Mass generation by fractional instantons in SU(n) chains

Kyle Wamer and Ian Affleck

Department of Physics and Astronomy and Stewart Blusson Quantum Matter Institute,
University of British Columbia, Vancouver, B.C., Canada, V6T1Z1

(Dated: January 27, 2020)

Recently, Haldane’s conjecture about SU(2) chains has been generalized to SU(n) chains in the symmetric representations. For a rank-p representation, a gapless phase is predicted when p and n are coprime; otherwise, a finite energy gap is present above the ground state. In this work, we offer an intuitive explanation of this behavior based on fractional topological excitations, which are able to generate a mass gap except when p and n have no common divisor. This is a generalization of an older work in SU(2), which explains the generation of the Haldane gap in terms of merons in the O(3) nonlinear sigma model.

Haldane’s conjecture is the prediction that antiferromagnetic spin chains with integer spin have a gap above the ground state, while those with half-odd integer spin are gapless [1,2]. The distinction between these two cases can be seen by taking a large spin limit, in which case the quantum fluctuations of the antiferromagnet are governed by the O(3) nonlinear sigma model, with topological angle of \( \theta = 2\pi \alpha \):

\[
\mathcal{L} = \frac{1}{2g} (\partial_\mu \vec{\phi})^2 + \mathcal{L}_{\text{top}} \quad \int d^2 x \mathcal{L}_{\text{top}} \in \theta \mathbb{Z}. \tag{1}
\]

This sigma model, in the absence of a topological term, is known to have a finite energy gap above the ground state [3]. In [4], it was argued that this energy gap can be thought of as being generated by half-quantized topological excitations called merons. Merons behave like planar vortices far from their centers, but lift out of the plane near their cores, and form finite action configurations when a collection of them has a net vorticity of zero.

Remarkably, when the O(3) sigma model has a topological angle of \( \pi \) (corresponding to a half-odd-integer spin chain), the merons that point up at their origin destructively interfere with the merons that point down at their origin (also known as antimerons), and the mass generating mechanism of the model is no longer effective. While this isn’t the full story, since larger-action configurations such as double-merons do not destructively interfere at \( \theta = \pi \), it offers an intuitive picture in favor of Haldane’s conjecture.

Recently, this paradigm of mapping spin chains to relativistic quantum field theories has been generalized to SU(n) chains in the rank-p symmetric representations [5,6]. In this case, the O(3) model is replaced with a SU(n)/[U(1)]^{n-1} flag manifold sigma model, which is described by \( n \) orthonormalized fields \( \vec{\phi}_\alpha \in \mathbb{C}^n \), and has \( n-1 \) independent topological angles, \( \theta_\alpha = 2\pi \alpha \). At low energies, this sigma model acquires Lorentz invariance, and has the form \( \mathcal{L} = \sum_{\alpha < \beta} \mathcal{L}_{\alpha\beta} + \mathcal{L}_{\text{top}} \), where

\[
\mathcal{L}_{\alpha\beta} = \frac{|\vec{\phi}_\alpha \cdot \partial_\mu \vec{\phi}_\beta|^2}{g_{\alpha,\beta}} + \epsilon_{\mu\nu\lambda_{\alpha,\beta}}(\vec{\phi}_\alpha \cdot \partial_\nu \vec{\phi}_\beta)(\vec{\phi}_\mu \vec{\phi}_\beta \cdot \vec{\phi}_\alpha) \tag{2}
\]

and \( \mathcal{L}_{\text{top}} = \sum_\alpha \theta_\alpha q_\alpha \), with

\[
q_\alpha = \frac{1}{2\pi i} \epsilon_{\mu\nu\lambda} \partial_\mu \vec{\phi}_\alpha \cdot \partial_\nu [\vec{\phi}_\alpha]^\lambda \quad \int d^2 x q_\alpha \in \mathbb{Z}. \tag{3}
\]

