

THE UNIVERSITY OF CHICAGO

DEEP LEARNING AND NON-INVERTIBLE SYMMETRIES IN GAUGE THEORIES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES DIVISION
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
ANUJ APTE

CHICAGO, ILLINOIS

JUNE 2025

To my teacher Dr. Vivek Wagh for nurturing my curiosity and a love for learning.

We dwell too infrequently upon the truth that under the heavens everything is interrelated,
even as each of us to the other.

- Max Adler

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	xi
ACKNOWLEDGMENTS	xii
ABSTRACT	xiv
1 INTRODUCTION	1
1.1 From Atoms to Quantum Theory of Fields	1
1.2 Symmetries in Quantum Field Theory	3
1.3 Symmetry Breaking and Phase Transitions	5
1.4 Anomalies and Renormalization Group Flow	6
1.5 Deep Learning	8
1.6 Neural Quantum States and Lattice Gauge Theory	10
2 OBSTRUCTIONS TO GAPPED PHASES FROM NON-INVERTIBLE SYMMETRIES	13
2.1 Introduction	13
2.1.1 Symmetry in Duality Invariant Theories	14
2.1.2 Phases of Duality Invariant Field Theories	16
2.2 Invertible Phases	19
2.2.1 Self-Dual SPT Phases	20
2.2.2 Physics of the Duality Defect	22
2.2.3 General Duality Invariant Invertible Phases	29
2.3 Topologically Ordered Phases	31
2.3.1 Examples	32
2.3.2 Constraints on TQFTs	34
2.3.3 General Duality Invariant TQFTs	36
2.4 Lattice Examples	37
2.5 Supplementary Material	41
2.5.1 Gauss Sums of Intersection Forms	41
2.5.2 Partition Function of $\mathbb{Z}_N$ 2-form Gauge Theory	45
2.5.3 Details of TQFT Argument	48
2.5.4 Duality Symmetry	48
2.5.5 Spontaneous Symmetry Breaking	55
3 DEEP LEARNING LATTICE GAUGE THEORIES	58
3.1 Introduction	58
3.2 $\mathbb{Z}_N$ lattice gauge theory	68
3.2.1 Wilson loop operators	71
3.2.2 Creutz ratio	73

3.2.3	't Hooft operators	74
3.2.4	Global symmetry and dual spin models	75
3.2.5	Ground-state degeneracy	76
3.2.6	Potential energy between charges	78
3.3	Phases and phase transitions in $\mathbb{Z}_N$ theories	80
3.3.1	First-order vs continuous phase transitions	81
3.3.2	Confined and deconfined phases	82
3.3.3	Confinement transition	83
3.4	Neural network quantum states	84
3.4.1	L-CNNs	89
3.4.2	NetKet implementation	93
3.4.3	Hilbert space sector of the NNQS	99
3.5	$\mathbb{Z}_2$ gauge theory in $2 + 1$ dimensions	101
3.5.1	Ground-state energies	103
3.5.2	Searching for the critical point and phase structure	103
3.5.3	Critical exponents and finite-size scaling	108
3.5.4	Potential energy between charges	114
3.6	$\mathbb{Z}_3$ gauge theory in $2 + 1$ dimensions	115
3.6.1	Ground-state energies	115
3.6.2	First-order phase transition	118
3.6.3	Static charges	123
3.7	Supplementary Material	125
3.7.1	Implementation of Wilson loop operators	125
3.7.2	Exponents from data collapse	127
3.7.3	BST extrapolation	128
4	CONCLUSION	131
	REFERENCES	135

LIST OF FIGURES

2.1	The definition of the duality defect $\mathcal{D}$ via gauging in half of spacetime. The left region couples to a background two-form gauge field B_L associated to the $\mathbb{Z}_N^{(1)}$ global symmetry. In the right region this symmetry is gauged with dynamical field b . The right region recovers the $\mathbb{Z}_N^{(1)}$ global symmetry through the Wilson surface operators of b which couple to the background field B_R	16
2.2	$\mathbb{Z}_N^{(1)} \times \mathbb{Z}_N^{(1)}$ symmetry of $\mathcal{D}$ arises from ending of one-form symmetry defects from $\mathcal{Q}$ (purple) and $S\mathcal{Q}$ (green).	23
2.3	The duality defect $\mathcal{D}$ in a bulk SPT phase. The defect $\mathcal{D}$ is a well-defined (2+1)d TQFT which is indentified with a minimal abelian TQFT.	24
2.4	$\mathbb{T}$ symmetry of the duality defect separating phases $\mathcal{Q}$ and $S\mathcal{Q}$. R_π combined with gauging of one-form symmetry in all of spacetime maps the boundary conditions to themselves, but reverses the orientation of $\mathcal{D}$. Hence this composite operation defines a time-reversal symmetry $\mathbb{T}$ of the duality defect worldvolume.	25
2.5	The $\mathbb{T}$ symmetry of the duality defect $\mathcal{D}$ is realized by intersecting duality defects (shown as black dots). At each such intersection the duality defect orientations reverse. Colliding two such junctions leads to the bulk fusion algebra of $\mathcal{D}$ with its orientation reversal $\overline{\mathcal{D}}$ resulting in a condensation of one-form symmetry surfaces shown in magenta. Within the duality defect worldvolume, this restricts to a condensation of abelian anyons that produces the charge conjugation symmetry $\mathcal{C}$	26
2.6	Schematic Phase Diagram of $\mathbb{Z}_N$ Lattice Gauge theory. Solid and dashed lines represent first order phase transitions. The vertical line at coupling $g_*^2 = N/2\pi$ is duality invariant. At weak coupling, the theory is in the Higgs phase <i>i.e.</i> it flows to a $\mathbb{Z}_N$ TQFT in the IR and at strong coupling, the theory is in the confining phase <i>i.e.</i> the vacuum is trivial. For $N \leq 4$, the Higgs and confining phases are separated by a first order phase transition at the self-dual coupling $g_*^2 = N/2\pi$ where the duality symmetry is spontaneously broken. For $N \geq 5$, the two phases are separated by an intermediate gapless coulomb phase which flows to a Maxwell theory and at the self-dual coupling $g_*^2 = 2\pi/N$ the IR Maxwell theory has $e^2 = 2\pi/N$	39
3.1	The left figure shows seconds per training iteration as a function of lattice size L . The depths of the networks are chosen so that, with kernel size $K = 2$ per layer, the largest Wilson loop is large enough to cover the entire lattice. All other hyperparameters are identical. The right figure shows the number of seconds per network parameter for a single iteration. The computations were performed on a single NVIDIA A100 GPU.	65
3.2	The left figure shows our convention for associating link variables to a lattice site at x . The middle figure shows the links (l_1, l_2, l_3, l_4) which make up a plaquette p when defining the magnetic term in the Hamiltonian (3.5). The right figure shows the links associated with the operator Θ_x acting at the lattice site at x , defined in (3.7), which implements Gauss's law on the lattice.	68

3.3	Order and disorder operators on the lattice. The blue path denotes a Wilson loop, $W_\Gamma = \prod_\Gamma Q$, given by the product of clock operators along the closed contour Γ . The shaded blue region shows the area enclosed by the Wilson loop. The dotted gray path denotes a ‘t Hooft string, $T_{\tilde{\Gamma}} = \prod_{\tilde{\Gamma}} P$, given by the product of shift operators on the links pierced by the contour $\tilde{\Gamma}$ between two points on the dual lattice.	72
3.4	Low-lying energy levels of $\mathbb{Z}_2$ lattice gauge theory computed using exact diagonalization on an $L = 2$ lattice. Away from $g = 0$, there is a unique ground state (teal), with an approximate fourfold degeneracy for small but non-zero couplings.	76
3.5	Phases of $\mathbb{Z}_N$ gauge theory in $2 + 1$ dimensions. The deconfined and confined phases are separated by a confinement transition as the critical coupling g_c	82
3.6	A schematic of an L-CNN with a single L-CB layer with $(N_{\text{out}}, K) = (2, 1)$ and no Dense-SELU layers. Here, the input is a configuration of link variables $\mathcal{U} = \{U_{x,\mu}\}$. The first layer, Plaq , converts the link variables into plaquette variables and stores them as $\mathcal{W} = \{W_x\}$ in a single channel. The L-Conv layer then parallel transports the plaquette variables by up to one lattice site in the positive x and y directions, giving a new set of plaquette-like variables $\mathcal{W}' = \{W_x, W_{x+\hat{x}}, W_{x+\hat{y}}\}$ with three channels. The sets $\mathcal{W}$ and $\mathcal{W}'$ are then multiplied by element-wise multiplication (denoted by the Hadamard product $\odot$), giving $W_x \odot W_x$ (squares of 1×1 loops), $W_x \odot W_{x+\hat{x}}$ (2×1 loops), and $W_x \odot W_{x+\hat{y}}$ (1×2 loops). Since $N_{\text{out}} = 2$, one then takes linear combinations of these (determined by the weights α_{ijk}) to give two output channels. Finally, a global pooling layer averages each of these channels over the lattice, giving two neurons which can be processed further. To help the reader follow variables through the network, we have highlighted three 1×1 plaquettes in blue, green and orange. Red dots indicate the lattice points to which the plaquette variables are associated. For example, the action of $W_x \odot W_{x+\hat{x}}$ equivariantly multiplies the blue plaquette by the orange plaquette (since they are associated to the same lattice point), resulting in a 2×1 plaquette.	90
3.7	Expectation value of $\langle H \rangle$ for $\mathbb{Z}_2$ gauge theory (3.5) on a 2×2 lattice as a function of coupling g . The ground state was computed using an L-CNN with six L-CB layers with hyperparameters $(N_{\text{out}}, K) = (3, 1)$. The networks were trained for up to 100 iterations with 4096 samples, though they converged within 25 or so iterations. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point. The solid teal line is the exact diagonalization result, computed using NetKet. The inset shows the fractional difference between the ground state energies calculated using the neural network and exact diagonalization.	102

3.8	Expectation value $\langle H \rangle$ for $\mathbb{Z}_2$ gauge theory (3.5) for varying lattice sizes as a function of coupling g , focused on the region around the critical coupling. The networks were trained for up to 2000 iterations with 2048 samples. Training was stopped early when the variance of the energy per site stabilised to a value of 1.25×10^{-4} or less. The expectation values were then calculated using 32×4096 Monte Carlo samples. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point. . . .	104
3.9	The difference between the calculated ground-state energy per lattice site for $L = 10$ and varying L , $\Delta e _L = E_{L=10}/10^2 - E_L/L^2 $, as a function of coupling g . Taking the $L = 10$ result to be our “continuum” values, we see that the energies quickly approach this, even for smaller lattice sizes.	104
3.10	The lattice average of the single-plaquette Wilson loop operator $W_\square$ in $\mathbb{Z}_2$ gauge theory as a function of coupling g near the critical point for varying lattice sizes. This corresponds to the expectation value of the magnetic flux energy per site. The dotted yellow line indicates the position of the critical coupling at $g_c = 0.756$, as determined by our analysis in Section 3.5.3.	106
3.11	String tension estimated from the Creutz ratio χ_l for $l = 2, 3$ on a 10×10 lattice using $2^{20} \approx 1\text{M}$ Monte Carlo samples. The error bars shown only take into account Monte Carlo errors. For $g > g_c$, Wilson loops decay rapidly with size. Estimates of σ are then challenging due to ratios of very small numbers, as can be seen from the noisier signal for $g > 1$ with $l = 2$ and $g > 0.85$ with $l = 3$. In particular, above $g = 0.9375$, the estimates of χ_3 are unreliable, with the standard error becoming larger than the observable. As such, we have not shown these points.	107
3.12	The lattice average of the ‘t Hooft string operator in $\mathbb{Z}_2$ gauge theory as a function of coupling g near the critical point for a 10×10 lattice. The string operator is of length five in the x direction.	107
3.13	Lattice average of the ‘t Hooft string for varying lattice size and distance $L/2$ between ends of ‘t Hooft string. The left figure shows the unscaled expectation values as a function of coupling, spaced equally over the range $g \in [0.74, 0.78]$ around the critical coupling. The right figure shows curve collapse for the rescaled observable $L^{2\beta/\nu}T$, with the rescaled coupling $\tilde{g} = L^{1/\nu}(g - g_c)/g_c$. The critical coupling and exponents are given by $g_c = 0.7546$, $\beta = 0.326$ and $\nu = 0.630$. Note that, since $\tilde{g}$ depends on L , a fixed range of g values is mapped to a varying range of $\tilde{g}$ values, as can be seen in the figure.	108

3.14	The left figure shows the linear potential $V(r)$ for $\mathbb{Z}_2$ gauge theory between two static charges on the lattice divided by its expected strong-coupling limit, $2g^2r$. The charges are placed at $(0,0)$ and $(r,0)$ for $r \in \{1, \dots, 5\}$ on a 10×10 lattice. The coefficient for each g is computed by fitting a linear function with zero intercept. The right figure shows the potential energy, $V(r) \equiv E_q - E_0$, defined as the difference between the ground-state energy without static charges, E_0 , and the energy E_q with charges a distance r apart on a 10×10 lattice for various values of the coupling. The lines are linear fits of the form $V(r) \propto r$. The $r = 0$ data corresponds to the state without charges. From the form of the Hamiltonian (3.5), in the strong-coupling limit, one expects the potential to approach $2g^2r$, so we have divided by $2g^2$ to see this. We have also plotted this behavior as the dashed diagonal black line. In all cases, after inserting the static charges, the networks were trained for a further 200 iterations before computing $\langle W_r^\dagger H W_r \rangle_\theta$ (though we observed negligible difference between this and using the untwisted sector ground state).	112
3.15	Expectation value of $\langle H \rangle$ for $\mathbb{Z}_3$ gauge theory (3.5) on a 2×2 lattice as a function of coupling g . The ground state was computed using an L-CNN with six L-CB layers with $(N_{\text{out}}, K) = (3, 1)$. The networks were trained for up to 100 iterations with 4096 samples. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point. The solid teal line is the exact diagonalization result, computed using NetKet. The inset shows the fractional difference between the ground state energies calculated using the neural network and exact diagonalization.	116
3.16	Expectation value $\langle H \rangle$ for $\mathbb{Z}_3$ gauge theory for varying lattice sizes as a function of coupling g , focused on the region around the critical coupling. The networks were constructed from $\lceil \log_2 N_{\text{sites}} \rceil$ L-CB layers with $(N_{\text{out}}, K) = (4, 2)$, and trained for up to 1000 iterations with 4096 samples. Training was stopped early when the variance of the energy per site stabilised to a value of 1.25×10^{-4} or less. The expectation values were then calculated using 32×4096 Monte Carlo samples. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point.	117
3.17	String tension estimated from Creutz ratio χ_l for $l = 2$ on a 10×10 lattice using $2^{20} \approx 1\text{M}$ Monte Carlo samples. For $g > g_c$, Wilson loops decay rapidly with size. Estimates of σ are then challenging due to ratios of very small numbers, as can be seen around $g > 0.95$.	119
3.18	The lattice average of the 't Hooft string operator in $\mathbb{Z}_3$ gauge theory as a function of coupling g near the critical point for a 10×10 lattice. The string operator is of length five in the x direction. The dotted yellow line indicates our identification of the transition at $g_c = 0.781$.	120

- 3.19 The left figure shows various expectation values per lattice site for $\mathbb{Z}_2$ gauge theory computed on a 10×10 lattice using an L-CNN as a function of coupling. These include the ground-state energy, the electric and magnetic energies, and the derivative of the energy with respect to g^2 . We have also labeled the position of the critical coupling $g_c = 0.756$ with a dashed gold line as calculated from BST extrapolation in Section 3.5.3. The right figure shows the same data for $\mathbb{Z}_3$ gauge theory, with the transition marked midway between 0.775 and 0.7875 at $g_c = 0.781$.120
- 3.20 The left figure shows the linear potential $V(r)$ for $\mathbb{Z}_3$ gauge theory between two static charges on the lattice divided by its expected strong-coupling limit, $\frac{3}{2}g^2$. The charges are placed at $(0, 0)$ and $(r, 0)$ for $r \in \{1, \dots, 5\}$ on a 10×10 lattice. The coefficient for each g is computed by fitting a linear function with zero intercept. The right figure shows the potential energy, $V(r) \equiv E_q - E_0$, defined as the difference between the ground-state energy without static charges, E_0 , and the energy E_q with charges a distance r apart on a 10×10 lattice for various values of the coupling. The lines are linear fits of the form $V(r) \propto r$. The $r = 0$ data corresponds to the state without charges. From the form of the Hamiltonian (3.5), in the strong-coupling limit, one expects the potential to approach $\frac{3}{2}g^2r$, so we have divided by $\frac{3}{2}g^2$ to see this. We have also plotted this behavior as the dashed diagonal black line. In all cases, after inserting the static charges, the networks were trained for a further 200 iterations before computing $\langle W_r^\dagger H W_r \rangle$. 121

LIST OF TABLES

2.1	Possible $N \leq 60$ for $\mathbb{Z}_N^{(1)}$ symmetric gapped phases invariant under gauging (S). The top line enumerates those N where an SPT exists. These are integers where -1 is a quadratic residue modulo N . The second line enumerates those N where a duality invariant non-trivial TQFT exists. In this case $N = k^2\ell$, where -1 is a quadratic residue modulo ℓ . A duality symmetry preserving theory with N that does not appear above is gapless. These include $N = 3, 6, 7, 11, 12, \dots$	18
3.1	Brief comparison of path-integral Monte Carlo (MC) and neural network quantum state (NNQS) approaches.	84
3.2	Selected theoretical determinations of 2 + 1-d Ising critical exponents.	113

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Clay Córdova for his guidance during the last five years. He embodies a spirit of inquiry and creativity that I hope to carry forward in my own research. He has been generous with his time and offered advice on personal and professional matters way beyond research. When I moved from pen-and-paper theory work to deep learning, he encouraged me to follow this interdisciplinary path with enthusiasm. The work presented in this thesis would not have been possible without my wonderful collaborators Ho Tat, Anthony, and Jimmy.

I am deeply grateful to my committee members for their invaluable support. I thank Michael for being on my committee and for explaining various ideas over the years in a manner that is easy to understand. I have learned a great deal from Luca's physically grounded and intuitive approach to physics. I would like to thank Karri for serving on my committee and providing helpful feedback.

My time in graduate school has been enriched by countless discussions with peers and collaborators. I have greatly enjoyed impromptu, freewheeling conversations with Sav on a wide variety of topics, both scientific and non-scientific. I thank Arvind and Kabir for teaching me how to do wet lab biology. By working with them, I have developed a deeper appreciation for the beauty of evolution.

I am deeply grateful to my undergraduate research advisors, Scott and Mingda, for their mentorship and for introducing me to the world of scientific research. I am deeply grateful to my Boston friends Debaditya, Arindam, Arsen, Tugsuu, Lay, Pawan, Dhyey, Keshav, Sapna, and Richa for supporting me throughout my graduate school journey. Special thanks to Janak, Jeet, Ileana, and Kanishk for for encouraging me to explore deep learning and for sharing their perspectives.

The intellectual environment of our research group was enriched by my fellow doctoral candidates: Diego, Nick, and Davi. Our regular discussions have significantly influenced the

development of ideas presented in this thesis. The Michelson Center for Physics provided a vibrant intellectual home, and I'm thankful to my friends and colleagues there: Matt, Yunjia, Sihan, Diane, Julian, Murtaza, Wanqiang, Anath, Ruchira, Anna, Tong, Han Wen, Robbie, Gabriel, Cheyne, Umang, Carolyn, Wen-Han, and Tai Wai.

My graduate school experience was enriched by the friendships formed in my cohort. Special thanks to Tommy, Henry, Jinwei, Ege, Harry, Juan Pablo, Alex, Elise, George, David, Vikram, Shreya, Lynn, Syrian, Mohit, Maxim, Chin Yi, Victor, Sravan, and Minzhao for their camaraderie, especially as we navigated the challenges of the pandemic together. To my friends across the broader University of Chicago community Agnish, Pranjali, Manish, James, Kunal, and Olga, thank you for the many conversations and connections that made my time here so meaningful.

This research was made possible through the financial support of a Nambu Fellowship from the Department of Physics and a subsequent grant from the Data Science Institute at the University of Chicago.

My parents, Shivangi and Shripad Apte, have always put our education first. Their love, sacrifices, and steady support made everything that followed possible. My late grandparents, Vasudha and Madhukar, were steady sources of warmth and wisdom throughout my life. I'm especially grateful to my older brother, Salil, who was the first to get me excited about science. By pursuing his own doctoral studies, he blazed a trail that made my path clearer. I would also like to express my sincere gratitude to Dr. Vivek Wagh, who fundamentally shaped the way I learn and think and to whom this thesis is dedicated.

Finally, to everyone mentioned here, and to those I may have missed, your support and encouragement have meant more than I can say and have shaped this journey in countless ways.

ABSTRACT

This thesis investigates gauge theories by combining theoretical analysis of non-invertible symmetries with computational simulations using neural network quantum states. We first investigate (3+1)d $\mathbb{Z}_N$ gauge theories with non-invertible duality symmetries that arise from gauging one-form symmetries. These symmetries impose strict constraints on renormalization group flows, often prohibiting symmetry-preserving gapped phases. Our proof demonstrates that a self-dual theory with $\mathbb{Z}_N$ one-form symmetry must either remain gapless or break self-duality, except when $N = k^2\ell$ with -1 as a quadratic residue modulo ℓ . Lattice gauge theories provide concrete realizations of these constraints.

We then focus on the development of neural network quantum states for studying lattice gauge theories. By implementing gauge-invariant neural networks, we compute ground-state wavefunctions for $\mathbb{Z}_N$ lattice gauge theories in (2+1)d. For $\mathbb{Z}_2$ gauge theory, we characterize the continuous confinement transition and identify critical exponents consistent with the Ising universality class. In the $\mathbb{Z}_3$ case we observe a weakly first-order transition between the confined and deconfined phases. For both theories, we compute order and disorder parameters and verify the linear confining potential in the confining phase. These results demonstrate that neural network methods provide a powerful tool for studying quantum field theories computationally. By combining these simulations with symmetry-based analysis, this work uncovers new insights into the nonperturbative dynamics and phase structure of gauge theories.

CHAPTER 1

INTRODUCTION

1.1 From Atoms to Quantum Theory of Fields

The quest to understand the composition of the universe has fascinated humans since antiquity. A remarkable turn occurred in philosophical discourse around 5th century B.C. when independently in Greece and India, Democritus and Kanada suggested that everything in the universe was made of small indivisible particles. These ancient thinkers were the first to develop a truly materialist philosophy, asserting that physical reality consisted of tiny, eternal particles moving through empty space. Democritus called these fundamental units "atomos" (meaning "uncuttable" in Greek), while Kanada's Vaisheshika school proposed "anu" (the smallest, indestructible particles of matter) [1, 2]. Their revolutionary ideas challenged prevailing notions of continuous matter and divine creation, establishing the intellectual foundation that would eventually inspire modern atomic theory.

While ancient atomists laid the philosophical groundwork, it would take over two millennia for these ideas to evolve into scientific theories. A profound conceptual leap occurred when Isaac Newton introduced the concept of gravitational fields—invisible forces acting across space. This field concept was later extended to explain electricity and magnetism, most notably through the work of Faraday and Maxwell in the 19th century [3]. As scientific understanding deepened, physicists realized that fields themselves carried energy and momentum, existing as fundamental entities alongside particles [4]. This dual nature of reality emerged as a cornerstone of modern physics: particles served as sources of fields, while the arrangement and movement of matter occurred as a consequence of particles responding to these fields. This interplay between discrete particles and continuous fields created a more nuanced understanding than either Democritus or Kanada could have imagined.

A major revolution in our collective understanding came with the rise of quantum theory

in the early 20th century. While Bohr's model still conceptualized atoms as miniature solar systems with particle-like electrons orbiting a nucleus, quantum mechanics revealed a far stranger reality where electrons behaved as probability clouds rather than discrete objects [5]. The atomic view that had persisted since Democritus was dethroned by the development of quantum field theory in the 1920s and 1930s by Paul Dirac, Werner Heisenberg, and others [6]. These pioneers recognized that fields and not particles were the primary building blocks, and particles were re-conceptualized as excitations in these underlying quantum fields [7].

This remarkable view of reality has found extensive empirical confirmation in particle physics experiments. While most of the mass of the atom resides in protons and neutrons, these particles themselves are composed of quarks held together by gluons, which are the carriers of the strong nuclear force. Strikingly, about 99% of a proton's mass originates not from its constituent quarks (which contribute only $\sim 1\%$), but from the energy of the gluon field binding them together. The rest mass of a proton is approximately 938 MeV, while the three quarks inside it account for merely 9-12 MeV [8]. The remaining mass ~ 925 MeV, emerges from the binding energy of the gluon field, confirming that what we perceive as solid matter is primarily the manifestation of energy trapped in quantum fields rather than discrete, substantive particles as the ancient atomists envisioned.

The Standard Model of particle physics which culminated in the 2012 discovery of the Higgs boson [9, 10], represents our most comprehensive framework for understanding fundamental particles and forces. This theory describes three of the four fundamental forces in nature through quantum fields whose excitations we perceive as particles. The electromagnetic force, carried by photons, governs interactions between charged particles. The weak nuclear force, responsible for radioactive decay and carried by W and Z bosons, enables flavor-changing interactions that transform one type of particle into another. Quantum Chromodynamics (QCD), the theory of the strong nuclear force, describes how gluons bind

quarks together to form protons and neutrons, and holds these nucleons together in atomic nuclei. Gravity is the fourth fundamental force, but a full quantum theory of gravity remains to be discovered.

Yet significant mysteries remain unsolved—neutrinos possess unexplained tiny masses, and we still don’t know what constitutes the dark matter that makes up roughly 27% of the universe. Perhaps most humbling is our limited understanding of strongly coupled systems. The vast majority of what we teach and apply in quantum field theory relies on perturbation theory, which treats interactions as small corrections to free, non-interacting fields. This perturbative framework has been extraordinarily successful for calculating processes in quantum electrodynamics and aspects of the weak nuclear force, where the coupling constants are small [11]. The textbook formulations that work brilliantly for calculating electron interactions fail spectacularly when confronting the physics of quarks and gluons in their confined state [12].

The extreme physics inside neutron stars exemplifies this challenge, and we cannot definitively state whether their cores contain exotic quark matter [13] or kaon condensates [14, 15]. Similarly, we struggle to describe the quark-gluon plasma that existed microseconds after the Big Bang [16–18], QCD at high baryon density, or supersolid nuclear phases theorized to exist in pulsars. It took over 30 years from the development of QCD in the 1970s [19] to the first ab initio calculations of the proton’s mass using lattice QCD [20], highlighting the computational challenges posed by strong coupling. These limitations remind us that while fields have replaced particles as fundamental entities, our understanding of how these fields behave under extreme conditions remains incomplete.

1.2 Symmetries in Quantum Field Theory

Symmetry has long been the guiding principle in our understanding of physics. In quantum field theory (QFT), symmetries come in two varieties: gauge and global. Gauge symmetries

represent redundancies in our description of nature rather than actual physical symmetries [21], while continuous global symmetries correspond to conserved quantities via Noether's theorem [22].

In recent years, our understanding of symmetry has changed substantially. The conventional notion of symmetry has been generalized beyond group theory to include higher-form symmetries, which act on extended objects rather than point particles, and non-invertible symmetries, which lack inverses and thus go beyond the traditional group structure. In 2014, Gaiotto, Kapustin, Seiberg, and Willett formalized the idea that global higher-form symmetries are associated with topological operators supported on extended submanifolds [23]. Their framework provides a unified way to describe and classify topological operators of various dimensions through the lens of generalized global symmetries. These symmetries act on charged objects of extended dimensionality; for instance, a 1-form symmetry acts on line operators such as Wilson lines in gauge theories.

Non-invertible symmetries represent another generalization of the symmetry concept. While conventional symmetries form groups, with each symmetry operation having an inverse, non-invertible symmetries follow more general algebraic structures called fusion categories [24]. These symmetries arise naturally in certain quantum field theories, particularly in the context of dualities. For example, the famous Kramers-Wannier duality of the 2D Ising model can be understood as a non-invertible symmetry [25]. When we act twice with this duality operation, instead of returning to the identity, we obtain a sum over conventional symmetry operations. This behavior is characteristic of non-invertible symmetries [26].

The discovery and classification of these generalized symmetries have had profound implications across theoretical physics. In condensed matter theory, they provide new tools for understanding topological phases of matter, including fractional quantum Hall states [27,28] and topological insulators [29,30]. In high-energy physics, they offer new perspectives on confinement [31], symmetry breaking, and anomalies [32]. The connection to quantum compu-

tation is particularly important: topological quantum error correction codes such as Kitaev's surface code [33] can be understood through the lens of higher-form symmetries [34], wherein logical qubits are encoded in topological degrees of freedom protected by these symmetries.

1.3 Symmetry Breaking and Phase Transitions

Phase transitions represent one of the key phenomena in physics, where small changes in parameters such as temperature can lead to dramatic changes in a system's properties [35]. A simple example is water, which undergoes a transformation from a liquid to a solid state when its temperature falls below zero degrees Celsius.

In the Landau-Ginzburg paradigm, phase transitions are characterized by symmetry breaking, where the system on either side of the phase transition has different pattern of symmetries [36]. This framework has been tremendously successful in describing a wide class of phase transitions, ranging from ferromagnetism to superconductivity.

The power of the symmetry-breaking paradigm lies in its universality, systems with entirely different microscopic details but the same symmetry-breaking pattern exhibit identical critical behavior [37]. This profound insight explains why seemingly unrelated physical systems can be described by the same mathematical framework near their critical points [38]. For example, the liquid-vapor transition in water and the ferromagnetic transition in iron both belong to the Ising universality class, characterized by the spontaneous breaking of a $\mathbb{Z}_2$ symmetry.

However, the discovery of topological phases of matter revealed the limitations of the Landau paradigm [39]. Topological phases are distinct quantum states that cannot be distinguished by local order parameters. These phases remain robust against local perturbations and exhibit phenomena such as quantized responses and protected edge states [40]. The fractional quantum Hall states first discovered by Tsui, Stormer, and Gossard [41], were the first to exhibit fractional excitations [42] and topological ground state degeneracy. From

the modern perspective, these systems possess emergent higher-form symmetries that protect their topological order [23], with transitions between different quantum Hall plateaus corresponding to topological phase transitions.

Topological phase transitions, which occur between different topologically ordered states, can be understood as the breaking or restoration of higher-form symmetries [43]. For example, the confinement-deconfinement transition in gauge theories corresponds to the spontaneous breaking of a 1-form symmetry [44]. In $\mathbb{Z}_2$ lattice gauge theory, the confined phase preserves the $\mathbb{Z}_2$ 1-form symmetry, while the deconfined phase breaks it.

1.4 Anomalies and Renormalization Group Flow

In quantum field theory, anomalies arise when a symmetry of the classical action fails to survive quantization [45]. These quantum symmetry violations have far-reaching implications, placing stringent constraints on the structure of both high-energy and condensed matter theories. The chiral anomaly, first discovered by Adler, Bell, and Jackiw [46] in the context of neutral pion decay, provides a striking example: although classical chiral symmetry would forbid the decay of a neutral pion into two photons, quantum effects allow it. In the Standard Model, the requirement that gauge anomalies cancel imposes tight constraints on the spectrum of fermions—forcing precise relationships between quark and lepton charges to ensure gauge consistency. In condensed matter physics, anomalies play a central role in phenomena such as the quantum Hall effect, where the presence of gapless edge modes on the boundary of a topological insulator compensates for an anomaly in the bulk.

The renormalization group (RG), developed by Kenneth Wilson [47], describes how physical theories change with scale. It establishes that systems at different length scales can be represented by effective field theories with scale-dependent parameters [48]. These theories are connected by RG flows that systematically integrate out microscopic degrees of freedom. At critical points where phase transitions occur, RG flows reach fixed points characterized

by scale invariance [49].

In RG language, "infrared" (IR) refers to the long-distance or low-energy behavior of a system, while "ultraviolet" (UV) refers to short-distance or high-energy behavior. For example, in a magnetic material, the UV description includes details about individual atomic spins and their local interactions, while the IR description captures only the collective behavior like overall magnetization. This distinction is crucial because many physical systems appear completely different at different scales.

Wilson's formulation revolutionized the understanding of critical phenomena, explaining universality in phase transitions and providing tools for calculating critical exponents. Near critical points, the correlation length diverges and the system becomes scale-invariant. The behavior is dominated entirely by fluctuations at all length scales with no characteristic size. At such points, microscopic details become irrelevant, and the system's properties are determined solely by a small number of relevant operators that survive in the IR limit. This explains the remarkable phenomenon of universality: systems with entirely different microscopic constituents can exhibit identical critical behavior characterized by the same set of universal exponents. The systems all flow under renormalization group transformations to the same IR fixed point, erasing their microscopic differences.

't Hooft's anomaly matching condition [50] places strict constraints on RG flows by requiring anomalies to match between microscopic degrees of freedom and infrared descriptions of a theory. If a symmetry has an anomaly in the microscopic theory, the IR theory must account for this anomaly regardless of how the physics reorganizes at low energies. This often rules out the possibility of a trivially gapped phase, forcing theories with certain anomalies to either break symmetry spontaneously or maintain gapless excitations [51]. Gapless excitations refer to low-energy modes that require arbitrarily small amounts of energy to excite, indicating the absence of an energy gap in the spectrum. The presence of gapless degrees of freedom fundamentally alters a system's response to perturbations and typically leads to

power-law correlations rather than the exponential decay characteristic of gapped phases.

The matching of anomalies during RG flow leads means that even if the theory is strong coupled in the IR, one can use the knowledge of the anomaly in the UV to get a better understanding of the dynamics of the theory. Thus, global symmetries and their anomalies offer one of the few tools for analytically studying the QFT at string coupling. In QCD, anomaly matching explains why massless pions must emerge when chiral symmetry is spontaneously broken.

Chapter 2 extends this powerful technique to non-invertible symmetries, a generalization that goes beyond the conventional group structure of symmetries [52]. While 't Hooft anomalies for ordinary symmetries can obstruct gapped phases, we demonstrate analogous obstructions arising from non-invertible duality symmetries in (3+1)d $\mathbb{Z}_N$ gauge theories. We classify which values of N permit symmetric gapped phases, and which values necessarily lead to gapless excitations or spontaneous symmetry breaking. A gapped phase has a finite energy difference (a gap) between its ground state and first excited state, even in the thermodynamic limit, implying that all excitations require a minimum finite energy to create. Symmetric gapped phases preserve the system's symmetries while maintaining this energy gap. When such phases are forbidden by non-invertible symmetry constraints, the system must either develop gapless excitations or spontaneously break the symmetry. This extends the Landau paradigm to encompass non-invertible symmetries and provides a systematic framework for understanding phase transitions in terms of these generalized symmetries.

1.5 Deep Learning

Deep learning has emerged as a powerful and versatile tool, transforming fields from computer vision and natural language processing to physics, biology, and materials science [53]. Its development started in the 1950s, but progress remained slow for decades due to limited computational resources and insufficient data. A key turning point came in the 2010s, when

three trends converged: (1) the availability of massive datasets, (2) advances in parallel hardware (especially GPUs and TPUs), and (3) the development of algorithms and architectures that scale effectively with data and compute.

The central task in deep learning is to learn conditional probability distributions from data. Given input variables x , the model learns to approximate $p(y|x)$: the probability distribution of outputs y conditioned on x . This paradigm is quite flexible and cover a wide range of tasks:

- In image classification, x is an image and y is a label.
- In language modeling, x is a sequence of text and y is the next word or token.
- In protein structure prediction models such as AlphaFold, x is an amino acid sequence and y is a 3D protein structure [54].

To learn these conditional distributions, deep learning models are trained to minimize a quantity called the loss function, which measures how far the model’s prediction $\hat{y}$ is from the true value y . The loss function encodes the learning objective such as cross-entropy loss for classification tasks, or the mean squared error for regression. For example, in image classification, the model outputs a probability distribution over possible labels, and cross-entropy penalizes deviations from the true label. In AlphaFold, the loss is more complex: it includes terms that penalize differences between predicted and experimentally determined 3D coordinates, as well as violations of physical or geometric constraints [54].

During training, the model parameters (weights and biases) are updated to reduce the loss using gradient-based optimization, most commonly stochastic gradient descent (SGD) [55] or one of its variants like Adam [56]. At each step, the model computes the gradient of the loss with respect to its parameters and updates them in the direction that most rapidly reduces the loss. This iterative process continues until the model converges to a function that approximates $p(y|x)$ well over the training data.

The power of this framework lies in its ability to generalize across domains, modeling complex, high-dimensional data without requiring explicit rules. Today’s most capable models e.g.: AlphaFold [54] and GPT-4 [57] are built on this principle. These neural networks with billions of parameters are trained on massive datasets, optimized by stochastic gradient descent accelerated with large-scale parallel processors.

1.6 Neural Quantum States and Lattice Gauge Theory

Neural quantum states (NQS) are a powerful class of variational wavefunctions where the quantum states is parameterized by neural networks [58]. The central task is to approximate the ground state of a quantum Hamiltonian by minimizing the expected energy. Unlike supervised learning problems with a fixed labeled dataset, this task is inherently unsupervised: data is generated on the fly by sampling from the basis elements of the Hilbert space. The network learns to assign amplitudes $\psi(x)$ to configurations x , and training proceeds by minimizing the expectation value of energy, which serves as the loss function [58]. Gradients of the energy with respect to the network parameters are computed using automatic differentiation, allowing for efficient optimization via stochastic gradient descent [59]. In this framework, learning the ground state becomes equivalent to learning a complex-valued probability distribution that minimizes a physically meaningful loss.

Neural quantum states have been successfully applied to a wide range of quantum many-body problems [60]. In quantum chemistry, they have been used to approximate the ground states of molecular electronic systems, capturing strong electron correlations beyond traditional mean-field methods. In lattice models, they have described frustrated magnets, topological phases, and systems with sign problems that are beyond the reach of traditional Monte Carlo methods [61]. Unlike tensor networks, which are efficient primarily for low-entanglement states in one dimension, NQS can scale to higher dimensions and represent states with more complex entanglement, including volume-law states. Global and gauge sym-

metries can be incorporated directly into the network architecture using group-equivariant neural networks, enabling computationally efficient representations [62]. Thus, NQS provide a computational tool to study strongly coupled quantum many-body systems, enabling the exploration of phenomena beyond traditional computational methods.

In Chapter 3 we use neural quantum states to study discrete lattice gauge theories in $(2+1)d$, which serve as simplified models capturing essential features of more complex gauge theories including confinement and symmetry breaking [62, 63]. For the first time, we show that gauge-equivariant NQS can not only approximate ground states, but also accurately extract critical exponents associated with second-order phase transitions. This demonstrates that NQS can characterize universality classes directly from real-space simulations, offering capabilities beyond those of standard Euclidean Monte Carlo or tensor network methods.

