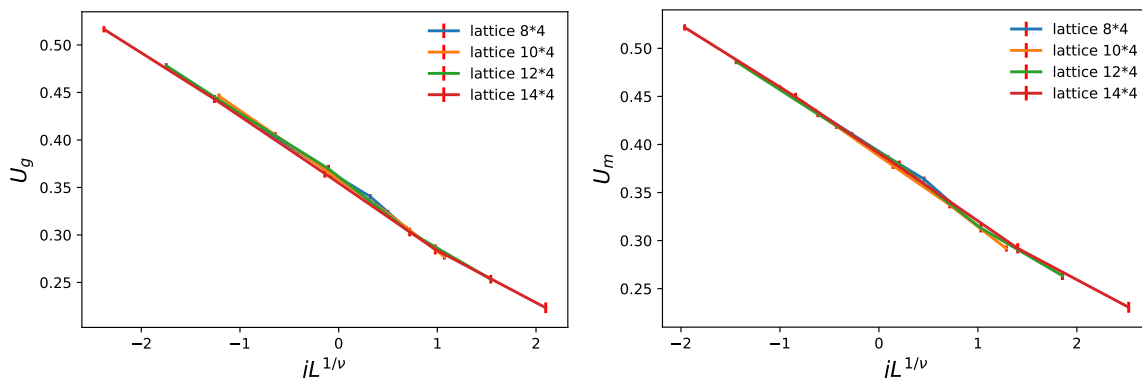


Figure 7. The Binder cumulant for the gauge field order parameter (lhs.) and the matter order parameter (rhs.). We plot the results at $\beta = \beta^*$ as a function of $jL^{1/\nu}$ with $j = J - J_2$ and $\nu = 1/2$. We compare the results for different volumes L^4 .

conjectured Gaussian fixed point would be $\gamma = 1$. Again we employ finite size scaling which for the susceptibilities takes the form (we define $j \equiv J - J_2$),

$$\chi \sim L^{\gamma/\nu} \left(g(jL^{1/\nu}) + \frac{f(jL^{1/\nu})}{\ln(L)} \right), \quad (3.23)$$

where here also the leading logarithmic corrections are taken into account. Considering only the constant term in the expansion of the scaling function f , i.e., setting $f = -a$, where a is some constant we find that $\chi L^{-\gamma/\nu} + a/\ln(L)$ should be a universal function of $jL^{1/\nu}$ for a suitably chosen parameter a . In figure 8 we plot the combination $\chi L^{-\gamma/\nu} + a/\ln(L)$ with $\gamma = 1$ as a function of $jL^{1/\nu}$ for different volumes L^4 . The lhs. plot is for the gauge field susceptibility, while the r.h.s. shows the results for the matter susceptibility. The parameter a was chosen such that the collapse of the data for the different volumes is optimized. We find that the collapse is not as good as for the Binder cumulants, but nevertheless confirm, that a value of $\gamma = 1$ is plausible. We may summarize the discussion of the endpoints as follows: at $J_1 = 0.518(2)$ we find a first order transition, while at $J_2 = 0.700(1)$ the transition is of second order with critical exponents that are compatible with the conjectured 4d Ising universality class, i.e., a Gaussian fixed point.

We finally explore the possibility expressed in subsection 2.3 that charge conjugation \mathcal{C} might be broken in the vicinity of the horizontal line of the phase diagram sketched in figure 2. For this study we consider the following two order parameters that are odd under \mathcal{C} , which in continuum language are defined as

$$T_1 \sim \sum_{\mu} j_{\mu}^e \partial_{\mu} \sum_{\alpha < \beta} F_{\alpha\beta} F_{\alpha\beta} \quad \text{and} \quad T_2 \sim \sum_{\mu} j_{\mu}^e \partial_{\mu} \sum_{\substack{\alpha < \beta \\ \rho < \sigma}} \epsilon_{\alpha\beta\rho\sigma} F_{\alpha\beta} F_{\rho\sigma}. \quad (3.24)$$

Here j_{μ}^e is the current of the electric matter which we discretize as $j_{x,\mu}^e = \text{Im} \phi_x^{e*} e^{iA_{x,\mu}} \phi_{x+\hat{\mu}}^e$. For the discretization of the field strength we use (2.3) and the partial derivative ∂_{μ} is discretized with a nearest neighbor difference.

In figure 9 we show our results for $\langle |T_1| \rangle$ and $\langle |T_2| \rangle$ in the top row of plots, while at the bottom the two expectation values are rescaled with L^2 . We work at a fixed matter

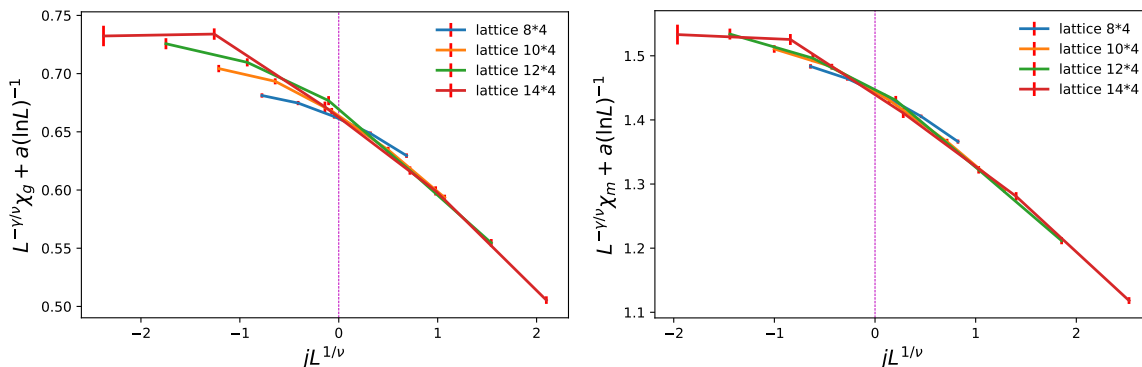


Figure 8. Finite size scaling analysis of the gauge field (lhs. plot) and matter (rhs.) susceptibilities.