The coupling constants \( g_{\alpha,\beta} \) and \( \lambda_{\alpha,\beta} \) are functions of the antiferromagnetic interaction strength at the bare level. This model possesses a [U(1)]^{n-1} gauge symmetry, corresponding to a local phase rotation for each of \( \vec{\phi}_\alpha \) (the \( n \)th field, \( \vec{\phi}_n \), can be written in terms of the first \( n-1 \) fields, and does not admit its own, independent gauge symmetry). According to [8], these sigma models have a gapless ground state for all \( p \), except in the special case when \( p \) and \( n \) are a common divisor greater than 1, in which case a gapped phase is possible. This is in accordance with the LSMA theorem, which predicts either a gapless phase or spontaneously broken translation symmetry when \( p \) is not a multiple of \( n \) [2,10]. A subtle feature of this result, which can only occur for \( n \geq 4 \), is that when \( p \) and \( n \) have a nontrivial divisor, and \( p \) is not a multiple of \( n \), translation symmetry is always spontaneously broken. This follows from an ‘t Hooft anomaly that forbids the flow of the above sigma model to SU(n)_1, the only stable SU(n) fixed point [11,12].

In this Letter, we offer an intuitive explanation of this gap-dependent behavior on \( p \) and \( n \), based on a mass generation mechanism due to fractionalized topological excitations. It is the SU(n) version of the meron story in the O(3) sigma model: for \( p \) a multiple of \( n \), we will argue that fractional instantons with charge \( \frac{1}{n} \) proliferate, and generate a mass gap in a similar fashion as the vortices in the Kosterlitz-Thouless transition [14]. Moreover, we will argue that when \( p \) and \( n \) have a common divisor less than \( n \), topological objects with fractional charge \( \frac{1}{\text{gcd}(p,n)} \) will be responsible for generating a mass gap. Only when \( p \) and \( n \) are coprime is a gapless phase possible.

Following [4], our strategy is to break the symmetry of the flag manifold down to U(1), where a phase transition is well understood in terms of vortex proliferation. We first add to [4] an anisotropic potential \( V_1 \) that breaks the SU(n) symmetry down to [U(1)]^{n-1}, while preserving the
discrete $\mathbb{Z}_n$ symmetry and the $[U(1)]^{n-1}$ gauge structure:

$$V_1 = m \sum_{\alpha=1}^{n} \sum_{\beta < \gamma} \left( |\phi^\beta_\alpha|^2 - |\phi^\gamma_\alpha|^2 \right).$$

(4)

In the limit $m \to \infty$, this restricts each of the $\vec{\phi}_\alpha$ to lie on an $n-1$ torus, parameterized by $n-1$ angles, $\sigma_\beta = \sigma_\beta(\alpha)$:

$$\vec{\phi}_\alpha = \frac{1}{\sqrt{n}} \left( e^{i\sigma_1} e^{i\sigma_2} \ldots e^{i\sigma_{n-1}} e^{i\sigma_n} \right)^T.$$

(5)

Throughout, we fix a gauge by choosing $\phi^1_\alpha$ to be real and positive, for all $\alpha$. Since the number of free parameters equals the number of orthogonality constraints (both equal $n(n-1)/2$), one might conclude that the there is a unique configuration of the $\{\vec{\phi}_\alpha\}$ in this limit (up to permutations). However, the orthogonality constraints are invariant under the following transformations,

$$\phi^\beta_\alpha \to e^{i\theta_\beta} \phi^\beta_\alpha \quad \forall \alpha, \beta,$$

(6)

which are true symmetries for each of the $n-1$ un-gauge-fixed directions, $\beta = 2, 3, \ldots, n$. Therefore, by introducing the potential $\mathcal{V}$, the continuous symmetry group of the sigma model has been broken down to $U(1)^{n-1}$. A typical configuration of the $\{\vec{\phi}_\alpha\}$ in this submanifold can be constructed from the $n$th root of unity, $\zeta := e^{i\pi/n}$:

$$\phi^\beta_\alpha = \frac{1}{\sqrt{n}} \zeta^{\rho_\alpha} e^{i\sigma_\beta} \quad \sigma_\beta \in [0, 2\pi].$$

(7)