Our architecture is based on lattice convolutional neural networks (L-CNNs), which are both local and gauge-equivariant by design [64]. What makes the use of CNNs in physics so striking is that they encode only a very minimal inductive bias: locality. The core assumption behind CNNs is that nearby inputs are more likely to interact than distant ones, and that the same local patterns can occur anywhere in the input space. This principle of local weight sharing is incredibly general and applies equally well to the pixels in an image as to spins on a lattice or degrees of freedom in a quantum field.

This minimalism is precisely what enables convolutional neural networks to generalize across domains. Their power lies not in hand-crafted, domain-specific heuristics, but in their ability to learn directly from data. The effectiveness of such general-purpose systems reflects what Sutton termed the "Bitter Lesson" [65], which encapsulates the idea that progress in AI has consistently come from methods that leverage computation and scale, rather than those built around human-designed structure or domain knowledge. CNNs exemplify this lesson: with locality as their only built-in assumption, they have proven capable of tackling tasks in image recognition, self-driving cars, and even quantum many-body physics.

The fact that a single architectural principle of locality can be applied across such diverse domains is a striking example of methodological unification. This convergence heralds a shift in scientific inquiry, with data-driven architectures emerging as powerful tools for discerning structure in systems once deemed too complex to model. A vivid example of this transformation is AlphaFold [54], a deep learning system that solved the long-standing protein folding problem and, within just a few years lead to a Nobel Prize in Chemistry. What was once an intractable biological puzzle became tractable through scalable neural architectures trained end-to-end on raw sequence data without human-designed rules.

This breakthrough exemplifies a broader transformation in scientific inquiry, where machine learning methods can tackle diverse challenges from particle tagging in high-energy physics [66] and real-time detection of gravitational waves [67] to the accelerated discovery of novel materials [68]. As machine learning systems grow ever more capable, they promise to reshape how we tackle intractable problems and uncover new questions. We are in the midst of a new computational paradigm, in which tools of artificial intelligence will accelerate the pace of discovery across the sciences.

CHAPTER 2

OBSTRUCTIONS TO GAPPED PHASES FROM NON-INVERTIBLE SYMMETRIES

This chapter is reprinted with permission from:

Apte, A., Córdova, C., and Lam, H. T. Obstructions to gapped phases from noninvertible symmetries. *Physical Review B* (Vol. 108, Issue 4). © 2023 American Physical Society
DOI:10.1103/PhysRevB.108.045134

2.1 Introduction

The study of renormalization group flows and phase transitions has been a central topic in quantum field theory for the past half century. A core conceptual idea is the organization of phases of field theories by their global symmetries and the realization of these symmetries on the ground state. This is the Landau paradigm for states of matter. In this paper, we explore this framework for novel non-invertible duality symmetries.

In its modern incarnation, the idea of symmetry has become intrinsically linked with topology. Each global symmetry of a theory may be understood as a co-dimension one defect operator (or as a domain wall in a spontaneously broken phase). These symmetry defects are topological: continuous deformations of their positions which do not cross other operators leave their correlation functions invariant.

Over the last decade this idea has radically broadened in its scope and applicability, encompassing higher-form symmetry corresponding to invertible topological operators of general dimension [69], higher-group symmetries which intertwine invertible topological operators of different dimensions [70–73], and finally non-invertible symmetries [71, 74–77]: the algebraic frontier where the defect operators are characterized by general higher fusion categories [78–84].

In this work, we focus on non-invertible symmetries of (3+1)d theories, with the basic goal of determining when these symmetries are compatible with a unique vacuum state and a gapped spectrum. As we describe below, in general we will find that such realizations of duality symmetries are impossible. These obstructions are similar in spirit to the Lieb-Shultz-Mattis theorem [85] which implies that certain (1+1)d lattice models are gapless or have degenerate ground states given the existence of certain types of symmetry. Our results extend these ideas to the arena of higher-dimensional field theories invariant under a novel class of symmetries.

2.1.1 Symmetry in Duality Invariant Theories

Examples of (3+1)d systems invariant under non-invertible symmetries may be constructed by starting from the class of quantum field theories which have a $\mathbb{Z}_N^{(1)}$ one-form global symmetry. Given any such theory $\mathcal{Q}$, one can construct others by the following operations:

- S : Gauging the $\mathbb{Z}_N^{(1)}$ symmetry by coupling to a dynamical $\mathbb{Z}_N$ two-form gauge field b . The resulting theory $S\mathcal{Q}$, then has an emergent dual $\mathbb{Z}_N^{(1)}$ global symmetry arising from the Wilson surface operators $\exp\left(\frac{2\pi i}{N} \oint b\right)$.
- T : Stacking the theory $\mathcal{Q}$ with a minimal invertible theory for the $\mathbb{Z}_N^{(1)}$ global symmetry. In terms of a $\mathbb{Z}_N^{(1)}$ background gauge field B this is expressed as

$$\begin{cases} \exp\left[\frac{2\pi i}{2N} \int_X \mathcal{P}(B)\right] & N \text{ even} , \\ \exp\left[\frac{2\pi i(N+1)}{2N} \int_X \mathcal{P}(B)\right] & N \text{ odd} , \end{cases} \quad (2.1)$$

where $\mathcal{P}(B)$ is a suitable quadratic function (see equation (2.6) below).

As an action on the set of all field theories with $\mathbb{Z}_N^{(1)}$ symmetry, S , and T realize a discrete analog of the modular group and obey the equations:

$$S^2 = C, \quad (ST)^3 = Y, \quad (2.2)$$

where C is charge conjugation acting as $B \leftrightarrow -B$, and Y represents stacking the original theory $\mathcal{Q}$ with an invertible field theory depending only on the spacetime manifold [69, 86]. By performing such a gauging procedure in half of space-time with Dirichlet boundary conditions separating the two halves, one obtains an interface separating two theories $\mathcal{Q}$ and $S\mathcal{Q}$ (see Figure 2.1). As observed in [76, 77] when the theory is self-dual under gauging i.e. $\mathcal{Q} \cong S\mathcal{Q}$, then this interface is a topological symmetry operator in $\mathcal{Q}$, a duality defect. Such duality defects generalize Kramers-Wannier duality lines from (1+1)d QFTs to this higher-dimensional setting. More generally, one may also consider theories invariant under gauging operations built from composites of S and T . For instance, invariance $T^{-1}S$ leads to a triality defect [81].

These defects generalize the ordinary concept of symmetry. In particular, their fusion is not defined by a group. For instance, the duality defect $\mathcal{D}$ is invertible up to a condensate of one-form symmetry defects. Specifically, wrapping the defect $\mathcal{D}$ on a three-manifold M and colliding it with its CPT conjugate $\overline{\mathcal{D}}$ one finds:

$$\mathcal{D}(M) \times \overline{\mathcal{D}}(M) \cong \frac{1}{N} \sum_{S \in H_2(M, \mathbb{Z}_N)} \exp \left(\frac{2\pi i}{N} \oint_S b \right), \quad (2.3)$$

For a complete list of the fusion rules for the duality and triality defects see [81].

A wide variety of non-trivial examples realizing these symmetries or related non-invertible defects have recently been constructed in the literature. These include [43, 76, 77, 79–83, 87–124], which build on extensive previous investigations of non-invertible symmetries in $(1+1)d$ theories [71, 75, 125–152].

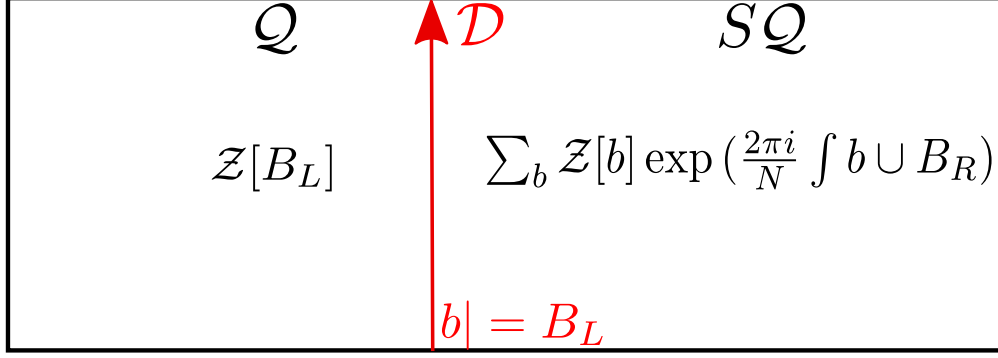


Figure 2.1: The definition of the duality defect $\mathcal{D}$ via gauging in half of spacetime. The left region couples to a background two-form gauge field B_L associated to the $\mathbb{Z}_N^{(1)}$ global symmetry. In the right region this symmetry is gauged with dynamical field b . The right region recovers the $\mathbb{Z}_N^{(1)}$ global symmetry through the Wilson surface operators of b which couple to the background field B_R .

2.1.2 Phases of Duality Invariant Field Theories

Like all notions of symmetry in quantum field theory, the duality defects described above can be used to constrain dynamics. Below our main focus is on characterizing the interplay between duality invariant phases and the mass gap.

In general, given any symmetry one may always realize it in a spontaneously broken phase. In the case of a discrete symmetry, such as the duality defects discussed here, a spontaneously broken phase means that there are multiple local vacua which are characterized by the presence of multiple topological local operators.¹ The local physics in a given ground state is unconstrained, but the various ground states are related by the gauging operation. This possibility can occur at a first order phase transition, where the duality defect forms the domain wall between different ground states. We discuss spontaneous symmetry breaking further in supplementary section 2.5.5, and highlight its occurrence in $\mathbb{Z}_N$ lattice gauge theory for small N below.

In contrast with the scenario outlined above, symmetry preserving phases are tightly constrained. In this situation there is a unique local topological operator (the identity).

1. Here and below, by a local vacuum state we mean a ground state on $\mathbb{R}^3$ or equivalently S^3 .

Scale invariant phases arising at the end of renormalization group flows can be classified as follows:

- Invertible Phases: The theory is gapped and invertible. Its partition function is a phase and there is a unique ground state on any spatial topology.
- Topologically Ordered Phases: The theory is gapped, but described by a non-trivial topological quantum field theory (TQFT). The ground state degeneracy depends on the spatial manifold.
- Gapless Phases: The mass gap vanishes and the theory is a free or interacting conformal field theory, with non-trivial power law correlation functions.

In the case of invertible symmetry (characterized by groups), the obstruction to the existence of an invertible realization is the 't Hooft anomaly. However, anomalies can also obstruct the existence of topologically ordered phase. This phenomenon is referred to as *symmetry enforced gaplessness* explored in condensed matter systems in [153–157], and through the lens of quantum field theory in [158–164].

We generalize these considerations to (3+1)d duality invariant field theories by providing the first examples of TQFTs which are duality invariant without spontaneous symmetry breaking, and classifying all possible N for which this can occur. This extends parallel results in (1+1)d [75] as well as the work of [77, 81], where duality invariant (3+1)d one-form symmetry protected topological (SPT) phases were classified. For the case of theories invariant under S with the duality defect $\mathcal{D}$, our findings are enumerated in Table 2.1.

Our basic method is to generalize the analysis of [162, 163] to the setting of non-invertible symmetries. The key idea is to examine the $\mathbb{Z}_N^{(1)}$ one-form symmetry action on line operators. A distinguished role is played by the subgroup of $\mathbb{Z}_N^{(1)}$ that does not act

SPT	2, 5, 10, 13, 17, 25, 26, 29, 34, 37, 41, 50, 53, 58, ...
TQFT	4, 8, 9, 16, 18, 20, 25, 32, 36, 40, 45, 49, 50, 52, ...

Table 2.1: Possible $N \leq 60$ for $\mathbb{Z}_N^{(1)}$ symmetric gapped phases invariant under gauging (S). The top line enumerates those N where an SPT exists. These are integers where -1 is a quadratic residue modulo N . The second line enumerates those N where a duality invariant non-trivial TQFT exists. In this case $N = k^2\ell$, where -1 is a quadratic residue modulo ℓ . A duality symmetry preserving theory with N that does not appear above is gapless. These include $N = 3, 6, 7, 11, 12, \dots$

faithfully, i.e. the subgroup of the symmetry with no associated charged objects.² In a gapped phase, these operators admit topological boundary conditions. If we further assume that there is no spontaneous symmetry breaking of the duality invariance, the effect of these unfaithfully acting surface operators can be reduced to insertions of the identity operator. Enforcing duality invariance then leads to constraints on N as elucidated in theorem 5. We also provide a converse to our constraints on N by explicitly constructing duality invariant TQFTs for all allowed values of N , generalizing the results of [164] to the setting of non-invertible symmetries.

Beyond simply reproducing the analysis of [77] for SPT phases, our analysis of duality defects also reveals their physical properties. In particular, we show that for a duality invariant SPT phase the world-volume theory of the defect $\mathcal{D}$ is a minimal Abelian TQFT [167] with a $\mathbb{T}$ invariant spectrum of anyons as analyzed in [168]. This is analogous to the analysis presented in [169–171], where $\mathbb{Z}_2$ symmetry defects were shown to admit $\mathbb{T}$ symmetry with a map relating the bulk and defect anomalies.

Finally, we illustrate our results by considering $\mathbb{Z}_N$ lattice gauge theories. In these theories, topological phase transitions for $\mathbb{Z}_2, \mathbb{Z}_3$ and $\mathbb{Z}_4$ occur by spontaneous breaking of non-invertible self-duality, while for larger N the duality invariant point is gapless. Thus, these first order phase transitions for small N can be viewed as part of a generalized

2. Note that a one form symmetry which acts unfaithfully by linking with line operators may still have non-trivial correlation functions for instance via triple linking or other more intricate configurations [165, 166].

Landau paradigm by incorporating non-invertible symmetries. Several appendices summarize more technical material about quadratic Gauss sums, partition functions of topological two-form gauge theories, details of our argument classifying duality invariant TQFTs, and comments about models that spontaneously break duality symmetry.

2.2 Invertible Phases

In this section, we classify invertible field theories with duality symmetry which are invariant under combinations of the gauging operations S and T following [77, 81]. We also highlight how self-duality leads to an anti-unitary time-reversal symmetry acting on the duality defect $\mathcal{D}$, and comment on other physical aspects of the resulting symmetry defects. For simplicity, throughout this paper we will work on a smooth simply-connected Euclidean spacetime four-manifold X . The manifold X has the bi-linear intersection pairing in the middle dimension

$$I : H^2(X; \mathbb{Z}) \times H^2(X; \mathbb{Z}) \rightarrow \mathbb{Z} , \quad (2.4)$$

defined by the cup product. Associated to I we introduce a quadratic function $\mathcal{P}(B)$ where $B \in H^2(X; \mathbb{Z}_N)$ is a background field for the one-form symmetry and $\mathcal{P}$ depends on the parity of N as:

$$\mathcal{P}(B) = \begin{cases} \text{Pontryagin square} \in H^4(X, \mathbb{Z}_{2N}) & N \text{ even} , \\ B \cup B \in H^4(X, \mathbb{Z}_N) & N \text{ odd} . \end{cases} \quad (2.5)$$

2.2.1 Self-Dual SPT Phases

The most general bosonic SPT phase for a $\mathbb{Z}_N^{(1)}$ one-form symmetry can be expressed as [172]:

$$Z[B] = \exp \left[\int_X \frac{2\pi i p}{2N} \mathcal{P}(B) \right] . \quad (2.6)$$

Here the integer p characterizes the phase and is identified as $p \sim p + 2N$. For N even there are thus $2N$ distinct phases, while for N odd, p is further constrained to be even leading to N distinct phases.

The phase (2.6) enjoys duality symmetry if it is invariant under the gauging operation S . This can be directly checked by computing the partition function after gauging:³

$$SZ[B] = \lambda_N^{-1} \sum_{b \in H^2(X, \mathbb{Z}_N)} Z[b] \exp \left[\int_X \frac{2\pi i}{N} b \cup B \right] , \quad (2.7)$$

where $\lambda_N = \sqrt{|H^2(X, \mathbb{Z}_N)|}$. The right-hand side above is in general a non-invertible TQFT, and hence can only be equivalent to (2.6) in the special case where all operators are trivialized by the equations of motion for b , given by

$$pb + B = 0 \quad \text{mod } N . \quad (2.8)$$

If p is co-prime to N , the above equation may be solved for b restricting it to be a non-fluctuating classical background B . Thus only in this case can gauging the SPT (2.6) result in an invertible theory. From now on we assume this condition. Substituting (2.8)

3. In (2.7) one may also consider a more general bi-character by modifying the weight of the $b \cup B$ term by an integer ℓ which is co-prime to N . This does not modify the conclusions below.

back into the action (2.7) leads to⁴

$$SZ[B] = G(I, p, N) \exp \left[- \int_X \frac{2\pi i (p)_{\gamma(N)N}^{-1}}{2N} \mathcal{P}(B) \right] . \quad (2.9)$$

In the above, $G(I, p, N)$ is an overall B independent phase defined by a Gauss sum for the intersection form I (see supplementary section 2.5.1):

$$G(I, p, N) = \frac{1}{N^{\text{rk}(I)/2}} \sum_{x=0}^{N-1} \exp \left(\frac{2\pi i p}{2N} x^T I x \right) , \quad (2.10)$$

where $\text{rk}(I)$ is the rank of I .

Let us focus first on the B dependence of (2.9). We see that the gauged action has a quadratic dependence on B with a map on p :

$$p \mapsto -(p)_{\gamma(N)N}^{-1} , \quad \gamma(N) \equiv \begin{cases} 1 & \text{odd } N \\ 2 & \text{even } N \end{cases} , \quad (2.11)$$

where the notation $(\alpha)_{\beta}^{-1}$ denotes the inverse of α in the ring $\mathbb{Z}_{\beta}$ for co-prime integers α and β . Invariance under the S transformation therefore means that in the discrete group classifying the SPT, this map is the identity.

Thus, a duality invariant SPT can exist only if there are solutions to the modular quadratic equation:

$$p^2 = -1 \pmod{\gamma(N)N} . \quad (2.12)$$

Moreover, when solutions to the above exist it is straightforward to check that the overall phase defined by the Gauss sum in (2.9) is unity (see (2.58).) As is well known, the existence of solutions to (2.12) depends on the prime factorization of N . Thus, we

4. When N is odd, we pick an even $(p)_N^{-1}$ so that the action is well-defined. If $(p)_N^{-1}$ is odd, we can redefine it by a shift of N to obtain an even $(p)_N^{-1}$.

reproduce the result of [77]:

Theorem 1. *There exist invertible bosonic phases realizing the duality defect $\mathcal{D}$, i.e. invariant under the gauging operation S if and only if N is odd and a product of Pythagorean primes:*

$$N = p_1^{a_1} p_2^{a_2} \dots p_n^{a_n} , \quad \forall i \quad p_i = 1 \bmod 4 . \quad (2.13)$$

For fermionic or equivalently spin theories, we can further refine the result above. When N is even, using the fact that,

$$\mathcal{P}(B) = B \cup w_2(X) \bmod 2 , \quad (2.14)$$

we deduce that on a spin manifold, $\mathcal{P}(B)$ is also even and hence there is an identification $p \sim p + N$. Then, the same steps as above imply that we have to solve the quadratic residue equation modulo N for both odd and even N

$$p^2 = -1 \bmod N . \quad (2.15)$$

Moreover in this case the overall phase defined by the Gauss sum is trivial due to Rokhlin's theorem (see below (2.57)). Therefore we again reproduce the result of [77]:

Theorem 2. *There exist invertible spin phases invariant under the gauging operation S if and only if N a product of Pythagorean primes up to a single factor of two.*

2.2.2 Physics of the Duality Defect

It is instructive to analyze the physics of the duality defect $\mathcal{D}$. We focus below on the bosonic case, though analogous considerations hold for duality defects of fermionic SPTs.

With general bulk dynamics, the symmetry defect has a $\mathbb{Z}_N^{(1)} \times \mathbb{Z}_N^{(1)}$ symmetry which arises

from bulk one-form symmetry defects ending on the duality defect (see Figure 2.2). This ending is possible because the duality defect is defined by Dirichlet boundary conditions for the dynamical gauge field (see Figure 2.1).

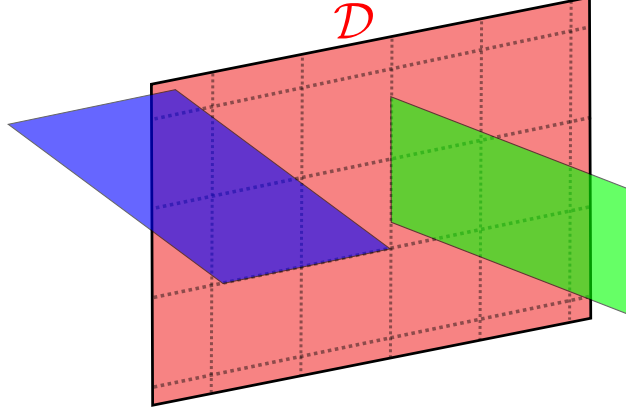


Figure 2.2: $\mathbb{Z}_N^{(1)} \times \mathbb{Z}_N^{(1)}$ symmetry of $\mathcal{D}$ arises from ending of one-form symmetry defects from $\mathcal{Q}$ (purple) and $S\mathcal{Q}$ (green).

Physically, the one-form symmetries of $\mathcal{D}$ define Abelian anyons within $\mathcal{D}$. The self-braiding of each $\mathbb{Z}_N^{(1)}$ factor is theory dependent, but the braiding between the generators of each $\mathbb{Z}_N^{(1)}$ factor is fixed by the pairing of b and B in the gauging defining the defect (See Figure 2.1).⁵

When the bulk dynamics are invertible the theory on the duality defect is a well-defined (2+1)d TQFT and the bulk SPT phase on the left and the right specify the anomaly of this system. Moreover the construction of the defect by gauging in a half space identifies the (2+1)d defect theory as a minimal Abelian TQFT $\mathcal{A}^{N,p}$ [167] (see Figure 2.3). Such a theory has a spectrum of lines which are exactly given by the abelian anyons defining the one-form symmetry.

The theory $\mathcal{A}^{N,p}$ has a single $\mathbb{Z}_N^{(1)}$ symmetry generated by an abelian anyon a . Thus, the left and right bulk symmetry surfaces must each end on these abelian anyons but with a

5. This generalizes an analogous statement in (1+1)d theories with Tambara-Yamagami fusion category symmetry where the bi-character encodes the anomaly between left and right symmetries of the defect [75, 143, 173].

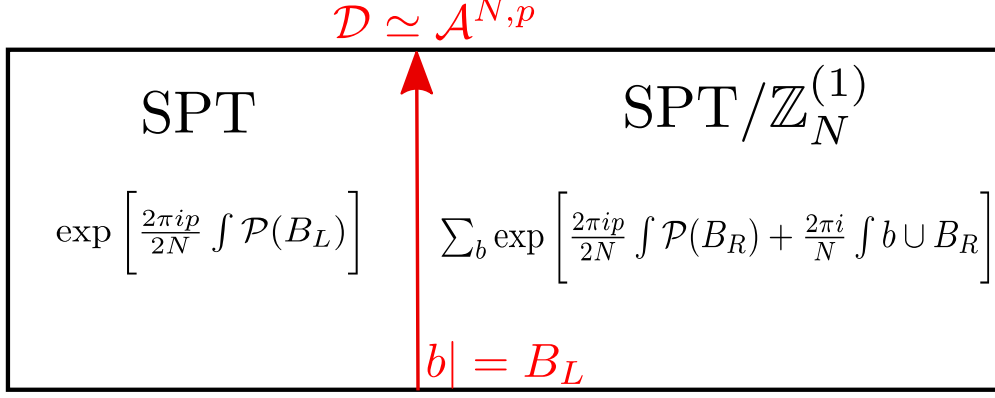


Figure 2.3: The duality defect $\mathcal{D}$ in a bulk SPT phase. The defect $\mathcal{D}$ is a well-defined (2+1)d TQFT which is identified with a minimal abelian TQFT.

possible difference in the choice of generating line. To deduce the relationship between left and right, we consider the equation of motion (2.8) and interpret B and b as the fields on the wall sourcing the right and left symmetries. When (2.8) is not solved the partition function of the total system including the defect vanishes. This means that the quantity appearing in the equations of motion sources charged abelian anyons wrapping non-trivial cycles in the defect. Correspondingly, when the equation of motions hold no charged lines are inserted. Hence, the left and right surfaces end on the lines:

$$\text{Left : } a \quad \text{Right : } a^{(p)_N^{-1}}. \quad (2.16)$$

This relation between the left and the right symmetries (2.16) may be further understood by observing that in an invertible phase, the duality defect $\mathcal{D}$ has an antiunitary time-reversal symmetry, $\mathbb{T}$. Note that this is true even though for general p the bulk SPT defined in (2.6) does not have $\mathbb{T}$ symmetry.

To argue for this conclusion we proceed analogously to [169–171]. Consider the theories $\mathcal{Q}$ and $S\mathcal{Q}$ separated by the duality defect as in Figure 2.4. A rotation by π along an axis in the defect, which we denote by R_π , swaps the theories $\mathcal{Q}$ and $S\mathcal{Q}$. By composing this operation with gauging of the one-form symmetry in all of space-time i.e. S we obtain a

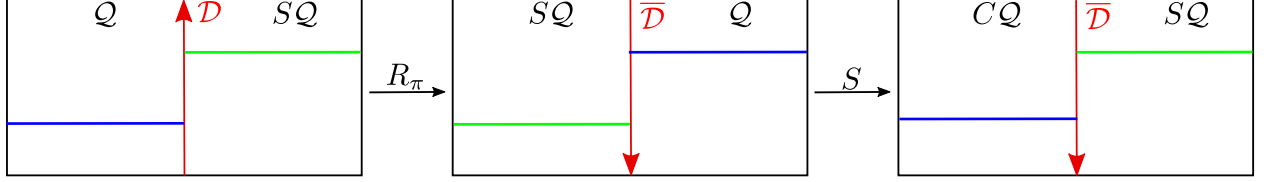


Figure 2.4: T symmetry of the duality defect separating phases $\mathcal{Q}$ and $S\mathcal{Q}$. R_π combined with gauging of one-form symmetry in all of spacetime maps the boundary conditions to themselves, but reverses the orientation of $\mathcal{D}$. Hence this composite operation defines a time-reversal symmetry T of the duality defect worldvolume.

symmetry T

$$\mathsf{T} \equiv S \cdot R_\pi . \quad (2.17)$$

By construction this leaves the boundary conditions invariant and therefore acts on the duality defect. Moreover since it reverses the defect orientation it is a time-reversal symmetry.⁶

We can also deduce the unitary symmetry T^2 from the fusion algebra of the duality defect (2.3). Within the wall the T symmetry defect is realized by intersecting $\mathcal{D}$ with itself.

Since this is a time-reversal symmetry, at each such intersection the orientations of the defects reverse (see Figure 2.5). If we now collide two such junctions we obtain the fusion of $\mathcal{D} \times \overline{\mathcal{D}}$ resulting in a condensation of one-form symmetry surfaces. Such surfaces come in two varieties:

- Surfaces parallel to the vertical duality defect $\mathcal{D}$ in Figure 2.5. These surfaces are simply absorbed by the duality defect. Indeed, $\mathcal{D}$ is characterized by Dirichlet boundary conditions (see Figure 2.3), and therefore absorbs one-form symmetry defects contained completely within it.
- One-form symmetry surfaces that intersect the vertical duality defect $\mathcal{D}$ in Figure 2.5.

These are non-trivial and give rise to abelian anyons within the duality defect.

6. More generally when the bulk dynamics is not invertible, we expect that the duality defect $\mathcal{D}$ still enjoys an antiunitary non-invertible symmetry analogous to those recently discussed in [99].

Taking into account the identification of generators, we see that a generator for the bulk one-form symmetry yields an abelian anyon $a^{1+(p)_N^{-1}}$ within $\mathcal{D}$.

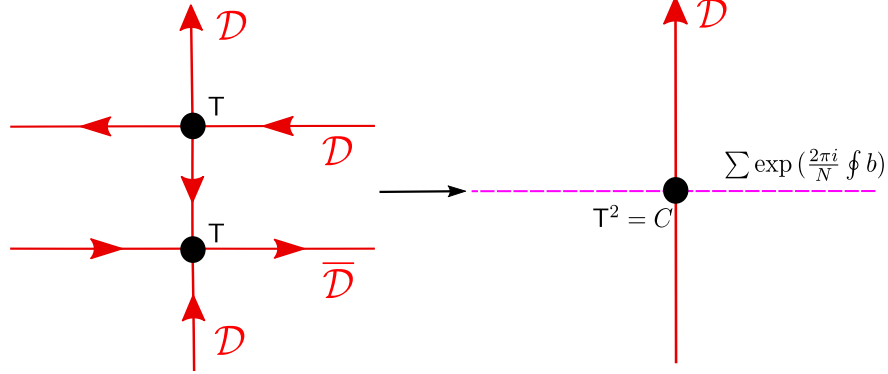


Figure 2.5: The $\mathbb{T}$ symmetry of the duality defect $\mathcal{D}$ is realized by intersecting duality defects (shown as black dots). At each such intersection the duality defect orientations reverse. Colliding two such junctions leads to the bulk fusion algebra of $\mathcal{D}$ with its orientation reversal $\overline{\mathcal{D}}$ resulting in a condensation of one-form symmetry surfaces shown in magenta. Within the duality defect worldvolume, this restricts to a condensation of abelian anyons that produces the charge conjugation symmetry C .

From these observations, we deduce that the operator $\mathbb{T}^2$ in $\mathcal{D}$ is defined by a condensate of abelian anyons [79]. For instance taking the surface supporting $\mathbb{T}$ to be a torus M we find:

$$\mathbb{T}^2 = \frac{1}{N} \sum_{\gamma \in H_1(M, \mathbb{Z}_N)} a^{1+(p)_N^{-1}}(\gamma) . \quad (2.18)$$

Although condensation operators like the one appearing above are frequently non-invertible, in fact in this special case the condensation is invertible and is identified with the charge conjugation symmetry C that maps a^r to a^{-r} .

To demonstrate this we note that as a consequence of (2.12), when N is odd, $1 + (p)_N^{-1}$ is co-prime to N .⁷ Therefore, the condensation defect above is equivalent to a sum over all

7. Using (2.12), we have $(1 + (p)_N^{-1})(2)_N^{-1}(1 - (p)_N^{-1}) = 1 \bmod N$ for odd N and therefore $1 + (p)_N^{-1}$ is co-prime to N .

abelian anyons:

$$\mathbb{T}^2 = \frac{1}{N} \sum_{\gamma \in H_1(M, \mathbb{Z}_N)} a(\gamma) . \quad (2.19)$$

Now we follow [79] and evaluate the action of $\mathbb{T}^2$ on a line a^s . Let $\alpha, \beta \in H_1(M, \mathbb{Z})$ denote a basis for the one-cycles on the torus with intersection form $\alpha \cap \beta = 1$. We can split a general line on a cycle γ as:

$$a(x\alpha + y\beta) = \exp\left(\frac{2\pi i p x y}{2N}\right) a^y(\beta) a^x(\alpha) , \quad (2.20)$$

where $x, y \in \mathbb{Z}$, and above we have used that the spin of the line a^r is fixed as:

$$h(a^r) = \frac{pr^2}{2N} . \quad (2.21)$$

We consider a geometry where the anyon a^s wraps the β cycle at the center of the solid torus whose boundary is the support M of $\mathbb{T}^2$. Then the lines wrapping α braid with the central anyon, while those wrapping β fuse with it. Whence using (2.20):

$$\begin{aligned} \mathbb{T}^2(a^s) &= \frac{1}{N} \sum_{x,y=0}^{N-1} \exp\left(\frac{2\pi i p x y}{2N}\right) a^y(\beta) a^x(\alpha) a^s(\beta) , \\ &= \frac{1}{N} \sum_{x,y=0}^{N-1} \exp\left(\frac{2\pi i p x y}{2N} + \frac{2\pi i p x s}{N}\right) (a^{y+s}) . \end{aligned} \quad (2.22)$$

The sum above over x now enforces

$$\frac{py}{2} + ps = 0 , \quad \text{mod } N . \quad (2.23)$$

And since N is odd and p and N are co-prime, we can solve for y to find simply:

$$\mathbb{T}^2(a^s) = a^{-s} . \quad (2.24)$$

This is exactly the expected action of charge conjugation on lines. In summary we have derived the algebra:⁸

$$\mathbb{T}^2 = C . \quad (2.25)$$

The symmetry $\mathbb{T}$ that we have identified provides another explanation of several of the features of the duality defect. As is evident from Figure 2.4, $\mathbb{T}$ exchanges the left and right bulk one-form symmetries. Since these couple differently to the defect as in (2.16), we conclude that $\mathbb{T}$ acts on the anyons of the defect theory $\mathcal{A}^{N,p}$ as:

$$\mathbb{T}(a) = a^{(p)N^{-1}} . \quad (2.26)$$

We can see that this is a consistent action of time-reversal by recalling that the spin of the anyon a^r in (2.21). Indeed, $\mathbb{T}$ changes the sign of the spin precisely when p obeys $p^2 = -1 \pmod{N}$. This reproduces the condition (2.8) necessary for the existence of a duality invariant SPT. However, now we see that it is a consequence of the $\mathbb{T}$ symmetry of the (2+1)d duality defect TQFT in agreement with analysis of [168].

Finally, let us comment on the chiral central charge c of the duality defect. This may be computed by summing over the spins of the distinct lines (2.21). Recalling that pN is even (and hence $\mathcal{A}^{N,p}$ is bosonic), the distinct lines correspond to a^r for $r = \{0, \dots, N-1\}$.

This yields:

$$\exp\left(\frac{2\pi ic}{8}\right) = \frac{1}{\sqrt{N}} \sum_{r=0}^{N-1} e^{\frac{2\pi ip}{2N} r^2} = G([+1], p, N) , \quad (2.27)$$

where the right-hand side is the quadratic Gauss sum for the intersection form of $\mathbb{CP}^2$.

Pragmatically, this is also the same phase factor that we encounter in (2.9) when we integrate out the dynamical field b .

8. For even N , charge conjugation is instead the condensation defect generated by a^2 . This matches (2.18) because (2.15) implies that $(1 + (p)_N^{-1})(1 - (p)_N^{-1}) = 2 \pmod{N}$ and therefore $1 + (p)_N^{-1}$ is even and $(1 + (p)_N^{-1})/2$ is co-prime to $N/2$.

As remarked above (and proven in (2.58)), when the conditions for the existence of the duality defect are met, namely when $p^2 = -1 \bmod N$ for odd N , the Gauss sum above is trivial and hence the chiral central charge vanishes modulo eight. In particular, this implies that up to possibly stacking with a properly quantized bosonic gravitational Chern-Simons term, we deduce that the chiral central charge of the duality defect $\mathcal{A}^{N,p}$ vanishes.

In fact, the condition $c = 0$ is also compatible with the time-reversal invariance of the duality defect, since $\mathbb{T}$ relates left and right moving chiral edge modes. Thus we see how the $\mathbb{T}$ symmetry of the duality defect explains and unifies many of its physical properties.

2.2.3 General Duality Invariant Invertible Phases

It is straightforward to extend the previous analysis to invertible phases invariant under compositions of gauging (S) and stacking (T) operations defined in section 2.1.1. Let us now consider the composite operation $T^{-p'}S$. In particular, the case $p' = \pm 1$ or $p' = \pm(1 + N)$, for even or odd N respectively give rise to the existence of triality defects studied in [81].

Acting on the general invertible theory defined in (2.6), the operation $T^{-p'}S$ leads to the action

$$\begin{aligned} (T^{-p'}SZ)[B] \\ = \lambda_N^{-1} \sum_{b \in H^2(X, \mathbb{Z}_N)} Z[b] \exp \left[\int_X \frac{2\pi i}{N} b \cup B - \frac{2\pi i p'}{2N} \mathcal{P}(B) \right]. \end{aligned} \quad (2.28)$$

Using the equations of motion for b , we can follow the same steps that we did to derive (2.9) to obtain

$$\begin{aligned} (T^{-p'}SZ)[B] \\ = G(I, p, N) \exp \left[- \int_X \frac{2\pi i \left(p' + (p)_{\gamma(N)N}^{-1} \right)}{2N} \mathcal{P}(B) \right], \end{aligned} \quad (2.29)$$

If the theory $T^{-p'}S\mathcal{Q}$ is equivalent to the original theory $\mathcal{Q}$ (up to a gravitational counter-term), then the dependence on B must match. Imposing this condition leads to the following theorem [81]:

Theorem 3. *There exists a bosonic SPT that is invariant under $T^{-p'}S$ gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry, if and only if there exists a solution p to*

$$p(p + p') = -1 \mod \gamma(N)N . \quad (2.30)$$

Analogously, for spin theories we have [81]:

Theorem 4. *There exists a spin SPT that is invariant under $T^{-p'}S$ gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry, if and only if there exists a solution p to*

$$p(p + p') = -1 \mod N . \quad (2.31)$$

We now restrict our attention to triality invariant SPTs. For even N , this means that the SPT is invariant under $T^{-1}S$ gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry. Such a (spin) SPT does not exist because there is no solution to the conditions (2.30) or (2.31) when $p' = 1$ and N is even. For odd N , $p' = -(1 + N)$ and such SPTs may exist.⁹

As in the case of SPTs invariant under the S operation, it is instructive to analyze the physics on the triality defect defined by doing $T^{-(1+N)}S$ gauging in half of the spacetime. The theory on the triality defect is the minimal Abelian TQFTs $\mathcal{A}^{N,p} \times \mathcal{A}^{N,N+1}$. The first minimal TQFT comes from S gauging the SPT in half of the spacetime while the second one is introduced to cancel the anomaly inflow from the $T^{-(1+N)}$ operation on the right. We can evaluate the chiral central charge of the TQFT on the defect by summing over lines

9. Inspecting equation (2.30), we deduce that solutions exist if and only if -3 is a quadratic residue modulo N [81].

as in (2.27) yielding:

$$\exp\left(\frac{2\pi ic}{8}\right) = G([+1], p, N)G([+1], 1 + N, N) . \quad (2.32)$$

The first Gauss sum is the same phase factor that shows up in (2.29). The Gauss sum in (2.32) does not equal to unity in general but it is independent of the integer p characterizing the triality invariant SPT. Using (2.61), we get

$$\exp\left(\frac{2\pi ic}{8}\right) = \begin{cases} 1 & N \equiv 1 \pmod{4} , \\ -1 & N \equiv 3 \pmod{4} . \end{cases} \quad (2.33)$$

Since the Gauss sum is always an eighth root of unity the chiral central charge of the symmetry defect may be cancelled by stacking with a well quantized $\text{spin}_{\mathbb{C}}$ gravitational Chern-Simons term.

Therefore, for triality invariant SPT phases, the symmetry defects implementing the gauging operation have a vanishing chiral central charge modulo the addition of invertible (2+1)d gravitational theories.