coupling of $J = 0.65$ and compare four different volumes L^4 with $L = 8, 10, 12$ and 14 . The results are again plotted as a function of $\ln(2\pi\beta)$. Note that the order parameters are not symmetrized under \mathcal{S} , such that here we do not expect symmetry under $\beta \leftrightarrow \tilde{\beta}$.

The top row of plots show that both expectation values $\langle |T_1| \rangle$ and $\langle |T_2| \rangle$ vanish in the thermodynamic limit. In the bottom plots we show $L^2 \langle |T_1| \rangle$ (lhs.) and $L^2 \langle |T_2| \rangle$ (rhs.). This rescaling collapses the data and establishes that the volume scaling is $\propto 1/L^2 = 1/\sqrt{V}$, including also the small remnant of the first order transition at $\ln(2\pi\beta) = 0$, i.e., at the self dual point $\beta = 1/2\pi$.

The $1/\sqrt{V}$ behavior may be understood by viewing the values for T_1 and T_2 at some space-time point as independent random variables. The distribution of the average of random variables with sample size N has a standard deviation scaling as $1/\sqrt{N}$. In our case the values of the order parameters T_1 and T_2 on the V space-time points are not fully independent, but since there is a mass-gap, N is roughly given by the volume $V = L^4$. This gives rise to the $1/\sqrt{V}$ scaling we observe. We conclude that the order parameters T_1 and T_2 have vanishing expectation value in the thermodynamic limit and \mathcal{C} remains unbroken.

4 Conclusion and future prospects

In this work we discussed the possible phase structure of self-dual U(1) lattice gauge theories based on a modified Villain action. We have seen that the space of such theories is large, allowing arbitrary matter to be coupled electrically and/or magnetically. When coupling multiplets of such matter fields the electric and magnetic flavor symmetry that arises often has 't Hooft anomalies, eliminating the possibility of a trivially gapped phase.

An interesting question is whether any of such theories have a self-dual CFT fixed point. The natural place to look for the new fixed points is the self-dual line $\beta = 1/2\pi$ as in that case the coupling is protected by self-duality. For a single electric and magnetic bosonic flavor $N_f^e = N_f^m = N_f = 1$ we argued, and numerically confirmed, that the phase structure along the self-dual line has two transitions as the bosonic matter condenses. The transition from the photon phase to the Higgs/confined phase is 1st order, continuing in the Higgs/monopole coexistence phase, which breaks the self-dual symmetry spontaneously.