Indeed, orthonormality of these states follows from the identity $\sum_{\alpha=1}^{n} \zeta^{\rho_\alpha} = 0$, which holds for all $\rho \neq 0 \mod n$. Since the Lagrangian $\mathcal{L}$ couples all of the $n-1$ copies of $U(1)$ together, it will be useful to break the symmetry down further. To this end, we introduce a second potential $V_2$, given by

$$V_2 = -m \sum_{\alpha=1}^{n} \sum_{\beta=2}^{n-1} \left( \phi^1_\alpha [\phi^{\beta}_\alpha]^* - (\phi^1_\alpha)^* \phi^{\beta}_\alpha \right)^2.$$ 

(8)

The factors of $n$ in the exponent are necessary in order to preserve the $\mathbb{Z}_n$ symmetry, which corresponds to translational invariance in the underlying lattice model. Setting $V_2$ is still minimized by the configurations $\mathcal{C}$, it can be rewritten in terms of the $U(1)$ fields, $\sigma_\beta$, yielding

$$V_2 = 4m \sum_{\alpha=1}^{n} \sum_{\beta=2}^{n-1} \sin^2(n(\sigma_1 - \sigma_\beta)).$$

(9)

It is clear that the effect of $V_2$ is to equate all but one of the $U(1)$ fields. In our fixed gauge, this amounts to setting $\sigma_\beta = 0$ for $\beta \neq n$, resulting in a theory that involves $\sigma_\alpha$ only. This is equivalent to the $O(2)$ model of a vector $\vec{n} \in \mathbb{R}^3$ restricted to the XY plane. By inserting this restricted form of $\vec{\phi}_\alpha$ into $\mathcal{V}$, it is easy to show that the $\lambda_{\alpha,\beta}$ terms vanish, and the resulting $O(2)$ coupling constant is $g$, defined by $g^{-1} = \sum_{\alpha,\beta} g_{\alpha,\beta}$.

In the absence of the $\theta_\alpha$, it is well known that a mass gap is generated in the $O(2)$ model, due to vortex proliferation. Moreover, if we view the $O(2)$ model as a special limit of the $O(3)$ sigma model with a large potential added to restrict the field $\vec{n}$ to the XY plane, this mass gap is still generated when the potential is weakened, and $\vec{n}$ can lift off the plane. Now, the gap is generated by nonplanar vortices, known as merons. This argument was used to identify merons as the mass-gap generating mechanism in the $O(3)$ sigma model which corresponds to the purely isotropic case of $m = 0$. Here, we would like to follow this same logic and argue that a mass gap is present in the $SU(n)$ flag manifold sigma model (without topological term), and that it is generated by topological excitations.

We begin by considering the case of $n = 2$, where $\mathcal{V}$ vanishes and the perturbation $\mathcal{V}$ is equivalent to adding a mass term $m(n_3)^2$ to the $O(3)$ model, restricting $\vec{n}$ to lie in the XY plane in the large $m$ limit. This follows from the equivalence equation $n^\circ = \vec{\phi} \sigma^i \vec{\phi}$, while the lattice version of this theory experiences a phase transition due to vortex proliferation, in the continuum limit, the vortices must become non-planar at their core in order to avoid a UV singularity. It is these nonplanar vortices that are the merons, since they have topological charge

$$Q = \frac{\pm 1}{2} \quad \text{(sign depending on whether } n_3 \neq 0 \text{ at their core)},$$

as compared to $Q = \pm 1$ for the more familiar instantons and antiinstantons of the $O(3)$ model.

To demonstrate the fractional nature of the meron charge, we return to the $\phi$ notation, since this will immediately generalize to the case of $SU(n)$. While there are two fields $\phi^1$, $\phi^2$, their mutual orthogonality allows us to restrict our attention to a single field, which we refer to as simply $\vec{\phi}$. The $U(1)$ degree of freedom corresponds to the phase difference between the two components of $\vec{\phi}$. The meron configuration for $\phi$ behaves like a traditional $U(1)$ vortex far from its core, and has the form

$$\vec{\phi}_1 = \frac{1}{\sqrt{2}} \left( \begin{array}{c} 1 \\ e^{i\omega} \end{array} \right),$$