2.3 Topologically Ordered Phases

In this section, we construct examples of (3+1)d duality symmetry preserving TQFTs which are invariant under combinations of S and T gauging operations of a $\mathbb{Z}_N^{(1)}$ one-form symmetry. Without loss of generality, we consider only TQFTs that have a unique local vacuum. After these explicit constructions, we prove that for values of N that are not realized by these explicit examples, such TQFTs do not exist.

2.3.1 Examples

As a starting example, consider the case when $N = k^2$ is a perfect square. Then, a $\mathbb{Z}_k$ gauge theory is invariant under gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry that couples to the background gauge field B as

$$\begin{aligned} Z[B] &= \lambda_k^{-1} \sum_{b \in H^2(X, \mathbb{Z}_k)} \exp \left[\frac{2\pi i}{k} \int_X b \cup B \right] \\ &= \lambda_k \delta(B \bmod k) , \end{aligned} \tag{2.34}$$

where $\lambda_k = \sqrt{|H^2(X, \mathbb{Z}_k)|}$. The $\mathbb{Z}_N^{(1)}$ one-form symmetry, in particular its $\mathbb{Z}_k^{(1)}$ subgroup, does not act faithfully on line operators. It is straightforward to check that gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry leads to the same partition function as the original $\mathbb{Z}_k$ gauge theory. Using the half space gauging construction, we find that the duality defect factorizes into the product of Dirichlet boundary conditions for the $\mathbb{Z}_k$ two-form gauge fields on the two sides.

More generally, consider the case when $N = k^2 \ell$ for some integers k and ℓ , and -1 is a quadratic residue of ℓ . Then, there exists a (spin) TQFT that is invariant under gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry. Let p be a solution to $p^2 + 1 = 0 \bmod \ell$. In particular, this implies that p and ℓ are co-prime.

Let us first consider the case when ℓ is odd. Then solutions exist if and only if ℓ is a product of Pythagorean primes. The TQFT invariant under gauging is a $\mathbb{Z}_k$ gauge theory

stacked with a $\mathbb{Z}_N^{(1)}$ one-form SPT described by the following partition function¹⁰

$$\begin{aligned}
Z[B] &= \lambda_k^{-1} \sum_{b \in H^2(X, \mathbb{Z}_k)} \exp \left[2\pi i \int_X \left(\frac{p}{2N} \mathcal{P}(B) + \frac{1}{k} b \cup B \right) \right] \\
&= \lambda_k \delta(B \bmod k) \exp \left[\frac{2\pi i p}{2\ell} \int_X \mathcal{P} \left(\frac{B}{k} \right) \right].
\end{aligned} \tag{2.35}$$

Gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry, we obtain a partition function identical to (2.35)

$$\begin{aligned}
SZ[B] &= \lambda_{k\ell}^{-1} \sum_{b \in H^2(X, \mathbb{Z}_{k\ell})} \exp \left[2\pi i \int_X \left(\frac{p}{2\ell} \mathcal{P}(b) + \frac{1}{k\ell} b \cup B \right) \right] \\
&= \lambda_k \delta(B \bmod k) G(I, p, \ell) \exp \left[-\frac{2\pi i (p)_\ell^{-1}}{2\ell} \int_X \mathcal{P} \left(\frac{B}{k} \right) \right] \\
&= \lambda_k \delta(B \bmod k) \exp \left[\frac{2\pi i p}{2\ell} \int_X \mathcal{P} \left(\frac{B}{k} \right) \right] = Z[B].
\end{aligned} \tag{2.36}$$

In the second equality, we used (2.9) for the odd ℓ case since p and ℓ are co-prime, and we pick an even $(p)_\ell^{-1}$. In the last equality, we simplified the expression using the fact that $G(I, p, \ell) = 1$ when $p^2 + 1 = 0 \bmod \ell$ and $(p)_\ell^{-1}/2 = -p/2 \bmod \ell$. Note that the TQFT (2.35), as a bosonic TQFT, is already invariant under gauging.

Now consider the case when ℓ is even. Then solutions exist if $\ell/2$ is odd and $\ell/2$ is a product of Pythagorean primes. Again, let us consider the theory (2.35). Using (2.9) for the even ℓ case, gauging leads to

$$\begin{aligned}
SZ[B] &= \lambda_k \delta(B \bmod k) G(I, p, \ell) \\
&\quad \times \exp \left[-\frac{2\pi i (p)_{2\ell}^{-1}}{2\ell} \int_X \mathcal{P} \left(\frac{B}{k} \right) \right].
\end{aligned} \tag{2.37}$$

10. For the expressions to be well-defined, we pick an even solution to $p^2 + 1 = 0 \bmod \ell$. If the solution p is odd, we can get an even solution by redefining p to $p + \ell$.

As a spin TQFT, the theory is exactly invariant under gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry since $(p)_{2\ell}^{-1} = -p \bmod \ell$ and $G(I, p, \ell) = 1$ on spin manifolds. However, as a bosonic TQFT, the theory cannot be invariant under gauging because there is no solution to $(p)_{2\ell}^{-1} = -p \bmod 2\ell$.

2.3.2 Constraints on TQFTs

So far, we have explicitly constructed (spin) TQFTs that have a unique local vacuum and are invariant under gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry for some particular N . We now prove that such TQFTs do not exist for the values of N that have not been realized in the last subsection. This means that

Theorem 5. *There exists a (spin) TQFT, that has a unique local vacuum and is invariant under gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry up to a gravitational counterterm $e^{i\Omega(X)}$, if and only if $N = k^2\ell$ for some integers k and ℓ and -1 is a quadratic residue of ℓ (or equivalently ℓ is a product of Pythagorean primes up to a factor of two). If such a (spin) TQFT exists, $e^{i\Omega(X)} = 1$ for every simply-connected smooth spin 4-manifold X .*

We now present a brief sketch of the proof, while the details are fleshed out in supplementary section 2.5.3.

In general, the $\mathbb{Z}_N^{(1)}$ one-form symmetry can act unfaithfully on line operators *i.e.* some symmetry surface operators link trivially with all line operators. Denote the unfaithful subgroup by $\mathbb{Z}_M^{(1)}$. Then, we have $N = Mk$ for some integers M and k . If the theory is a TQFT, after gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry, the $\hat{\mathbb{Z}}_k^{(1)}$ subgroup of the dual $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry necessarily act unfaithfully on line operators because their symmetry surface operators admit topological boundary conditions provided by the topological line operators charged under the original $\mathbb{Z}_N^{(1)}$ one-form symmetry. Using these topological boundary conditions, we can unlink the dual $\hat{\mathbb{Z}}_k^{(1)}$ symmetry operators and any line operator. Since the TQFT is invariant under gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry, the

unfaithful subgroup of the dual $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry should be $\hat{\mathbb{Z}}_M^{(1)}$, which should include the unfaithful $\hat{\mathbb{Z}}_k^{(1)}$ one-form symmetry as a subgroup. This implies that M is divisible by k and $N = k^2 \ell$ for some integer ℓ .

In a TQFT, symmetry operators of an unfaithful one-form symmetry admit topological boundary conditions [162]. In our case, this means that we can cut open the symmetry surface operators of the unfaithful $\mathbb{Z}_M^{(1)}$ one-form symmetry and shrink them to local operators. Consider placing the TQFT on $X = S^2 \times S^2$ and turning on the background gauge field $B = (B_1, B_2) \in H^2(S^2 \times S^2, \mathbb{Z}_M) = \mathbb{Z}_M^2$ for the unfaithful $\mathbb{Z}_M^{(1)}$ one-form symmetry. By shrinking the symmetry operators to local operators at their intersections, we argue that the partition function takes the form

$$Z[B_1, B_2] = Z[0, 0] \exp \left(\frac{2\pi i p}{M} B_1 B_2 \right) . \quad (2.38)$$

The partition function $Z[0, 0]$ on $S^2 \times S^2$ is a positive number because of unitarity [162]. Using the fact that the TQFT is invariant under gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry (upto a gravitational counterterm), we show that the partition function (2.38) obeys a constraint

$$\begin{aligned} & \frac{1}{k\ell} \sum_{b_{1,2}=1}^{k\ell} Z[b_1, b_2] \exp \left(\frac{2\pi i}{\ell} (b_1 C_2 + b_2 C_1) \right) \\ &= \frac{1}{k} \sum_{b_{1,2}=1}^k Z[\ell b_1 + C_1, \ell b_2 + C_2] \times e^{i\Omega(S^2 \times S^2)} , \end{aligned} \quad (2.39)$$

where $C = (C_1, C_2) \in H^2(S^2 \times S^2, \mathbb{Z}_\ell) = \mathbb{Z}_\ell^2$ and $e^{i\Omega(S^2 \times S^2)}$ is a phase independent of (C_1, C_2) . First line of the equality is the partition function of the TQFT after gauging the unfaithful $\mathbb{Z}_M^{(1)}$ one-form symmetry with the dual $\tilde{\mathbb{Z}}_\ell^{(1)}$ background gauge field C turned on. The same partition function can also be obtained from the $\mathbb{Z}_N^{(1)}$ -gauged theory by gauging the $\hat{\mathbb{Z}}_k^{(1)}$ subgroup of the dual $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry with the background gauge field C

for the dual $\hat{\mathbb{Z}}_M^{(1)}/\hat{\mathbb{Z}}_k^{(1)}$ one-form symmetry turned on. Since the $\mathbb{Z}_N^{(1)}$ -gauged TQFT share the same partition function as the original TQFT up to a gravitation counterterm $e^{i\Omega(X)}$,

$$Z[B] = (SZ)[B]e^{-i\Omega(X)} , \quad (2.40)$$

we have the equality (2.39).

In the end, the partition function (2.38) together with the constraint (2.39) implies that -1 is a quadratic residue of ℓ . We also learned that $e^{i\Omega(S^2 \times S^2)} = 1$. Similar reasoning holds on every simply-connected smooth spin 4-manifold X . This leads to the same constraint on ℓ and gives $e^{i\Omega(X)} = 1$. This concludes the proof for theorem 5.

2.3.3 General Duality Invariant TQFTs

It is straightforward to extend the results to more general non-invertible symmetries associated to the invariance of $T^{-p'}S$ gauging, including the triality symmetry.

Generalizing the argument used in the proof of theorem 5, in supplementary section 2.5.3 we prove the following theorem:

Theorem 6. *There exists a (spin) TQFT, that has a unique local vacuum and is invariant under $T^{-p'}S$ gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry up to a gravitational counterterm $e^{i\Omega(X)}$, if and only if $N = k^2\ell$ for some integers k and ℓ and there exists a solution p to*

$$p(p + p') + 1 = 0 \text{ mod } \ell . \quad (2.41)$$

If such a (spin) TQFT exists, $e^{i\Omega(X)} = 1$ for every simply-connected smooth spin 4-manifold X .

When N obeys the condition, we can explicitly construct examples of spin TQFTs that are invariant under $T^{-p'}S$ gauging of the $\mathbb{Z}_N^{(1)}$ one-form symmetry. This spin TQFT is a $\mathbb{Z}_k$

gauge theory stacked with a $\mathbb{Z}_N^{(1)}$ one-form SPT as in (2.35), where the parameter p is given by the solution of (2.41). Using (2.41), it is straightforward to check that the theory as a spin TQFT is invariant under gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry and $e^{i\Omega(X)} = 1$ on simply-connected smooth spin 4-manifold X .

2.4 Lattice Examples

As an application of the analysis of the previous section, let us study (3+1)d $\mathbb{Z}_N$ lattice gauge theory in the Villain formulation [174], which is defined on a 4d Euclidean hyper-cube lattice. The action is

$$S_{\mathcal{Q}} = \frac{1}{2g^2} \sum_{\text{plaquette}} \left(\Delta m^{(1)} - N n^{(2)} \right)^2, \quad (2.42)$$

where $m^{(1)}$ is the integer one-form gauge field on each link, while $n^{(2)}$ is the integer two-form gauge field on each plaquette. There is an integer gauge symmetry

$$\begin{aligned} m^{(1)} &\rightarrow m^{(1)} + \Delta k^{(0)} + N k^{(1)}, \\ n^{(2)} &\rightarrow n^{(2)} + \Delta k^{(1)}, \end{aligned} \quad (2.43)$$

that effectively makes the one-form gauge field a $\mathbb{Z}_N$ gauge field. This theory has an electric $\mathbb{Z}_N^{(1)}$ one-form global symmetry that shifts $m^{(1)}$ by a flat integer gauge field. By performing a Poisson resummation on $n^{(2)}$ in (2.42), we obtain a dual description of the theory in terms of an integer gauge field $\hat{n}^{(2)}$ on the plaquettes of the dual lattice. After introducing the Stueckelberg fields $\tilde{m}^{(1)}$ and $\tilde{n}^{(2)}$, we obtain

$$\begin{aligned} S_{S\mathcal{Q}} &= \frac{2\pi i}{N} \sum_{\text{link}} m^{(1)} \Delta \tilde{n}^{(2)} + \\ &\quad \frac{4\pi^2 g^2}{2N^2} \sum_{\text{plaquette}} \left(\Delta \tilde{m}^{(1)} - N \tilde{n}^{(2)} - \hat{n}^{(2)} \right)^2. \end{aligned} \quad (2.44)$$

By comparing the second term in (2.44) with (2.42), we see that the duality maps the theory at coupling g^2 to the one at coupling $N^2/4\pi^2 g^2$ with the electric $\mathbb{Z}_N^{(1)}$ one-form symmetry being gauged [175–180]. At the self-dual coupling $g_*^2 = N/2\pi$, this implies that the theory is invariant under gauging the electric $\mathbb{Z}_N^{(1)}$ one-form symmetry and therefore has a non-invertible duality symmetry [77]. Consequently, we can apply theorem 5 to infer that the $\mathbb{Z}_N$ lattice gauge theory at the self-dual coupling should either be gapless or spontaneously break the duality symmetry unless N is of the form $N = k^2\ell$, with -1 being a quadratic residue modulo ℓ . All values of $N \leq 60$ where duality invariant SPTs or TQFTs are not ruled out are listed in Table 2.1.

The schematic phase diagram of the $\mathbb{Z}_N$ lattice gauge theory as a function of N and coupling g is summarized in Figure 2.6, with the vertical line at $g_*^{-2}N = 2\pi$ demarcating the self-dual coupling. This phase diagram is supported by Monte-Carlo simulations [181]. When $N \leq 4$, at the self-dual coupling, the duality symmetry is spontaneously broken leading to two local vacua: one is trivial and preserves the $\mathbb{Z}_N^{(1)}$ electric one-form symmetry, and the other one supports a deconfined $\mathbb{Z}_N$ TQFT that spontaneously breaks the electric $\mathbb{Z}_N^{(1)}$ one-form symmetry. See supplementary section 2.5.5 for more discussion regarding the spontaneous breaking of non-invertible symmetries. On the other hand, when $N \geq 5$, the duality symmetry is preserved and the theory flows to the gapless Maxwell theory ($U(1)$ gauge theory) at a particular coupling

$$S_{\text{Maxwell}} = \frac{1}{2e^2} \int F \wedge \star F, \quad e^2 = \frac{2\pi}{N}. \quad (2.45)$$

The coupling above is fixed by matching the duality symmetry across the RG flow. The Maxwell theory at $e^2 = 2\pi/N$ is invariant under gauging the $\mathbb{Z}_N^{(1)}$ subgroup of the $U(1)^{(1)}$ electric one-form symmetry thanks to the electromagnetic duality [77]. Thus, the advocated phase diagram is consistent with the constraints stated in theorem 5.

More generally, one can consider a broader class of lattice gauge theories by introducing a θ

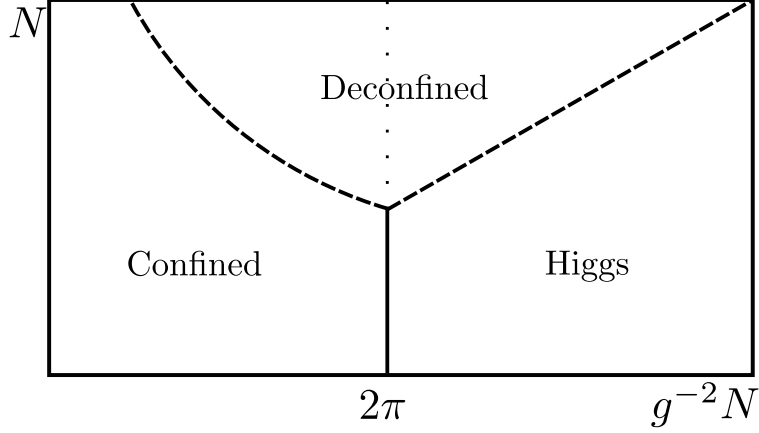


Figure 2.6: Schematic Phase Diagram of $\mathbb{Z}_N$ Lattice Gauge theory. Solid and dashed lines represent first order phase transitions. The vertical line at coupling $g_*^2 = N/2\pi$ is duality invariant. At weak coupling, the theory is in the Higgs phase *i.e.* it flows to a $\mathbb{Z}_N$ TQFT in the IR and at strong coupling, the theory is in the confining phase *i.e.* the vacuum is trivial. For $N \leq 4$, the Higgs and confining phases are separated by a first order phase transition at the self-dual coupling $g_*^2 = N/2\pi$ where the duality symmetry is spontaneously broken. For $N \geq 5$, the two phases are separated by an intermediate gapless coulomb phase which flows to a Maxwell theory and at the self-dual coupling $g_*^2 = 2\pi/N$ the IR Maxwell theory has $e^2 = 2\pi/N$.

term on the lattice. One concrete realization is the Cardy-Rabinovici model [182, 183] or its generalization [184, 185], which is a $U(1)$ lattice gauge theory that couples to charge N electric matter and charge 1 magnetic matter.¹¹ The lattice model is parameterized by a complex coupling τ . We will work with a formal continuum description of the Cardy-Rabinovici model given in [187].¹² In the formal continuum description, the Cardy-Rabinovici model is described as a continuum $U(1)$ gauge theory

$$\frac{1}{2e^2} \int F \wedge \star F + \frac{iN\theta}{8\pi^2} \int F \wedge F, \quad (2.46)$$

at coupling $\tau = 2\pi i/Ne^2 + \theta/2\pi$ with a sum over insertions of charge N Wilson lines and

11. For simplicity, we consider only a version of the Cardy-Rabinovici model where both bosonic and fermionic matter is present. Otherwise, the duality group in general differs from $SL(2, \mathbb{Z})$ [186].

12. We can also work with the self-dual lattice models constructed in [184, 185] which have the $SL(2, \mathbb{Z})$ duality as an exact duality on the lattice.

charge 1 't Hooft lines. The sum is weighted by the matter action. In the continuum approximation, this theory has a $SL(2, \mathbb{Z})$ duality generated by the $\mathbb{S}$ and $\mathbb{T}$ duality transformation. The $\mathbb{S}$ duality maps $\tau \rightarrow -1/\tau$ and swaps the electric and magnetic matter. Since the charges of electric and magnetic matter are different, the duality maps the theory at coupling τ to the one at $-1/\tau$ with the $\mathbb{Z}_N^{(1)}$ electric one-form symmetry gauged. The $\mathbb{T}$ duality maps $\tau \rightarrow \tau + 1$ and turns the magnetic matter to dyonic matter with magnetic charge 1 and electric charge N .

The model has a duality symmetry at $\tau = i$, which leads to the same constraint as in the (3+1)d $\mathbb{Z}_N$ lattice gauge theory at the self-dual coupling.

More interestingly, at $\tau_* = e^{\pi i/3}$, because of the $\mathbb{S}$ duality the model is invariant under the ST^{-1} gauging of the electric $\mathbb{Z}_N^{(1)}$ one-form symmetry and therefore has a triality symmetry.¹³ According to theorem 6, the model either is gapless or spontaneously breaks the triality symmetry unless $N = k^2 \ell$ for some integers k and ℓ such that there exists a solution to $p(p+1) + 1 = 0 \pmod{\ell}$. Even if N obeys the condition, we can still exclude the possibility of triality invariant TQFTs using the mixed triality-gravity anomaly. In [188], Hayashi and Tanizaki showed that on a spin manifold X , the model at the complex coupling $\tau_* = e^{i\pi/3}$ has a mixed triality-gravity anomaly given by

$$Z_{ST^{-1}\mathcal{Q}}[X] = Z_{\mathcal{Q}}[X] \exp \left[-\frac{i\pi}{3} \sigma(X) \right]. \quad (2.47)$$

In particular, we can pick X to be the simply-connected smooth spin 4-manifold $K3$ whose signature is -16 . We then have $Z_{ST^{-1}\mathcal{Q}}[K3]/Z_{\mathcal{Q}}[K3] = e^{-2\pi i/3}$. However, theorem 6 states that any triality invariant TQFT with a unique local vacuum has

$Z_{ST^{-1}\mathcal{Q}}[X]/Z_{\mathcal{Q}}[X] = 1$ on simply-connected smooth spin 4 manifold X . Therefore, we can deduce that the theory at $\tau_* = e^{i\pi/3}$ cannot flow to a triality symmetry preserving TQFT. Note that this is a stronger result than that of [188], wherein it is shown that the theory

13. If the theory $\mathcal{Q}$ is invariant under ST^{-1} operation, the theory $T^{-1}\mathcal{Q}$ is invariant under $T^{-1}S$ operation.

cannot be in a SPT phase.

A heuristic calculation of the free-energy of dyons [182] suggests the following scenario for the Cardy-Rabinovici model at $\tau_* = e^{\pi i/3}$. When N is small, the triality symmetry is spontaneously broken leading to three vacua: a Higgs vacuum (a $\mathbb{Z}_N$ TQFT that spontaneously breaks the $\mathbb{Z}_N^{(1)}$ one-form symmetry), a confined vacuum (a trivial vacuum that preserves the $\mathbb{Z}_N^{(1)}$ one-form symmetry) and an oblique confined vacuum (a $\mathbb{Z}_N^{(1)}$ one-form SPT) [189]. The triality symmetry permutes these three vacua. When N is large, the theory flows to the triality invariant gapless Maxwell theory. By matching the triality symmetry, we can fix the coupling of the Maxwell theory to be the triality symmetric point [81]

$$S_{\text{Maxwell}} = \frac{N}{4\pi} \frac{\sqrt{3}}{2} \int F \wedge \star F + \frac{N}{4\pi} \frac{i}{2} \int F \wedge F . \quad (2.48)$$

These two scenarios are both consistent with the constraint stated in theorem 6.

Acknowledgements

We thank K. Ohmori, and R. Thorngren for discussions. AA is supported by Yoichiro Nambu Graduate Fellowship courtesy of Department of Physics, University of Chicago. CC is supported by the US Department of Energy DE-SC0021432 and the Simons Collaboration on Global Categorical Symmetries. HTL is supported in part by a Croucher fellowship from the Croucher Foundation, the Packard Foundation and the Center for Theoretical Physics at MIT.

2.5 Supplementary Material

2.5.1 Gauss Sums of Intersection Forms

Let X be a smooth simply connected 4-manifold, with the intersection form

$I : H_2(X; \mathbb{Z}) \times H_2(X; \mathbb{Z}) \rightarrow \mathbb{Z}$. The intersection form I is a bi-linear symmetric form. We

are interested in computing its normalized Gauss sum given by

$$G(I, p, N) = \frac{1}{N^{\text{rk}(I)/2}} \sum_{x_i=1}^N \exp \left(\frac{2\pi i p}{2N} x_i I_{ij} x_j \right), \quad (2.49)$$

where $\text{rk}(I)$ is the rank of I . The normalized Gauss sum can be interpreted as the partition function of a $\mathbb{Z}_N$ two-form gauge theory when pN is even (see supplementary section 2.5.2 for more details). In this supplementary section, we will focus on the case that has even pN and p co-prime to N .

Note that the intersection form I may be definite or indefinite. Let E_8 be the intersection form of the E_8 manifold and H be the intersection form of $S^2 \times S^2$, then by combining Donaldson's theorem with Hasse-Minkowski theorem we have the following result [190]

Theorem. *The only bi-linear symmetric forms that can be realized as intersection forms of a smooth simply connected 4-manifold are*

$$\oplus a[+1], \oplus b[-1], \oplus \pm mE_8 \oplus nH; \quad (2.50)$$

where a, b, m, n are positive integers.

By the theorem above, the task of computing the Gauss sum (2.49) for a general intersection form reduces to computing the Gauss sum for $[+1]$, $[-1]$, E_8 and H .

To prepare for the calculation below, define

$$\varepsilon_N = \begin{cases} 1 & \text{if } N \equiv 1 \pmod{4} \\ i & \text{if } N \equiv 3 \pmod{4} \end{cases}. \quad (2.51)$$

Let us also review the Jacobi symbol $(\frac{a}{n})$, which is defined for any integer a and any positive odd integer n . The Jacobi symbol factorizes into the product of the Legendre

symbols

$$\left(\frac{a}{n}\right) = \left(\frac{a}{s_1}\right)^{r_1} \left(\frac{a}{s_2}\right)^{r_2} \cdots \left(\frac{a}{s_k}\right)^{r_k}, \quad (2.52)$$

where $n = s_1^{r_1} s_2^{r_2} \cdots s_k^{r_k}$ is the prime factorization of n . The Legendre symbols is defined for any integer a and any odd prime s as

$$\left(\frac{a}{s}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{s} \\ 1 & \text{if } a \not\equiv 0 \pmod{s} \text{ and } a \equiv x^2 \pmod{s} \text{ for some integer } x \\ -1 & \text{if } a \not\equiv 0 \pmod{s} \text{ and there is no such } x \end{cases} \quad (2.53)$$

The Gauss sum for $I = [+1]$ is [191]

$$G([+1], p, N) = \begin{cases} \varepsilon_N \left(\frac{p/2}{N}\right) & \text{odd } N \\ e^{\frac{\pi i}{4}} \varepsilon_p^{-1} \left(\frac{2N}{p}\right) & \text{even } N \end{cases}. \quad (2.54)$$

It is valued in 4th root of unity when N is odd and 8th root of unity when N is even.

Taking its complex conjugate we obtain the Gauss sum for $I = [-1]$. For H , the Gauss sum is unity

$$G(H, p, N) = \frac{1}{N} \sum_{x_1, x_2=1}^N \exp\left(\frac{2\pi i p}{2N} 2x_1 x_2\right) = 1. \quad (2.55)$$

For E_8 , the Gauss sum is also unity [188]

$$G(E_8, p, N) = 1. \quad (2.56)$$

Since $[+1]$, $[-1]$, H and E_8 have signature $+1$, -1 , 0 and $+8$ respectively, the normalized Gauss sum for the intersection form I of a smooth simply connected 4-manifold X can be

summarized a function of the signature $\sigma(X)$ of the 4-manifold

$$G(I, p, N) = \begin{cases} \left[\varepsilon_N \left(\frac{p/2}{N} \right) \right]^{\sigma(X)} & \text{odd } N \\ \left[e^{\frac{\pi i}{4}} \varepsilon_p^{-1} \left(\frac{2N}{p} \right) \right]^{\sigma(X)} & \text{even } N \end{cases} . \quad (2.57)$$

Since the normalized Gauss sum is valued in 8th root of unity and by Rokhlin's theorem the signature of the intersection form I of a spin manifold is divisible by 16, the Gauss sum is always unity on a spin-manifold.

We now study two special cases. First, consider the case when $p^2 \equiv -1 \pmod{N}$ and N is odd. This situation arises only if all the prime factors of N are 1 modulo 4, and thus $\varepsilon_N = 1$. Since $p/2 \equiv \left[(2)_N^{-1}(p+1) \right]^2 \pmod{N}$, we have $\left(\frac{p/2}{N} \right) = 1$. Combining these two facts, we find that the normalized Gauss sum is unity:

$$G(I, p, N) = 1 \text{ when } p^2 \equiv -1 \pmod{N} \text{ and } N \text{ is odd} . \quad (2.58)$$

Next, consider the case when $p(p+1) \equiv -1 \pmod{N}$. This is possible only when N is odd.

Using the following properties of the Jacobi symbol

$$\left(\frac{ab}{n} \right) = \left(\frac{a}{n} \right) \left(\frac{b}{n} \right) , \quad \left(\frac{a}{n} \right)^2 = 1 , \quad \left(\frac{1}{n} \right) = 1 , \quad (2.59)$$

we find that

$$\left(\frac{p/2}{N} \right) = \left(\frac{2}{N} \right) \left(\frac{p}{N} \right) = \left(\frac{2}{N} \right) \left(\frac{(p+1)^2}{N} \right) = \left(\frac{2}{N} \right) . \quad (2.60)$$

Therefore,

$$G(I, p, N) = \left[\varepsilon_N \left(\frac{2}{N} \right) \right]^{\sigma(X)} \text{ when } p(p+1) \equiv -1 \pmod{N} . \quad (2.61)$$

2.5.2 Partition Function of $\mathbb{Z}_N$ 2-form Gauge Theory

In this supplementary section, we compute the partition function of a (3+1)d $\mathbb{Z}_N$ 2-form gauge theory in the presence of the background gauge field B for the $\mathbb{Z}_N^{(1)}$ one-form symmetry. For simplicity, we assume the underlying manifold X is simply connected and thus $H^2(X, \mathbb{Z})$ has no torsion classes.

The partition functions of a (3+1)d $\mathbb{Z}_N$ 2-form gauge theory is

$$Z[B] = \frac{1}{\sqrt{|H^2(X, \mathbb{Z}_N)|}} \sum_{b \in H^2(X, \mathbb{Z}_N)} \exp \left[2\pi i \int_X \left(\frac{p}{2N} \mathcal{P}(b) + \frac{1}{N} b \cup B \right) \right] . \quad (2.62)$$

When N is even, $p \sim p + 2N$ and $\mathcal{P}(b) : H^2(X, \mathbb{Z}_N) \rightarrow H^4(X, \mathbb{Z}_{2N})$ is the Pontryagin square map. When N is odd, $p \sim p + 2N \in 2\mathbb{Z}$ and $\mathcal{P}(b) = b \cup b : H^2(X, \mathbb{Z}_N) \rightarrow H^4(X, \mathbb{Z}_N)$. In both cases pN is even. Let us define

$$L \equiv \gcd(p, N), \quad q \equiv \frac{p}{L}, \quad K \equiv \frac{N}{L}, \quad J \equiv \frac{qKL}{2} . \quad (2.63)$$

If q and K are both odd, L has to be even because $pN = qKL^2$ is even. Thus, J is always an integer. To evaluate the partition function, we split b into a $\mathbb{Z}_L$ cochain b_0 and $\mathbb{Z}_K$ cocycle b_1 :

$$b = Kb_0 + b_1 . \quad (2.64)$$

The gauge symmetries are

$$b_0 \rightarrow b_0 + \delta\alpha_0 + L\beta_0 - \beta_1, \quad b_1 \rightarrow b_1 + \delta\alpha_1 + K\beta_1 . \quad (2.65)$$

The cocycle conditions are

$$\delta b_1 = 0 \bmod K, \quad \delta b_0 = -\delta b_1 / K \bmod L \equiv -\text{Bock}(b_1) \bmod L ; \quad (2.66)$$

where $\text{Bock}(b_1)$ is the Bockstein homomorphism applied to the cocycle b_1 . The second cocycle condition ensures that $\delta b = \delta(Kb_0 + b_1) = 0 \bmod N$. Now substituting (2.64) into (2.62), we obtain

$$Z[B] = \frac{1}{\sqrt{|H^2(X, \mathbb{Z}_N)|}} \times \sum_{b_1 \in H^2(X, \mathbb{Z}_K)} \exp \left[2\pi i \int_X \left(\frac{q}{2K} \mathcal{P}(b_1) + \frac{1}{N} b_1 \cup B \right) \right] \times \sum_{\substack{b_0 \in C^2(X, \mathbb{Z}_L) \\ \delta b_0 = -\text{Bock}(b_1)}} \exp \left[2\pi i \int_X \left(\frac{qK}{2} \mathcal{P}(b_0) + \frac{1}{L} b_0 \cup B \right) \right]. \quad (2.67)$$

The second sum can be simplified using the Wu formula, $\mathcal{P}(b) = b \cup \nu_2(X) \bmod 2$, where $\nu_2(X) = w_2(X) + w_1(X) \cup w_1(X)$ is the second Wu class of the underlying manifold X and $w_i(X)$ is the i th Stiefel-Whitney class. Summing over b_0 then gives

$$Z[B] = \frac{|H^2(X, \mathbb{Z}_L)|}{\sqrt{|H^2(X, \mathbb{Z}_N)|}} \times \sum_{b_1 \in H^2(X, \mathbb{Z}_K)} \exp \left[2\pi i \int_X \left(\frac{q(1-K)}{2K} \mathcal{P}(b_1) + \frac{1}{K} b_1 \cup \frac{1}{L} (B + J\nu_2(X)) \right) \right] \delta((B + J\nu_2(X)) \bmod L). \quad (2.68)$$

We reorganized the exponents using the Wu formula so that each term is independently well-defined. The first term is well-defined because $qK(1-K)$ is always even. The second term is well-defined when the delta function is non-zero. We can complete the square in the exponent by a change of variable

$$\tilde{b}_1 = b_1 + (1+K)(q)_{\gamma(K)K}^{-1} \frac{B + J\nu_2}{L}, \quad (2.69)$$

where the notation $(\alpha)_\beta^{-1}$ denotes the inverse of α in the ring $\mathbb{Z}_\beta$ for co-prime integers α

and β , and

$$\gamma(K) = \begin{cases} 1, & \text{odd } K \\ 2, & \text{even } K \end{cases}. \quad (2.70)$$

When $q = 0$, we define $(q)_{\gamma(K)K}^{-1} = 0$. The exponents in the sum now becomes

$$\exp \left[2\pi i \int_X \left(\frac{q(1-K)}{2K} \mathcal{P}(\tilde{b}_1) - \frac{(1-K)(q)_{\gamma(K)K}^{-1}}{2K} \mathcal{P}\left(\frac{B + J\nu_2}{L}\right) \right) \right]; \quad (2.71)$$

The remaining sum over $\tilde{b}_1$ is related to the normalized Gauss sum for the intersection form I via

$$\sum_{\tilde{b}_1 \in H^2(X, \mathbb{Z}_K)} \exp \left[2\pi i \int_X \frac{q(1-K)}{2K} \mathcal{P}(\tilde{b}_1) \right] = \sqrt{|H^2(X, \mathbb{Z}_K)|} G(I, q(1-K), K). \quad (2.72)$$

Summing over b_1 then leads to

$$Z[B] = \sqrt{|H^2(X, \mathbb{Z}_L)|} G(I, q(1-K), K) \times \exp \left[-2\pi i \int_X \frac{(1-K)(q)_{\gamma(K)K}^{-1}}{2K} \mathcal{P}\left(\frac{B + J\nu_2}{L}\right) \right] \delta((B + J\nu_2) \bmod L). \quad (2.73)$$

We now discuss two situations when the formula simplifies. When the underlying manifold X has a trivial second Wu class $\nu_2(X)$, (2.73) simplifies to

$$Z[B] = \sqrt{|H^2(X, \mathbb{Z}_L)|} G(I, q, K) \delta(B \bmod L) \exp \left[-2\pi i \int_X \frac{(q)_K^{-1}}{2K} \mathcal{P}\left(\frac{B}{L}\right) \right]. \quad (2.74)$$

In the special case of $L = 1$, (2.73) simplifies to

$$Z[B] = G(I, p, N) \exp \left[- \int_X \frac{2\pi i (p)_{\gamma(N)N}^{-1}}{2N} \mathcal{P}(B) \right]. \quad (2.75)$$

When N is odd, we pick an even $(p)_N^{-1}$ for the expression to be well-defined. This can always be achieved by redefining $(p)_N^{-1}$ to $(p)_N^{-1} + N$.

2.5.3 Details of TQFT Argument

In this supplementary section, we provide more details of the proof of theorem 5 and theorem 6.

2.5.4 Duality Symmetry

Consider a (3+1)d (spin) TQFT $\mathcal{Q}$ that has a unique vacuum and a $\mathbb{Z}_N^{(1)}$ one-form symmetry. We will prove that the TQFT cannot be invariant under gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry up to a gravitational counterterm $e^{i\Omega(X)}$ unless $N = k^2\ell$ with $k, \ell \in \mathbb{Z}$ and -1 is a quadratic residue of ℓ . Furthermore, if the TQFT is invariant under gauging, the gravitational counterterm $e^{i\Omega(X)} = 1$ on every simply-connected smooth spin 4-manifold X .

The $\mathbb{Z}_N^{(1)}$ one-form symmetry is generated by surface operators U_g . To detect the one-form charge of a line operator $\mathcal{L}$, we link the surface operators U_g with the line operator $\mathcal{L}$. If the correlation function, where U_g and $\mathcal{L}$ form a Hopf link, differs from the one, where the link is trivial, we say the surface operator U_g links nontrivially with the line operator $\mathcal{L}$. Otherwise, we say U_g and $\mathcal{L}$ link trivially with each other.

In general, the $\mathbb{Z}_N^{(1)}$ one-form symmetry can act unfaithfully on line operators in the sense that some symmetry operators U_g link trivially with all line operators.¹⁴ The unfaithful one-form symmetry forms a $\mathbb{Z}_M^{(1)}$ subgroup of the $\mathbb{Z}_N^{(1)}$ one-form symmetry where $M \in \mathbb{Z}$ and $k = N/M \in \mathbb{Z}$. In this case, all line operators transform with a k th root of unity, instead of a N th root of unity, under the $\mathbb{Z}_N^{(1)}$ one-form symmetry.

14. An unfaithful symmetry operator can still have non-trivial correlation functions associated to other topological invariants, such as intersections, triple linkings and quadruple linkings of surfaces and etc. [165, 166].

We now gauge the $\mathbb{Z}_N^{(1)}$ one-form symmetry by coupling the theory to a dynamical $\mathbb{Z}_N$ two-form gauge field $b \in H^2(X, \mathbb{Z}_N)$. This leads to a dual $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry, generated by the Wilson surface operator $\hat{U} = \exp(2\pi i \oint b/N)$, in the gauged theory $\mathcal{Q}/\mathbb{Z}_N^{(1)}$. Here, we add a hat to the dual symmetry group to distinguish it from the original symmetry group.

Line operators charged under the $\mathbb{Z}_N^{(1)}$ one-form symmetry become non-gauge-invariant in the gauged theory $\mathcal{Q}/\mathbb{Z}_N^{(1)}$. To make them gauge invariant, we attach an open Wilson surface operator to them. Recall that all line operators transform only with a k th root of unity under the $\mathbb{Z}_N^{(1)}$ one-form symmetry. Therefore, only the Wilson surface operators generated by $\hat{U}^M = \exp(2\pi i \oint b/k)$ can end on these charged lines. The charged lines are topological, based on the assumption that $\mathcal{Q}$ is a TQFT, so they provide a topological boundary condition to the Wilson surface operators ending on them. Using these topological boundary conditions, we can unlink any Hopf link between the surface operators generated by $\hat{U}^M = \exp(2\pi i \oint b/k)$ and any line operator. Thus, the $\hat{\mathbb{Z}}_k^{(1)}$ subgroup, generated by $\hat{U}^M = \exp(2\pi i \oint b/k)$, of the dual $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry acts unfaithfully in the gauged theory $\mathcal{Q}/\mathbb{Z}_N^{(1)}$.