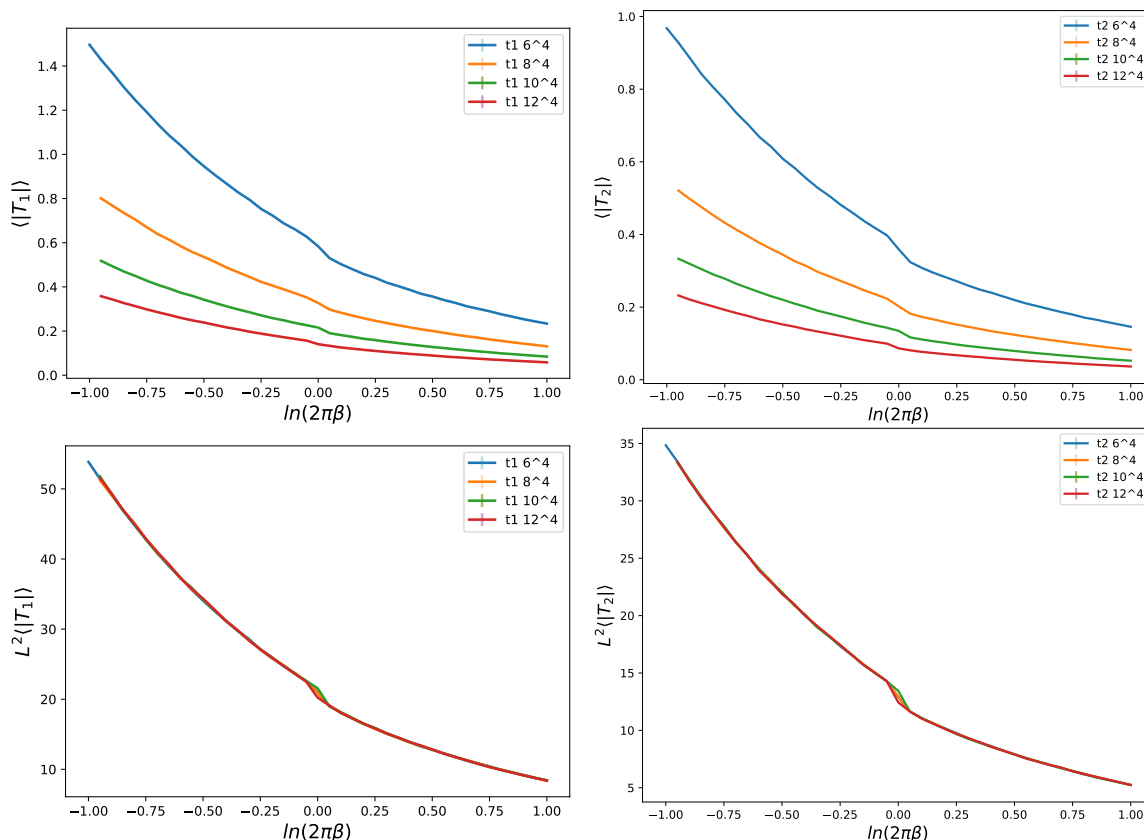


Figure 9. Results for the C -breaking order parameters T_1 and T_2 . In the top row of plots we show $\langle |T_1| \rangle$ (lhs.) and $\langle |T_2| \rangle$ (rhs.) for different volumes L^4 with $L = 8, 10, 12, 14$. The results are plotted as a function of $\ln(2\pi\beta)$ and we use a matter field coupling of $J = 0.65$. In the bottom row of plots we show the rescaled expectation values $L^2 \langle |T_1| \rangle$ (lhs.) and $L^2 \langle |T_2| \rangle$ (rhs.), illustrating that the expectation values vanish as $1/L^2 = 1/\sqrt{V}$. All error bars are smaller than the line width.

The coexistence phase then disappears in a 2nd order 4d Ising (i.e. gaussian) transition, and a trivially gapped phase ensues.

For the case of $N_f > 1$ the bosonic theory has a mixed 't Hooft anomaly between the two flavor symmetry groups $\text{PSU}(N_f) \times \text{PSU}(N_f)$ and a trivial phase is not allowed. Still, the transition between the photon phase and the self-duality broken phase is likely 1st order, because away from the self-dual point for a weak electric (magnetic) coupling this can be shown by perturbative RG flow equations. However, for $N_f > 183$ the RG equations allow for a 2nd order transition, which may persist all the way to the self-dual line, resulting in an interacting fixed point. If this is the case the fixed point has to be interacting, because the electric coupling is fixed exactly by self-duality.

We here also add a comment on fermionic self-dual theories. To begin with, let us consider QED without monopoles and with N_f massive electric flavors of mass m_e . The flavor group is $\text{PSU}(N_f)$ and if the electric coupling is sufficiently weak, the theory flows to a free photon phase. When the mass of the electric flavors is exactly zero, the theory flows to a theory of free photons and N_f fermions.

Moreover, the above conclusion must follow even when the coupling is strong, as long as N_f is sufficiently large. Namely if the coupling is strong, the RG iteration will generate a term $\propto N_f F_{\mu\nu}^2$ and render the electric coupling weak. It is then natural to conjecture that for any flavor N_f the massless theory flows to a phase non-interacting photons and N_f fermions. Note also that the massless limit has an $SU(N_f)$ axial symmetry, which has a 't Hooft anomaly. The anomaly must be saturated, and it is saturated by the free fermions. One might wonder whether $SU(N_f)$ could spontaneously break and saturate the anomaly in this way. This option is certainly not eliminated, but intuitively we expect to need confinement for spontaneous symmetry breaking, and to generate confinement we need magnetically charged dynamically matter, which is absent.

Now consider coupling N_f fermions to the magnetic gauge field, and endow the electric and magnetic fermions with masses m_e and m_m respectively. Then, a natural conjecture is that when m_e and m_m are non-zero, the theory flows to free photons. If one of the masses is zero, then by arguments above the theory flows to free massless fermions and photons (see figure 10). The question is what happens when both electric and magnetic fermions are massless. In this case, there exists a vector symmetry $PSU(N_f) \times PSU(N_f)$ and an axial vector symmetry $SU(N_f) \times SU(N_f)$. As we discussed in the main text, the vector symmetries have mixed 't Hooft anomalies. However, the axial vector symmetries have 't Hooft anomalies individually, which must both be satisfied.

The 't Hooft anomalies are typically saturated by either spontaneously breaking the symmetries, or by a CFT. The spontaneous breaking of the symmetry seems unlikely to us. The reason is that a spontaneously broken phase is typically robust against small perturbations.³⁴ Yet the massless point is surrounded by free CFTs, which suggests that the massless point is a CFT itself. Unfortunately this lattice theory cannot be simulated easily, because of the fermionic sign-problem.

However, for both bosonic and fermionic self-dual theories, one can try to bootstrap the tentative CFT, imposing the symmetries and 't Hooft anomalies that we found. The bootstrap approach may yield insights into whether such CFT phases are expected at small numbers of flavors, as well as into the properties of these CFTs. Another interesting approach is to attempt to build fermionic models which evade the sign problem and simulate them.

Acknowledgments

We thank Madalena Lemos, John McGreevy, Mithat Ünsal and Yuya Tanizaki for discussions. TS is supported by the Royal Society University Research Grant. NI is supported in part by the STFC under consolidated grant ST/L000407/1

A The RG equations

A.1 Setup

In this section, we discuss the RG flow of N_f complex scalars coupled to a $U(1)$ abelian gauge field, i.e., the theory described by the action (2.51), which we reproduce here for

³⁴Save for maybe lifting goldstones if the perturbation breaks the broken symmetry.