(10)

where $\omega$ is the polar angle in the XY plane. However, at the object’s centre, this dependence on $\omega$ leads to a UV singularity, which can only be avoided by sending one of the two components of $\vec{\phi}$ to zero (which corresponds to $n_3 \to \pm 1$ in the $O(3)$ model):

$$\vec{\phi} \to \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \quad \text{or} \quad \vec{\phi} \to \left( \begin{array}{c} 0 \\ e^{i\omega} \end{array} \right).$$

(11)

The phase remains in the second case due to our gauge choice; however, at the meron’s core, it can be removed by a pure gauge transformation, and thus does not lead to a true UV singularity. Since the topological charge, $Q$, of $\vec{\phi}$ is the integral of a total derivative, we may rewrite $Q$ as the difference
of two contour integrals, one around the origin and one around infinity:
\[
Q = \frac{1}{2\pi i} \left[ \oint_0 dx^\mu \tilde{\phi} \cdot \partial_\mu \tilde{\phi}^* - \oint_\infty dx^\mu \tilde{\phi} \cdot \partial_\mu \tilde{\phi}^* \right].
\] (12)

The contour is necessary to avoid any gauge singularity that may be present at the origin (see Supplemental Material). For both possibilities in (11), the contour about infinity contributes a charge of \( \frac{1}{2} \). Meanwhile, the contribution from the contour about the origin is trivial in the first possibility, and -1 in the second, leading to \( Q = \pm \frac{1}{2} \), which proves the half-quantized charge of the meron.

Extending this calculation to SU\((n)\) is straightforward. Now, the topological charge is a vector \( \vec{Q} = (Q_1, Q_2, \cdots, Q_n) \), corresponding to the \( n \) fields \( \vec{\phi}_\alpha \). A configuration far from its core has \( \vec{\phi}_\alpha = \frac{1}{\sqrt{n}} e^{i\omega} \) for all \( \alpha \), and receives a contribution to the topological charge of \( \frac{1}{n} \) from the contour at infinity. Meanwhile, in order to be UV finite at its core, the phase \( e^{i\omega} \) in each of the \( n \)th components of the \( \vec{\phi}_\alpha \) must either vanish or become a pure gauge transformation. This leads to an \( n \)-fold family of topological excitations, distinguished by which one of the \( \vec{\phi}_\alpha \) tends to \((0, 0, 0, \cdots, e^{i\omega})^T \). For that particular \( \vec{\phi} \), a contribution of -1 is added to its topological charge from the contour about the origin in (12), while the remaining \( n - 1 \) fields receive no contribution, as they are topologically trivially there. See the Supplemental Material for an explicit solution in the case of \( n = 3 \). Thus, we obtain \( n \) species of topological configurations, with charges
\[
\vec{Q}_\beta = \frac{1}{n} (1, 1, \cdots, \underbrace{-(n - 1), \cdots, 1}_{\text{position } \beta}).
\] (13)

And so for all \( n \), we find fractionalized topological excitations, or “fractional instantons”, in our symmetry broken model. And while the number of such configurations has increased, the original argument from SU(2) for mass generation carries over to this more general case: for each species of topological excitation in this \( n \)-fold family, we have a species of particle in the Coulomb gas formalism [4]. That is, each particle has a partition function that is represented (at large distances) by a sine-Gordon (sG) model,
\[
\mathcal{L}_{SG} = \frac{1}{2} (\partial_\mu \sigma)^2 + \gamma \cos\left( \frac{2\pi}{g} \sigma \right),
\] (14)
in the limit of large \( \gamma \), which represents the fugacity, or density, of the fractional instanton gas. This expression is derived in detail in [10]. Here, \( g \) is the O(2) coupling constant found above, and plays the role of temperature in the sG model. Since all of the \( n \) species arise from the same action, each will have the same fugacity and critical \( g \), so that the above model (14) is merely copied \( n \) times, and the SU(2) analysis from [4] can be applied directly:

For large \( m \), the fractional instantons are dilute and we are in a massless boson phase. As \( m \) is lowered, the effective critical temperature is increased until the topological excitations condense and a mass gap is produced. Thus, we conclude that fractional instantons are responsible for generating a mass gap in the flag manifold sigma model [2], in the absence of topological angles.