The unfaithful subgroup of the dual $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry can be larger than $\hat{\mathbb{Z}}_k^{(1)}$. If the TQFT is invariant under gauging, the unfaithful subgroup in the gauged theory $\mathcal{Q}/\mathbb{Z}_N^{(1)}$ should be $\hat{\mathbb{Z}}_M^{(1)}$, which should include $\hat{\mathbb{Z}}_k^{(1)}$ as a subgroup. It implies that $N = k^2 \ell$ for some integer ℓ .

We now derive a constraint on ℓ . Consider placing the theory $\mathcal{Q}$ on $S^2 \times S^2$, which has $H_2(S^2 \times S^2, \mathbb{Z}) = H^2(S^2 \times S^2, \mathbb{Z}) = \mathbb{Z}^2$ and an intersection form H . We can turn on the background gauge field $(B_1, B_2) \in H^2(S^2 \times S^2, \mathbb{Z}_M) = \mathbb{Z}_M^2$ for the unfaithful $\mathbb{Z}_M^{(1)}$ subgroup one-form symmetry. It amounts to inserting B_1 and B_2 number of basic $\mathbb{Z}_M^{(1)}$ symmetry operators wrapping around the two S^2 respectively. According to Proposition 1 of [162], any surface operator in a TQFT that links trivially with all line operators admits

a topological boundary condition. In our TQFT $\mathcal{Q}$, it implies that the symmetry operators of the unfaithful $\mathbb{Z}_M^{(1)}$ subgroup one-form symmetry can be opened up topologically without changing the correlation functions. On $S^2 \times S^2$, opening up the (B_1, B_2) number of symmetry operators and shrinking them leaves us with $B_1 B_2$ number of identical local operators $\mathcal{O}$ at the intersections. Since the TQFT has a unique local vacuum, these local operators must be a multiple of the identity operator, i.e. $\mathcal{O} = \lambda \mathbf{1}$ and therefore the partition function is $Z_{\mathcal{Q}}[B_1, B_2] = \lambda^{B_1 B_2} Z_{\mathcal{Q}}[0, 0]$. Because of the identification $B_{1,2} \sim B_{1,2} + M$, the coefficient λ obeys

$$\frac{Z_{\mathcal{Q}}[B_1 + M, B_2]}{Z_{\mathcal{Q}}[B_1, B_2]} = \lambda^{M B_2} = 1 . \quad (2.76)$$

Solving for λ , we get

$$Z_{\mathcal{Q}}[B_1, B_2] = Z_{\mathcal{Q}}[0, 0] \exp \left[\frac{2\pi i p}{M} B_1 B_2 \right] , \quad p \in \mathbb{Z} . \quad (2.77)$$

The fact that $\mathcal{Q}$ is invariant under gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry imposes a constraint on $Z_{\mathcal{Q}}[B_1, B_2]$. Starting from the theory $\mathcal{Q}$, we can gauge the $\mathbb{Z}_M^{(1)}$ subgroup one-form symmetry and obtain the theory $\mathcal{Q}/\mathbb{Z}_M^{(1)}$, which has a dual $\tilde{\mathbb{Z}}_M^{(1)}$ and a $\tilde{\mathbb{Z}}_k^{(1)}$ one-form symmetry. We put a tilde on the symmetry group of the theory $\mathcal{Q}/\mathbb{Z}_M^{(1)}$. The dual $\tilde{\mathbb{Z}}_M^{(1)}$ one-form symmetry is generated by the Wilson surface operators of the $\mathbb{Z}_M^{(1)}$ two-form gauge field. Turning on the background gauge field $(C_1, C_2) \in H^2(S^2 \times S^2, \mathbb{Z}_M)$ for the $\tilde{\mathbb{Z}}_\ell^{(1)}$ subgroup of the $\tilde{\mathbb{Z}}_M^{(1)}$ one-form symmetry leads to the following partition function for the $\mathcal{Q}/\mathbb{Z}_M^{(1)}$ theory on $S^2 \times S^2$ that can be computed with the formula for

partition functions (2.77) and (2.74)

$$\begin{aligned}
Z_{\mathcal{Q}/\mathbb{Z}_M^{(1)}}[C_1, C_2] &= \frac{1}{M} \sum_{b_{1,2}=1}^M Z_{\mathcal{Q}}[b_1, b_2] \exp \left[\frac{2\pi i}{\ell} (b_1 C_2 + b_2 C_1) \right] \\
&= \frac{1}{M} \sum_{b_{1,2}=1}^{k\ell} Z_{\mathcal{Q}}[0, 0] \exp \left[\frac{2\pi i p}{k\ell} b_1 b_2 + \frac{2\pi i}{\ell} (b_1 C_2 + b_2 C_1) \right] \\
&= Z_{\mathcal{Q}}[0, 0] \times L \exp \left[-\frac{2\pi i(q)^{-1}_K}{K} \frac{kC_1}{L} \frac{kC_2}{L} \right] \delta(kC_{1,2} \bmod L) .
\end{aligned} \tag{2.78}$$

where $L = \gcd(p, k\ell)$, $K = k\ell/L$, $q = p/L$. Note that the Gauss sum equals unity for the intersection pairing on $S^2 \times S^2$ as shown in the supplementary section on Gauss sums. On the other hand, we can also start from the gauged theory $\mathcal{Q}/\mathbb{Z}_N^{(1)}$ and gauge the $\hat{\mathbb{Z}}_k^{(1)}$ subgroup of the $\hat{\mathbb{Z}}_N^{(1)}$ one-form symmetry to obtain the theory $\mathcal{Q}/\mathbb{Z}_M^{(1)}$. The $\tilde{\mathbb{Z}}_k^{(1)}$ one-form symmetry of $\mathcal{Q}/\mathbb{Z}_M^{(1)}$ is the dual symmetry generated by the Wilson surface operators of the $\hat{\mathbb{Z}}_k^{(1)}$ two-form gauge field. The $\hat{\mathbb{Z}}_M^{(1)}/\hat{\mathbb{Z}}_k^{(1)} = \mathbb{Z}_\ell^{(1)}$ part of the $\hat{\mathbb{Z}}_M^{(1)}$ one-form symmetry of $\mathcal{Q}/\mathbb{Z}_N^{(1)}$ becomes the $\tilde{\mathbb{Z}}_\ell^{(1)} \subset \tilde{\mathbb{Z}}_M^{(1)}$ one-form symmetry of $\mathcal{Q}/\mathbb{Z}_M^{(1)}$. We then have another expression for $Z_{\mathcal{Q}/\mathbb{Z}_M^{(1)}}[C]$

$$Z_{\mathcal{Q}/\mathbb{Z}_M^{(1)}}[C_1, C_2] = \frac{1}{\sqrt{k}} \sum_{b_{1,2}=1}^k Z_{\mathcal{Q}/\mathbb{Z}_N^{(1)}}[\ell b_1 + C_1, \ell b_2 + C_2] . \tag{2.79}$$

Recall that the partition functions before and after gauging are identical up to a gravitational counterterm $e^{i\Omega(X)}$ which is a pure phase. On $S^2 \times S^2$, this gives

$$Z_{\mathcal{Q}}[B_1, B_2] = Z_{\mathcal{Q}/\mathbb{Z}_N^{(1)}}[B_1, B_2] e^{-i\Omega(S^2 \times S^2)} . \tag{2.80}$$

Together with (2.77) and the formula (2.74), we obtain

$$\begin{aligned}
Z_{\mathcal{Q}/\mathbb{Z}_M^{(1)}}[C_1, C_2] &= \frac{1}{k} \sum_{b_{1,2}=1}^k Z_{\mathcal{Q}}[0, 0] e^{i\Omega(S^2 \times S^2)} \exp \left[\frac{2\pi i p \ell}{k} b_1 b_2 + \frac{2\pi i p}{k} (b_1 C_2 + b_2 C_1) + \frac{2\pi i p}{k \ell} C_1 C_2 \right] \\
&= Z_{\mathcal{Q}}[0, 0] e^{i\Omega(S^2 \times S^2)} \times L \exp \left[\frac{2\pi i p}{k \ell} C_1 C_2 - \frac{2\pi i (\tilde{q})^{-1}}{\tilde{K}} \frac{p C_1}{\tilde{L}} \frac{p C_2}{\tilde{L}} \right] \delta(p C_{1,2} \bmod \tilde{L}) ,
\end{aligned} \tag{2.81}$$

where $\tilde{L} = \gcd(p\ell, k)$, $\tilde{K} = k/\tilde{L}$, $\tilde{q} = p\ell/\tilde{L}$.

On $S^2 \times S^2$, the partition function $Z_{\mathcal{Q}}[0, 0]$ is positive [162].¹⁵ Hence, equating (2.78) and (2.81), we get the following constraints. First of all, to match the magnitudes of the two partition functions, we have

$$L \delta(k C_{1,2} \bmod L) = \tilde{L} \delta(p C_{1,2} \bmod \tilde{L}) , \quad L \equiv \gcd(p, k\ell) , \quad \tilde{L} \equiv \gcd(p\ell, k) . \tag{2.82}$$

It implies that $L = \tilde{L} = \gcd(p, k)$ and therefore $\ell, \hat{p} \equiv p/\gcd(p, k), \hat{k} \equiv k/\gcd(p, k)$ are all co-prime to each other. Once these conditions are obeyed, the delta functions in (2.78) and (2.81) both become trivial. Next, we match the phases of the two partition functions. Since $e^{i\Omega(S^2 \times S^2)}$ is independent of (C_1, C_2) , we have

$$e^{i\Omega(S^2 \times S^2)} = 1 . \tag{2.83}$$

Matching the remaining phases that depend on (C_1, C_2) leads to the constraint

$$-\frac{(\hat{p})^{-1}}{\hat{k}\ell} \hat{k}^2 = \frac{\hat{p}}{\hat{k}\ell} - \frac{(\hat{p}\ell)^{-1}}{\hat{k}} \hat{p}^2 \bmod 1 \tag{2.84}$$

15. This is because $S^2 \times S^2$ is the double of an open four-manifold χ which may be embedded inside S^4 . In particular this means that the partition function of $\mathcal{Q}$ on S^4 which is necessarily non-zero by unitarity, may be viewed as the inner-product of the state defined by χ and the state defined by the complement of χ . Therefore χ defines a non-zero state and the partition function on $S^2 \times S^2$ is the norm of a non-zero vector and so positive.

We can multiply the equation by $\hat{k}\ell$ and treat it as an equation modulo ℓ .¹⁶ The second term on the right-hand side dropped out and the constraint simplifies to

$$(\hat{p})_{\hat{k}\ell}^{-1} \hat{k}^2 + \hat{p} = 0 \bmod \ell \quad (2.85)$$

If we further multiply the equality by $(\hat{p})_{\hat{k}\ell}^{-1}$ and still treat the equality as an equality modulo ℓ , we get

$$(\hat{k}(\hat{p})_{\hat{k}\ell}^{-1})^2 + 1 = 0 \bmod \ell . \quad (2.86)$$

The constraint implies that -1 is a quadratic residue of ℓ . The condition holds if and only if every odd prime factor x_i in the prime factorization of ℓ

$$\ell = 2^{y_0} x_1^{y_1} x_3^{y_3} \cdots x_m^{y_m} , \quad (2.87)$$

is 1 modulo 4 and the power of 2 in the prime factorization is $y_0 = 0, 1$. This completes the first part of the proof.

The foregoing derivation remains unmodified if we replace the spacetime manifold $S^2 \times S^2$ by any other simply-connected smooth spin 4-manifold X . On such manifolds, by opening up the unfaithful $\mathbb{Z}_M^{(1)}$ symmetry operator and shrinking them to the intersection points, we can fix the partition function to

$$Z_{\mathcal{Q}}[B] = Z_{\mathcal{Q}}[0] \exp \left[\frac{2\pi i p}{M} \frac{B^T I B}{2} \right] , \quad p \in \mathbb{Z} , \quad (2.88)$$

where $B \in H^2(X, \mathbb{Z}_M) = \mathbb{Z}_M^n$ is the background gauge field for the unfaithful $\mathbb{Z}_M^{(1)}$ one-form symmetry and I is the intersection form of the 4-manifold X . Here, p is an integer because the intersection form of a spin manifold is even. Note that $Z_{\mathcal{Q}}[0]$ is positive

16. If we multiply the equation (2.84) by $\hat{k}\ell$ and treat it as an equation modulo $\hat{k}$, we get a trivial equality. Since $\hat{k}$ and ℓ are co-prime, the equality (2.84) is equivalent to (2.85).

on any simply-connected smooth spin 4-manifold [163]. Generalizing (2.78) and (2.81), we have the equality

$$\frac{1}{M^{\text{rk}(I)/2}} \sum_{b=1}^M Z_{\mathcal{Q}}[b] \exp \left[\frac{2\pi i}{\ell} b^T I C \right] = \frac{1}{k^{\text{rk}(I)/2}} \sum_{b=1}^k Z_{\mathcal{Q}/\mathbb{Z}_N^{(1)}}[\ell b + C] . \quad (2.89)$$

Substituting (2.88) and

$$Z_{\mathcal{Q}}[B] = Z_{\mathcal{Q}/\mathbb{Z}_N^{(1)}}[B] e^{-i\Omega(X)} \quad (2.90)$$

into (2.89) leads to the same constraint on ℓ as before and since the Gauss sum for the intersection forms of X is unity, we have $e^{i\Omega(X)} = 1$ on every simply-connected smooth spin 4-manifolds. This concludes the proof. Since the proof only uses spin 4-manifolds, which have a trivial second Stiefel-Whitney class, we cannot determine whether the TQFT is a spin TQFT or a non-spin TQFT.

Triality Symmetry and More General Non-invertible Symmetry

We now generalize the constraint on duality invariant TQFTs to TQFTs that preserves more general non-invertible symmetry including the triality symmetry. Consider a (3+1)d (spin) TQFT $\mathcal{Q}$ that has a unique local vacuum and a $\mathbb{Z}_N^{(1)}$ one-form symmetry. We will prove that the TQFT cannot be invariant under $T^{-p'}S$ gauging the $\mathbb{Z}_N^{(1)}$ one-form symmetry upto a gravitational counterterm $e^{i\Omega(X)}$ unless $N = k^2\ell$ with $k, \ell \in \mathbb{Z}$ and there exists a solution $p \in \mathbb{Z}$ such that

$$p(p + p') + 1 = 0 \bmod \ell . \quad (2.91)$$

Furthermore, if the TQFT is invariant under the $T^{-p'}S$ gauging, the gravitational counterterm $e^{i\Omega(X)} = 1$ on every simply-connected smooth spin 4-manifold X . When N is even and $p' = \pm 1$ or when N is odd and $p' = \pm(1 + N)$, the non-invertible defects

constructed by half-gauging the theory leads to the triality defects.

The proof for these more general $T^{-p'}S$ symmetries follows the same reasoning as in the duality case. We will directly work on simply connected smooth spin manifolds X . (2.88) and (2.89) still holds but now due to the invariance under $T^{-p'}S$ gauging we have

$$Z_{\mathcal{Q}/\mathbb{Z}_N^{(1)}}[B] = Z_{\mathcal{Q}}[B] \exp \left[\frac{2\pi i p'}{N} (kB)^T I(kB) \right] e^{i\Omega(X)} . \quad (2.92)$$

Combining this relation with (2.88) and (2.89), we get the same constraint as in (2.82) so ℓ , $\hat{p} \equiv p/\gcd(p, k)$, $\hat{k} \equiv k/\gcd(p, k)$ are all co-prime to each other and we further derived

$$\tilde{p}(\tilde{p} + p') + 1 = 0 \bmod \ell . \quad (2.93)$$

where $\tilde{p} = \hat{k}(\hat{p})_{\hat{k}\ell}^{-1}$. Again, because the Gauss sum for the intersection forms of a simply connected smooth spin 4-manifold X is unity, we learned that $e^{i\Omega(X)} = 1$. This concludes the proof.

2.5.5 Spontaneous Symmetry Breaking

While the main focus of our analysis is on theories with a unique vacuum state on S^3 , it is also possible to spontaneously break duality symmetries leading to multiple vacuum states. In this case there is no constraint on N . Here we briefly summarize some simple examples of this phenomenon. We also note that spontaneous duality symmetry breaking is realized in $\mathbb{Z}_N$ lattice gauge theory for small N as discussed in section 2.4.

We can construct an elementary gapped example of spontaneously breaking the duality defect $\mathcal{D}$ by starting from $\mathbb{Z}_N$ topological gauge theory (i.e. Dijkgraaf-Witten theory).

This theory has a $\mathbb{Z}_N^{(1)}$ one-form symmetry under which the Wilson lines are charged. In the presence of a background gauge field B the partition function on a simply connected

four-manifold X is:

$$Z_{\text{DW}}[B] = \frac{1}{\sqrt{|H^2(X, \mathbb{Z}_N)|}} \sum_{c \in H^2(X, \mathbb{Z}_N)} \exp \left[\frac{2\pi i}{N} \int_X c \cup B \right] = \sqrt{|H^2(X, \mathbb{Z}_N)|} \delta(B \bmod N) . \quad (2.94)$$

Performing the S gauging operation replaces the partition function above by

$$SZ_{\text{DW}}[B] = \frac{1}{|H^2(X, \mathbb{Z}_N)|} \sum_{b, c \in H^2(X, \mathbb{Z}_N)} \exp \left[\frac{2\pi i}{N} \int_X b \cup (c + B) \right] . \quad (2.95)$$

The equation of motion for b forces $c = -B$, and as a consequence the theory obtained after gauging is trivial

$$SZ_{\text{DW}}[B] = Z_{\text{trivial}}[B] = 1 . \quad (2.96)$$

This gauging procedure is reversible: gauging a $\mathbb{Z}_N^{(1)}$ one-form symmetry in the trivial theory takes us back to the original Dijkgraaf-Witten theory.

We can now use these results to construct a theory which is self-dual under gauging $\mathbb{Z}_N^{(1)}$ one-form symmetry for all values of N . We consider a theory $\mathcal{Q}$ which is the direct sum of the Dijkgraaf-Witten theory and the trivial theory:¹⁷

$$Z_{\mathcal{Q}} \equiv Z_{\text{DW}} \oplus Z_{\text{trivial}} . \quad (2.97)$$

By construction, $\mathcal{Q}$ has two local ground states and correspondingly two topological local operators corresponding to the identity operator in each of the sectors above. According to the discussion above the S operation exchanges these two summands. This implies that the theory $\mathcal{Q}$ has a duality defect $\mathcal{D}$ which permutes the two topological local operators.

Physically, the construction above can occur at a first order phase transition with a

17. The direct sum of d -dimensional theories assigns to a $d-1$ manifold a Hilbert space which is the direct sum of the Hilbert space assigned to each summand, and assigns to a closed d manifold the sum of the partition functions. See references [192, 193] for additional details.

spontaneously broken duality symmetry. In this case the symmetry defect $\mathcal{D}$ describes a domain wall connecting the two local vacua which are hence related by the gauging operation S . In particular, this is exactly the phase transition which occurs in self-dual $\mathbb{Z}_N$ lattice gauge theory for $N \leq 4$. (See section 2.4.)

The construction above may be straightforwardly generalized to produce myriad examples of spontaneously broken duality symmetry. Indeed, given any theory $\mathcal{G}$ with $\mathbb{Z}_N^{(1)}$ one-form symmetry, we can consider the direct sum:

$$Z_Q \equiv Z_{\mathcal{G}} \oplus Z_{S\mathcal{G}} , \quad (2.98)$$

which again realizes spontaneously broken duality symmetry. Similarly, we can construct gapped or gapless phases realizing spontaneous symmetry breaking for the more general non-invertible symmetries arising from invariance under $T^{-p'}S$. For instance, a spontaneously broken triality symmetry leads in general to three local vacua connected by sequential gauging and stacking operations. This occurs in the Cardy-Rabinovici model at $\tau_* = e^{\pi i/3}$ for small N . (See section 2.4.)

CHAPTER 3

DEEP LEARNING LATTICE GAUGE THEORIES

This chapter is reprinted with permission from:

Apte, A., Ashmore, A., Córdova, C., and Huang, T. C. Deep learning lattice gauge theories. Physical Review B (Vol. 110, Issue 16). © 2024 American Physical Society
DOI:10.1103/PhysRevB.110.165133

3.1 Introduction

Since their introduction by Wilson [194, 195], lattice gauge theories have been pivotal in understanding fundamental physics, particularly quantum chromodynamics (QCD) [196, 197], which governs the strong nuclear force. QCD, a non-Abelian gauge theory, describes interactions between quarks and gluons, the elementary constituents of hadrons. Despite its apparent simplicity, the non-perturbative regime of this theory exhibits remarkably complex phenomena, including confinement [198] and chiral symmetry breaking [199, 200], often resisting analytical tools. As such, Monte Carlo-based numerical methods have become invaluable [201], providing insights into QCD phases [202] and enabling first-principles calculations of hadron masses [203].

Discrete lattice gauge theories – lattice gauge theories with *finite* gauge groups – serve as important models for studying the non-perturbative behavior and emergent phenomena of gauge theories. While simpler than their continuous counterparts, they still capture key features such as gauge invariance, confinement and topological excitations. This makes them valuable testing grounds for developing and benchmarking new theoretical and computational techniques.

The present work focuses on finding the ground-state wavefunction of discrete lattice gauge theories in $2 + 1$ dimensions. Since the wavefunction contains rich information about

quantum correlations and topological properties of the system, there are many reasons to directly compute the wavefunction itself instead of simply calculating expectation values of observables via traditional (Euclidean) Monte Carlo. For example, entanglement measures such as Rényi entropies can be computed with this approach, since the wavefunction encodes the full correlation structure of the lattice system [204, 205].

Obtaining accurate ground-state wavefunctions for lattice gauge theories is challenging due to the complexity of the non-perturbative regime and the exponential growth of the Hilbert space with system size. In particular, traditional analytical and numerical methods often struggle to capture the intricate correlations and entanglement present in the ground state of these systems. Recent years have shown that deep neural networks are adept at learning and sampling from classical probability distributions across a variety of domains, including text, images and speech [206]. The problem of finding a ground-state wavefunction can be viewed as a generalization of this problem, namely, computing the complex-valued probability amplitude associated to the basis elements of the underlying Hilbert space, so it is not surprising that neural networks are well suited to this task.

The use of neural networks to represent wavefunctions was pioneered by Carleo and Troyer [58], and has since become known as a “neural network quantum state” (NQS or NNQS). The idea behind a neural network quantum state is to model the wavefunction as a deep neural network with hidden layers, leveraging the expressive power of these networks to construct a highly flexible variational ansatz for the ground state. One then “trains” the network to minimize the energy of the variational state, with expectation values calculated using Markov chain Monte Carlo. Key to this process is the ability to efficiently compute gradients of expectation values with respect to the parameters of the network, so that the energy can be minimized via (stochastic) gradient descent [207].

Thanks to differentiable programming and backpropagation [208, 209], one can compute these gradients to machine precision without resorting to numerical interpolation. At the

end of the training process, one hopes to have a NNQS that can accurately capture the physics of the ground state and enable calculation of other observables.

There are two fundamental problems in simulating quantum many-body systems: the problems of representation and computation. Focusing on a discrete lattice system for concreteness, a wavefunction associates a complex number to every lattice configuration. Since the number of configurations grows exponentially with the system size, naively one might need an ansatz with a similarly large number of parameters. This is the problem of representation. There is no reason to expect that a randomly chosen state of the full Hilbert space can be represented efficiently, but fortunately physical states – and ground states in particular – are not generic, instead exhibiting features such as area-law entanglement [210].

NNQSs provide an efficient and flexible parametrisation of these states with a sub-exponential number of variational parameters.¹

Given such an efficient encoding of a quantum state, one still needs to be able to compute probability amplitudes in a reasonable time. Furthermore, physical observables are still operators on an exponentially large Hilbert space, so computing expectation values may be challenging. These comprise the problem of computation. For NNQSs, it is simple and fast to compute the amplitude for an input lattice configuration, with observables then calculated by restricting to a finite sum over Hilbert space via Monte Carlo importance sampling [215].

Much like traditional lattice Monte Carlo, NNQSs can be used to study finite-temperature

1. NNQSs should also be compared with tensor networks for $1+1$ -dimensional systems. Tensor networks are known to be complete (any pure state can be described by increasing the bond dimension sufficiently) and efficient for states with area-law entanglement (the cost of computing expectation values grows polynomially with the number of parameters). Similarly, thanks to universal representation theorems [211], NNQSs can describe arbitrary quantum states, including ground states with volume-law entanglement [212]. Although there is no bound on the size of the network required to encode these states, it is known that NNQSs can exactly reproduce tensor network or projected entangled pair states, and there are examples where NNQSs provide efficient representations of physically relevant states where these other approaches fail to be efficient [213, 214].

systems [216]. There are also a number of questions that are difficult to tackle with Euclidean methods but which are within reach of NNQSs, including real-time and dissipative dynamics, fermionic systems with finite chemical potential, quantum-state reconstruction, and open quantum systems [217]. It is possible to impose both global and gauge symmetries exactly in NNQS by using invariant neural networks [218].

The particular neural network architecture that we employ is based on “lattice gauge-equivariant convolutional neural networks” (L-CNNs), introduced by Favoni et al. [64]. L-CNNs possess the usual locality and weight sharing properties of traditional convolutional networks. Crucially, they are also manifestly gauge equivariant and so one does not have to impose gauge symmetry via inexact methods, such as energy penalties. Their uses to date have focused on predicting gauge-invariant observables via supervised learning [219–224]. We instead use an L-CNN as a NNQS, i.e. a variational ansatz for the ground-state wavefunction, similar to the approach taken by Luo et al. [225, 226]. We have implemented this architecture using NetKet [227], a flexible machine-learning framework for many-body quantum physics built on top of JAX [228], enabling automatic differentiation and GPU acceleration.

Previous work

Machine learning methods have already made an impact in many areas of physics, including astronomy and cosmology, condensed-matter physics, fluid dynamics, quantum chemistry, and nuclear and particle physics, and have shown potential for studying lattice gauge theories [229].

Since our approach is based on an existing network architecture, it is useful to outline how our work adds to the state of the art. Favoni et al. [64] introduced L-CNNs to incorporate gauge symmetries directly into the network structure while maintaining the ability to approximate any gauge-covariant function on the lattice. The authors showed that L-CNNs

outperform conventional CNNs in regression tasks involving Wilson loops of different sizes and shapes in pure $SU(2)$ gauge theory, with the performance gap increasing as the loop size grows. This supervised learning problem demonstrated that L-CNNs can learn the physical information encoded in Wilson loops and other gauge-invariant observables. There have been a number of follow-up works which explore the generalization ability of L-CNNs and their geometric interpretation [222, 230], but no attempt has been made to use this architecture as a variational ansatz for the ground state of a lattice theory.

Closer to our approach is the work by Luo et al. [225], where gauge-equivariant neural network quantum states were introduced as a technique to efficiently model the wavefunction of lattice gauge theories while maintaining exact local gauge invariance. For the specific case of $\mathbb{Z}_2$ gauge theory on a square lattice, the authors combined this architecture with variational Monte Carlo to study the ground state and the confinement-deconfinement transition. This was extended to include matter in [231] and to $U(1)$ gauge theory in [226]. The aims of the present paper include greatly improving upon the precision of these results and studying $\mathbb{Z}_3$ gauge theory. For example, these studies did not determine the order of the phase transition, calculate the critical exponents, investigate static charges, or provide a precise value for the critical coupling. In filling these gaps, our goal is to demonstrate that deep learning can be used for precise calculation, and in doing so enhance our understanding of the phases and phase transitions in discrete gauge theories.

Main results

For $\mathbb{Z}_2$ lattice gauge theory, we compute the ground-state energy as a function of coupling for various lattice sizes (Figure 3.8) and confirm the existence of confined and deconfined phases via calculations of order and disorder parameters and the string tension. We show that the confinement phase transition between the two phases is continuous, with a critical

coupling and critical exponents given by

$$g_c = 0.7546(8), \beta = 0.326(4), \nu = 0.630(3) . \quad (3.1)$$

These values are calculated via “curve collapse” from finite-scaling theory. The high quality of the resulting collapse can be seen in Figure 3.13. These values imply that the conformal dimensions of the σ and ϵ primary operators in the 2 + 1-d Ising CFT are:

$$\Delta_\sigma = 0.518(4), \Delta_\epsilon = 1.412(8) , \quad (3.2)$$

in excellent agreement with previous results from perturbative calculations, Monte Carlo simulations and the conformal bootstrap. We further refine our estimate of the critical coupling using BST extrapolation [232–234] and find $g_c = 0.756431$. We also confirm the linear potential between static charges and compute it as a function of coupling, finding it vanishes for small couplings and smoothly increases around the critical coupling, approaching the strong-coupling prediction for large couplings.

For $\mathbb{Z}_3$, we compute the ground-state energies for varying lattice size (Figure 3.16) and, for small lattices, confirm that they agree with results from exact diagonalization. As with $\mathbb{Z}_2$, we show the existence of confined and deconfined phases via Wilson loop and ‘t Hooft string operators, and the string tension. We investigate the nature of the phase transition by examining the electric and magnetic energies, and derivatives of the energy as a function of coupling. We plot these in Figure 3.19, finding that $\mathbb{Z}_3$ displays a sharp change at the transition compared with a smooth change for $\mathbb{Z}_2$, indicating the presence of a first-order transition. This transition occurs within the range $g \in [0.775, 0.7875]$, giving a critical coupling of

$$g_c = 0.781(6) . \quad (3.3)$$

Due to finite system size and the weakly first-order nature of the transition, we do not

observe a discontinuity in either the energy derivatives or the order and disorder parameters.

We also quantify how well our networks have learned the charge conjugation symmetry of the $\mathbb{Z}_3$ theory by measuring the imaginary component of Wilson loop expectations, which should vanish for the exact ground state. Depending on the value of the coupling, we find that this holds to better than one part in 10^4 to 10^6 . Finally, we compute the static potential between charges, recovering the expected linear potential and finding agreement with perturbative results in the small and large coupling regimes.

The NNQs that we use to obtain these results have up to 10k parameters, depending on the depth and width of the networks. To illustrate the computational resources and time needed for training, in Figure 3.1 we plot the number of seconds per iteration and the number of seconds per network parameter for a single iteration for various lattice sizes.

The networks are the same size as those used to obtain the results in the main text, with $(N_{\text{out}}, K) = (4, 2)$ (see Section 3.4.1) and depths given by $\lceil \log_2 L^2 \rceil$ for an $L \times L$ lattice, so that the largest Wilson loops constructed by the network can cover the full lattice. We train our networks for 500 to 1000 iterations, with smaller networks converging well before this and larger networks benefitting from longer training. The networks used for studying $L = 12$ systems train in approximately five hours on a single NVIDIA A100 GPU [235]. In fact, the networks often converge well before this if one utilises transfer learning when scanning over couplings (see Section 3.4.2).

Outline

We begin in Section 3.2 by recalling $\mathbb{Z}_N$ gauge theory on the lattice, including a discussion of Wilson loops, the string tension and 't Hooft strings. We also comment on the global symmetry of these theories and their relation to $\mathbb{Z}_N$ spin models, their ground-state degeneracies and the potential between static charges. In Section 3.3, we give more details

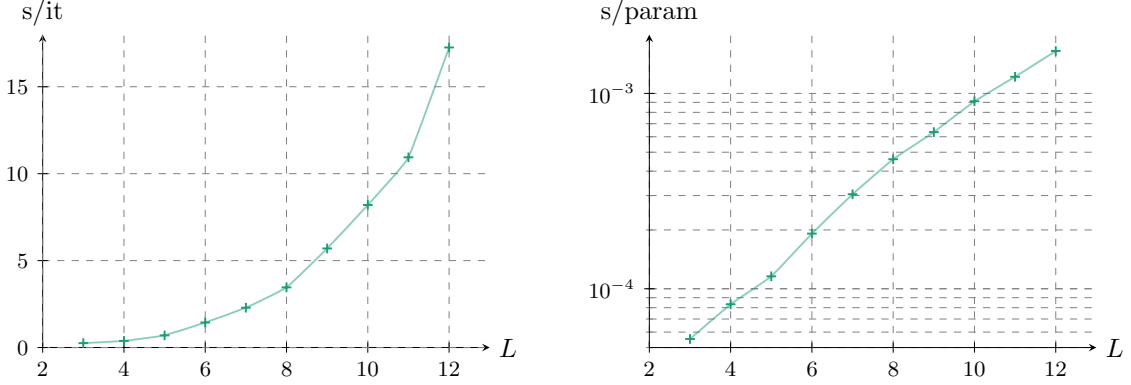


Figure 3.1: The left figure shows seconds per training iteration as a function of lattice size L . The depths of the networks are chosen so that, with kernel size $K = 2$ per layer, the largest Wilson loop is large enough to cover the entire lattice. All other hyperparameters are identical. The right figure shows the number of seconds per network parameter for a single iteration. The computations were performed on a single NVIDIA A100 GPU.

on the phases of zero-temperature $\mathbb{Z}_N$ gauge theories in $2 + 1$ dimensions, including an overview of confined and deconfined phases, and the expected nature of the transition for varying N . In Section 3.4, we introduce the ideas behind neural network quantum states and compare them with traditional Euclidean Monte Carlo. Following this, we review the network architecture that we use in this work, namely lattice gauge-equivariant convolutional neural networks [64]. We describe our implementation of this and give details of sampling, initialization, training, transfer learning and calculation of observables.

Thereafter, we explain why the states represented by the networks are always in the trivial superselection sector of the $\mathbb{Z}_N^{(1)}$ one-form global symmetry. With this groundwork set, in Section 3.5 we analyze $\mathbb{Z}_2$ gauge theory. Our exploration includes computing ground-state energies as a function of coupling for varying lattice size, and expectation values of order and disorder operators and the string tension. Using finite-size scaling theory and curve collapse, we compute the critical coupling and critical exponents characterising the theory's continuous phase transition, finding excellent agreement with previous results. We then study the potential between static charges as a function of coupling, recovering the expected linear dependence on distance. We move to $\mathbb{Z}_3$ gauge theory in Section 3.6. Here,

we compute ground-state energies and look at order and disorder operators to identify the location of the phase transition. By looking at derivatives of the energy as a function of coupling, we argue that the transition is weakly first-order in nature. We finish with a check that the network has correctly learned charge conjugation symmetry and analyze the static charge potential. The appendices contain an overview of our implementation of Wilson loop operators, a review of calculating critical exponents from curve collapse, and a brief summary of the BST extrapolation technique which we use to identify the critical coupling.

Future directions

Obvious next steps include examining $\mathbb{Z}_N$ theories for $N > 3$, extending our networks to continuous gauge groups and including matter fields. In particular, there is renewed interest in $\mathbb{Z}_N$ and $U(1)$ theories under the guise of studying flux tubes using confining strings [236, 237], and it would be interesting to investigate if mass gaps can be computed using NNQSs. We also intend to study theories in $3 + 1$ dimensions; unlike approaches such as matrix product states, no conceptual changes to the network architecture are needed to move to higher dimensions.

Our identification of critical exponents for $\mathbb{Z}_2$ relies on curve collapse via a leading-order finite-scaling analysis. It would be interesting to see if incorporating subleading corrections gives more precise estimates [238], even without going to larger system sizes. In another direction, it would be encouraging to observe signs of the discontinuities that characterise a first-order transition for $\mathbb{Z}_3$ by studying larger lattices. Moreover, since second Rényi entropies can be computed efficiently from a NNQS [239, 240], it would be fruitful to study the entanglement structure of the ground state and to determine the central charge of the CFTs governing the continuous phase transitions that occur for $N \neq 3$. Note that the replica method is another approach to computing the entanglement entropy [241, 242], and it will be insightful to compare the two techniques.

We have not fully explored how to optimize the training process. For example, we have used transfer learning when scanning over couplings. Since the form of an L-CNNs is system-size agnostic, one could also use transfer learning to move from smaller to larger lattices as demonstrated in [231, 243]. It has also been observed that for ground states with non-trivial sign structure, it is often helpful to first learn the phase of the wavefunction, with the amplitude learned towards the end of training [244]. In a different direction, our ground-state wavefunctions were found by minimizing their energy, however one might also try to minimize the variance of the energy. Empirically, variance minimization alone seems ill-suited for finding ground states [245], though some combination of the two appears to improve performance [246]. It may also be useful to monitor other observables such as expectation values of 't Hooft strings during training. Since the systematic error in a general observable depends linearly on the difference between the variational state and the true ground state (while the energy is quadratic in the difference), tracking other observables may provide a more sensitive measure of whether the network has converged. By construction, L-CNNs give wavefunctions which are invariant under both gauge transformations and lattice translations. In the future, we plan to investigate whether including further symmetries improves the accuracy of the variational state. For example, the wavefunction should also be invariant under rotations and reflections, which can be implemented using group convolutional neural networks (G-CNNs) [247]. Rather than relying on the network learning this information, imposing these symmetries exactly has been found to improve results [248]. Furthermore, one can impose exact charge conjugation invariance, so that the amplitude for a lattice configuration $\mathcal{U}$ is the same as for $\mathcal{U}^\dagger$. This can be done simply by averaging the network over $\mathcal{U}$ and $\mathcal{U}^\dagger$.

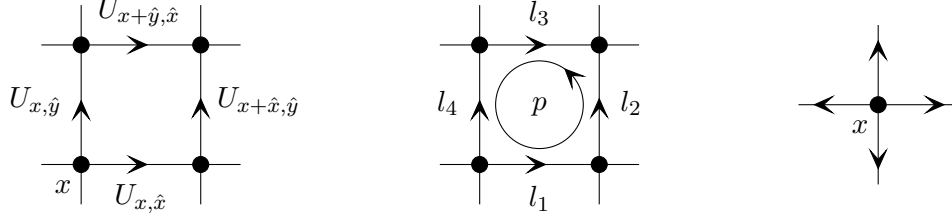


Figure 3.2: The left figure shows our convention for associating link variables to a lattice site at x . The middle figure shows the links (l_1, l_2, l_3, l_4) which make up a plaquette p when defining the magnetic term in the Hamiltonian (3.5). The right figure shows the links associated with the operator Θ_x acting at the lattice site at x , defined in (3.7), which implements Gauss's law on the lattice.

3.2 $\mathbb{Z}_N$ lattice gauge theory

Following [64, 249], we first review pure $\mathbb{Z}_N$ gauge theory on the lattice. In the Hamiltonian approach to lattice gauge theory, space is discretized while time remains continuous. Consider a spacelike two-dimensional lattice corresponding to a lattice gauge theory in $2 + 1$ dimensions, with periodic boundary conditions. The lattice spacing is taken to be one, and the size of the lattice is $L \times L$.