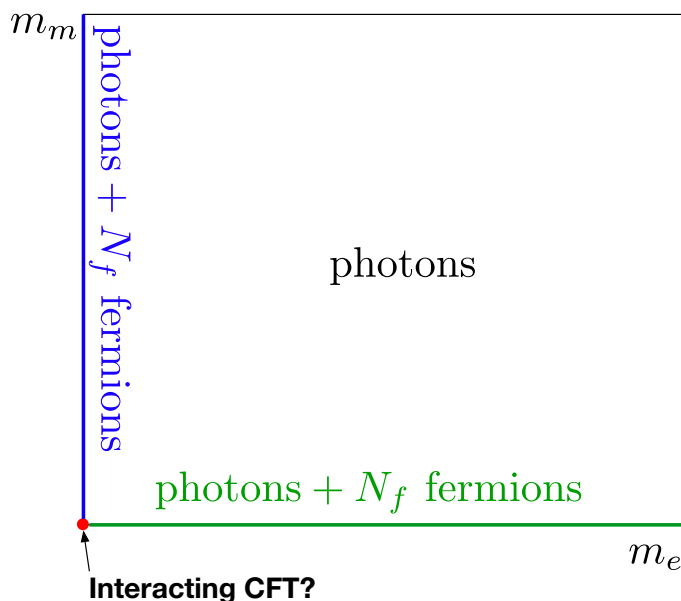


Figure 10. The conjectured phase diagram of the theory at the self-dual coupling, with N_f electric and N_f magnetic fermions. The point when fermions are exactly massless is likely to be an interacting CFT.

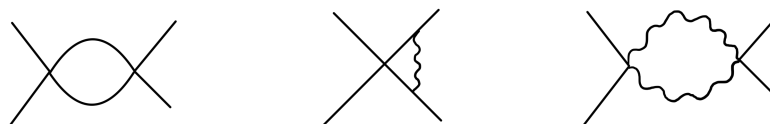


Figure 11. The topology of the diagrams contributing to the running of the ϕ^4 -coupling.

convenience:

$$S[\phi, A] = \int d^4x \left(\frac{1}{4e^2} F^2 + \sum_{i=1}^{N_f} (|\partial\phi_i|^2 + m^2|\phi_i|^2) + \frac{\lambda}{4} \left(\sum_{i=1}^{N_f} |\phi_i|^2 \right)^2 \right) \quad (\text{A.1})$$

We are interested in understanding the order of the phase transition at $m^2 \sim 0$ that separates the Higgsed and Coulomb phases. If we can generically arrive at an RG fixed point by tuning only the single parameter m^2 , then this means that the phase transition is of second order. On the other hand, if we generically don't arrive at such a fixed point, then the transition will be first order.

To that end, we tune the mass of the scalars to zero, and look at the RG flow equations for $g = e^2$ and λ . On general grounds they have the form

$$\frac{dg}{d\ell} = -ag^2, \quad (\text{A.2a})$$

$$\frac{d\lambda}{d\ell} = -b\lambda^2 - c\lambda g - dg^2, \quad (\text{A.2b})$$

where ℓ parametrizes the RG flow, i.e., if Λ is the mass-dimension UV cutoff of the theory, and we flow to $\Lambda' < \Lambda$, then $\ell = \log(\Lambda/\Lambda')$.

Let us discuss these equations a bit. The first equation is just the running of the gauge coupling which has no contributions from ϕ^4 interactions to this order. Famously $a > 0$, so that the coupling $g = e^2$ becomes more negative in the IR, i.e., QED is IR free.

The diagrams contributing to the beta-function for λ are given in the top of figure 11. The first of these diagrams is of order λ^2 , and should have $b > 0$ (i.e., λ is marginally irrelevant). The last diagram is similar, except that a photon runs in the loop. Finally the middle diagram contributes to the coefficient c in (A.2b) and turns out to be positive, such that $c < 0$, although the right-hand side is still a negative-definite quadratic form in the (λ, g) space.

Let us now discuss the space of solutions to these equations. The only fixed point is at $(\lambda, g) = (0, 0)$. Do we approach this fixed point from generic initial data $(\lambda, g) > 0$? As the right-hand sides of both equations are negative, both couplings (λ, g) become more negative in the IR. If there was only one coupling (e.g., λ), then this would imply that λ flows to zero and the theory becomes IR free. However as there are two couplings, it is quite possible for the RG flow to *miss* the origin and flow in the infrared towards increasingly negative (λ, g) (presumably corresponding to a phase where the scalars are condensed). Whether or not this generically happens requires a more careful study of the RG equations.

To do this, we will seek to construct *separatrices* in solution space, i.e., one-dimensional lines $L \subset (\lambda, g)$ running through the origin such that if the system is started with initial conditions on L , under RG flow the system remains on L for all RG time. To find L , we just simultaneously solve the following equations:

$$g = \alpha\lambda \quad \frac{dg}{d\ell} = \alpha \frac{d\lambda}{d\ell} \tag{A.3}$$

with α a constant that is the slope of the line. These have the solutions

$$\alpha = 0 \quad \text{and} \quad \alpha = \frac{1}{2d} (a - c \pm K), \quad K \equiv \sqrt{(a - c)^2 - 4db}. \tag{A.4}$$

Thus, if K is positive, we may construct real separatrices along which the RG flow definitely hits the origin. Because solutions to a first order differential equation cannot cross, this means that all points inside a funnel between two separatrices will also hit the origin. We conclude that if $K > 0$ there exists an open set of initial data from which we reach the free fixed point by tuning only a single parameter m^2 , and thus that the phase transition is second order. Otherwise, the RG flows cannot be bounded, the couplings flow to negative infinity, and we expect the transition to be first order. This is borne out by an exact solution below (see figure 12).