We now restore the topological angles \( \theta_\alpha \), and study how the mass generating mechanism changes. For large \( m \), we are in the O(2) model and the \( \theta \)-terms do not play a role. However, as \( m \) is lowered towards zero, the fugacity \( \gamma \) in the Sine-Gordon model is modified to
\[
\gamma \sum_{\beta=1}^n e^{i\theta} Q_\beta,
\] (15)
where the sum is over the \( n \) species of fractional instanton, and \( \theta_\alpha = \alpha \theta \), with \( \theta = \frac{2\pi}{n} \). Using (13), one easily finds that this sum equals \( \gamma \sum_k \xi^{pk} \). So long as \( p \) is not a multiple of \( n \), this sum vanishes. In other words, for such values of \( p \), the fugacity of the Sine-Gordon model vanishes, suggesting that the Coulomb gas is always in its massless phase. At first glance, this appears to be inconsistent with the conjecture put forward in [8], which also predicts a gap when \( p \) and \( \theta \) are coprime, mass-generation terms do not having cancelling contributions. However, the conclusion arrived at in that paper also applies here: since these higher vorticity objects have larger action, we can at least conclude that the critical value of \( m \), above which occurs a gapless phase, is lower at \( \theta = \frac{2\pi}{n} \) than at \( \theta = 0 \).

Now, if gcd\((p,n)\) \( \neq 1 \), then not only do events with winding number \( n \) contribute to the fugacity, but so do events with winding number \( d := n / \text{gcd}(p,n) \). Since these contributions will have a much smaller action than the \( n \)-winding events, this shows that whenever \( p \) and \( n \) have a nontrivial common divisor, the critical value \( m \) is larger than at \( \theta = \frac{2\pi}{n} \) (although still lower than at \( \theta = 0 \)). Thus, it is possible to interpret the generalized Haldane conjecture of [8] in terms of these fractional topological excitations: When \( p \) and \( n \) are coprime, mass-generation only starts to occur for configurations that have winding number \( \pm n \). In SU(2), these events are not strong
enough to open a gap at the isotropic point $m = 0$, and we predict that this holds for general $n$. In other words, we are claiming that the critical value of $m$ is zero when $p$ and $n$ are coprime. When $p$ and $n$ have a nontrivial common divisor different from $n$, configurations that have much less action begin to contribute to mass generation, starting with objects that have winding number $d$. The simplest example of this is in SU(4), with $p = 2$.

In this case, configurations that have winding number $\pm 2$ successfully generate a mass gap, while for $p = 1$ these configurations destructively interfere, and a mass gap is not generated by the subleading configurations, with winding $\pm 4$. Finally, when $p$ is a multiple of $n$, the least action configurations that have winding $\pm 1$ produce a mass gap, just like the merons in SU(2).

We would like to thank Pedro Lopes for helpful discussion. This work was supported by NSERC through Discovery Grant 04033-2016 and an NSERC PGS-D Scholarship, as well as by the QuEST Program of SBQMI.

[1] F. D. M. Haldane, Phys. Rev. Lett. 50, 1153 (1983)

SUPPLEMENTAL MATERIAL

Gauge Invariance of the Topological Charge

In (12), we rewrite the topological charge $Q$ as a difference of two contour integrals. It is interesting to note that while their sum is necessarily invariant under the $[U(1)]^{n-1}$ gauge symmetries (6), an individual contour integral is not. Indeed, under $\tilde{\phi} \to e^{i\theta} \tilde{\phi}$,

$$\oint dx^\mu \tilde{\phi} \cdot \partial_\mu \tilde{\phi}^* \mapsto \oint dx^\mu \tilde{\phi} \cdot \partial_\mu \tilde{\phi}^* + i \oint dx^\mu \partial_\mu \theta.$$  \hfill (17)

The second term, which breaks gauge invariance of the contour, is necessary in order to account for gauge transformations that alter the U(1) winding number of $\tilde{\phi}$ along the contour. This feature implies that it is not meaningful to ask where the topological charge of a configuration $\tilde{\phi}$ is localized. For instance, two gauge-equivalent versions of the SU(2) meron have the following behaviors at infinity:

$$\vec{\phi} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ e^{i\omega} \end{pmatrix} \quad \vec{\phi}' = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-i\omega} \\ 1 \end{pmatrix}.$$  \hfill (18)