The gauge field degrees of freedom live on the links between the lattice sites. The link variables $U_{x,\mu}$ are valued in the gauge group. Here our notation is that $U_{x,\mu}$ determines the parallel transport from a lattice site at x to a neighboring site $x + \mu$, with $\mu = \hat{e}_\mu$, where the lattice spacing is taken to be one and $\hat{e}_\mu$ is a unit vector pointing in the positive μ direction. This is shown in Figure 3.2. We denote a particular configuration of the $2L^2$ link variables by $\mathcal{U} = \{U_{x,\mu}\}$. Under a gauge transformation, the link variables transform *non-locally* as

$$U_{x,\mu} \mapsto T_\Omega U_{x,\mu} = \Omega_x U_{x,\mu} \Omega_{x+\mu}^\dagger, \quad (3.4)$$

where Ω_x are group elements.

The Hamiltonian of pure $\mathbb{Z}_N$ gauge theory can then be written as [249–251]

$$\begin{aligned}
H &= H_E + H_B , \\
&= \frac{g^2}{2} \sum_l \left[2 - (P_l + P_l^\dagger) \right] \\
&\quad + \frac{1}{2g^2} \sum_{p=(l_1, l_2, l_3, l_4)} \left[2 - (Q_{l_1}^\dagger Q_{l_2}^\dagger Q_{l_3} Q_{l_4} + \text{h.c.}) \right] .
\end{aligned} \tag{3.5}$$

The first term is a sum over all links l of the lattice, while the second term is a sum over all plaquettes p . We refer to these as the electric and magnetic terms, respectively. The links (l_1, l_2, l_3, l_4) make up the plaquette p , as shown in Figure 3.2. The operators Q_l and P_l are the standard “clock” and “shift” operators on the link l . These unitary operators are generalizations of the Pauli operators and satisfy a $\mathbb{Z}_N$ algebra:

$$\begin{aligned}
P_l^\dagger P_l &= Q_l^\dagger Q_l = \mathbf{1}, \quad P_l^N = Q_l^N = \mathbf{1} , \\
P_l^\dagger Q_l P_l &= e^{2\pi i/N} Q_l ,
\end{aligned} \tag{3.6}$$

where operators that act on different links commute with one another.

The discretized version of Gauss’s law for a $\mathbb{Z}_N$ gauge theory is encoded by a set of local unitary vertex operators Θ_x . For $d = 2$, these operators are defined as

$$\Theta_x = P_{x,\mu} P_{x,\nu} P_{x-\mu,\mu}^\dagger P_{x-\nu,\nu}^\dagger , \tag{3.7}$$

where $\mu = \hat{x}$ and $\nu = \hat{y}$, so that the links correspond to those in Figure 3.2. The lattice Hamiltonian commutes with all of these operators,

$$[\Theta_x, H] = 0 \quad \forall x , \tag{3.8}$$

which implies the local gauge invariance of the theory. The vertex operators have

eigenvalues $e^{2\pi i n/N}$, with operators for different sites commuting. For a state $|\Psi\rangle$, Gauss's law is the statement that

$$\Theta_x |\Psi\rangle = |\Psi\rangle \quad \forall x . \quad (3.9)$$

A state which satisfies this constraint is gauge invariant. The ground state of the theory is always gauge invariant. There are, however, other superselection sectors of the Hilbert space of the theory, classified by the eigenvalues of Θ_x . For example, the twisted or charged sector where $\Theta_x = e^{2\pi i n/N}$ and $\Theta_{x'} = e^{2\pi i n'/N}$ for integer n and n' can be interpreted as the gauge-invariant states with charges n and n' located at the lattice sites x and x' respectively.

The link variables $U_{x,\mu}$ are elements of $\mathbb{Z}_N$, and so can be thought of simply as phases of the form $e^{2\pi i q/N}$, where $q \in \{0, \dots, N-1\}$. With this, we define states $|q\rangle_l$ which span the Hilbert space on the link l . These states are assumed to give an orthonormal basis for the Hilbert space, $\langle q|q'\rangle = \delta_{q,q'}$ and are eigenstates of the clock operator:

$$Q_l |q\rangle_l = e^{2\pi i q_l/N} |q\rangle_l . \quad (3.10)$$

Moreover, the shift operator P acts as a periodic lowering operator,

$$P_l |q\rangle_l = |q-1\rangle_l , \quad (3.11)$$

with $P_l |0\rangle_l = |N-1\rangle_l$. We also define plaquette variables $U_{x,\mu\nu}$ as 1×1 untraced Wilson loops of the form

$$U_{x,\mu\nu} = U_{x,\mu} U_{x+\mu,\nu} U_{x+\nu,\mu}^\dagger U_{x,\nu}^\dagger . \quad (3.12)$$

Under (3.4), the plaquette variables transform *locally* as

$$U_{x,\mu\nu} \mapsto T_\Omega U_{x,\mu\nu} = \Omega_x U_{x,\mu\nu} \Omega_x^\dagger . \quad (3.13)$$

For our example of a $d = 2$ lattice, we take $\mu = \hat{x}$ and $\nu = \hat{y}$ so that the loop is traversed anticlockwise, in agreement with the convention in Figure 3.2.

At a given time, a wavefunction Ψ for the lattice system takes in a configuration $\mathcal{U} \in \mathbb{Z}_N^{2L^2} \subset \mathbb{C}^{2L^2}$ of the link variables and returns a complex number:

$$\Psi: \mathbb{C}^{2L^2} \rightarrow \mathbb{C}, \quad \mathcal{U} \mapsto \Psi(\mathcal{U}). \quad (3.14)$$

We will be particularly interested in those wavefunctions which can approximate the ground state of the system. Since the Hamiltonian is time independent, the ground state is time independent, so it is sufficient to consider only the two-dimensional spatial lattice for a $2 + 1$ -dimensional theory. The ground state must also be gauge invariant. Denoting the ground state by $|\Psi_0\rangle$, this means that under the gauge transformation (3.4), the wavefunction Ψ_0 is invariant, i.e. $\Psi_0(T_\Omega \mathcal{U}) = \Psi_0(\mathcal{U})$.²

3.2.1 Wilson loop operators

Naively, one might expect that, in analogy to the magnetization of a spin system, the phases of a lattice gauge theory should be distinguished by a spontaneous alignment of the link degrees of freedom. This would manifest as a non-zero value of $\sum_l \langle Q_l \rangle$. In the analogous spin system, a non-zero magnetization indicates a spontaneous breaking of a global symmetry. In the gauge theory, a non-zero expectation value would instead imply the spontaneous breaking of a local symmetry. This is impossible by Elitzur's theorem [252], and so the phases of the theory cannot be distinguished by these expectation values (which vanish by gauge invariance). Instead, one must look at non-local, gauge-invariant observables, the simplest of which is the Wilson loop.

A Wilson loop operator W_Γ is defined given Γ , a closed oriented path on the lattice [194].

2. More generally, a gauge equivariant function f obeys $f(T_\Omega \mathcal{U}) = T_\Omega f(\mathcal{U})$.

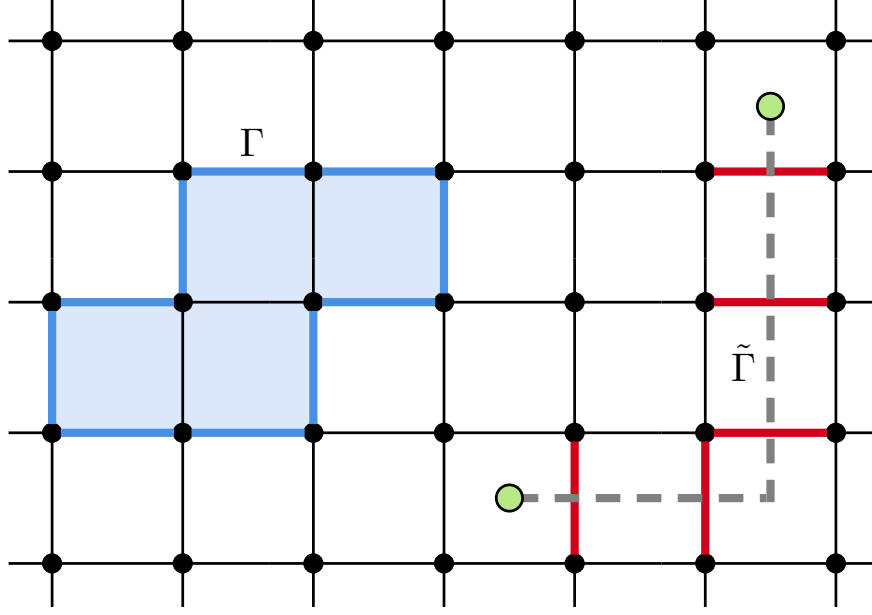


Figure 3.3: Order and disorder operators on the lattice. The blue path denotes a Wilson loop, $W_\Gamma = \prod_\Gamma Q$, given by the product of clock operators along the closed contour Γ . The shaded blue region shows the area enclosed by the Wilson loop. The dotted gray path denotes a ‘t Hooft string, $T_{\tilde{\Gamma}} = \prod_{\tilde{\Gamma}} P$, given by the product of shift operators on the links pierced by the contour $\tilde{\Gamma}$ between two points on the dual lattice.

Explicitly, we define

$$W_\Gamma = \prod_{l \in \Gamma} Q_l, \quad (3.15)$$

where our convention is that if the link l is oriented to agree with the arrows in Figure 3.2, one has Q_l , whereas if the link has the opposite orientation, one instead takes the Hermitian conjugate, $Q_l^\dagger$. Note that the magnetic term H_B in the Hamiltonian is simply $(1 - \text{re } W_\square)$, summed over the 1×1 plaquettes $\square$. An example of a Wilson loop on the lattice is shown in Figure 3.3.

In a pure $\mathbb{Z}_N$ gauge theory, the expectation values of Wilson loop operators can detect the phase of the system [194, 253, 254]. In particular, whether the system is in a confined or deconfined phase is determined by the scaling of the expectation value of W_Γ as the size of the loop Γ is varied. Therefore, the Wilson loop serves as an (unconventional) order operator for the lattice gauge theory. In a deconfined phase, the magnetic term H_B in the

Hamiltonian dominates and the expectation value of the Wilson loop decays exponentially with the length of the perimeter P_Γ of the loop as

$$\langle W_\Gamma \rangle \sim e^{-\kappa P_\Gamma} , \quad (3.16)$$

where κ is coupling dependent. In a confined phase, the electric term H_E dominates, leading to an additional confining area-law scaling:

$$\langle W_\Gamma \rangle \sim e^{-\kappa P_\Gamma - \sigma A_\Gamma} , \quad (3.17)$$

where A_Γ is the area of the loop and σ is known as the string tension. Again, σ is expected to depend on the coupling g . Note that the Wilson loop does not follow a strict area law in the confined phase since there is always a perimeter law contribution.

3.2.2 Creutz ratio

In the thermodynamic limit, one expects an order parameter to be positive in an ordered phase and zero in a disordered phase. The expectation values of Wilson loop operators do not display this behavior, and their interpretation is more nuanced. Instead, as we discussed above, it is the scaling with the size of the corresponding loop that is different in each phase. However, this behavior can be used to extract the string tension, which gives a conventional order parameter. It is simple to check that the so-called Creutz ratio [255]

$$\chi_l = -\ln \frac{\langle W_{l \times l} \rangle \langle W_{(l-1) \times (l-1)} \rangle}{\langle W_{(l-1) \times l} \rangle \langle W_{l \times (l-1)} \rangle} , \quad (3.18)$$

where $W_{l \times m}$ is an $l \times m$ Wilson loop operator on the lattice, removes the perimeter scaling and so computes the string tension, $\chi_l = \sigma$. Due to this, the Creutz ratio behaves like a standard order parameter in the thermodynamic limit: positive in the confining phase and

vanishing in the deconfined phase [256]. Note that the perimeter- and area-law scalings given in (3.16) and (3.17) are only the leading contributions. The Creutz ratio is constructed to remove some subleading finite-size corrections and also corrections that come from edge/corner effects. The leftover corrections are smaller for larger Wilson loops, so that χ_l is a better estimate of the continuum string tension for larger values of l . Unfortunately, computing χ_l using Monte Carlo sampling can be challenging. In the confined phase, where χ_l is expected to be non-zero, Wilson loops decay both with the area of the loop and increasing coupling. The string tension is then estimated by computing a ratio of very small numbers, which can be very sensitive to Monte Carlo errors – we will see this later in Figure 3.11. Consequently, one often restricts the computation to smaller values of l .

3.2.3 ‘t Hooft operators

A string of P operators, $T = \prod_{\tilde{\Gamma}} P$, along an open path $\tilde{\Gamma}$ between two points on the *dual* lattice defines a magnetic line operator known as a ‘t Hooft string. These ‘t Hooft strings serve as disorder operators for the gauge theory, and can diagnose the confined-deconfined phases. An example of a ‘t Hooft string operator is shown in Figure 3.3.

The insertion of a Gauss’s law operator Θ_x acts on a ‘t Hooft string simply by shifting the path $\tilde{\Gamma}$ without moving its endpoints. Furthermore, since the ground state is gauge invariant, insertions of Θ_x are “free” and do not change expectation values. Thus, $\langle T_{\tilde{\Gamma}} \rangle$ depends only the endpoints of the string and not on the path itself. This can be interpreted as the string operator creating a pair of quasi-particles (magnetic monopoles) which reside on the plaquettes at either end of the string, corresponding to the green dots in the example in Figure 3.3. Note that the distance between the end points must scale with the lattice size L in order to obtain a line operator in the limit as $L \rightarrow \infty$. In a deconfined phase, the expectation value of a ‘t Hooft string decays exponentially with the

distance between its endpoints. In a confined phase, $\langle T_{\tilde{F}} \rangle$ is independent of distance, since the monopoles living at the ends of the ‘t Hooft string are condensed.

3.2.4 Global symmetry and dual spin models

$\mathbb{Z}_N$ gauge theory in $2 + 1$ d can be obtained by gauging the $\mathbb{Z}_N^{(0)}$ zero-form global symmetry of a $\mathbb{Z}_N$ spin model. The $\mathbb{Z}_N^{(0)}$ symmetry corresponds to transforming all spins of the model by a global $\mathbb{Z}_N$ factor. In the case of the Ising model, we have a $\mathbb{Z}_2$ symmetry that corresponds to flipping all the spins. By Poincaré duality, we expect that the theory after gauging will have a dual $\mathbb{Z}_N^{(1)}$ one-form global symmetry in $2 + 1$ dimensions [69]. The phase transitions that occur in $\mathbb{Z}_N$ gauge theories are characterised by breaking or restoration of this one-form global symmetry. In the following subsection, we will use this one-form global symmetry to study the ground-state degeneracy of these theories at small coupling. On a torus, these global symmetry operators are magnetic ‘t Hooft *loops* wrapping non-contractible cycles. Similar to ‘t Hooft strings, these operators are topological and depend on only the homotopy class of the loops. However, unlike the open string operators, the closed loops intersect each plaquette exactly twice and, as a result, they commute with the magnetic term of the Hamiltonian. Since they are built from the P operators, they manifestly commute with the electric term of the Hamiltonian and thus commute with the entire Hamiltonian for all values of the coupling g [257, 258].

As a result of the gauging procedure, the *local* operators that transform non-trivially under the $\mathbb{Z}_N^{(0)}$ zero-form global symmetry are projected out. In contrast, extended line operators with these charges at their end points are a part of the spectrum of the gauge theory [75]. Examples of such extended operators include the Wilson and ‘t Hooft line operators discussed earlier. Since local gauge symmetries cannot break spontaneously on a finite lattice, these extended operators can serve as order parameters for diagnosing phase transitions.

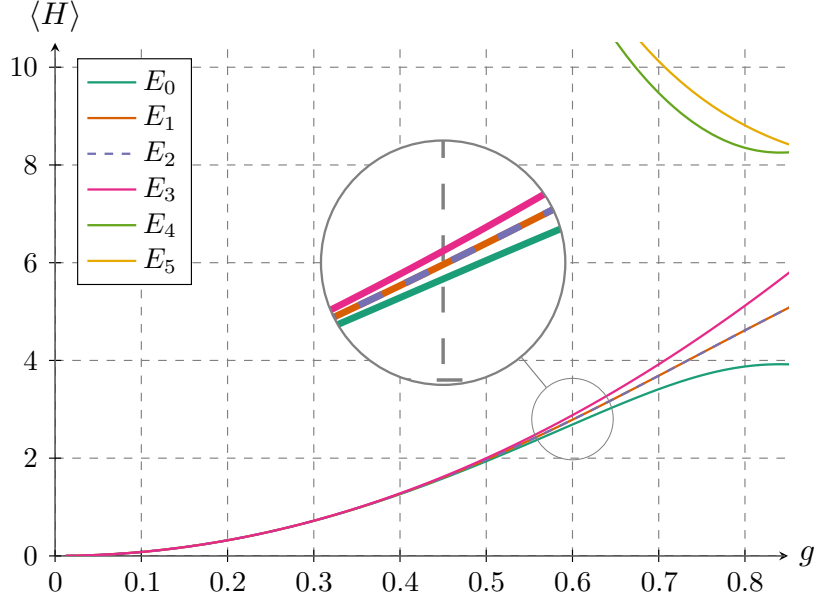


Figure 3.4: Low-lying energy levels of $\mathbb{Z}_2$ lattice gauge theory computed using exact diagonalization on an $L = 2$ lattice. Away from $g = 0$, there is a unique ground state (teal), with an approximate fourfold degeneracy for small but non-zero couplings.

3.2.5 Ground-state degeneracy

$\mathbb{Z}_N$ gauge theory has interesting low-lying state structure on lattices with non-trivial topology, such as the torus defined by our periodic boundary conditions. Focusing on $\mathbb{Z}_2$ for concreteness, these states can be discerned from the global symmetry operators wrapping the two non-contractible cycles of the torus [257]. The magnetic ‘t Hooft loops which wrap the x and y directions of the torus are denoted by V_x and V_y respectively. As with ‘t Hooft strings, these operators are topological and depend on only the homotopy class of the loops. Since these two operators commute with the Hamiltonian, even when the eigenstates of H are degenerate one can find a basis of states which are simultaneously eigenstates of the Hamiltonian and $V_{x,y}$.

In the $g \rightarrow \infty$ limit, the ground state $|\Psi_0\rangle$ is given by all links in the $+1$ eigenstate of the shift operator (X for $\mathbb{Z}_2$), and so the ground state has eigenvalue $+1$ for both V_x and V_y . States with different $V_{x,y}$ eigenvalues must have at least one spin in the -1 eigenstate of

the X operator, which will cost an energy $\sim g^2$. As a result, these states cannot be degenerate with the ground state, so that $|\Psi_0\rangle$ is unique even in the $L \rightarrow \infty$ limit. As $g \rightarrow 0$, we can express the ground state in terms of the basis of the clock operator which is Z for $\mathbb{Z}_2$. By taking linear combinations, we can express the four lowest energy states as $(1 \pm V_x)(1 \pm V_y)|0\rangle^{\otimes 2L^2}$ with eigenvalues $V_x = \pm 1$ and $V_y = \pm 1$. At $g = 0$, these states have the same energy, giving a fourfold degeneracy for the ground state. For small but non-zero couplings, these $V_{x,y}$ eigenstates are no longer degenerate, but separated by a splitting that scales as g^{4L} with the size L of the lattice [257]. As a result, the splitting becomes exponentially small deep in the deconfined phase, $g \ll 1$. This effect occurs since there is non-zero tunneling amplitude between states with distinct values of $\mathbb{Z}_2$ flux. In the general case of $\mathbb{Z}_N$ gauge theory on a closed compact oriented surface Σ with first homology group $H_1(\Sigma, \mathbb{Z}_N)$, the ground state degeneracy for $g = 0$ is given by [259]

$$\text{Ground-state degeneracy on } \Sigma = N^{|H_1(\Sigma, \mathbb{Z}_N)|}, \quad (3.19)$$

where $|H_1(\Sigma, \mathbb{Z}_N)| = 2$ if Σ is a torus. This degeneracy is one of the characteristic features of $\mathbb{Z}_N$ topological order [260–266].

The behaviors we have discussed actually persist throughout the deconfined and confined phases. In general, for $g \neq 0$ the ground state is unique and has eigenvalue $+1$ for both V_x and V_y . Below the critical coupling, for $\mathbb{Z}_2$, there is an approximate fourfold degeneracy with a splitting governed by the size of the lattice that goes to zero as $g \rightarrow 0$. This can be seen in Figure 3.4, which shows the lowest-lying states of $\mathbb{Z}_2$ gauge theory on a 2×2 lattice computed using exact diagonalization. One observes the approximate fourfold degeneracy at small couplings, with the gap growing with g .

3.2.6 Potential energy between charges

As reviewed earlier, the uncharged sector of the Hilbert space, describing the gauge field degrees of freedom, is the subsector which satisfies $\Theta_x|\Psi\rangle = |\Psi\rangle$ for all lattice sites x . We now discuss the subsectors with static charges placed at the lattice sites. Since the insertion of a single charge at a lattice site is not gauge invariant, the simplest setup one can consider is an open Wilson line with opposite charges at each end. Starting from a gauge-invariant state $|\Psi\rangle$, one can introduce charges by acting with a string of clock operators that stretch between lattice sites on which the charges are placed. For example, the following string operator can be used to place the charges a distance r apart along the x -direction:

$$W_r = Q_{x,\mu} Q_{x+\hat{x},\mu} \cdots Q_{x+r\hat{x},\mu} . \quad (3.20)$$

Note that the charges in the $\mathbb{Z}_N$ theory are constrained to be $\{0, 1, \dots, N-1\}$. Using the $\mathbb{Z}_N$ commutation relations, one finds

$$\begin{aligned} \Theta_x(W_r|\Psi) &= e^{-2\pi i/N} (W_r|\Psi) , \\ \Theta_{x+r\hat{x}}(W_r|\Psi) &= e^{2\pi i/N} (W_r|\Psi) . \end{aligned} \quad (3.21)$$

From these phases, one deduces that the Wilson line connects static charges at x and $x + r\hat{x}$, with charges $N-1$ and 1 respectively. To compute the potential between the charges, one can compare the ground-state energies with and without the charges. In practice, this means one compares

$$\langle H \rangle_0 \quad \text{vs} \quad \langle W_r^\dagger H W_r \rangle_{0'} , \quad (3.22)$$

where the expectation values are evaluated using the ground-state wavefunction in each sector, denoted by 0 and $0'$ respectively.

Let us now analyze the potential energy $V(r)$ needed to separate the charges in the limit of very small and large values of g . When $g \ll 1$, the magnetic term in the Hamiltonian (3.5) dominates and as a result the ground state is an eigenvector of Q_l with eigenvalue $+1$ for all links l . Consequently, the Wilson line operator W_r leaves the ground state unchanged. Therefore, pulling the charges out of the vacuum and separating them does not take any energy, thus the charges are deconfined.

In the opposite limit, when $g \gg 1$, the electric term in the Hamiltonian (3.5) dominates and the ground state is an eigenvector of P_l with eigenvalue $+1$ for all links l . Employing the commutation relations (3.6), we observe that the Q operators forming the Wilson line (3.20) modify the eigenvalues of P_l and $P_l^\dagger$ along the Wilson line to $e^{2\pi i/N}$ and $e^{-2\pi i/N}$, respectively. To first-order in perturbation theory, the ground-state energy in the twisted sector is given by evaluating $W_r^\dagger H W_r$ using the untwisted ground-state wavefunction.³ Plugging this back into the Hamiltonian, and neglecting the magnetic term which is very small in this limit, one finds that the energy difference is linear in the distance r

$$\begin{aligned} V(r) &\approx \langle W_r^\dagger H W_r \rangle_0 - \langle H \rangle_0 , \\ &= g^2 r (1 - \cos(2\pi/N)) . \end{aligned} \tag{3.23}$$

We see that it costs a great deal of energy to increase the separation between the charges, implying that the charges are confined. Note that one can instead place charges k and

3. Experimentally, we have observed negligible difference between evaluating the expectation of $W_r^\dagger H W_r$ using the “true” ground state in the twisted sector and the ground state of the sector without static charges, at least for couplings in the range 0.5 to 1.1. Using the $\mathbb{Z}_N$ operator identities, the result of conjugating the Hamiltonian by the Wilson string operator is $W_r^\dagger H W_r = H + \frac{g^2}{2} \sum_{i=0}^{r-1} (P_{x+i,\mu} (1 - e^{-2\pi i/N}) + \text{h.c.})$, where the second term can be thought of as a perturbation. Since the original Hamiltonian is given by a sum over $2L^2$ links, while the perturbation is a sum over only r links, the parameter controlling the size of the perturbation is $r/2L^2$ which will be small for large enough lattices. Providing the gap between the original ground state and the first excited state is large compared with the energy shift due to the perturbation, one will get a good estimate of the perturbed ground-state energy using first-order perturbation theory, where one computes the correction by evaluating the expectation value of the perturbation using the unperturbed ground state.

$N - k$ on the two ends by acting with the operator W_r^k . In this case, the potential energy in the confined phase is given by $V(r) \approx g^2 r (1 - \cos(2\pi k/N))$.

In summary, the energy needed to separate the charges differs in the two phases. In the confined phase, the potential energy rises linearly with the distance between the charges, corresponding to the formation of a confining flux tube or string between the charges. In the deconfined phase, the electric flux lines are condensed and the potential between the charges is very small, becoming independent of distance in the limit of large L .

It is useful to compare this perspective with Euclidean lattice gauge theory, where a timelike Polyakov loop (a Wilson loop extended in imaginary time) serves as an order parameter for confinement. The expectation value of this non-local operator is then related to the free energy of an isolated static charge. In particular, it should be zero in the confined phase, indicating an infinite free energy cost for isolating a single charge.

Conversely, a non-zero Polyakov loop expectation value signals deconfinement, where isolated charges can exist. To directly study confinement, one instead considers the correlation function of two timelike Polyakov loops separated in space. If the theory is in the confined phase, this correlation function should decay exponentially with the separation, reflecting the linear confining potential between the static charges represented by the Polyakov loops. Note that it is only in the continuum limit where Lorentz invariance emerges that one can directly relate spatial and timelike Wilson loops, and thus identify the string tension σ appearing in (3.17) with the coefficient of the linear potential between static charges. Away from this limit, these two quantities do not necessarily agree.

3.3 Phases and phase transitions in $\mathbb{Z}_N$ theories

The central aim of this paper is to use neural network quantum states to model the ground-state wavefunction of $\mathbb{Z}_N$ lattice gauge theories. Beyond simply computing the energy of the ground state, one can use this approach to study the phases and phase

transitions as the coupling g is varied. With this in mind, we now review some useful concepts for studying transitions and critical behavior on the lattice.

3.3.1 First-order vs continuous phase transitions

In quantum mechanics, a first-order phase transition is characterised by level-crossing in the energy spectrum. This means that, as a parameter in the Hamiltonian (such as the coupling g) is varied, the energy levels of two distinct states intersect. At the point of intersection, the ground state of the system changes abruptly from one state to another. The ground-state energy of the system

$$E_0 = \langle H \rangle_0 \equiv \frac{\langle \Psi_0 | H | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \quad (3.24)$$

has a kink at the transition point, while the derivative of the energy with respect to the coupling is discontinuous. This abrupt change is the hallmark of a first order phase transition. For the case of the first order transitions, the energy gap between the lowest energy state and the first excited state remains finite even as the ground state energy levels cross and as a result of this gap the system has finite correlation length.

For a lattice gauge theory, in the limit where the lattice size is much greater than the correlation length of the system, one expects a sharp change between an ordered and a disordered state. Moreover, the order parameters which characterise the order and disordered states also display a discontinuity at the transition. For finite L , the ordered and disordered states coexist in the vicinity of the transition. The sharp kink in the energy is smoothed somewhat, with the derivative of the energy displaying a large gradient at the transition point.

Unlike a first-order transition, a first continuous quantum phase transition occurs without level-crossing. Here, the ground state of the system evolves smoothly as a function of the coupling. There is no abrupt change in the ground state, but rather a gradual change in its

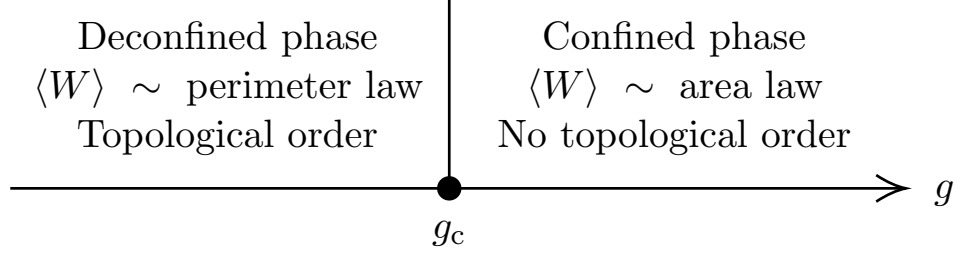


Figure 3.5: Phases of $\mathbb{Z}_N$ gauge theory in $2 + 1$ dimensions. The deconfined and confined phases are separated by a confinement transition as the critical coupling g_c .

properties. The smooth change of the state is mirrored in the smooth change of the ground-state energy, in contrast to the kink/discontinuity at a first-order transition. When such a continuous phase transition occurs, the spectral gap between the ground state and the first excited state closes. Consequently, the correlation length of the system diverges, leading to the onset of long-range correlations across the system [267]. Due to the resulting scale invariance of the system, continuous critical phenomenon can be characterised using critical exponents, which capture the behavior of physical quantities near the phase transition. For many quantum-mechanical lattice models, such as lattice gauge theory, the system at the critical coupling has an enhanced conformal symmetry in addition to scale invariance leading to a conformal field theory (CFT) [268].

3.3.2 Confined and deconfined phases

In $2 + 1$ dimensions, discrete gauge theories with a Hamiltonian given by (3.5) have two possible phases. They are in a deconfined phase at weak coupling and in a confined phase at strong coupling. As explained in Section 3.2.6, in the deconfined phase it takes very little energy to separate the charges, while in the confined phase the energy required to pull charges apart grows linearly with the length of the flux string. The expectation values of Wilson loops follow a perimeter law in the deconfined phase and an area law in the confined phase, as described in Section 3.2.1. Since the magnetic monopoles are in condensed in the deconfined phase, the value of a 't Hooft string operator is independent of

distance in this phase. The deconfined phase is topologically ordered, with a ground state degeneracy that depends on the topology of the system. This topologically ordered state has long-range quantum entanglement [269]. The $\mathbb{Z}_N^{(1)}$ one-form symmetry is preserved in the confined phase, while it is broken in the deconfined phase. These characteristics of the two phases are summarised in Figure 3.5.

3.3.3 Confinement transition

The nature of the confinement phase transition depends on N , with a first-order transition for $N = 3$ and a continuous transition occurring for all other values of N [270–272].

Evidence for this comes from Monte Carlo simulation of spin systems, which are dual to these lattice gauge theories. Furthermore, high-order perturbative expansions confirm the Monte Carlo simulation results [273, 274]. As explained in Section 3.2.4, this duality corresponds to gauging the global $\mathbb{Z}_N$ zero-form symmetry in the spin-system to obtain the lattice gauge theory. As a result, in the case of a continuous transition the theory at the critical coupling g_c is described by the gauged version of the CFT that describes the continuous transition in the corresponding spin system.

The second-order phase transition for the transverse-field Ising model is in the universality class of $2 + 1$ -d Ising Wilson–Fisher CFT. Thus, the confinement phase transition for the $\mathbb{Z}_2$ lattice gauge theory belongs to the $2 + 1$ -d gauged Ising universality class [257]. This theory is also sometimes referred to as the Ising* theory. In Section 3.5, we calculate the critical exponents for this transition using neural network quantum states. For $\mathbb{Z}_3$, we observe the first-order nature of the phase transition and determine the approximate location of the critical coupling in Section 3.6.

Remarkably, the $\mathbb{Z}_4$ theory decomposes into two decoupled $\mathbb{Z}_2$ theories for every value of the coupling, with the critical point hosting two copies of the $2 + 1$ -d Ising* CFT [275]. Monte Carlo simulations suggest that the continuous transitions for $N > 4$ belong to the

2 + 1-d XY (O(2) model) universality class [272, 276]. The critical coupling at which the confinement transition occurs goes to zero as the value of N increases. The $N \rightarrow \infty$ limit recovers U(1) lattice gauge theory, which is confined for all values of the coupling [270]. As shown by Polyakov, this is a result of the fact that certain monopole operators are relevant, leading to a phase of unbroken $\mathbb{Z}_N^{(1)}$ one-form global symmetry [277, 278]. The phase diagram is much richer in 3 + 1 dimensions, with two phase transitions occurring for $N \geq 5$ [52, 279]. Investigating these phase transitions in 3 + 1 dimensions with neural network quantum states will be a fruitful direction for future work.

Note that the critical couplings (or inverse temperatures) at which the transition occurs computed with Monte Carlo methods are not directly comparable to the critical couplings that one finds from a Hamiltonian approach. The reason for this is that the time direction of the $d + 1$ -dimensional spacetime is also discretized. In order to recover the Hamiltonian perspective, one must take a certain anisotropic limit of the lattice [280, 281], which comes with a non-trivial rescaling of the couplings of the theory. For this reason, other than for simple examples, it is difficult to directly compare the location of phase transitions in the two approaches.

3.4 Neural network quantum states

Monte Carlo	Neural network quantum states
Computes path integral	Computes ground state
$d + 1$ -dimensional spacetime lattice	d -dimensional spatial lattice
Sign problem	No sign problem
Asymptotic convergence guarantee	No convergence guarantee
Difficult sampling problem	Difficult optimization problem
Few hyperparameters	Many hyperparameters

Table 3.1: Brief comparison of path-integral Monte Carlo (MC) and neural network quantum state (NNQS) approaches.

There is a long history of using the variational method to find approximations to the

ground-state wavefunction of quantum-mechanical systems [282]. This relies on the observation that, given the system's Hamiltonian H and a normalisable wavefunction Ψ_θ , where θ are parameters that specify the particular wavefunction from a family of variational states, the functional⁴

$$E_\theta = \langle H \rangle_\theta \equiv \frac{\langle \Psi_\theta | H | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle}, \quad (3.25)$$

is bounded from below by the true ground-state energy E_0 of the system, with $E_\theta = E_0$ if and only if Ψ_θ is the exact ground-state wavefunction of the system. Thus, by minimizing E_θ with respect to the parameters θ , one obtains an approximation to both the ground-state energy and wavefunction itself.

The more flexible the variational ansatz, the better the resulting approximation should be. Moreover, if the form of Ψ_θ is particularly well suited to the system, one expects that fewer parameters are needed to obtain a good approximation. For example, for the simple harmonic oscillator, an ansatz of a Gaussian in the square of the displacement needs only a single parameter to describe the exact ground state, while an ansatz in terms of Fourier modes would need more parameters (and an infinite number to capture the exact ground state). Finding a good approximation to the ground state of a complicated quantum system then requires variational states which are well suited for the system under consideration and flexible enough to capture the relevant physics.

Neural network quantum states (NNQS) were developed in 2016 by Carleo and Troyer as a new kind of variational ansatz for quantum systems [58]. Their idea was to use a neural network as the trial wavefunction of a quantum system, with the parameters of the network chosen to minimize the expectation value of the Hamiltonian. In their most basic form, neural networks give a map from inputs to outputs, with the map given by compositions of

4. Here we are explicitly dividing by the norm of the variational state. For ease of notation, we often suppress this and assume that the state is correctly normalized, though in practice it is simpler to work with unnormalized NNQSs.

linear transformations and non-linear activation functions. In our case, we are interested in networks which map from a configuration of link variables to a single complex number, $\mathbb{C}^{2L^2} \rightarrow \mathbb{C}$, so that the network can be interpreted as assigning a probability amplitude to a given lattice configuration. The linear transformations can be thought of as acting with matrices whose entries are known as “weights”, which are then interpreted as variational parameters. The non-linear activation functions, often alternated with the linear transformations, result in the network output being a complicated non-linear function of the variational parameters. Thanks to this, NNQSs can capture a wide range of phenomena, including ground-state wavefunctions.

A NNQS gives a variational ansatz Ψ_θ for the wavefunction parametrised by the choice of weights θ . The next question is how to choose these weights so that Ψ_θ provides an approximation to the ground state of a given quantum system. Again, one can use the variational method by trying to minimize the expectation value of the Hamiltonian with respect to the weights of the neural network. Since the network is a complicated non-linear function of these weights, and these weights often number in the thousands or tens of thousands, it is not possible to solve this minimization problem exactly. Instead, one resorts to numerical methods to iteratively reduce E_θ by varying the weights. The way to do this is (stochastic) gradient descent. The key to this is the fact that modern neural network packages allow for automatic differentiation, so that one can differentiate E_θ with respect to the weights θ and evaluate the resulting gradient exactly (that is, without using finite differences). Given this gradient, one adjusts the weights to move in the direction of steepest descent. By repeatedly iterating, one hopes to “train” the network and eventually find the set of weights which minimize E_θ . At the end of training, one not only has an estimate of the ground-state energy, but also an approximation Ψ_θ for the exact ground-state wavefunction Ψ_0 .

MC vs NNQS

Before describing the gauge-invariance networks we will use in this paper, let us quickly compare traditional Monte Carlo (MC) methods with neural network quantum states (NNQS) for studying lattice theories:

- Dimensionality of the lattice: In NNQS, the trial wavefunction is defined on the d -dimensional spatial lattice. In MC, fields are defined on a $d + 1$ -dimensional lattice that includes the imaginary time direction.
- Variational principle: NNQS relies on a variational principle, wherein the trial wavefunction is optimized to minimize the energy expectation value. MC does not involve any optimisation, but instead aims to directly evaluate the path integral on the lattice using Monte Carlo sampling.
- Sampling problem: In NNQS, one generates configurations of gauge fields on a d -dimensional lattice weighted according to the trial wavefunction. In MC, one instead samples $d + 1$ -dimensional lattice configurations weighted according to the Euclidean action e^{-S} .
- Sign problem: In Monte Carlo (MC) simulations, a sign problem occurs because the fermion determinants in the path integral can turn complex in the presence of a finite chemical potential or a non-zero theta angle. This complexity makes it difficult to treat e^{-S} as a straightforward probability measure, which prevents the use of standard Monte Carlo methods. Moreover, even when one can absorb the complex phase, evaluating expectation values often requires extremely high accuracies due to possible cancellations. This sign problem is absent in NNQS, since one samples from the Born distribution $|\Psi|^2$, meaning that standard Monte Carlo sampling can be

used.⁵

- **Convergence:** In the absence of sign problem, MC is guaranteed to converge asymptotically when a sufficient number of samples have been drawn. Nevertheless, number of steps required to adequately sample the system increases as the critical point is approached due to critical slowing, making simulations near criticality computationally expensive [283]. In contrast, a fixed network has limited expressivity [284] and may not be capable of representing the true ground state of a system. Furthermore, even if a network is capable of representing the ground state, there is no guarantee that gradient based optimization methods will efficiently find the optimal set of parameters.
- **Systematic improvements:** In MC, systematic improvements come from improving the Monte Carlo sampling and summing over more lattice configurations. The accuracy of NNQS can be systematically improved by using more sophisticated trial wavefunctions with more variational expressivity. However, there is a trade-off between expressivity and the efficiency of optimization.
- **Observables:** In MC, observables are calculated as ensemble averages over the generated lattice configurations, which are related to the path integral. In NNQS, observables are calculated directly as expectation values of operators with respect to the trial wavefunction.
- **Hyperparameters:** NNQS involve a large number of hyperparameters including the number of layers in the network, the number of neurons per layer, the type of activation functions, the initialization methods and the learning rate. The selection and tuning of these hyperparameters can greatly influence the learning dynamics and

5. Note that one can still run into sign problems when considering operators with non-trivial phase structure, so that expectation values rely on many cancellations, leading to large variances in Monte Carlo estimates.

the accuracy of the models. The key hyperparameters in the MC approach are the number of samples and the choice of update algorithm. Although the update algorithm affects the rate of convergence, the overall method is quite robust.