To proceed, we need values for (a, b, c, d) , which are given by (2.52) repeated here for convenience

$$\frac{dg}{d\ell} = -\frac{N_f g^2}{3}, \tag{2.52a}$$

$$\frac{d\lambda}{d\ell} = -(N_f + 4)\lambda^2 + 6\lambda g - 6g^2. \tag{2.52b}$$

Let us briefly discuss the terms entering the equations above. The beta function appearing on the r.h.s. of the equation for g is the vacuum polarization, which is enhanced

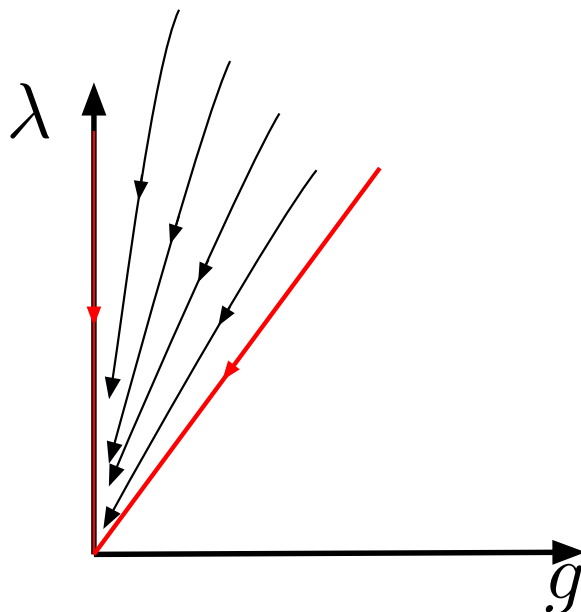


Figure 12. The RG flow is bounded by the separatrix solutions $g = \alpha\lambda$ with α given in (A.4) and drawn as red lines. The separatrices form a funnel which guaranties that the fixed point is reached as long as the flow is started within the funnel.

by N_f . The beta function of λ has three contributions given in figure 11. The latter two diagrams do not get an N_f enhancement, but the first diagram has an enhancement by $N_f + 4$ because the flavors can run in a loop.

Putting in these values we see that $K > 0$ if $N_f > 182.95$. Thus only if we have more than 183 flavors is the transition second order.

Finally, we note that in this case we can also exactly solve the RG flow equations. In the remainder of this appendix we write out the exact solution.

A.2 The RG analysis of scalar QED

Beginning with the general equations (A.2), we can define a new coupling $\bar{\lambda} = \lambda + Ag$, where A is a constant chosen such that the g^2 term on the rhs. of (A.2b) vanishes. A simple calculation yields

$$\bar{\lambda} = -b\bar{\lambda}^2 - (c - 2Ab)\bar{\lambda}g - (bA^2 + (a - c)A + d)g^2. \tag{A.5}$$

The g^2 term vanishes if we set

$$bA^2 + (a - c)A + d = 0, \tag{A.6}$$

such that

$$A_{\pm} = \frac{-(a - c) \pm \sqrt{(a - c)^2 - 4db}}{2b}. \tag{A.7}$$

The RG equations then take the form

$$\frac{d\bar{g}}{d\ell} = -\bar{a}\bar{g}^2, \quad (\text{A.8a})$$

$$\frac{d\bar{\lambda}}{d\ell} = -\bar{b}\bar{\lambda}^2 - \bar{c}\bar{\lambda}\bar{g}, \quad (\text{A.8b})$$

where

$$\bar{g}(\ell) = g(\ell), \quad \bar{\lambda}(\ell) = \lambda(\ell) + A_{\pm}g(\ell), \quad (\text{A.9})$$

$$\bar{a} = a, \quad \bar{b} = b, \quad (\text{A.10})$$

$$\bar{c} = c - 2Ab. \quad (\text{A.11})$$

The solution of (A.8) can be found explicitly (see appendix A), which we can use to write down the general solution for $\lambda(\ell)$ and $g(\ell)$.

$$g(\ell) = \frac{g_0}{1 + ag_0\ell}, \quad (\text{A.12a})$$

$$\lambda(\ell) = -\frac{(c - a - 2A_{\pm}b)g(\ell)^{\frac{c-2A_{\pm}b}{a}}}{b \left(g(\ell)^{\frac{c-a-2bA_{\pm}}{a}} - g_0^{\frac{c-a-2bA_{\pm}}{a}} \right) - (c - a - 2A_{\pm}b)\frac{g_0^{\frac{c-2A_{\pm}b}{a}}}{\lambda_0 + A_{\pm}g_0}} + A_{\pm}g(\ell). \quad (\text{A.12b})$$

The equation for $\lambda(\ell)$ can be rewritten as

$$g(\ell) = \frac{g_0}{1 + ag_0\ell}, \quad (\text{A.13a})$$

$$\lambda(\ell) = -g(\ell) \frac{(\mp K)g(\ell)^{\frac{\mp K}{a}}}{b \left(g(\ell)^{\frac{\mp K}{a}} - g_0^{\frac{\mp K}{a}} \right) \pm K \frac{g_0^{\frac{\mp K}{a}}}{\lambda_0 + A_{\pm}}}} + A_{\pm}g(\ell), \quad (\text{A.13b})$$

where as above,

$$K = \sqrt{(a - c)^2 - 4db}. \quad (\text{A.14})$$

Note that both signs used above give the same solution.