Their behavior at zero is given by the first possibility in (11). The contour around infinity in (12) equals $-\frac{1}{2}$ for the case of $\vec{\phi}$, while it equals $+\frac{1}{2}$ for the case of $\vec{\phi}'$. Meanwhile, the contour around zero is trivial in the case of $\vec{\phi}$, and equals $+1$ for the case of $\vec{\phi}'$. The full expression in (12) yields $Q = \frac{1}{2}$ in both cases, as it must. Finally, we note that by choosing the gauge $\vec{\phi}' = (e^{-i\omega/2}, e^{i\omega/2})^T / \sqrt{2}$, the contour around infinity vanishes, which is what one finds in the O(3) version of the meron [4].

An Explicit Fractional Instanton in SU(3)

Here, we prove the existence of a fractional instanton in SU(3), with topological charge $\vec{Q} = (-\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$. While the most general complex unit vector in $\mathbb{C}^3$ depends on five real parameters, we make a symmetric ansatz that reduces
this number down to two – the minimum number of parameters necessary to construct a fractional instanton. For our three orthogonal vectors $\phi_\alpha$, we take:

$$
\phi_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \sin \alpha \\ \sin \alpha \cos \alpha e^{i\omega} \end{pmatrix}, \quad \phi_2 = \frac{1}{2} \begin{pmatrix} \cos \alpha + i \\ \cos \alpha - i \end{pmatrix}, \quad \phi_3 = \frac{1}{2} \begin{pmatrix} \cos \alpha - i \\ \cos \alpha + i \end{pmatrix}.
$$

(19)

Here $\omega$ is the polar angle coordinate, and $\alpha \in [0, \pi/2]$ depends only on the radial coordinate: $\alpha = \alpha(r)$. Inserting this ansatz into the flag manifold sigma model Lagrangian (2), one obtains

$$
\mathcal{L} = \frac{1}{g} \left[ (\partial_\alpha \phi)^2 + \frac{1}{r^2} U_1(\alpha) + mgU_2(\alpha) \right],
$$

(20)

where

$$
U_1(\alpha) = \sin^2 \alpha \cos^2 \alpha + \frac{1}{4} \sin^4 \alpha,
$$

(21)

and

$$
U_2(\alpha) = \frac{5}{2} \left[ \cos^2 \alpha - \frac{1}{2} \sin^2 \alpha \right]^2 + \frac{1}{128} \cos^2 \alpha (3 \sin^4 \alpha - 4 \cos^2 \alpha)^2.
$$

(22)

Note that there is a unique coupling constant $g = g_{12} = g_{13} = g_{23}$ in SU(3), and the terms proportional to $\lambda_{\alpha\beta}$ have vanished. This is the classical Lagrangian for a particle at position $\alpha$ at time $r$, in an time-dependent effective potential

$$
V_{\text{eff}} = -\frac{1}{r^2} U_1(\alpha) - mgU_2(\alpha).
$$

(23)

Now, at radii $r$ such that $r^2 mg \ll 1$, $V_{\text{eff}}$ is maximized by minimizing $U_1(\alpha)$ on $[0, \pi/2]$, which is achieved with $\alpha = 0$. On the other hand, for $r \to \infty$, the effective potential is maximized by minimizing $U_2(\alpha)$, which is done with $\alpha = \alpha_0 = \arccos(1/\sqrt{3})$. Since $\frac{d}{d\alpha} U_1 \geq 0$ on $[0, \alpha_0]$, and $\frac{d}{d\alpha} U_2 \leq 0$ on the same interval, we may conclude that a solution exists with $\alpha(r)$ increasing monotonically from 0 at $r = 0$ to $\alpha_0$ at $r = \infty$. As expected, the vectors $\vec{\phi}_{\alpha}$ in (19) tend to (7), up to a gauge transformation, as $r \to \infty$. Moreover, as $r \to 0$, the UV singularities $e^{i\omega}$ in the $\vec{\phi}_2$ and $\vec{\phi}_3$ vanish, and $\vec{\phi}_1 \to (0, 0, e^{i\omega})^T$. 