Another point to emphasise are the trade-offs between MC and NNQS in situations where both can be used. One swaps a higher-dimensional sampling problem for a lower-dimensional sampling problem with an additional optimisation problem. Naively, this optimisation problem may eat up any efficiency gain in reducing the dimension of the lattice. However, this optimisation problem can be put on hardware designed for large-scale machine learning tasks, such as clusters of GPUs, while utilising existing software libraries which implement automatic differentiation, etc. In addition, one also benefits from “transfer learning” – the wavefunction for a lattice system found for one value of a coupling will give a good starting point for the wavefunction at a nearby coupling. Thanks to this, one does not have to solve the optimisation problem from scratch when scanning over couplings or probing phase diagrams. The differences between the two methods discussed above are summarised in Table 3.1.

3.4.1 L-CNNs

The neural network architecture that we will use to approximate the ground state of a $\mathbb{Z}_N$ lattice gauge theory is a so-called “lattice gauge-equivariant convolutional neural network”, or L-CNN. This particular architecture was introduced by Favoni et al. in [64] as a way to approximate a large class of gauge-equivariant or gauge-invariant functions of a lattice system.⁶ In our case, since the network should approximate the ground-state wavefunction of the system, we want the network output to be gauge invariant.

An L-CNN is made up of constituent “layers”. Depending on their construction, the layers can do a variety of operations, including gauge-equivariant convolutions and

6. See also [225, 226] for an alternative gauge-equivariant architecture.

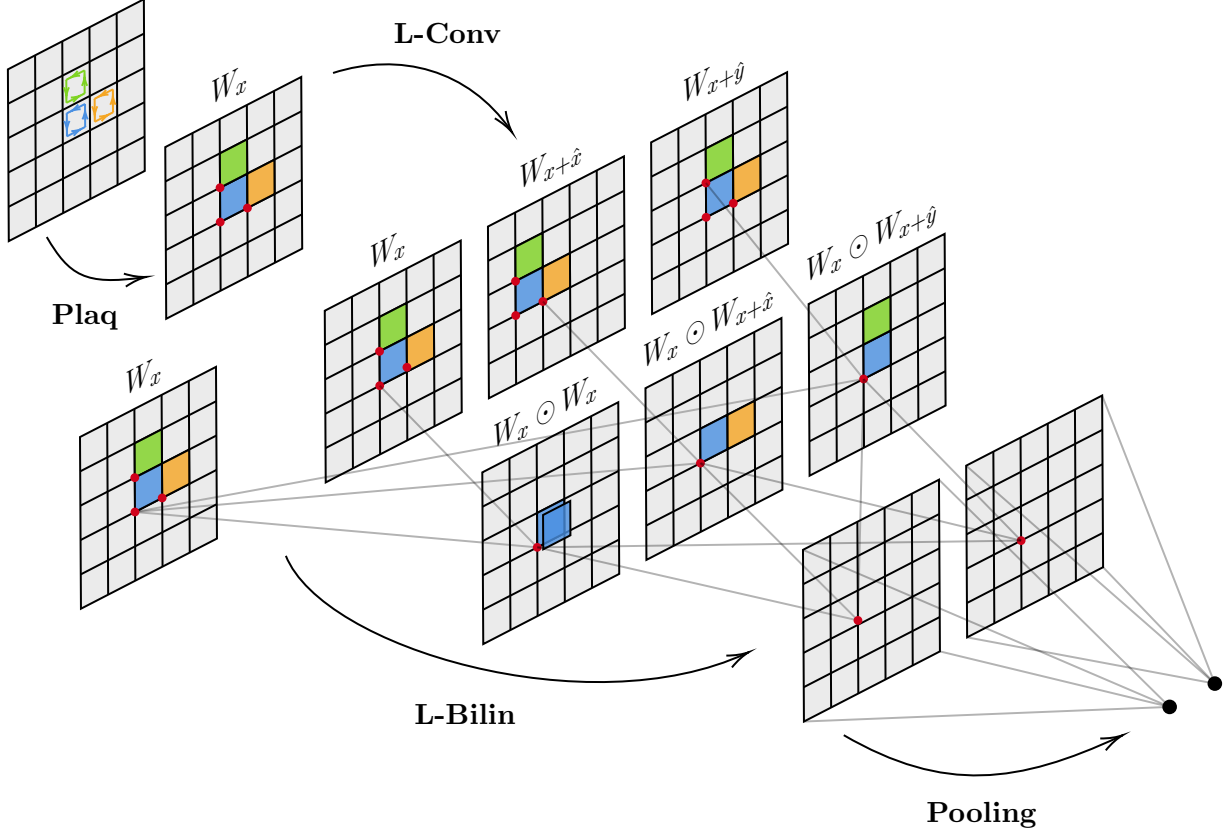


Figure 3.6: A schematic of an L-CNN with a single **L-CB** layer with $(N_{\text{out}}, K) = (2, 1)$ and no **Dense-SELU** layers. Here, the input is a configuration of link variables $\mathcal{U} = \{U_{x,\mu}\}$. The first layer, **Plaques**, converts the link variables into plaquette variables and stores them as $\mathcal{W} = \{W_x\}$ in a single channel. The **L-Conv** layer then parallel transports the plaquette variables by up to one lattice site in the positive x and y directions, giving a new set of plaquette-like variables $\mathcal{W}' = \{W_x, W_{x+\hat{x}}, W_{x+\hat{y}}\}$ with three channels. The sets $\mathcal{W}$ and $\mathcal{W}'$ are then multiplied by element-wise multiplication (denoted by the Hadamard product $\odot$), giving $W_x \odot W_x$ (squares of 1×1 loops), $W_x \odot W_{x+\hat{x}}$ (2×1 loops), and $W_x \odot W_{x+\hat{y}}$ (1×2 loops). Since $N_{\text{out}} = 2$, one then takes linear combinations of these (determined by the weights α_{ijk}) to give two output channels. Finally, a global pooling layer averages each of these channels over the lattice, giving two neurons which can be processed further. To help the reader follow variables through the network, we have highlighted three 1×1 plaquettes in blue, green and orange. Red dots indicate the lattice points to which the plaquette variables are associated. For example, the action of $W_x \odot W_{x+\hat{x}}$ equivariantly multiplies the blue plaquette by the orange plaquette (since they are associated to the same lattice point), resulting in a 2×1 plaquette.

multiplications, and acting with activation functions. For our purposes, we will need three of these layers, namely a plaquette layer, a convolution layer and a bilinear layer. At each layer, we keep track of two sets of data. The first is the set of link variables $\mathcal{U} = \{U_{x,\mu}\}$, which transforms non-locally under gauge transformations, as in (3.4). The second set transforms locally under gauge transformations, as in (3.13). We refer to these collectively as $\mathcal{W} = \{W_{x,i}\}$. Here, i is a “channel” index which allows us to associate multiple $\mathcal{W}$ elements to the same lattice site at x . The plaquette variables $\{U_{x,\mu\nu}\}$ are an example of these with a single channel; an L-CNN will naturally construct more general quantities which transform in the same way. We give a schematic diagram of an L-CNN in Figure 3.6. Each layer of the L-CNN can be thought of as acting on the pair $(\mathcal{U}, \mathcal{W})$. The initial input to the network is the set of link variables $\mathcal{U}$, describing a gauge field configuration on the lattice, while $\mathcal{W}$ is initially empty. The first layer carries out “preprocessing”, generating the plaquette variables from the link variables. This is done via a **Plaq** layer:

$$\mathbf{Plaq}: U_{x,\mu} \mapsto U_{x,\mu\nu} . \quad (3.26)$$

The plaquette variables are then stored in $\mathcal{W}$. As in [64], to reduce redundancy, we generate only the anticlockwise plaquettes (those with positive orientation in higher dimensions).

The next layer allows the parallel transport of objects stored in $\mathcal{W}$ from one lattice site to a neighboring site. This is implemented as a convolutional layer, **L-Conv**, given explicitly by

$$\mathbf{L-Conv}: \quad (3.27)$$

$$(U_{x,\mu}, W_{x,i}) \mapsto \sum_{j,\mu,k} \omega_{ij\mu k} U_{x,k\mu} W_{x+k\mu,j} U_{x,k\mu}^\dagger ,$$

where $\omega_{ij\mu} \in \mathbb{C}$ are the “weights” or parameters of the convolutional layer, i and j run over the number of output and input channels respectively, and $\mu \in \{\hat{x}, \hat{y}\}$ for a $d = 2$ lattice.

The index k runs over $0, \dots, K$, where K is an integer which determines the maximum

lattice distance to translate the $\mathcal{W}$ quantities, or equivalently the kernel size of the convolution.⁷

Finally, we need a layer which multiplies two sets $\mathcal{W}$ and $\mathcal{W}'$ in an equivariant manner.

This is done via an **L-Bilin** layer:

$$\mathbf{L-Bilin}: (W_{x,i}, W'_{x,i'}) \mapsto \sum_{j,k} \alpha_{ijk} W_{x,j} W'_{x,k} , \quad (3.28)$$

where $\alpha_{ijk} \in \mathbb{C}$ are weights, j and k run over the number of input channels for $\mathcal{W}$ and $\mathcal{W}'$ respectively, and i runs over the number of output channels. Since this layer multiplies locally transforming variables at the same position x , the output is again locally transforming, so that the layer output is gauge equivariant. For $\mathbb{Z}_N$ gauge theories, the gauge group is abelian and so the link variables are simply phases. Thanks to this, the traced and untraced Wilson loops are equivalent, and any function of the $\mathcal{W}$ variables is automatically gauge invariant.

In practice, prior to multiplying, $\mathcal{W}$ and $\mathcal{W}'$ are extended by including the Hermitian conjugate of all their elements, and the unit matrix $\mathbf{1}$ at each lattice site. A little thought should convince the reader that this allows the layer to include a bias and act as a residual module (i.e. the output also contains a linear combination of the inputs) [285]. As in [64], we combine **L-Conv** and **L-Bilin** into a single **L-CB** layer with a single set of trainable weights. A choice of **L-CB** layer is then fixed by a choice of (N_{out}, K) , i.e. the number of output channels and the kernel size of the convolution.

The power of an L-CNN is in the fact that by stacking **L-CB** layers, one can construct arbitrary (untraced) Wilson loops, and so approximate any gauge equivariant function. For example, after a **Pla** layer, one has all 1×1 Wilson loops. Following this with an **L-CB** layer, the output includes linear combinations of 1×2 and 2×1 Wilson loops (and 1×1

7. We restrict to non-negative shifts along the lattice.

loops and squares of 1×1 loops when the $\mathcal{W}$ variables are extended by the unit matrix). If the number of output channels is large enough, in principle, one can capture all possible Wilson loops of area two and below. With another **L-CB** layer, the output includes loops up to area four. This also makes clear that L-CNNs exploit locality on the spatial lattice, with nearby Wilson loops more likely to be multiplied together compared with distant loops.

Clearly, the number of possible Wilson loops grows very quickly with area, so quickly that one cannot hope to optimize a variational ansatz constructed by simply taking combinations of all loops. Instead, by restricting the size of the output channels, an L-CNN works with a much smaller number of combinations of loop variables. During training, the network then determines which combinations to keep within this much smaller subspace. It is this restriction that ensures an L-CNN variational ansatz has a sub-exponential number of parameters. For example, for $\mathbb{Z}_2$ gauge theory on a 10×10 lattice, there are $2^{2 \times 10^2} \approx 10^{60}$ possible lattice configurations, and so modeling the wavefunction as a “look-up table” that assigns an amplitude to each of these is clearly intractable.⁸ Instead, using an L-CNN with seven **L-CB** layers with $(N_{\text{out}}, K) = (4, 2)$ leads to a variational ansatz with approximately 10^4 parameters. As we will see, this network is sufficient to accurately capture the physics of the ground state, and so an L-CNN clearly gives an efficient encoding of the wavefunction.

3.4.2 NetKet implementation

We have implemented this network architecture using NetKet [227], a machine-learning framework for many-body quantum physics. NetKet is built on top of JAX [228], a framework for Python which allows automatic differentiation and GPU acceleration, with

8. This is obviously an overcounting, as the ground-state wavefunction depends only on gauge-invariant data and should also be invariant under translations, etc., but the rough level of complexity is what the reader should take away.

its neural network components implemented using Flax [286]. The network output is taken to be $\log \Psi_\theta$. Thanks to the complex weights of the network, the output is complex, so can accommodate a wavefunction with non-trivial phase structure. Schematically, as a functional, the wavefunction is given by the composition of the following layers

$$\log \Psi_\theta \equiv \mathbf{Dense} \circ \dots \circ \mathbf{Dense-SELU} \circ \mathbf{Pooling} \circ \dots \circ \mathbf{L-CB} \circ \mathbf{Plaq} , \quad (3.29)$$

where the final layers are fully connected dense layers, each followed by a scaled exponential linear unit layer (**Dense-SELU**), and a dense layer with a *single* neuron (**Dense**). Each **L-CB** layer is labeled by a choice of (N_{out}, K) , the number of output channels and the kernel size, while each **Dense-SELU** layer is fixed by N_{feat} , a choice of the number of neurons or “features” in the dense layer.

The network is trained using the in-built features of NetKet. Specifically, training attempts to minimize E_θ , the expectation value of the Hamiltonian in the variational state. It does this using automatic differentiation to compute the derivative (gradient) of E_θ with respect to the parameters (weights) of the network, and then stochastic gradient descent to move in the steepest direction of lower energy. In addition, NetKet includes the option to precondition the gradient – we use the quantum geometric tensor in all numerical experiments in this paper, with the resulting dynamics known as stochastic reconfiguration.⁹ We further elucidate the training process below.

Sampling

Gradient descent requires the calculation of E_θ and its gradient with respect to the parameters of the network at each training step. In principle, computing these quantities requires summing over the Hilbert space of the system. Since the Hilbert space of gauge field configurations is too large to sum over exactly, one must instead use stochastic

9. See, for example, [227, Section 4.1] for a discussion of this.

gradient descent with an estimate of the gradient. The expectation values of observables and their gradients with respect to the network parameters are computed using a representative sample of configurations which approximate the full sum over the Hilbert space. Here, representative means that a configuration $\mathcal{U}$ is sampled according to its probability $|\Psi_\theta(\mathcal{U})|^2$. These are selected via a standard local Metropolis algorithm. As Ψ_θ converges to the ground-state wavefunction, the sampling becomes better at reproducing the sum over the Hilbert space.

Obviously, in order to actually move in the direction of decreasing energy, one needs reasonably accurate estimates of the gradient of the energy. Since this gradient is calculated approximately by summing over a sample of lattice configurations, one might worry that this gradient (or the energy itself) cannot be estimated to sufficient accuracy without using a very large number of samples. However, the Hamiltonian is a particularly well-behaved observable as it satisfies the “zero-variance property”.¹⁰ Given a trial wavefunction Ψ_θ , one can consider both the variational error in the energy, $E_\theta - E_0$, and the variance, $\langle (H - E_\theta)^2 \rangle_\theta$. One can show that both the variational error and the variance are second order in the difference between the trial wavefunction Ψ_θ and the true ground-state wavefunction Ψ_0 . In the limit where $\Psi_\theta = \Psi_0$, both the error and the variance vanish, so that $E_\theta = E_0$. Moreover, since the variance determines the statistical error, the error in computing the gradient of the variational energy due to Monte Carlo sampling also decreases as Ψ_θ approaches Ψ_0 . Thanks to this, one does not need a large number of samples when estimating the energy or its gradient.

For the experiments described in this paper, each stochastic gradient descent step is evaluated using 4096 configurations. The number of sweeps (the number of Metropolis steps taken before returning a sample, i.e. the subsampling factor of the Markov chain) is chosen to be equal to the number of degrees of freedom of the Hilbert space. This is simply

10. See, for instance, the discussion in [287]

the number of links, so that for an $L \times L$ lattice, $2L^2$ sweeps are made. Combined with a warm-up phase, we found this to be sufficient to ensure reasonably small correlation of the Monte Carlo chains and an acceptable autocorrelation time. When approaching critical couplings, an increase in the number of sweeps was needed.

initialization

Since **L-CB** layers are multiplicative, when using deep networks, it is essential to properly initialize the network weights. A little thought should convince the reader that for a deep network, if the weights are initially too small or too large, one will quickly run into vanishing or exploding gradients. For fully connected or convolutional networks with standard activation functions, there are analytic results for choosing a good initialization. Without similar results for multiplicative networks, such as an L-CNN, the best one can do is to choose the initialization empirically. Following [288], one can do this via “layer-sequential unit-variance” initialization. The idea is that one starts with some distribution of weights for each layer, with known standard deviations. One then proceeds, layer by layer, changing the standard deviation so that the output of each layer has the same variance as the previous layer. In this way, the gradient should not vanish nor explode. We have implemented this for all the networks that we discuss. We find that this is essential for ensuring that deep networks do not immediately diverge, nor take a long time to begin training.

Training details

For all the networks in this paper, we used stochastic gradient descent together with a preconditioning of the gradient via the quantum geometric tensor. In the variational Monte Carlo community, this is known as stochastic reconfiguration [289, 290]. The update rule

for the weights θ of a variational state $|\Psi_\theta\rangle$ is

$$\theta \rightarrow \theta - \eta G^{-1} \nabla_\theta E_\theta , \quad (3.30)$$

where η is the learning rate and G^{-1} is the (pseudo-)inverse of the quantum geometric tensor. This tensor (also known as the quantum Fisher matrix) is the metric tensor induced by the Fubini–Study distance between pure quantum states. The resulting dynamics is akin to what is known as “natural” gradient descent in the machine-learning community [291], where G^{-1} takes into account that the space of states has non-trivial geometry [292]. In practice, G is often ill-conditioned, and so a small diagonal shift proportional to the identity matrix is often added before inversion. For large diagonal shifts, the identity matrix will dominate G , leading to standard stochastic gradient descent with the Euclidean metric on weight space. This should still converge to the ground state, though may be much slower than choosing an optimally small value of the shift.

In our experiments, the learning rate was initially set to 0.03 and reduced via cosine decay over 500 iterations. Stochastic reconfiguration was implemented following the “MinSR” algorithm [293, 294] using NetKet’s experimental `VMC_SRt` driver with a diagonal shift of 10^{-4} . Double (FP64) precision was used for all experiments, as we found this led to more stable training of the multiplicative layers that make up an L-CNN.

As discussed in [295], the variance of the energy is useful for tracking the convergence of a variational method.¹¹ Given an approximate eigenstate with energy $E = \langle H \rangle$ and $\Delta E = \sqrt{\text{var } H} = \sqrt{\langle H^2 \rangle - \langle H \rangle^2}$, there will be an exact energy eigenvalue E_{exact} within ΔE of the energy E , $|E - E_{\text{exact}}| \leq \Delta E$ [297, 298]. The variance thus gives an upper bound on how far an approximate state is from an exact energy eigenstate, and thus can be

11. See also [296] for the “V-score”, which can be thought of as a system-size agnostic definition of the accuracy of a variational state.

used as a stopping condition.¹² Training is stopped once the network has converged, which is indicated by E_θ no longer changing and the variance of H becoming sufficiently small.¹³ Upon convergence, the resulting trained network should then approximate the ground-state wavefunction.

Transfer learning

One of the advantages of computing the wavefunction of a quantum system, rather than using path-integral methods to compute observables directly, is the possibility of employing transfer learning. This relies on the fact that on a finite lattice, small changes in the parameters of the Hamiltonian should lead to only small changes in the ground-state wavefunction. We exploit this when performing scans over the coupling: by starting with the network weights corresponding to a nearby, previously learned wavefunction, the new NNQS is already relatively close to the sought-for ground state. This also helps with stability near to critical values of the of coupling, since the NNQS starts from a nearby wavefunction (in state space), rather than trying to converge from a generic state. We find this form of transfer learning helps the network to converge and greatly decreases overall training time.

There is also a second kind of transfer learning that we could take advantage of (though we did not in our experiments). Since an L-CNN is a convolutional network, it implements weight sharing for different lattice sites. Said differently, the trainable weights in an **L-CB** layer are encoded in a rank-three tensor with entries $\alpha_{ijk} \in \mathbb{C}$, where the indices run over the output and input channels. The number of these channels does not depends on the hyperparameters (N_{out}, K) for each layer, but does not depend on the size of the

12. One might worry that training might become stuck at an excited state, and that the variance cannot be used to discriminate between this and the true ground state. Fortunately, stochastic reconfiguration is excellent at driving the network to the ground state, after which the variance stopping condition can be trusted.

13. In practice, we use the sample variance of E_θ to estimate the variance of H while training.

underlying lattice. This means that one can train an L-CNN on a small lattice, and then transfer the network weights to an L-CNN for a larger lattice. Presumably, this would provide a good starting point for learning the wavefunction on the larger lattice.

Calculating observables

At the end of training, one has a NNQS which approximates the ground-state wavefunction of the system. With this in hand, one would like to compute other observables in order to probe various aspects of its physics. Unlike the Hamiltonian, general observables do not enjoy the zero-variance property. This has important consequences for computing accurate expectation values. Given an observable $\mathcal{O}$, the variational error $\langle \mathcal{O} \rangle_\theta - \langle \mathcal{O} \rangle_0$ is no longer quadratic in the difference $\Psi_\theta - \Psi_0$, but changes linearly. Moreover, the variance $\text{var}_\theta \mathcal{O} = \langle (\mathcal{O} - \langle \mathcal{O} \rangle_\theta)^2 \rangle_\theta$ remains order one (effectively because the exact ground-state does not have to be an eigenstate of $\mathcal{O}$). Thus, the statistical fluctuations of $\langle \mathcal{O} \rangle_\theta$ can be large, requiring many samples to reduce the standard error in the estimate, which naively goes as the square root of $\text{var}_\theta \mathcal{O} / n_{\text{samples}}$ modulo the effects of autocorrelation.

3.4.3 Hilbert space sector of the NNQS

The ‘t Hooft loop operators along the two one-cycles of the torus are the generators of the $\mathbb{Z}_N^{(1)}$ one-form global symmetry and commute with the Hamiltonian. Therefore, the Hilbert space can be decomposed into selection sectors based on the eigenvalues of the ‘t Hooft loop operators [299]. Within each sector, the states can be transformed from one to another by acting with contractible Wilson loop operators comprised of gauge field on links. This is in contrast to states belonging to different sectors that cannot be transformed into each other in this manner. Given our discussion of ground-state degeneracies in the deconfined phase in Section 3.3.3, one might wonder which sector the NNQS belongs to. Is the neural network finding some superposition of the approximately degenerate ground

states? In fact, as we now show, the network architecture ensures that one is always in the trivial ‘t Hooft loop sector. Therefore, the NNQS recovers the true ground states in both the confined and deconfined phases. To see this, as we review in supplementary section 3.7.1, recall that the expectation value of an observable $\mathcal{O}$ can be expressed as

$$\langle \mathcal{O} \rangle = \sum_{\mathcal{U}} \frac{|\Psi(\mathcal{U})|^2}{\langle \Psi | \Psi \rangle} \left(\sum_{\mathcal{U}'} \frac{\Psi(\mathcal{U}')}{\Psi(\mathcal{U})} \langle \mathcal{U} | \mathcal{O} | \mathcal{U}' \rangle \right) . \quad (3.31)$$

The summation over all possible field configurations $\mathcal{U}$ is usually approximated by sampling using Markov chain Monte Carlo (MCMC). The sum over $\mathcal{U}'$ inside the parentheses receives contributions only from “connected configurations”, that is, configurations where $|\mathcal{U}\rangle$ and $\mathcal{O}|\mathcal{U}'\rangle$ have non-zero overlap. In the case where $\mathcal{O} = \Theta_x$ is a local generator of gauge transformations on the lattice, there is only one connected configuration, and $\Psi(\mathcal{U})$ and $\Psi(\mathcal{U}')$ must be equal since the network used to calculate Ψ is gauge invariant by construction. The expectation value then reduces to an average of 1’s, giving $\langle \mathcal{O} \rangle = 1$, implying that the wavefunction is gauge invariant.

A similar argument applies when $\mathcal{O}$ is taken to be any of the ‘t Hooft loop operators that wrap the torus. Focusing on $N = 2$ for concreteness, one has $\mathcal{O} \in \{V_x, V_y\}$. These operators commute with all contractable Wilson loops, and so they do not change the expectation values of Wilson loops at all:

$$\langle \mathcal{U}' | W | \mathcal{U}' \rangle = \langle \mathcal{U} | \mathcal{O}^\dagger W \mathcal{O} | \mathcal{U} \rangle = \langle \mathcal{U} | W | \mathcal{U} \rangle . \quad (3.32)$$

Since our variational wavefunction Ψ is built from contractable Wilson loops of all sizes, it is invariant under V_x and V_y by construction. Therefore, we always have $(V_x, V_y) = (1, 1)$, which are the quantum numbers of the true ground state (and not of the approximately degenerate states). In order to explore other sectors, one can simply conjugate the Hamiltonian with appropriate combinations of the V_x and V_y operators.

3.5 $\mathbb{Z}_2$ gauge theory in $2 + 1$ dimensions

We now turn to the study of $\mathbb{Z}_2$ lattice gauge theory in $2 + 1$ dimensions [250, 253, 281, 300–302]. There are surprisingly few direct numerical studies of the ground state of $\mathbb{Z}_2$ lattice gauge theories, mainly due to the requirement of gauge invariance. Instead, the literature has focused either on path-integral Monte Carlo or simulating the dual spin system. Our focus will be on finding the ground-state wavefunction itself, allowing us to compute the ground-state energy as a function of coupling, to identify the critical coupling and the confined-deconfined phases, and to calculate estimates for the critical exponents that characterise the conformal field theory governing the phase transition. We also investigate the potential between two static charges on the lattice. As reviewed in [258], $\mathbb{Z}_2$ gauge theory at zero temperature is known to have two phases: an ordered (deconfined) phase and a disordered (confined) phase. These phases are distinguished by the behavior of the order and disorder parameters, given by expectation values of Wilson loop and ‘t Hooft string operators. $\mathbb{Z}_2$ gauge theory is dual to a classical Ising model in $2 + 1$ dimensions, and therefore the second-order phase transition at g_c is in the universality class of the three-dimensional gauged Ising CFT. The ordered phase is expected to appear for couplings below a critical coupling g_c , and is characterised by perimeter-law decay for Wilson loops. The $\mathbb{Z}_2^{(1)}$ one-form symmetry is broken in this phase. The slow decay of Wilson loops indicates that the ground state of the theory is dominated by these operators. One then says that the corresponding electric flux lines are “condensed” and the theory is deconfined. Above the critical coupling, the theory is in a disordered phase, with the Wilson loop expectations following an area-law decay. The ‘t Hooft string expectations are constant (independent of distance) due to the condensation of magnetic monopoles. The $\mathbb{Z}_2^{(1)}$ one-form symmetry is unbroken and the theory is confined.

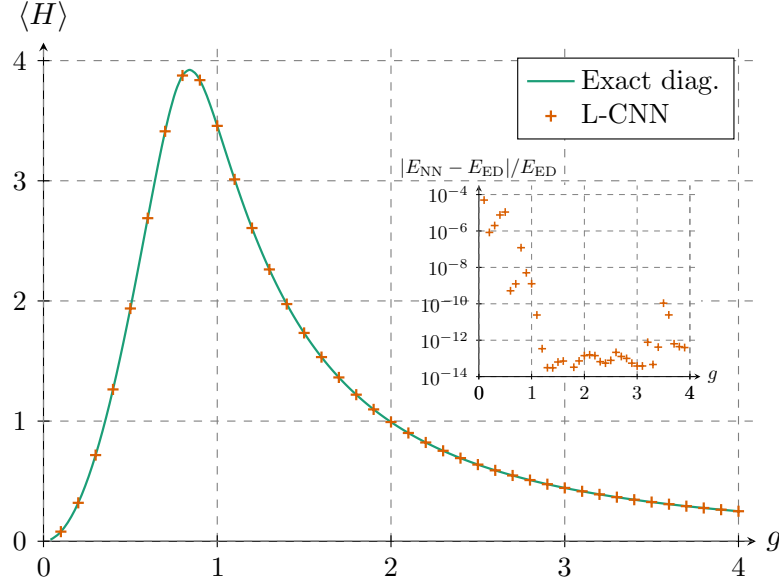


Figure 3.7: Expectation value of $\langle H \rangle$ for $\mathbb{Z}_2$ gauge theory (3.5) on a 2×2 lattice as a function of coupling g . The ground state was computed using an L-CNN with six **L-CB** layers with hyperparameters $(N_{\text{out}}, K) = (3, 1)$. The networks were trained for up to 100 iterations with 4096 samples, though they converged within 25 or so iterations. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point. The solid teal line is the exact diagonalization result, computed using NetKet. The inset shows the fractional difference between the ground state energies calculated using the neural network and exact diagonalization.

3.5.1 Ground-state energies

As a first test of the L-CNN, we compute the ground-state energy of pure $\mathbb{Z}_2$ lattice gauge theory in $2 + 1$ dimensions. In Figure 3.7, we plot the expectation value of the lattice Hamiltonian (3.5) in the ground state as a function of coupling g for a 2×2 lattice. We see excellent agreement with the energy calculated by exact diagonalization.

In Figure 3.8, we plot the ground-state energy per lattice site, $\langle H \rangle / L^2$, for couplings in the range $g \in [0.5, 1.1]$ with lattice sizes $L = 2, \dots, 10$. By eye, one sees that the $L = 10$ values are likely already very close to the continuum energy per site and that even relatively small lattices provide a good estimate, with only $L = 2$ showing large deviations. We quantify this further in Figure 3.9, which shows the difference between the ground-state energy per lattice site for $L = 10$ and smaller values of L . Other than for $L = 2$, away from $g \in [0.7, 0.8]$, the estimates agree to better than 10^{-4} to 10^{-6} . It is also interesting to observe that the differences are maximised for $g \approx 0.76$ which, as we will see, is in the vicinity of the phase transition.

3.5.2 Searching for the critical point and phase structure

The critical coupling g_c for the $\mathbb{Z}_2$ gauge theory has not been directly computed previously. From [225, Figure 5], the behavior of the string tension on a 12×12 lattice suggests $g_c \approx 0.74$.¹⁴ However, this comes from eyeballing the area-law scaling of Wilson loops and is accurate to, at best, one significant figure. Later work suggested a transition around $g_c \approx 0.76$ from considerations of ‘t Hooft string expectations and derivatives of the energy [231]. Instead, the most accurate identifications of the critical coupling come from a Monte Carlo analysis of the dual spin system, the quantum transverse-field Ising model. The authors of [280] simulate an anisotropic limit of the Ising model on a $2 + 1$ -dimensional

14. The critical coupling is given in [225] as $h \approx 0.3$. Examining their Hamiltonian, one finds that the h coupling is related to ours via $h_{\text{there}}^{1/4} = g_{\text{here}}$.

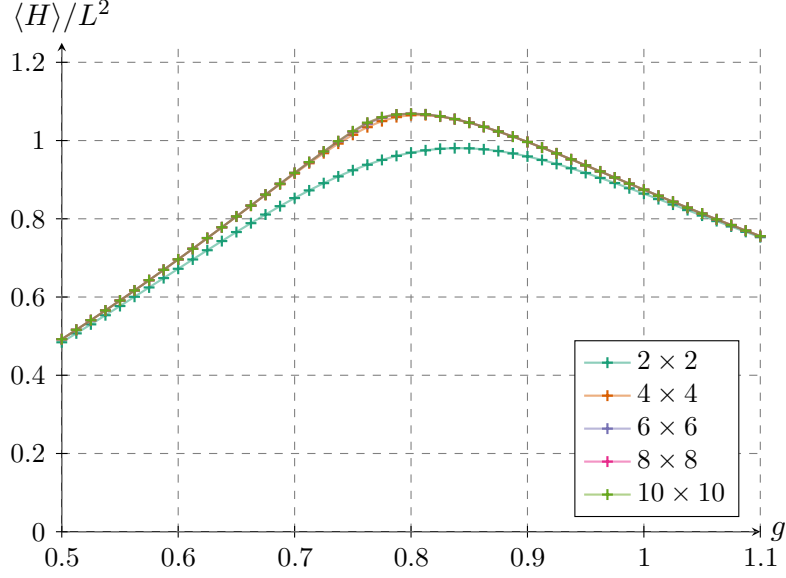


Figure 3.8: Expectation value $\langle H \rangle$ for $\mathbb{Z}_2$ gauge theory (3.5) for varying lattice sizes as a function of coupling g , focused on the region around the critical coupling. The networks were trained for up to 2000 iterations with 2048 samples. Training was stopped early when the variance of the energy per site stabilised to a value of 1.25×10^{-4} or less. The expectation values were then calculated using 32×4096 Monte Carlo samples. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point.

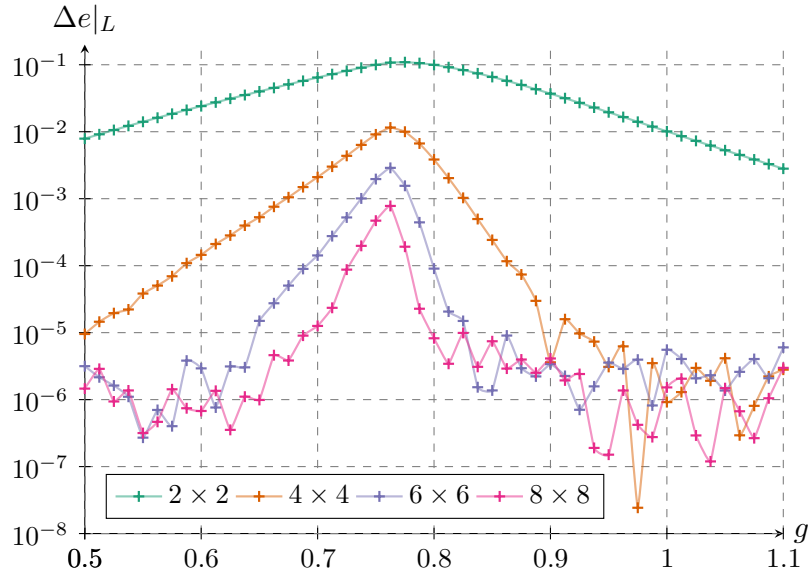


Figure 3.9: The difference between the calculated ground-state energy per lattice site for $L = 10$ and varying L , $\Delta e|_L = |E_{L=10}/10^2 - E_L/L^2|$, as a function of coupling g . Taking the $L = 10$ result to be our “continuum” values, we see that the energies quickly approach this, even for smaller lattice sizes.

lattice, equivalent to the two-dimensional quantum transverse-field Ising model on a square lattice. Their result is that $g_c = 0.757051$.¹⁵

A first attempt at locating the phase transition might proceed by looking for signs of the confinement-deconfinement transition using the order parameter of the theory. Recall that though the order parameter is the Wilson loop operator, confinement is not cleanly characterised by its expectation value, but by a change from perimeter- to area-law decay. Indeed, as can be seen in Figure 3.10, $\langle W_\square \rangle$ is non-zero for all finite couplings on a finite-size lattice. Instead, it is the string tension that displays the usual behavior of an order parameter – zero in the in the area-law decay of the loops. As discussed in Section 3.3.3, given the ground-state wavefunction, the string tension can be estimated using the Creutz ratio. In Figure 3.11, we show the string tension σ estimated using χ_2 on a 10×10 lattice. In the small-coupling regime, sufficiently far from the phase transition, the string tension is zero to within Monte Carlo errors. As we approach $g \approx 0.76$, the expected critical coupling, we see that the string tension becomes non-zero and has approximately linear growth with coupling. Using a rough linear fit to this regime, one finds $\sigma(g) = -3.48 + 4.59g$. Extrapolating back to zero string tension implies $g_c \approx 0.758$, which is already in good agreement with Monte Carlo results from the dual spin system. For larger couplings, $g > 0.95$, Monte Carlo errors become of the same order as the quantities appearing in the Creutz ratio, leading to a noisier signal. Figure 3.11 also shows the string tension estimated using χ_3 . Since larger Wilson loops should be less affected by finite-size corrections, this should give a better estimate of σ , though we also observe that the signal quickly becomes dominated by Monte Carlo errors in the confined phase.

Next, we can look for evidence of the phase transition in the disorder parameter, i.e. the ‘t Hooft string. In the ordered/deconfined phase, expectation values of these string operators decay with the distance between the ends of the string, while in the

15. Our coupling is related to theirs via $t_{\text{there}}^{-1/4} = g_{\text{here}}$.

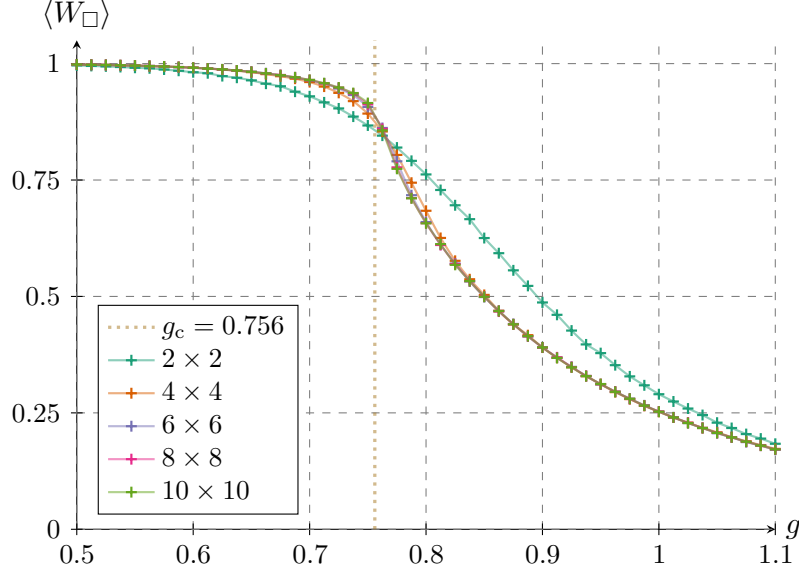


Figure 3.10: The lattice average of the single-plaquette Wilson loop operator $W_{\square}$ in $\mathbb{Z}_2$ gauge theory as a function of coupling g near the critical point for varying lattice sizes. This corresponds to the expectation value of the magnetic flux energy per site. The dotted yellow line indicates the position of the critical coupling at $g_c = 0.756$, as determined by our analysis in Section 3.5.3.

disordered/confined phase, they should be constant. In Figure 3.12, we plot the disorder parameter as a function of coupling for monopoles a distance five apart on a 10×10 lattice. We see that below the critical coupling, the expectation values decays quickly to zero, while near to the transition it grows rapidly, eventually approaching one. These results should be compared with those in [231, Figure 11 (b)], which display a similar, though much noisier, transition.