If $(a - c)^2 - 4db < 0$ then $K = i\kappa$, with $\kappa \in \mathbb{R}$, and the solution must be oscillatory in ℓ , because it depends on the combination $g(\ell)^{K/a} = e^{i\frac{\kappa}{a} \log(g(\ell))}$. On the other hand we expect $\lambda(\ell)$ to be monotonically decreasing. Indeed we can write

$$\frac{d\lambda}{d\ell} = -b \left(\lambda + \frac{c}{2b} \right)^2 + \frac{c^2 - 4bd}{4b} g^2, \quad (\text{A.15})$$

and as long as $c^2 - 4bd < 0$ holds λ must be monotone decreasing. This condition is satisfied for all N_f in equation (A.2).

Since the function $\lambda(\ell)$ has to be both oscillatory and monotone, it must be singular at finite time ℓ and λ flows to minus infinity at finite ℓ , which means that the fixed point is not reached and the transition is 1st order.

On the other hand if $(a - c)^2 - 4db > 0$, then K is real and the solution $\lambda(\ell)$ is no longer oscillatory, hence there is no obstruction to the flow reaching the fixed point

$\lambda = g = 0$. This condition is fulfilled if $N_f > 182.95$.³⁵ What needs to be checked is that the denominator appearing in the solution does not go to zero for any value of ℓ .

We take the lower sign in (A.13) and examine the possibility that the denominator is zero, i.e.

$$b \left(g(\ell)^{\frac{K}{a}} - g_0^{\frac{K}{a}} \right) - K \frac{g_0^{\frac{K}{a}}}{\frac{\lambda_0}{g_0} + A_-} = 0, \quad (\text{A.16})$$

such that

$$g(\ell)^{\frac{K}{a}} = g_0^{\frac{K}{a}} \left(1 + \frac{K}{b} \frac{1}{\frac{\lambda_0}{g_0} + A_-} \right). \quad (\text{A.17})$$

Now since the left hand side is bounded from above by $g_0^{\frac{K}{a}}$, and from below by zero, the condition that there is no pole is

$$\frac{K}{b} \frac{1}{\frac{\lambda_0}{g_0} + A_-} > 0 \quad \text{or} \quad \frac{K}{b} \frac{1}{\frac{\lambda_0}{g_0} + A_-} < -1. \quad (\text{A.18})$$

For the first condition to be satisfied we must have

$$\frac{\lambda_0}{g_0} > -A_-. \quad (\text{A.19})$$

The second condition in (A.18) can be satisfied only if the first is not, i.e., for $\frac{\lambda_0}{g_0} + A_- < 0$. Upon multiplication of the second equation in (A.18) by $\frac{\lambda_0}{g_0} + A_-$ we would find the condition

$$-A_- > \frac{\lambda_0}{g_0} > -A_- - \frac{K}{b}. \quad (\text{A.20})$$

However, since $K/b > 0$, the above equation is inconsistent because $-A_- < -a_- - \frac{K}{b}$. Hence only the first condition makes sense and we find that the critical ratio of the bare couplings $\frac{\lambda_0}{g_0}$ is given by

$$r_c = -A_- = \frac{N_f + \sqrt{(N_f - 180)N_f - 540 + 18}}{6(N_f + 4)}, \quad (\text{A.21})$$

and for $\lambda_0/g_0 > r_c$ the transition is 2nd order. This is depicted in figure 13.

A.2.1 Solving the equations (A.8)

We wish to solve the following RG equations

$$\frac{dg_1}{d\ell} = -ag_1^2, \quad (\text{A.22})$$

$$\frac{dg_2}{d\ell} = -bg_1^2 - cg_1g_2. \quad (\text{A.23})$$

³⁵Note that this is precisely the condition for the existence of an interacting Wilson-Fischer fixed point in the ϵ -expansion [37].

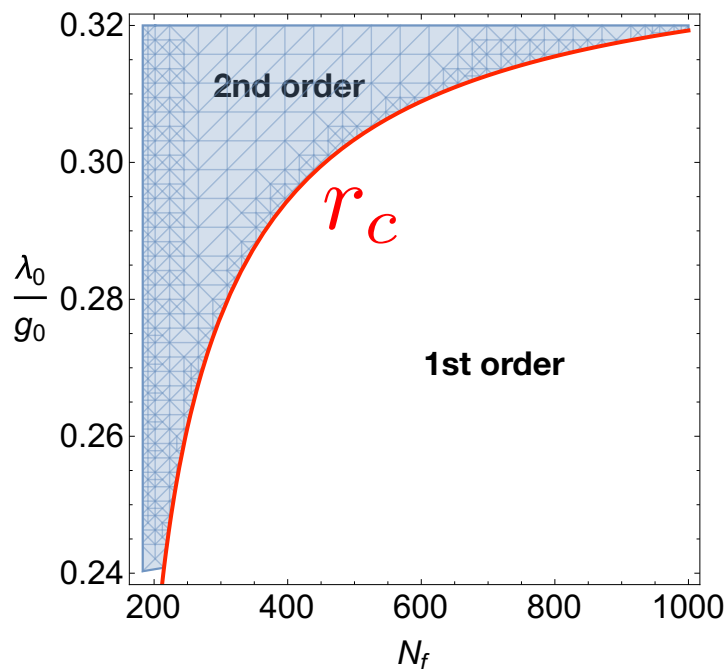


Figure 13. The plot of the critical ratio r_0 as a function of N_f defined in (A.21). For the ratio of the bare couplings λ_0/g_0 in the shaded region the transition is 2nd order.

Obviously $g_1(\ell)$ can be determined easily as

$$g_1(\ell) = \frac{g_1(0)}{1 + ag_1(0)\ell} . \tag{A.24}$$

Since $g_1(\ell)$ is monotonic, we can view g_2 as depending on ℓ through g_1 , i.e.,

$$\frac{dg_2}{d\ell} = \frac{dg_1}{d\ell} \frac{dg_2}{dg_1} = -ag_1^2 \frac{dg_2}{dg_1} , \tag{A.25}$$

where we used the RG equation for g_1 . Now we see that the RG equation for g_2 becomes

$$-ag_1^2 \frac{dg_2}{dg_1} = -bg_2^2 - cg_1g_2 , \tag{A.26}$$

or after rewriting

$$\left(-ag_1^2 \frac{d}{dg_1} + cg_1 \right) g_2 = -bg_2^2 . \tag{A.27}$$

Now, replacing

$$g_2(g_1) = g_1^{\frac{c}{a}} G(g_1) , \tag{A.28}$$

we find that $G(g_1)$ satisfies the equation

$$ag_1^2 G'(g_1) = bg_1^{\frac{c}{a}} G(g_1)^2 , \tag{A.29}$$

which is solved by

$$G(g_1) = -\frac{c-a}{bg_1^{\frac{c}{a}-1} + C} , \tag{A.30}$$

where C is a constant. Hence $g_2(\ell)$ is given by

$$g_2(\ell) = -(c - a) \frac{g_1(\ell)^{\frac{c}{a}}}{bg_1(\ell)^{\frac{c}{a}-1} + C}, \quad (\text{A.31})$$

where $g_1(\ell)$ is given by (A.24). By setting $\ell = 0$ we find C in terms of $g_1(0), g_2(0)$,

$$C = -(c - a) \frac{g_1(0)^{\frac{c}{a}}}{g_2(0)} - bg_1(0)^{\frac{c}{a}-1}. \quad (\text{A.32})$$

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] F.D.M. Haldane, *Continuum dynamics of the 1D Heisenberg antiferromagnetic identification with the $O(3)$ nonlinear sigma model*, *Phys. Lett. A* **93** (1983) 464 [[INSPIRE](#)].
- [2] A.M. Polyakov, *Quark confinement and topology of gauge groups*, *Nucl. Phys. B* **120** (1977) 429 [[INSPIRE](#)].
- [3] T. Senthil, A. Vishwanath, L. Balents, S. Sachdev and M.P.A. Fisher, *Deconfined quantum critical points*, *Science* **303** (2004) 1490 [[cond-mat/0311326](#)] [[INSPIRE](#)].
- [4] A. Vishwanath, L. Balents and T. Senthil, *Quantum criticality and deconfinement in phase transitions between valence bond solids*, *Phys. Rev. B* **69** (2004) 224416 [[cond-mat/0311085](#)] [[INSPIRE](#)].
- [5] S.S. Pufu, *Anomalous dimensions of monopole operators in three-dimensional quantum electrodynamics*, *Phys. Rev. D* **89** (2014) 065016 [[arXiv:1303.6125](#)] [[INSPIRE](#)].
- [6] E. Dyer, M. Mezei, S.S. Pufu and S. Sachdev, *Scaling dimensions of monopole operators in the CP^{N_b-1} theory in $2+1$ dimensions*, *JHEP* **06** (2015) 037 [*Erratum ibid.* **03** (2016) 111] [[arXiv:1504.00368](#)] [[INSPIRE](#)].
- [7] H. Shao, W. Guo and A.W. Sandvik, *Quantum criticality with two length scales*, *Science* **352** (2016) 213.
- [8] C. Montonen and D.I. Olive, *Magnetic monopoles as gauge particles?*, *Phys. Lett. B* **72** (1977) 117 [[INSPIRE](#)].
- [9] T. Sulejmanpasic and C. Gattringer, *Abelian gauge theories on the lattice: θ -terms and compact gauge theory with(out) monopoles*, *Nucl. Phys. B* **943** (2019) 114616 [[arXiv:1901.02637](#)] [[INSPIRE](#)].
- [10] C. Gattringer, D. Göschl and T. Sulejmanpasic, *Dual simulation of the 2d $U(1)$ gauge Higgs model at topological angle $\theta = \pi$: critical endpoint behavior*, *Nucl. Phys. B* **935** (2018) 344 [[arXiv:1807.07793](#)] [[INSPIRE](#)].
- [11] D. Göschl, C. Gattringer and T. Sulejmanpasic, *The critical endpoint in the 2D $U(1)$ gauge-Higgs model at topological angle $\theta = \pi$* , *PoS LATTICE2018* (2018) 226 [[arXiv:1810.09671](#)] [[INSPIRE](#)].
- [12] T. Sulejmanpasic, D. Göschl and C. Gattringer, *First-principles simulations of $1+1D$ quantum field theories at $\theta = \pi$ and spin chains*, *Phys. Rev. Lett.* **125** (2020) 201602 [[arXiv:2007.06323](#)] [[INSPIRE](#)].