At this point, we have demonstrated the presence of confined and deconfined phases, with the transition occurring at around 0.76. We would now like to identify the critical couplings more accurately and to establish whether the transition is first-order or continuous. We will see that it is continuous, in agreement with the literature, and then go on to calculate the critical exponents that characterise the conformal field theory that governs the critical point.

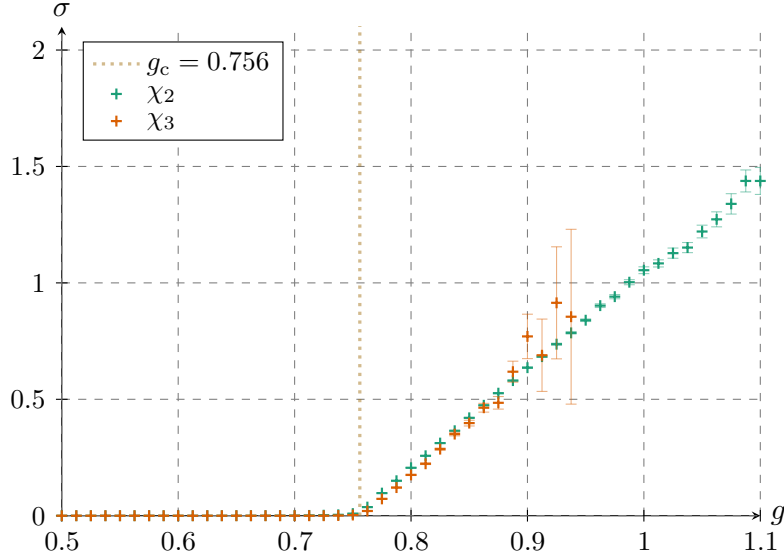


Figure 3.11: String tension estimated from the Creutz ratio χ_l for $l = 2, 3$ on a 10×10 lattice using $2^{20} \approx 1\text{M}$ Monte Carlo samples. The error bars shown only take into account Monte Carlo errors. For $g > g_c$, Wilson loops decay rapidly with size. Estimates of σ are then challenging due to ratios of very small numbers, as can be seen from the noisier signal for $g > 1$ with $l = 2$ and $g > 0.85$ with $l = 3$. In particular, above $g = 0.9375$, the estimates of χ_3 are unreliable, with the standard error becoming larger than the observable. As such, we have not shown these points.

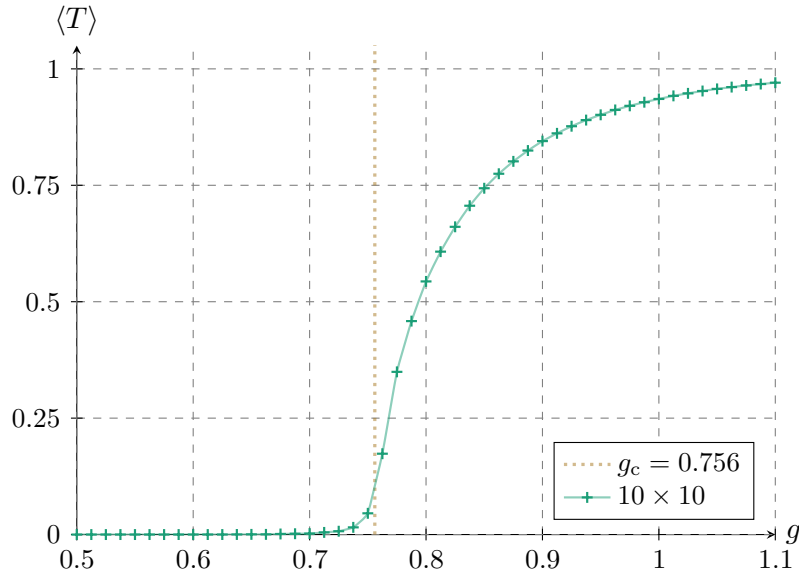


Figure 3.12: The lattice average of the ‘t Hooft string operator in $\mathbb{Z}_2$ gauge theory as a function of coupling g near the critical point for a 10×10 lattice. The string operator is of length five in the x direction.

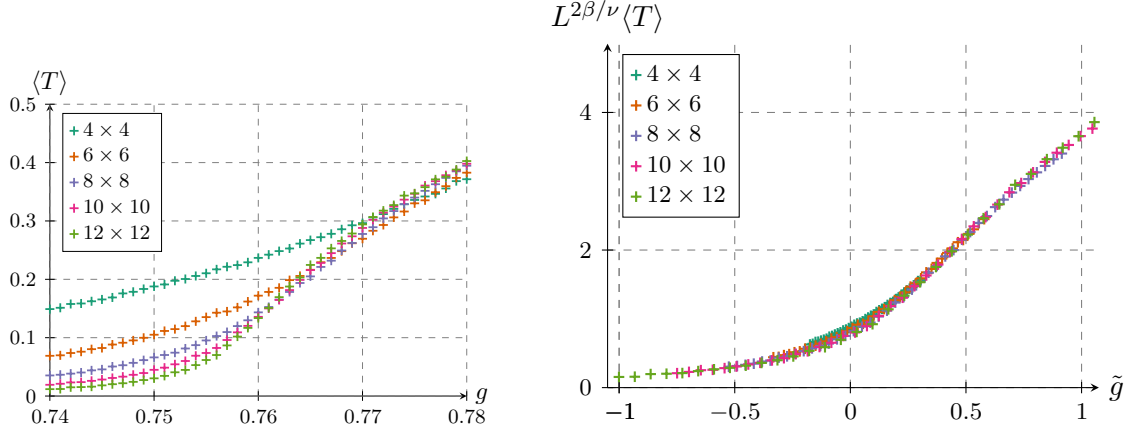


Figure 3.13: Lattice average of the ‘t Hooft string for varying lattice size and distance $L/2$ between ends of ‘t Hooft string. The left figure shows the unscaled expectation values as a function of coupling, spaced equally over the range $g \in [0.74, 0.78]$ around the critical coupling. The right figure shows curve collapse for the rescaled observable $L^{2\beta/\nu}T$, with the rescaled coupling $\tilde{g} = L^{1/\nu}(g - g_c)/g_c$. The critical coupling and exponents are given by $g_c = 0.7546$, $\beta = 0.326$ and $\nu = 0.630$. Note that, since $\tilde{g}$ depends on L , a fixed range of g values is mapped to a varying range of $\tilde{g}$ values, as can be seen in the figure.

3.5.3 Critical exponents and finite-size scaling

The discontinuities and singular behavior that characterise phase transitions appear only in the thermodynamic limit, wherein the volume of the system (the lattice size) approaches infinity [303, 304]. This is a result of the fact that for finite system size, all calculations involving the ground state and path integrals are finite and lead to smooth functions.

Away from the critical coupling for the phase transition, there is an exponential decay of two-point correlation functions as $e^{-r/\xi}$, where r is the distance between the operator insertions and ξ is the correlation length. For continuous phase transitions, the correlation length diverges as the system approaches the critical point $\xi = \xi_0|g - g_c|^{-\nu}$ with a positive critical exponent $\nu > 0$. For a finite-size system of size L , the theory of finite-size scaling describes how observables behave in the vicinity of the critical point as a function of the length scale L and the distance from the critical point [305]. The key insight of this theory is that, close to the critical coupling g_c , the correlation length ξ is comparable to the

system size and therefore the microscopic length given by the lattice spacing which governs the range of interaction no longer affects the correlation functions on scales larger than the lattice spacing.

The Ising model has a $\mathbb{Z}_2$ global symmetry, which corresponds to flipping all the spins on the lattice. Recall that in the Ising model, one has to tune the temperature to the critical value $T = T_c$ and the external field to zero to obtain the Ising CFT, which inherits this $\mathbb{Z}_2$ global symmetry. Consequently, all operators are either $\mathbb{Z}_2$ odd (e.g. the spin field σ) or $\mathbb{Z}_2$ even (e.g. the energy density field ϵ) and there are exactly two relevant scalar operators in the CFT [306]. The $\mathbb{Z}_2$ -even scalar with the lowest scaling dimension is the energy density field ϵ , and its scaling dimension Δ_ϵ is related to the critical exponent ν , which characterises the divergence of the correlation length as $T \rightarrow T_c$. The $\mathbb{Z}_2$ -odd scalar with the lowest scaling dimension is the spin field σ , which transforms as $\sigma \rightarrow -\sigma$ under the $\mathbb{Z}_2$ global symmetry and serves as the order parameter for the $\mathbb{Z}_2$ symmetry breaking phase transition. The scaling dimension Δ_σ is related to the magnetization critical exponent β , which characterises the vanishing of magnetization as temperature is increased to the critical value. Renormalization group (RG) analysis reveals the precise relation between scaling dimensions of these operators and the critical exponents in d -dimensions [49]:

$$\nu = \frac{1}{d - \Delta_\epsilon}, \quad \beta = \frac{\Delta_\sigma}{d - \Delta_\epsilon} . \quad (3.33)$$

As $\mathbb{Z}_2$ lattice gauge theory can be obtained by gauging the $\mathbb{Z}_2$ global symmetry of the Ising model, the theory at the critical coupling g_c is described by the $\mathbb{Z}_2$ gauged Ising CFT (Ising* CFT). Since the two theories are related by a discrete gauging, the corresponding scaling dimensions and critical exponents are identical. The $\mathbb{Z}_2$ odd local operators get projected out during the gauging procedure. However, the $\mathbb{Z}_2$ odd σ field that is attached to the 't Hooft string operator remains in the spectrum. Consequently, the order parameter that we employ to diagnose the confinement phase transition and compute the critical

exponents β and ν is the lattice average of the expectation value of a ‘t Hooft string operator.

The ‘t Hooft string is taken to lie in the x -direction with length $L/2$, corresponding to one-half of the lattice size. This choice numerically leads to the smallest variance and, since the length of the string operator scales with the system, it ensures that one obtains a line operator in the $L \rightarrow \infty$ limit. Since there are two σ fields (monopole operators) at the end of the string operator, the CFT then predicts that in the vicinity of the critical point the expectation value of the operator should vary with L as

$$\langle T \rangle \sim L^{-2\Delta_\sigma} , \quad (3.34)$$

where Δ_σ is the scaling dimensions of the σ operator attached to the ‘t Hooft string in the Ising* CFT [75].

Combining this with the finite-size scaling hypothesis, the functional dependence can be expressed solely in terms of the dimensionless ratio ξ/L and system size L as

$$\langle T \rangle = L^{-2\Delta_\sigma} f(\xi/L) , \quad (3.35)$$

where f is an L -independent scaling function. Using the relations $\xi(g) \sim |g - g_c|^{-\nu}$ and $\beta/\nu = \Delta_\sigma$, this can be expressed in terms of g using another scaling function F :

$$\langle T \rangle = L^{-2\beta/\nu} F(L^{1/\nu}(g - g_c)) . \quad (3.36)$$

Thus, plotting $L^{2\beta/\nu} \langle T \rangle$ against $\tilde{g} \equiv L^{1/\nu}(g - g_c)/g_c$ should give curves that are independent of L and so “collapse” onto each other. One can use this curve collapse to fix the values of g_c , β and ν directly from data. Following the approach of [307], reviewed in supplementary section 3.7.2, one can automate this by defining a measure of how well the

curves collapse and then numerically minimizing the measure as a function of (g_c, β, ν) . The data for this procedure is generated by finding L-CNNs which approximate the ground-state wavefunction for $g \in [0.74, 0.78]$ in steps of 0.001 and for lattice sizes $L = 4, \dots, 12$, and then computing the lattice average of the expectation value of a ‘t Hooft string of length $L/2$ in the x direction. The raw data is shown in the left plot of Figure 3.13. We note that these curves do not collapse on each other for the range of couplings considered. Following the procedure outlined in supplementary section 3.7.2, we find the values of g_c , β and ν that provide the best collapse across the range $\tilde{g} \in [-1, 1]$ for $L = 8, 10, 12$,¹⁶ giving

$$g_c = 0.7546(8), \quad \beta = 0.326(4), \quad \nu = 0.630(3) . \quad (3.37)$$

Here, the errors are computed by bootstrapping the fits over the original data. Using these values, in the right plot of Figure 3.13 we show that the curves of $L^{2\beta/\nu}\langle T \rangle$ do indeed collapse onto each other as expected. The collapse for $L = 6, \dots, 12$ is excellent. The collapse holds even for $L = 4$, which will have larger corrections in L^{-1} not captured by the leading-order analysis. These results predict that the conformal dimensions of the σ and ϵ primary operators in the critical 2 + 1-d Ising model are $\Delta_\sigma = 0.518(4)$ and $\Delta_\epsilon = 1.412(8)$. We can compare our results with previous calculations of the critical exponents. Since the scaling dimensions of operators are identical for Ising and Ising* CFT, our results should match the ones computed for all systems in the Ising universality class. The three principal theoretical methods of computing critical exponents are Monte Carlo study of lattice systems, perturbative RG flow and the conformal bootstrap. The conformal bootstrap computations are the most precise, but all three techniques are in agreement with each other. From the conformal bootstrap program, state-of-the-art bounds on conformal

16. Since our finite-size scaling ansatz includes only leading-order corrections, we have used only the three largest lattice sizes to compute the fits – these should be least affected by subleading corrections in the inverse lattice size.

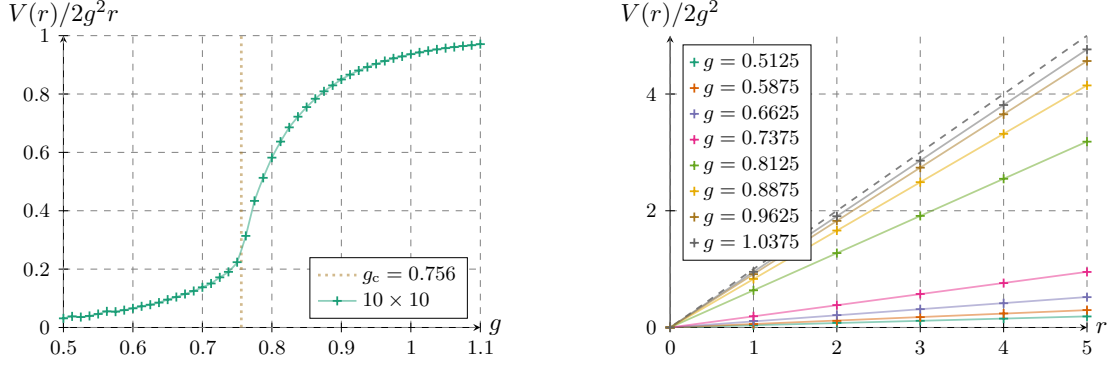


Figure 3.14: The left figure shows the linear potential $V(r)$ for $\mathbb{Z}_2$ gauge theory between two static charges on the lattice divided by its expected strong-coupling limit, $2g^2r$. The charges are placed at $(0, 0)$ and $(r, 0)$ for $r \in \{1, \dots, 5\}$ on a 10×10 lattice. The coefficient for each g is computed by fitting a linear function with zero intercept. The right figure shows the potential energy, $V(r) \equiv E_q - E_0$, defined as the difference between the ground-state energy without static charges, E_0 , and the energy E_q with charges a distance r apart on a 10×10 lattice for various values of the coupling. The lines are linear fits of the form $V(r) \propto r$. The $r = 0$ data corresponds to the state without charges. From the form of the Hamiltonian (3.5), in the strong-coupling limit, one expects the potential to approach $2g^2r$, so we have divided by $2g^2$ to see this. We have also plotted this behavior as the dashed diagonal black line. In all cases, after inserting the static charges, the networks were trained for a further 200 iterations before computing $\langle W_r^\dagger H W_r \rangle_\theta$ (though we observed negligible difference between this and using the untwisted sector ground state).

dimensions yield $\Delta_\sigma = 0.5181489(10)$ and $\Delta_\epsilon = 1.412625(10)$ [308, 309], which in turn fix

$$\beta_{\text{CB}} = 0.326419(3) \ , \quad \nu_{\text{CB}} = 0.629971(4) \ . \quad (3.38)$$

The most precise Monte Carlo results for exponents of the $2 + 1$ -d Ising model are due to Ferrenberg et al. [310], extending previous results of Hasenbusch [311], leading to $\nu_{\text{MC}} = 0.62960(15)$ and $\beta_{\text{MC}} = 0.32620(31)$.

These results are summarised in Table 3.2. We see that our estimates for β and ν are in excellent agreement with high-precision conformal bootstrap and Monte Carlo values. The value of the critical coupling determined by our scaling collapse analysis, $g_c = 0.7546(8)$, is within the 95% confidence interval (two standard deviations) of the value $g_c = 0.757051$ determined by Monte Carlo calculations [280]. This small deviation can be attributed to

the corrections from irrelevant operators in the conformal field theory, which shift the computed value of the critical coupling from the exact value. It is also important to note that our analysis considers only the leading-order effects in the inverse lattice size L^{-1} , whereas the Monte Carlo results incorporate subleading corrections, which may further explain the discrepancy in the values of g_c .

Method	ν	β
NNQS	0.630(3)	0.326(4)
Conformal Bootstrap [309]	0.629971(4)	0.326419(3)
Monte Carlo [312]	0.62960(15)	0.32620(31)
ϵ expansion [313]	0.6305(25)	0.3265(15)
Large N [313]	0.6304(13)	0.3258(14)

Table 3.2: Selected theoretical determinations of 2 + 1-d Ising critical exponents.

Critical coupling from BST extrapolation

So far, we have seen that both the order and disorder parameters display changes in behavior in the vicinity of critical coupling predicted by Monte Carlo studies of the dual spin system, and that curve collapse can be used to identify the critical coupling and exponents. Here, we provide an alternative estimate for the critical coupling directly from our results.

Our curve collapse approach relied on the leading-order corrections from finite-size scaling theory. In fact, one can do better than this by taking into account subleading corrections in L^{-1} and extrapolating to the continuum limit. One way to do this is via BST extrapolation [232–234], a technique which uses rational functions to accelerate convergence. In practice, one looks at the pairs (L, g_{cross}) , where the coupling g_{cross} is where the curve for lattice size L and the next largest lattice size cross. One then follows an iterative procedure to find the continuum critical coupling, as outlined in supplementary section 3.7.3.

Applying this to the crossing points of $\langle T \rangle$ for a 't Hooft string operator of length $L/2$ gives the estimate¹⁷

$$g_c = 0.756431 , \quad (3.39)$$

where the error solely due to convergence of the BST procedure is negligible. This is slightly larger than the coupling we found via curve collapse, though a small difference is not surprising – curve collapse considers only leading-order corrections in L^{-1} while BST extrapolation can also account for higher-order corrections, including an L -dependent shift of g_c .

3.5.4 Potential energy between charges

By using the ground-state wavefunctions obtained after training, we can compute the expectation value of the Wilson string to find the potential between charges on the lattice. Since the gauge group is $\mathbb{Z}_2$, the possible charges are $\{0, 1\}$. The Wilson string stretching between two non-zero charges lying along the x direction of length r is simply a product of Z (clock) operators

$$W_r = Z_{x,\mu} Z_{x+\hat{x},\mu} \cdots Z_{x+r\hat{x},\mu} . \quad (3.40)$$

At small couplings $g \ll 1$, the theory is deconfined and the potential per unit length of the Wilson string vanishes. On the other hand, for large values of the coupling, the potential energy can be obtained by setting $N = 2$ in (3.23):

$$\langle W_r^\dagger H W_r \rangle_0 - \langle H \rangle_0 = 2g^2 r . \quad (3.41)$$

The expected linear dependence of the potential energy as a function of the distance between charges is confirmed on the right plot in Figure 3.14. The potential energy per

17. The error solely from BST extrapolation can be estimated from $\varepsilon_{N-1}^{(0)}$ (see Equation (3.58)). The optimal choice of ω leads to $\varepsilon_{N-1}^{(0)} \sim 10^{-11}$, implying that the extrapolation has converged well.

unit length increases rapidly near the phase transition and saturates to $2g^2$ at strong coupling, as can be seen from the left plot of Figure 3.14.

3.6 $\mathbb{Z}_3$ gauge theory in $2 + 1$ dimensions

We now turn to studying $\mathbb{Z}_3$ gauge theory. This theory is expected to have a first-order phase transition separating the deconfined and confined phase [250, 314]. For a first-order transition, the spectral gap does not vanish even in the thermodynamic limit, and therefore the correlation length is always finite. Since the ground state of the system changes abruptly at the point of transition, physical observables such as the energy and its derivatives as a function of coupling display kinks or discontinuities. For finite lattice sizes, the ordered and disordered phases coexist, smoothing out these features.

By computing the energy and its derivatives, we can determine the critical coupling g_c at which the transition occurs. The computation of other observables such as the Creutz ratio, 't Hooft string operator and the potential energy between charges provides insight into the deconfined and confined phases. The theory is in a deconfined phase below the critical coupling, with perimeter-law decay for Wilson loops and a broken $\mathbb{Z}_3^{(1)}$ one-form symmetry. In the confined phase, the Wilson loops obey an area-law decay and the $\mathbb{Z}_3^{(1)}$ one-form symmetry is preserved. As with $\mathbb{Z}_2$, the expectation value of the 't Hooft string operators become constant due to condensation of magnetic monopoles.

3.6.1 Ground-state energies

As a test of our approach, we compare it with previous work using tensor networks to numerically investigate the ground state of a pure $\mathbb{Z}_3$ lattice gauge theory in $2 + 1$ dimensions [249]. In Figure 3.15, we plot the expectation value of the lattice Hamiltonian (3.5) in the ground state, i.e. the ground state energy, as a function of coupling g for a 2×2 lattice. We see excellent agreement with the energy calculated by exact diagonalization,

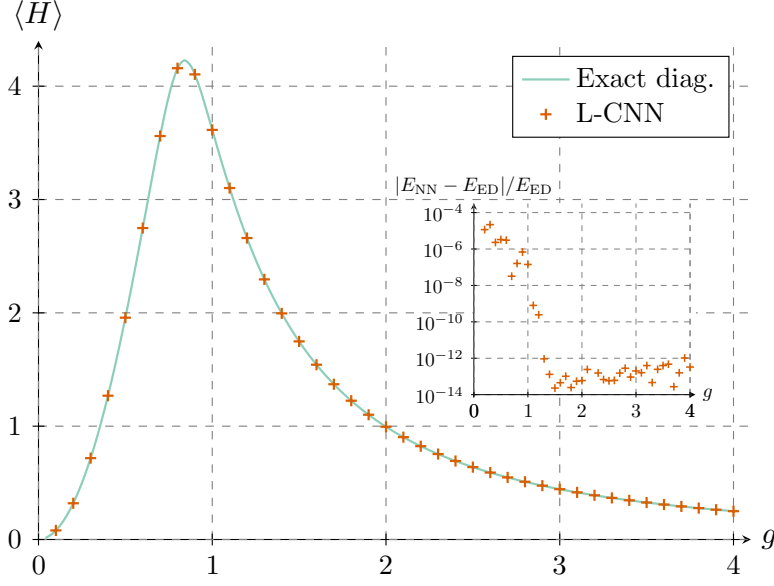


Figure 3.15: Expectation value of $\langle H \rangle$ for $\mathbb{Z}_3$ gauge theory (3.5) on a 2×2 lattice as a function of coupling g . The ground state was computed using an L-CNN with six **L-CB** layers with $(N_{\text{out}}, K) = (3, 1)$. The networks were trained for up to 100 iterations with 4096 samples. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point. The solid teal line is the exact diagonalization result, computed using NetKet. The inset shows the fractional difference between the ground state energies calculated using the neural network and exact diagonalization.

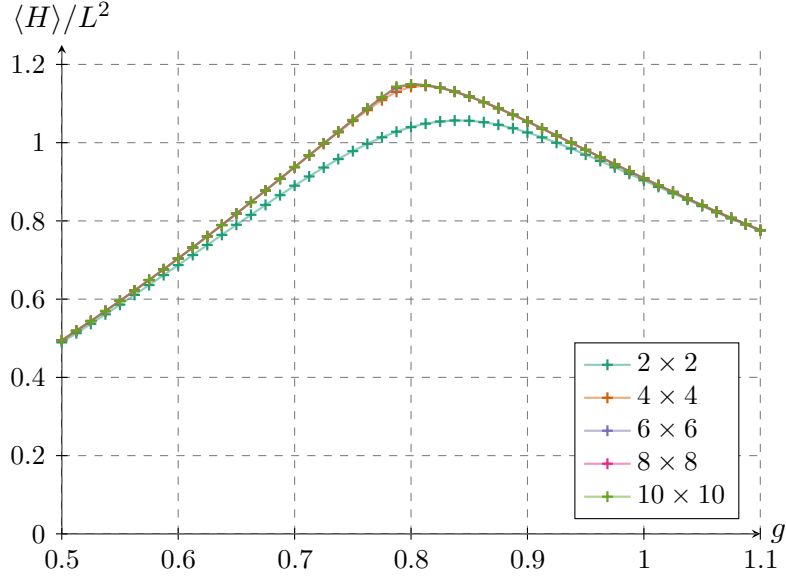


Figure 3.16: Expectation value $\langle H \rangle$ for $\mathbb{Z}_3$ gauge theory for varying lattice sizes as a function of coupling g , focused on the region around the critical coupling. The networks were constructed from $\lceil \log_2 N_{\text{sites}} \rceil$ **L-CB** layers with $(N_{\text{out}}, K) = (4, 2)$, and trained for up to 1000 iterations with 4096 samples. Training was stopped early when the variance of the energy per site stabilised to a value of 1.25×10^{-4} or less. The expectation values were then calculated using 32×4096 Monte Carlo samples. We trained the networks from larger to smaller values of the coupling g , using the previously trained network as the starting point.

and find more accurate results than those recently obtained using tensor networks [249].

In Figure 3.16, we show the ground-state energy on lattices of size $L \in \{2, 4, 6, 8, 10\}$ as a function of coupling in the region $g \in [0.5, 1.1]$. Compared with the results for $\mathbb{Z}_2$ in Figure 3.8, the curves for $\mathbb{Z}_3$ appear to be somewhat sharper at the transition around $g \approx 0.8$. As we will see in a moment, this can be seen more clearly by computing the derivative of the energy with respect to the coupling.

3.6.2 First-order phase transition

The zero-temperature $\mathbb{Z}_3$ theory is expected to undergo a first-order phase transition at some critical coupling g_c . There are few works in the literature which determine this coupling directly from the gauge theory. From [249, Figure 7], the behavior of the string tension on a 6×6 lattice suggests $g_c \in [0.75, 0.90]$. However, the string tension data is rather noisy, and the authors do not give an explicit estimate of g_c . A somewhat cleaner calculation is given in [315] using an infinite projected entangled-pair state (iPEPS) ansatz. The ground-state energy displays a kink at the critical coupling, and a fit to the area-law coefficient agrees with the location of the kink, giving $g_c = 0.818$.¹⁸

Before locating the critical coupling, we first look for evidence of the confined and deconfined phases. Following a similar discussion in Section 3.5, in Figure 3.17, we plot an estimate of the string tension, computed using the Creutz ratio for $l = 2$. At small couplings, the string tension is zero, indicating the presence of an ordered, deconfined phase. At larger couplings, the string tension is non-zero and increasing with coupling, corresponding to a disordered, confined phase. The phase transition appears to occur in the region $g \in [0.7755, 0.7875]$. For completeness, we also compute the lattice average of the disorder operator, i.e. the ‘t Hooft string. As for $\mathbb{Z}_2$, one expects this operator to decay

18. The critical coupling is given in [315] as $g_c^2 = 1.159$. However, their Hamiltonian is not the same as ours. Specifically, working in a basis where the shift operator is diagonal, they have resummed the electric term in the Hamiltonian and absorbed various constants into the coupling. This results in the identifications $g_{\text{there}}^2 = \sqrt{3}g_{\text{here}}^2$ and $\sqrt{3}H_{\text{there}}(g_{\text{there}}) = H_{\text{here}}(g_{\text{here}}) - g_{\text{here}}^{-2}L^2$.

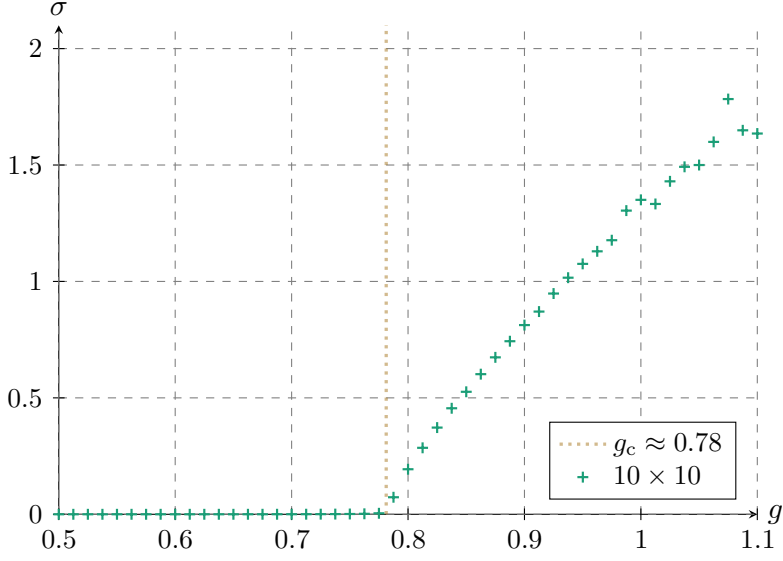


Figure 3.17: String tension estimated from Creutz ratio χ_l for $l = 2$ on a 10×10 lattice using $2^{20} \approx 1\text{M}$ Monte Carlo samples. For $g > g_c$, Wilson loops decay rapidly with size. Estimates of σ are then challenging due to ratios of very small numbers, as can be seen around $g > 0.95$.

exponentially with the string length in the ordered phase, while it should be independent of length (but coupling dependent) in the disordered phase. Our results in Figure 3.18 confirm these expectations, with the transition region agreeing with that suggested by the string tension.

Note that one would expect a first-order transition to be signalled by discontinuities in order and disorder parameters. For the ‘t Hooft string, the $\mathbb{Z}_3$ theory displays a steeper increase than $\mathbb{Z}_2$ at the transition, however it does not display clear signs of a discontinuity. In fact, this behavior is not unexpected, and is often referred to as being “weakly” first order. First-order transitions are referred to as weakly first order when the (finite) correlation length at the transition is larger than the system size. If the lattice system is smaller than the correlation length ξ , it is difficult to differentiate between first-order and continuous transitions, with the expected discontinuities smoothed out, the critical coupling shifted, and finite-size scaling analysis failing to recover trivial critical

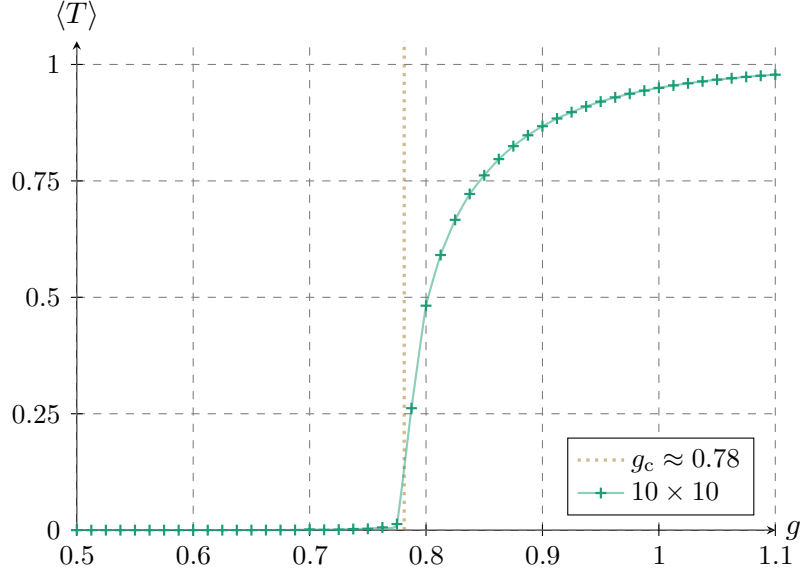


Figure 3.18: The lattice average of the ‘t Hooft string operator in $\mathbb{Z}_3$ gauge theory as a function of coupling g near the critical point for a 10×10 lattice. The string operator is of length five in the x direction. The dotted yellow line indicates our identification of the transition at $g_c = 0.781$.

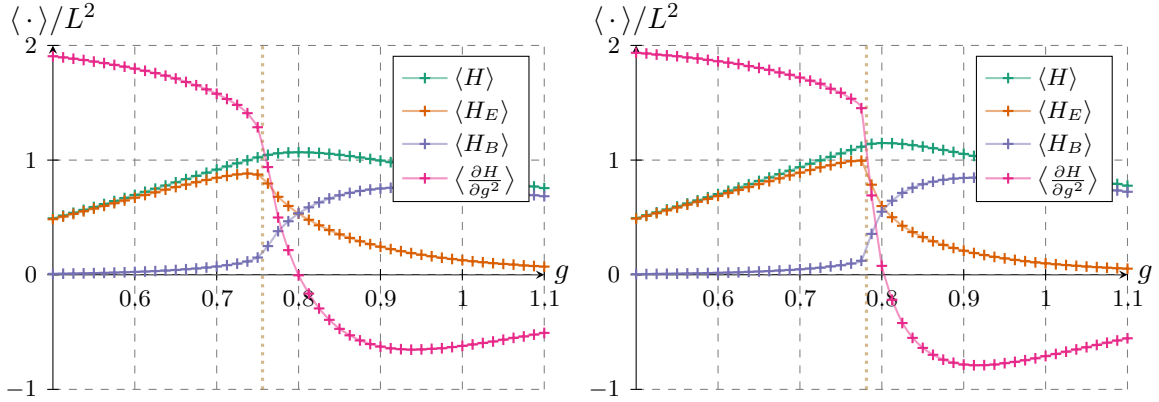


Figure 3.19: The left figure shows various expectation values per lattice site for $\mathbb{Z}_2$ gauge theory computed on a 10×10 lattice using an L-CNN as a function of coupling. These include the ground-state energy, the electric and magnetic energies, and the derivative of the energy with respect to g^2 . We have also labeled the position of the critical coupling $g_c = 0.756$ with a dashed gold line as calculated from BST extrapolation in Section 3.5.3. The right figure shows the same data for $\mathbb{Z}_3$ gauge theory, with the transition marked midway between 0.775 and 0.7875 at $g_c = 0.781$.

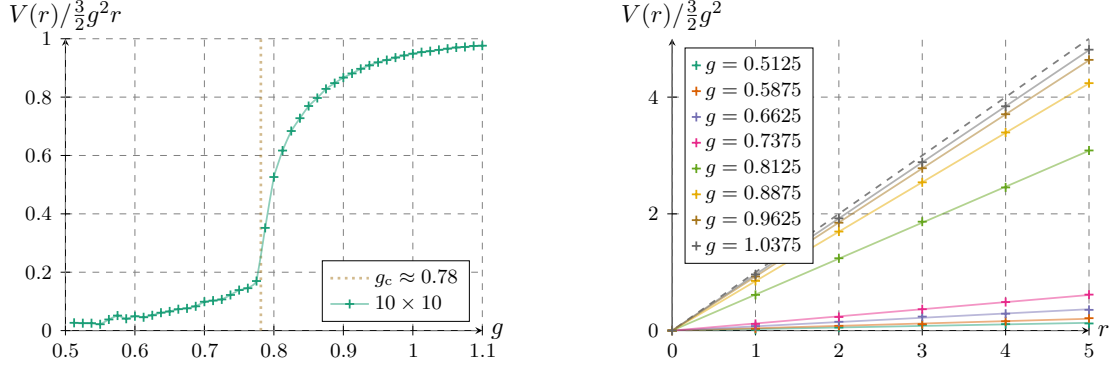


Figure 3.20: The left figure shows the linear potential $V(r)$ for $\mathbb{Z}_3$ gauge theory between two static charges on the lattice divided by its expected strong-coupling limit, $\frac{3}{2}g^2$. The charges are placed at $(0, 0)$ and $(r, 0)$ for $r \in \{1, \dots, 5\}$ on a 10×10 lattice. The coefficient for each g is computed by fitting a linear function with zero intercept. The right figure shows the potential energy, $V(r) \equiv E_q - E_0$, defined as the difference between the ground-state energy without static charges, E_0 , and the energy E_q with charges a distance r apart on a 10×10 lattice for various values of the coupling. The lines are linear fits of the form $V(r) \propto r$. The $r = 0$ data corresponds to the state without charges. From the form of the Hamiltonian (3.5), in the strong-coupling limit, one expects the potential to approach $\frac{3}{2}g^2r$, so we have divided by $\frac{3}{2}g^2$ to see this. We have also plotted this behavior as the dashed diagonal black line. In all cases, after inserting the static charges, the networks were trained for a further 200 iterations before computing $\langle W_r^\dagger H W_r \rangle$.

exponents (i.e. $\nu = 1/d$) [316].¹⁹ Indeed, it is known that the spin system dual to $\mathbb{Z}_3$ gauge theory, the three-state Potts model, has a correlation length $\xi \sim 10$ [318], and so we may need to look at larger lattices to see a clear signal of first-order behavior.