- [13] T. Sulejmanpasic, *Ising model as a U(1) lattice gauge theory with a θ -term*, *Phys. Rev. D* **103** (2021) 034512 [[arXiv:2009.13383](#)] [[INSPIRE](#)].
- [14] M. Anosova, C. Gattringer, N. Iqbal and T. Sulejmanpasic, *Numerical simulation of self-dual U(1) lattice field theory with electric and magnetic matter*, in *38th International symposium on lattice field theory*, (2021) [[arXiv:2111.02033](#)] [[INSPIRE](#)].
- [15] M. Anosova, C. Gattringer and T. Sulejmanpasic, *Self-dual U(1) lattice field theory with a θ -term*, *JHEP* **04** (2022) 120 [[arXiv:2201.09468](#)] [[INSPIRE](#)].
- [16] P. Gorantla, H.T. Lam, N. Seiberg and S.-H. Shao, *A modified Villain formulation of fractons and other exotic theories*, *J. Math. Phys.* **62** (2021) 102301 [[arXiv:2103.01257](#)] [[INSPIRE](#)].
- [17] Y. Choi, C. Cordova, P.-S. Hsin, H.T. Lam and S.-H. Shao, *Non-invertible duality defects in 3 + 1 dimensions*, [arXiv:2111.01139](#) [[INSPIRE](#)].
- [18] J. Villain, *Theory of one- and two-dimensional magnets with an easy magnetization plane. II. The planar, classical, two-dimensional magnet*, *J. Phys. (France)* **36** (1975) 581.
- [19] S. Elitzur, R.B. Pearson and J. Shigemitsu, *The phase structure of discrete Abelian spin and gauge systems*, *Phys. Rev. D* **19** (1979) 3698 [[INSPIRE](#)].
- [20] J.L. Cardy, *Duality and the theta parameter in Abelian lattice models*, *Nucl. Phys. B* **205** (1982) 17 [[INSPIRE](#)].
- [21] J.L. Cardy and E. Rabinovici, *Phase structure of Z_p models in the presence of a θ parameter*, *Nucl. Phys. B* **205** (1982) 1 [[INSPIRE](#)].
- [22] A. Kapustin and N. Seiberg, *Coupling a QFT to a TQFT and duality*, *JHEP* **04** (2014) 001 [[arXiv:1401.0740](#)] [[INSPIRE](#)].
- [23] J. Kaidi, K. Ohmori and Y. Zheng, *Kramers-Wannier-like duality defects in (3 + 1)D gauge theories*, *Phys. Rev. Lett.* **128** (2022) 111601 [[arXiv:2111.01141](#)] [[INSPIRE](#)].
- [24] S.M. Kravec and J. McGreevy, *A gauge theory generalization of the fermion-doubling theorem*, *Phys. Rev. Lett.* **111** (2013) 161603 [[arXiv:1306.3992](#)] [[INSPIRE](#)].
- [25] D. Gaiotto, A. Kapustin, N. Seiberg and B. Willett, *Generalized global symmetries*, *JHEP* **02** (2015) 172 [[arXiv:1412.5148](#)] [[INSPIRE](#)].
- [26] E.H. Fradkin and S.H. Shenker, *Phase diagrams of lattice gauge theories with Higgs fields*, *Phys. Rev. D* **19** (1979) 3682 [[INSPIRE](#)].
- [27] S.R. Coleman and E.J. Weinberg, *Radiative corrections as the origin of spontaneous symmetry breaking*, *Phys. Rev. D* **7** (1973) 1888 [[INSPIRE](#)].
- [28] S. Kolnberger and R. Folk, *Critical fluctuations in superconductors*, *Phys. Rev. B* **41** (1990) 4083.
- [29] R. Folk and Y. Holovatch, *Critical fluctuations in normal to superconducting transition*, in *1st Winter workshop on cooperative phenomena in condensed matter*, (1998) [[cond-mat/9807421](#)] [[INSPIRE](#)].
- [30] B. Ihrig, N. Zerf, P. Marquard, I.F. Herbut and M.M. Scherer, *Abelian Higgs model at four loops, fixed-point collision and deconfined criticality*, *Phys. Rev. B* **100** (2019) 134507 [[arXiv:1907.08140](#)] [[INSPIRE](#)].
- [31] D.M. Hofman and N. Iqbal, *Goldstone modes and photonization for higher form symmetries*, *SciPost Phys.* **6** (2019) 006 [[arXiv:1802.09512](#)] [[INSPIRE](#)].

- [32] W. Kerler, C. Rebbi and A. Weber, *Critical properties and monopoles in U(1) lattice gauge theory*, *Phys. Lett. B* **392** (1997) 438 [[hep-lat/9612001](#)] [[INSPIRE](#)].
- [33] G. Damm and W. Kerler, *Critical exponents in U(1) lattice gauge theory with a monopole term*, *Nucl. Phys. B Proc. Suppl.* **63** (1998) 703 [[hep-lat/9709061](#)] [[INSPIRE](#)].
- [34] A.M. Somoza, P. Serna and A. Nahum, *Self-dual criticality in three-dimensional Z_2 gauge theory with matter*, *Phys. Rev. X* **11** (2021) 041008 [[arXiv:2012.15845](#)] [[INSPIRE](#)].
- [35] Y. Delgado Mercado, C. Gattringer and A. Schmidt, *Surface worm algorithm for Abelian gauge-Higgs systems on the lattice*, *Comput. Phys. Commun.* **184** (2013) 1535 [[arXiv:1211.3436](#)] [[INSPIRE](#)].
- [36] Y. Delgado Mercado, C. Gattringer and A. Schmidt, *Dual lattice simulation of the Abelian gauge-Higgs model at finite density: an exploratory proof of concept study*, *Phys. Rev. Lett.* **111** (2013) 141601 [[arXiv:1307.6120](#)] [[INSPIRE](#)].
- [37] B.I. Halperin, T.C. Lubensky and S.-K. Ma, *First order phase transitions in superconductors and smectic-A liquid crystals*, *Phys. Rev. Lett.* **32** (1974) 292 [[INSPIRE](#)].