A first-order quantum phase transition can also be diagnosed by a discontinuity in the first derivative of the energy with respect to the coupling [315]. Using the Hellmann–Feynman theorem, this derivative can be calculated as

$$\frac{\partial \langle H \rangle_\theta}{\partial g^2} \equiv \left\langle \frac{\partial H}{\partial g^2} \right\rangle_\theta = g^{-2} \langle \Psi_\theta | H_E - H_B | \Psi_\theta \rangle. \quad (3.42)$$

In Figure 3.19, we plot the derivative of the energy per lattice site as a function of coupling

¹⁹ For example, the correlation lengths of the two-dimensional q -state Potts model were computed by Buffenoir and Wallon [317]. The theory has a continuous $\xi \rightarrow \infty$ transition for $q = 4$, and a first-order transition for $q \geq 5$. However, for $q = 5, \dots$ the correlation lengths $\xi \approx 2500, 150, 50, 25, 15, 10, \dots$ indicate that one will not see discontinuities in observables for smaller values of q without going to very large lattices.

for both $\mathbb{Z}_2$ and $\mathbb{Z}_3$ on a 10×10 lattice. In the case of $\mathbb{Z}_2$, which is known to have a continuous phase transition, the transition region near to the critical coupling ($g_c \approx 0.756$) shows a smooth decrease. For $\mathbb{Z}_3$, the derivative instead shows a steep decrease in the region $[0.775, 0.7875]$, indicating a first-order transition. This is also apparent in the curves for $\langle H_B \rangle$, the magnetic term in the Hamiltonian, which display a sharper, more sudden increase for $\mathbb{Z}_3$.²⁰ From this, and the order and disorder parameters, we identify the critical coupling to be

$$g_c = 0.781(6) . \quad (3.43)$$

where the error corresponds to the half-width of the interval $[0.775, 0.7875]$. If one plots the same expectations for varying lattice size, one sees that the curve for $\mathbb{Z}_2$ does not change much between $L = 6$ and $L = 10$, suggesting that it is not becoming discontinuous in the continuum limit. The curves for $\mathbb{Z}_3$ instead show the pattern of getting steeper and steeper with increasing L , exactly as one would expect for a first-order phase transition.

We can also check whether our variational state has correctly learned the symmetries of the ground state. For example, we know that the true ground state should be invariant under lattice translations, rotations and reflections, and charge conjugation. L-CNNs are manifestly translationally invariant, so this symmetry is exact. The remaining symmetries will be learned as part of the training process, and so they will approximate. One can check this by comparing the amplitude $\Psi(\mathcal{U})$ with $\Psi(\mathcal{U}')$, where $\mathcal{U}'$ is related to $\mathcal{U}$ by one of the symmetry transformations. For example, charge conjugation acts by sending all link variables to their conjugates, $\mathcal{U}' = \mathcal{U}^\dagger$. Charge conjugation invariance then implies that expectation values of Wilson loops for the ground state should be *real*.²¹ Denoting the

20. Note that H_B/L^2 is effectively one minus the single-plaquette Wilson loop $W_\square$, so we have not plotted this separately.

21. In fact, since the Hamiltonian commutes with time reversal, if the ground-state is unique, one can always pick the ground-state to be real (see, for instance, [319, Equation 6.4]). Then, if it is also charge-conjugation invariant, one also has $\Psi(\mathcal{U}) = \Psi(\mathcal{U}^\dagger)$. This symmetry could be imposed in the network architecture, though we have not done so in this work.

charge conjugation operator by C , this follows from

$$\langle W \rangle_0 = \langle \Psi_0 | C^\dagger W C | \Psi_0 \rangle = \langle W \rangle_0^\dagger, \quad (3.44)$$

where we have used $C|\Psi_0\rangle = |\Psi_0\rangle$ for the ground state. Again, we can check to what accuracy this holds for our trained networks by comparing the size of the imaginary part of the expectation value with its magnitude. On a 10×10 lattice, evaluating the lattice average of the single-plaquette Wilson loop operator for our NNQSSs with $2^{18} \approx 250\text{k}$ samples gives values of $\text{im}\langle W_\square \rangle_\theta / |\langle W_\square \rangle_\theta|$ in the range 10^{-6} to 10^{-4} for $g \in [0.5, 1.1]$. From this, we conclude that our networks have learned the charge conjugation symmetry to high accuracy.

3.6.3 Static charges

Finally, we turn to the potential between static charges in the $\mathbb{Z}_3$ theory, computed using the expectation value of the Hamiltonian conjugated by a Wilson string. In this case, the possible charges are $\{0, 1, 2\}$ and thus there are two distinct Wilson strings. The following Wilson string operator places the charges 1 at x and 2 at $x + r\hat{x}$, a distance r apart along the x -direction:

$$W_r = Q_{x,\mu} Q_{x+\hat{x},\mu} \cdots Q_{x+r\hat{x},\mu}. \quad (3.45)$$

The square of this operator $W_r^2 = W_r^\dagger$ places charges 2 at x and 1 at $x + r\hat{x}$, so the potential energy associated to it is identical to that of W_r .

The theory is deconfined at very small values of the coupling, and so the potential per unit length should vanish. Since there is a first-order phase transition for $\mathbb{Z}_3$, the potential energy per unit distance should increase drastically near the phase transition, with this behavior becoming singular in the thermodynamic limit. For very large values of g , the

potential energy can be estimated by setting $N = 3$ in (3.23)

$$V(r) \approx \langle W_r^\dagger H W_r \rangle_0 - \langle H \rangle_0 = \frac{3}{2} g^2 r . \quad (3.46)$$

The dramatic increase in the potential at the phase transition and the predicted weak-/strong-coupling behavior is demonstrated in the left plot of Figure 3.20, while the linear dependence of the potential energy on the distance between charges is illustrated in the right plot. Comparing these with the analogous results for $\mathbb{Z}_2$ in Figure 3.14, we see that both display potentials that increase linearly with distance, with the increase at the phase transition being smooth for $\mathbb{Z}_2$ and much sharper for $\mathbb{Z}_3$, in agreement with their continuous and first-order nature respectively.

Acknowledgements

We thank Miranda Cheng, Edward Mazenc, Matija Medvidović and Semon Rezchikov for stimulating discussions and useful insights, and Filippo Vicentini for patiently answering the authors’ questions about NetKet. The work of A. Apte is supported by the Data Science Institute at the University of Chicago. A. Ashmore is supported by the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 838776, and acknowledges previous support from NSF Grant PHY-2014195 and the Kadanoff Center for Theoretical Physics. Part of this work was carried out at the Aspen Center for Physics, which is supported by NSF grant PHY-2210452. C. Córdova is partially supported by the US Department of Energy Grant 5-29073 and the Sloan Foundation. C. Córdova and T.-C. Huang also acknowledge support from the Simons Collaboration on Global Categorical Symmetries. We are grateful to the University of Chicago’s Research Computing Center and AI+Science Research Initiative for assistance with the calculations carried out in this work.

3.7 Supplementary Material

3.7.1 Implementation of Wilson loop operators

In NetKet, Wilson loop operators can be constructed either by chaining together products of `LocalOperator`'s, or by constructing a custom operator. The first approach often runs into memory problems, as large Wilson loops need products of many local operators (e.g. an 8×8 Wilson loop will need a product of 32 local operators). Instead, we constructed a custom operator. The implementation is simplified by the fact that we are interested only in the expectation value of the Wilson loop operators, and not in their gradients with respect to the network parameters (which would require a more complicated implementation). Given a state $|\Psi\rangle$, one can compute the expectation value of an operator $\mathcal{O}$ via

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi | \mathcal{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \sum_{\mathcal{U}} \frac{|\Psi(\mathcal{U})|^2}{\langle \Psi | \Psi \rangle} \left(\sum_{\mathcal{U}'} \frac{\langle \mathcal{U} | \mathcal{O} | \mathcal{U}' \rangle \langle \mathcal{U}' | \Psi \rangle}{\langle \mathcal{U} | \Psi \rangle} \right) = \mathbb{E}_{\mathcal{U} \sim |\Psi(\mathcal{U})|^2} [\mathcal{O}_{\text{loc}}(\mathcal{U})] , \quad (3.47)$$

where $\mathcal{U}$ and $\mathcal{U}'$ are gauge field configurations, $\mathbb{E}_{\mathcal{U} \sim |\Psi(\mathcal{U})|^2}$ means take the expectation value over $\mathcal{U}$, with $\mathcal{U}$ distributed according to the PDF $|\Psi(\mathcal{U})|^2$, and we have defined the “local estimator”

$$\mathcal{O}_{\text{loc}}(\mathcal{U}) = \frac{\langle \mathcal{U} | \mathcal{O} | \Psi \rangle}{\langle \mathcal{U} | \Psi \rangle} = \sum_{\mathcal{U}'} \frac{\langle \mathcal{U} | \mathcal{O} | \mathcal{U}' \rangle \langle \mathcal{U}' | \Psi \rangle}{\langle \mathcal{U} | \Psi \rangle} = \sum_{\mathcal{U}'} \langle \mathcal{U} | \mathcal{O} | \mathcal{U}' \rangle \frac{\Psi(\mathcal{U}')}{\Psi(\mathcal{U})} . \quad (3.48)$$

In order to evaluate this expectation value via variational Monte Carlo, one needs the following:

1. Some way of restricting the sum over $\mathcal{U}$ to finitely many representative values by sampling the PDF, $|\Psi(\mathcal{U})|^2$. This is provided by the Monte Carlo variational state

interface of NetKet.

2. A method to take the samples $\{\mathcal{U}\}$ and compute the “connected elements”, i.e. those for which the matrix element $\langle \mathcal{U} | \mathcal{O} | \mathcal{U}' \rangle \neq 0$ is non-zero. This means that for each $\mathcal{U}$, one does not have to sum over all of the sampled configurations again, but can instead restrict to a (usually much) smaller set of configurations, labeled by $\{\mathcal{U}'\}$. One then also needs to compute the matrix elements themselves.
3. A calculation of the local estimators $\mathcal{O}_{\text{loc}}(\mathcal{U})$ given the matrix elements, the sets $\{\mathcal{U}\}$ and $\{\mathcal{U}'\}$, and the state $|\Psi\rangle$. Again, this implementation is provided by NetKet.
4. The statistical average of the local energies weighted by $|\Psi(\mathcal{U})|^2$.

For a single clock operator Q_l , it is simple to find the connected elements $\{\mathcal{U}'\}$ and the matrix elements $\langle \mathcal{U} | \mathcal{O} | \mathcal{U}' \rangle$. Given a configuration of link variables

$|\mathcal{U}\rangle = |q\rangle_1 \otimes \cdots \otimes |q\rangle_l \otimes \dots$ on the lattice, Q_l acts as

$$Q_l |\mathcal{U}\rangle = e^{2\pi i q_l / N} |\mathcal{U}\rangle . \quad (3.49)$$

Since we are working in a basis which diagonalises the action of Q_l , this preserves the configuration up to a phase, so that the connected set $\{\mathcal{U}'\}$ has only one element, namely $\mathcal{U}$ itself. That is

$$\langle \mathcal{U}' | Q_l | \mathcal{U} \rangle = e^{2\pi i q_l / N} \delta_{\mathcal{U}, \mathcal{U}'} . \quad (3.50)$$

The same holds for products of clock operators, such as those used for constructing Wilson loop operators. Using this, it is relatively straightforward to implement a custom Wilson loop operator in NetKet.

3.7.2 Exponents from data collapse

Finite-size scaling is one of the key pillars in the theory of critical phenomena [320].

Consider a system of characteristic length scale L close to the critical point g_c . The behavior of a physical quantity m in the vicinity of the transition can be captured by a scaling relation [321]

$$m(t, L) = L^d f(t/L^c) , \quad (3.51)$$

where $t = (g - g_c)/g_c$ is the reduced coupling, and depending on the physical system m may refer to quantities such as specific heat, magnetization, or in our case the expectation value of the ‘t Hooft string operator. The remarkable aspect of this relation is that it holds for different physical systems in the same universality class. The scaling relation predicts that when the data for mL^{-d} is plotted against tL^{-c} , the curves for different system sizes (and even for different materials belonging to the same universality class) should collapse onto a single curve. See, for example, the first figure in [320] for a striking illustration of curve collapse in magnetization data for five different materials belonging to the $2 + 1$ -d Heisenberg universality class.

The scaling relation can be used to extract the values of the exponents c and d , and the critical coupling g_c based on numerical data for $m(t, L)$. Given a measure for the data collapse, numerical minimization techniques can be used to automatically search for the exponents. Consider a collection of data for a set of lengths $\{L_j\}$. The tabulated values of m and t can be represented as $\{m_{ij}\}$ and $\{t_{ij}\}$ respectively, where ij labels the i ’th value of t for length L_j . The central obstacle to overcome in defining a measure of collapse is the absence of the knowledge of the scaling function f . Closely following [307], we can use a polynomial interpolation based on any length $L_p \in \{L_j\}$ to determine f and compute the deviation from this curve. Since any length can be used for this purpose, we repeat the

procedure for all lengths and average to obtain a measure of collapse:

$$P = \frac{1}{N_{\text{over}}} \sum_p \sum_{i,j} \left| m_{ij} L_j^{-d} - \mathcal{E}_p(t_{ij} L_j^{-c}) \right|, \quad (3.52)$$

where $\mathcal{E}_p(x)$ is the interpolation function based on L_p . The sum over i is performed only in the overlapping region to avoid extrapolation, and N_{over} is the total number of points used to compute P . We use a cubic spline interpolation scheme to find $\mathcal{E}_p(x)$. Since $P \geq 0$ and vanishes only for perfect data collapse, minimization of P can be used to extract the optimal values of the parameters. Since each of the terms that appears in the sum is non-negative, we can use the method of least-squares [322] to compute the optimal values of (c, d, g_c) that lead to the best-fit. The uncertainties in the computed values are estimated by using statistical bootstrapping [323], wherein the computation of optimal parameters is carried out multiple times using data that is re-sampled from the original dataset. This method of computing uncertainties in critical exponents is routinely employed in the Monte Carlo simulations of statistical systems [238, 324].

3.7.3 BST extrapolation

Finite-size scaling analysis has long been used to extrapolate finite-size lattice data to the thermodynamic limit, $L \rightarrow \infty$. In this supplementary section, we give a brief review of one such technique for extrapolating a sequence of finite-size data which empirically has been found to converge rapidly. This is known as BST extrapolation [232–234].

The general set-up is as follows. Certain continuum observables, such as the critical coupling g_c , admit an asymptotic form

$$g_c(L) = g_c + a_1 L^{-\omega_1} + a_2 L^{-\omega_2} + \dots, \quad (3.53)$$

where $g_c(L)$ is computed via a finite lattice simulation and g_c is the continuum critical

coupling that one is interested in predicting to high accuracy.

Let h_i , $i = 0, \dots, N$, denote a sequence of positive numbers which converges to zero for $i \rightarrow \infty$. In practice, one often takes the h_i to be given by the inverse lattice size. As an example, consider even lattices with $h_i = (1/2, 1/4, 1/6, 1/8)$ with $N = 3$ and $i = 0, \dots, 3$. Assume that the continuum observable of interest is T , which has a finite-size expansion

$$T(h) = T + a_1 h^{\omega_1} + a_2 h^{\omega_2} + \dots, \quad (3.54)$$

where $\omega_2 > \omega_1$, and so on. One then computes the following quantities

$$\begin{array}{cccc} m=0 & T_0^{(0)} & T_0^{(1)} & T_0^{(2)} & T_0^{(3)} \\ m=1 & T_1^{(0)} & T_1^{(1)} & T_1^{(2)} & \\ m=2 & T_2^{(0)} & T_2^{(1)} & & \\ m=3 & T_3^{(0)} & & & \end{array}$$

using the iterative rules

$$\begin{aligned} T_{-1}^{(i)} &= 0, \\ T_0^{(i)} &= T(h_i), \\ T_m^{(i)} &= T_{m-1}^{(i+1)} + (T_{m-1}^{(i+1)} - T_{m-1}^{(i)}) \left[\left(\frac{h_i}{h_{i+m}} \right)^\omega \left(1 - \frac{T_{m-1}^{(i+1)} - T_{m-1}^{(i)}}{T_{m-1}^{(i+1)} - T_{m-2}^{(i+1)}} \right) - 1 \right]^{-1}, \end{aligned} \quad (3.55)$$

where the final line is evaluated for $m \geq 1$. These rules come from approximating $T(h)$ via a sequence of rational functions.

The point of this is that the original sequence of finite-size data $T_0^{(i)}$ is expected to converge slowly to the continuum limit, with $T_0^{(3)}$ in our example being the best estimate that one can obtain from the four lattice sizes considered. Comparing with the expansion

of $T(h)$ in (3.54), this estimate obeys

$$T_0^{(3)} = T + \mathcal{O}(h_3^{\omega_1}) . \quad (3.56)$$

Using the rules above to calculate the $m = 1$ row, our new best estimate will be $T_1^{(2)}$. It then follows straightforwardly from the form of (3.55) that, upon picking $\omega = \omega_1$, this estimate obeys

$$T_1^{(2)} = T + \mathcal{O}(h_2^\eta) , \quad (3.57)$$

where $\eta = \min(\omega_2, 2\omega_1)$. Provided that the h_i are chosen appropriately (i.e. close enough), this correction term will be smaller than the naive one in (3.56). In other words, in one step of the BST algorithm, one can remove the leading-order correction in $T(h)$. One then iterates to eventually find $T_3^{(0)}$, which gives the “best” estimate of T .

It then remains to choose an appropriate value of ω given the data $T(h_i)$. In practice, this is done by calculating

$$\varepsilon_m^{(i)} = 2|T_m^{(i+1)} - T_m^{(i)}| . \quad (3.58)$$

Taking $i = 0$ and $m = N - 1$, $\varepsilon_{N-1}^{(0)}$ measures the difference between the extrapolants in the penultimate row. One should then choose ω to minimize this quantity. The idea behind this is that, in the $N \rightarrow \infty$ limit of many data points, one expects $|T_m^{(i)} - T| \leq \varepsilon_m^{(i)}$, and so minimizing the difference between the penultimate estimates will lead to a prediction $T_N^{(0)}$ for T with the smallest extrapolation error.

CHAPTER 4

CONCLUSION

In this thesis, we explored two complementary approaches to studying quantum field theory. The first part of our work examined non-invertible symmetries in gauge theories, extending the Landau paradigm for phase transitions beyond the traditional framework of invertible symmetries. The second part applied deep learning techniques to simulate lattice gauge theories, showcasing how neural network quantum states can efficiently capture ground-state wavefunctions and precisely characterize phase transitions.

Our investigation of non-invertible symmetries revealed that these generalized symmetries place significant constraints on the possible gapped phases of a quantum field theory. By studying $\mathbb{Z}_N$ gauge theories in 3+1 dimensions, we provided a complete classification of when symmetry-preserving gapped phases can exist. We proved that a duality invariant theory can be in a gapped phase if and only if $N = k^2\ell$ where -1 is a quadratic residue modulo ℓ . For other values of N , the theory must either spontaneously break the duality symmetry or flow to a gapless phase in the infrared. Our work thus provides a generalization of 't Hooft anomaly matching to the context of non-invertible symmetries.

The development of a complete algebraic theory of these symmetries remains an important frontier for future research. Current understanding suggests that non-invertible symmetries in quantum field theories are described by higher fusion categories [325]. However, the classification and systematic study of these higher categorical structures is still in its early stages [326–328]. Further work is needed to develop a comprehensive theory that can fully characterize the possible symmetries in arbitrary dimensions and understand their dynamical consequences.

A particularly exciting direction is the exploration of anomalies of non-invertible symmetries [329–331]. Traditional 't Hooft anomalies for group-like symmetries have a well-established mathematical description in terms of group cohomology and bordism

theory. Extending these frameworks to non-invertible symmetries would provide powerful tools for constraining the possible phases and renormalization group flows of quantum field theories. The mathematical structures needed for such an extension such as higher categorical versions of cohomology theories are active areas of research in both physics and mathematics.

The second part of this thesis demonstrated how deep learning techniques can be applied to study lattice gauge theories from first principles. By implementing gauge-equivariant convolutional neural networks, we constructed neural network quantum states that respect the exact gauge invariance of the theory while providing an efficient representation of the ground-state wavefunction. Our approach differs fundamentally from traditional Monte Carlo methods, which directly evaluate the path integral. Instead, we compute the ground-state wavefunction on a spatial lattice, which provides access to quantum mechanical quantities such as entanglement entropy. Our application to $\mathbb{Z}_2$ lattice gauge theory produced remarkably precise results for the critical coupling and critical exponents. For $\mathbb{Z}_3$ lattice gauge theory, our neural network approach confirmed the existence of a first-order phase transition.

Building on our success with discrete gauge theories in 2+1 dimensions, several natural extensions present themselves for future work. Firstly, it will be good to extend our analysis of $\mathbb{Z}_N$ gauge theories to 3+1-d dimensions. This should be straightforward and will require us to incorporate the z -direction in the generation of the Wilson loops. The phase diagram is more interesting in this case and locating the critical couplings and identifying phases and phase transition should be possible [181, 332]. Furthermore, it has been suggested that deforming the model with higher order plaquette terms while maintaining self-duality can lead to a new interacting fixed point [333]. In 3+1-d dimensions, one can include a θ -term which leads to a sign problem and thus cannot be simulated with Monte-Carlo techniques. There is way to write the Hamiltonian for this on

the lattice [334] and investigate the phases as elucidated by Cardy-Rabinovici [182].

The inclusion of matter even in the $\mathbb{Z}_2$ gauge theory in 2+1-dimensions leads to an interesting phase diagram with Higgs-Confinement continuity. Parts of the phase diagram, especially the transitions along the self-dual line are still not fully understood [281, 335].

Incorporating fermionic matter will allow us to investigate many systems of physical interest that suffer from a sign problem. As a simple example, the celebrated Hubbard model which can be used to study high-temperature superconductivity has a sign problem away from half-filling [336].

Simulating systems at finite temperature is challenging and this can also be tackled with neural quantum states. QCD undergoes a chiral phase transition and such a chiral transition also occurs in the NJL model, which is arguably the simplest model displaying such a transition for a finite chemical potential and temperature [337]. The thermal states can also be used to compute transport coefficients for gauge theories [338]. The eventual goal of simulating systems at finite temperature would be to precisely map out the phase diagram of QCD, which is largely conjectural [339].

The ability to accurately compute excited states is crucial for obtaining masses of hadrons built from quarks [340]. Even before computing hadron masses, the masses of lightest glue-balls in Non-Abelian gauge theories without matter will be a important milestone [341]. Using the scaling of the spectral gap with the size of the system, one can also compute the dynamical critical exponent z for quantum phase transitions [342].

Neural quantum states can also be used for real time evolution [343], although maintaining coherence for long times is still quite challenging. Once the real time evolution techniques are sufficiently developed, they will enable effective computation of scattering processes.

The dynamics of quenches and scars observed in Rydberg atom based quantum simulators can also be simulated with this method [344–346]. In terms of questions of concerning NQS themselves, it will be worthwhile to investigate if there is a scaling law of the accuracy of

the states with the number of parameters similar to scaling laws in large language models [347, 348].

The combination of analytical insights from symmetry principles and numerical precision from deep learning techniques represents a promising direction for the study of quantum field theories. Non-invertible symmetries provide powerful constraints on possible phases, while neural network methods allow us to explore the detailed structure of these phases and the transitions between them. Together, these approaches form a complementary toolkit for understanding the rich landscape of quantum field theories. Moreover, our findings indicate that neural networks with minimal inductive bias as exemplified by convolutional architectures, are nevertheless capable of learning representations of quantum states. This suggests that the simulations of quantum systems may be more accessible to deep learning approaches than previously thought.

In conclusion, this thesis has expanded our understanding of quantum field theories through both abstract symmetry principles and concrete numerical simulations. By studying the anomalies of non-invertible symmetries and leveraging the power of deep learning, we have provided new insights into the phases and phase transitions of gauge theories. These approaches not only deepen our theoretical understanding but also enhance our ability to make quantitative predictions. As mathematical foundations of symmetries become more fully developed and computational techniques advance, we anticipate significant progress towards a deeper understanding of quantum field theory.

REFERENCES

- [1] J. D. Fowler, *Perspectives of reality*. Sussex Academic Press, Brighton, England, Jan., 2002.
- [2] J. Barnes, *Early Greek philosophy*. Classics. Penguin Books, Harlow, England, Feb., 1987.
- [3] E. McMullin, *The origins of the field concept in physics*, *Physics in Perspective* **4** (Feb., 2002) 13–39.
- [4] D. J. Griffiths, *Resource letter em-1: Electromagnetic momentum*, *American Journal of Physics* **80** (Dec., 2011) 7–18.
- [5] D. Greenberger, K. Hentschel, and F. Weinert, eds., *Compendium of quantum physics*. Springer, Berlin, Germany, 2009 ed., July, 2009.
- [6] T. Y. Cao, *Conceptual developments of 20th century field theories*. Cambridge University Press, Cambridge, England, May, 1998.
- [7] S. Y. Auyang, *How is Quantum Field Theory Possible?* Oxford University Press, New York, NY, July, 1995.
- [8] Y.-B. Yang, J. Liang, Y.-J. Bi, Y. Chen, T. Draper, K.-F. Liu, and Z. Liu, *Proton mass decomposition from the qcd energy momentum tensor*, *Physical Review Letters* **121** (Nov., 2018).
- [9] ATLAS Collaboration, *Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012), no. 1 1–29.
- [10] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Physics Letters B* **716** (2012), no. 1 30–61.

- [11] A. Deur, S. Brodsky, and G. de Teramond, *The QCD running coupling*, *Progress in Particle and Nuclear Physics* **90** (2016) 1–74.
- [12] N. Brambilla, S. Eidelman, P. Foka, S. Gardner, A. S. Kronfeld, M. G. Alford, R. Alkofer, M. Butenschoen, T. D. Cohen, J. Erdmenger, *et al.*, *QCD and strongly coupled gauge theories: challenges and perspectives*, *The European Physical Journal C* **74** (2014), no. 10 2981.
- [13] E. Annala, T. Gorda, A. Kurkela, and A. Vuorinen, *Evidence for quark-matter cores in massive neutron stars*, *Nature Physics* **16** (2020), no. 9 907–910.
- [14] G. E. Brown, K. Kubodera, D. Page, and P. Pizzochero, *Composition, structure and evolution of neutron stars with kaon condensates*, *Physics Reports* **267** (1996), no. 5-6 377–466.
- [15] J. M. Lattimer and M. Prakash, *The physics of neutron stars*, *Science* **304** (2004), no. 5670 536–542.
- [16] E. V. Shuryak, *Quark-gluon plasma and hadronic production of leptons, photons and psions*, *Physics Letters B* **78** (1978), no. 1 150–153.
- [17] J. Adams *et al.*, *Experimental and theoretical challenges in the search for the quark-gluon plasma: The STAR collaboration’s critical assessment of the evidence from RHIC collisions*, *Nuclear Physics A* **757** (2005), no. 1-2 102–183.
- [18] K. Adcox *et al.*, *Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration*, *Nuclear Physics A* **757** (2005), no. 1-2 184–283.
- [19] W. Greiner, S. Schramm, and E. Stein, *Quantum chromodynamics*. Springer, Berlin, Germany, 3 ed., Dec., 2006.

- [20] Z. Fodor and C. Hoelbling, *Light hadron masses from lattice QCD*, *Reviews of Modern Physics* **84** (2012) 449.
- [21] S. Weinberg, *The Quantum Theory of Fields, Volume II: Modern Applications*. Cambridge University Press, Cambridge, 1996.
- [22] J. McGreevy, *Generalized symmetries in condensed matter*, *Annual Review of Condensed Matter Physics* **13** (2022) 247–265. arXiv:2204.03045.
- [23] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, *Generalized global symmetries*, *Journal of High Energy Physics* **2015** (Feb., 2015).
- [24] P. Etingof, D. Nikshych, and V. Ostrik, *On fusion categories*, *Annals of Mathematics* **162** (Sept., 2005) 581–642.
- [25] S.-H. Shao, *What’s done cannot be undone: TASI lectures on non-invertible symmetries*, *arXiv preprint arXiv:2308.00747* (2023).
- [26] O. Diatlyk, C. Luo, Y. Wang, and Y. Zheng, *Gauging non-invertible symmetries: topological interfaces and generalized orbifold groupoid in 2d qft*, *Journal of High Energy Physics* **2024** (2024), no. 3 127.
- [27] E. J. Bergholtz and Z. Liu, *Fractional quantum hall physics in topological flat bands*, *International Journal of Modern Physics B* **27** (2013), no. 24 1330017.
- [28] B. A. Bernevig, *Toward fractional quantum hall physics with cold atoms*, *Physics* **6** (2013) 49.
- [29] X.-L. Qi and S.-C. Zhang, *The quantum spin hall effect and topological insulators*, *Physics Today* **63** (2010), no. 1 33–38.
- [30] A. Stern, *Fractional topological insulators*, *Annual Review of Condensed Matter Physics* **7** (2016) 349–368.

- [31] F. Apruzzi, M. van Beest, D. S. Gould, and S. Schäfer-Nameki, *Holography, 1-form symmetries, and confinement*, *Physical Review D* **104** (Sept., 2021).
- [32] L. Bhardwaj, M. Bullimore, A. E. V. Ferrari, and S. Schäfer-Nameki, *Anomalies of generalized symmetries from solitonic defects*, *SciPost Physics* **16** (Mar., 2024).
- [33] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Topological quantum memory*, *Journal of Mathematical Physics* **43** (2002), no. 9 4452–4505.
- [34] B. Yoshida, *Topological phases with generalized global symmetries*, *Physical Review B* **93** (2016), no. 15 155131.
- [35] L. D. Landau, *On the theory of phase transitions*, *Zh. Eksp. Teor. Fiz.* **7** (1937) 19–32.
- [36] V. L. Ginzburg and L. D. Landau, *On the theory of superconductivity*, *Zh. Eksp. Teor. Fiz.* **20** (1950) 1064–1082.
- [37] L. P. Kadanoff, *Scaling laws for ising models near t_c* , *Physics Physique Fizika* **2** (1966), no. 6 263–272.
- [38] K. G. Wilson, *Renormalization group and critical phenomena. i. renormalization group and the kadanoff scaling picture*, *Physical Review B* **4** (1971), no. 9 3174–3183.
- [39] X.-G. Wen, *Topological orders in rigid states*, *International Journal of Modern Physics B* **4** (1990), no. 02 239–271.
- [40] A. Y. Kitaev, *Fault-tolerant quantum computation by anyons*, *Annals of Physics* **303** (2003), no. 1 2–30.
- [41] D. C. Tsui, H. L. Stormer, and A. C. Gossard, *Two-dimensional magnetotransport in the extreme quantum limit*, *Physical Review Letters* **48** (1982), no. 22 1559–1562.

- [42] R. B. Laughlin, *Anomalous quantum hall effect: An incompressible quantum fluid with fractionally charged excitations*, *Physical Review Letters* **50** (1983), no. 18 1395–1398.
- [43] H. Moradi, S. F. Moosavian, and A. Tiwari, *Topological Holography: Towards a Unification of Landau and Beyond-Landau Physics*, [[arXiv:2207.10712](#)].
- [44] J. Kogut and M. Stephanov, *The phases of quantum chromodynamics: From confinement to extreme environments*, *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology* **21** (2004).
- [45] S. L. Adler, *Axial-vector vertex in spinor electrodynamics*, *Physical Review* **177** (1969), no. 5 2426–2438.
- [46] J. S. Bell and R. Jackiw, *A pcac puzzle: $\pi^0 \rightarrow \gamma\gamma$ in the σ -model*, *Il Nuovo Cimento A* **60** (1969), no. 1 47–61.
- [47] K. G. Wilson, *The renormalization group: Critical phenomena and the kondo problem*, *Reviews of Modern Physics* **47** (1975), no. 4 773–840.
- [48] K. G. Wilson and J. Kogut, *The renormalization group and the ϵ expansion*, *Physics Reports* **12** (1974), no. 2 75–199.
- [49] J. Cardy, *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, 1996.
- [50] G. 't Hooft, *Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking*, *NATO Advanced Study Institutes Series* **59** (1981) 135–157.
- [51] C. Córdova and K. Ohmori, *Anomaly obstructions to symmetry preserving gapped phases*, 2019.

- [52] A. Apte, C. Cordova, and H. T. Lam, *Obstructions to gapped phases from noninvertible symmetries*, *Phys. Rev. B* **108** (2023), no. 4 045134, [arXiv:2212.14605].
- [53] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, *Nature* **521** (2015), no. 7553 436–444.
- [54] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, *Highly accurate protein structure prediction with alphafold*, *Nature* **596** (2021), no. 7873 583–589.
- [55] H. Robbins and S. Monro, *A stochastic approximation method*, *The annals of mathematical statistics* (1951) 400–407.
- [56] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *arXiv preprint arXiv:1412.6980* (2015).
- [57] OpenAI, *Gpt-4 technical report*, *arXiv preprint arXiv:2303.08774* (2023).
- [58] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, *Science* **355** (2017), no. 6325 602–606, [arXiv:1606.02318].
- [59] A. Harju, B. Barbiellini, S. Siljamäki, R. M. Nieminen, and G. Ortiz, *Stochastic gradient descent training of neural networks as quantum monte carlo sampling*, *Physical Review Letters* **79** (1997), no. 7 1173–1177.
- [60] A. Nagy and V. Savona, *Variational quantum monte carlo method with a neural-network ansatz for open quantum systems*, *Physical Review Letters* **122** (2019), no. 25 250501.
- [61] K. Choo, T. Neupert, and G. Carleo, *Two-dimensional frustrated j_1 - j_2 model studied with neural network quantum states*, *Physical Review B* **100** (2019), no. 12 125124.

- [62] D. Luo, J. Carrasquilla, G. Carleo, and B. K. Clark, *Gauge equivariant neural networks for quantum lattice gauge theories*, *Physical Review Letters* **127** (2021), no. 27 276402.
- [63] A. Apte, C. Córdova, T.-C. Huang, and A. Ashmore, *Deep learning lattice gauge theories*, *Physical Review B* **110** (Oct., 2024).
- [64] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Lattice Gauge Equivariant Convolutional Neural Networks*, *Phys. Rev. Lett.* **128** (2022), no. 3 032003, [arXiv:2012.12901].
- [65] R. S. Sutton, *The bitter lesson*, *Incomplete Ideas (blog)* (2019). Accessed: 2025-04-10.
- [66] M. D. Schwartz, *Modern Machine Learning and Particle Physics*, *Harvard Data Science Review* **3** (may 13, 2021). <https://hdsr.mitpress.mit.edu/pub/xqle7lat>.
- [67] D. George and E. A. Huerta, *Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data*, *Physics Letters B* **778** (2018) 64–70.
- [68] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C. W. Park, A. Choudhary, A. Agrawal, S. J. Billinge, *et al.*, *Recent advances and applications of deep learning methods in materials science*, *npj Computational Materials* **8** (2022), no. 1 59.
- [69] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, *Generalized Global Symmetries*, *JHEP* **02** (2015) 172, [arXiv:1412.5148].
- [70] A. Kapustin and R. Thorngren, *Higher symmetry and gapped phases of gauge theories*, arXiv:1309.4721.

- [71] Y. Tachikawa, *On gauging finite subgroups*, *SciPost Phys.* **8** (2020), no. 1 015, [[arXiv:1712.09542](#)].
- [72] C. Córdova, T. T. Dumitrescu, and K. Intriligator, *Exploring 2-Group Global Symmetries*, *JHEP* **02** (2019) 184, [[arXiv:1802.04790](#)].
- [73] F. Benini, C. Córdova, and P.-S. Hsin, *On 2-Group Global Symmetries and their Anomalies*, *JHEP* **03** (2019) 118, [[arXiv:1803.09336](#)].
- [74] L. Bhardwaj and Y. Tachikawa, *On finite symmetries and their gauging in two dimensions*, *Journal of High Energy Physics* **2018** (Mar., 2018).
- [75] C.-M. Chang, Y.-H. Lin, S.-H. Shao, Y. Wang, and X. Yin, *Topological Defect Lines and Renormalization Group Flows in Two Dimensions*, *JHEP* **01** (2019) 026, [[arXiv:1802.04445](#)].
- [76] J. Kaidi, K. Ohmori, and Y. Zheng, *Kramers-Wannier-like Duality Defects in (3+1)D Gauge Theories*, *Phys. Rev. Lett.* **128** (2022), no. 11 111601, [[arXiv:2111.01141](#)].
- [77] Y. Choi, C. Córdova, P.-S. Hsin, H. T. Lam, and S.-H. Shao, *Noninvertible duality defects in 3+1 dimensions*, *Phys. Rev. D* **105** (2022), no. 12 125016, [[arXiv:2111.01139](#)].
- [78] D. Gaiotto and T. Johnson-Freyd, *Condensations in higher categories*, [[arXiv:1905.09566](#)].
- [79] K. Roumpedakis, S. Seifnashri, and S.-H. Shao, *Higher Gauging and Non-invertible Condensation Defects*, [[arXiv:2204.02407](#)].
- [80] L. Bhardwaj, L. E. Bottini, S. Schafer-Nameki, and A. Tiwari, *Non-Invertible Higher-Categorical Symmetries*, [[arXiv:2204.06564](#)].

- [81] Y. Choi, C. Córdova, P.-S. Hsin, H. T. Lam, and S.-H. Shao, *Non-invertible Condensation, Duality, and Triality Defects in 3+1 Dimensions*, [[arXiv:2204.09025](#)].
- [82] D. S. Freed, G. W. Moore, and C. Teleman, *Topological symmetry in quantum field theory*, [[arXiv:2209.07471](#)].
- [83] J. Kaidi, K. Ohmori, and Y. Zheng, *Symmetry TFTs for Non-Invertible Defects*, [[arXiv:2209.11062](#)].
- [84] D. S. Freed, *Introduction to topological symmetry in QFT*, [[arXiv:2212.00195](#)].
- [85] E. H. Lieb, T. Schultz, and D. Mattis, *Two soluble models of an antiferromagnetic chain*, *Annals Phys.* **16** (1961) 407–466.
- [86] L. Bhardwaj, Y. Lee, and Y. Tachikawa, *$SL(2, \mathbb{Z})$ action on QFTs with $\mathbb{Z}_2$ symmetry and the Brown-Kervaire invariants*, *JHEP* **11** (2020) 141, [[arXiv:2009.10099](#)].
- [87] M. Koide, Y. Nagoya, and S. Yamaguchi, *Non-invertible topological defects in 4-dimensional $\mathbb{Z}_2$ pure lattice gauge theory*, *PTEP* **2022** (2022), no. 1 013B03, [[arXiv:2109.05992](#)].
- [88] C. Córdova, K. Ohmori, and T. Rudelius, *Generalized symmetry breaking scales and weak gravity conjectures*, *JHEP* **11** (2022) 154, [[arXiv:2202.05866](#)].
- [89] F. Benini, C. Copetti, and L. Di Pietro, *Factorization and global symmetries in holography*, [[arXiv:2203.09537](#)].
- [90] G. Arias-Tamargo and D. Rodriguez-Gomez, *Non-Invertible Symmetries from Discrete Gauging and Completeness of the Spectrum*, [[arXiv:2204.07523](#)].
- [91] Y. Hayashi and Y. Tanizaki, *Non-invertible self-duality defects of Cardy-Rabinovici model and mixed gravitational anomaly*, *JHEP* **08** (2022) 036, [[arXiv:2204.07440](#)].

- [92] J. Kaidi, G. Zafrir, and Y. Zheng, *Non-invertible symmetries of $\mathcal{N} = 4$ SYM and twisted compactification*, *JHEP* **08** (2022) 053, [[arXiv:2205.01104](#)].
- [93] Y. Choi, H. T. Lam, and S.-H. Shao, *Noninvertible Global Symmetries in the Standard Model*, *Phys. Rev. Lett.* **129** (2022), no. 16 161601, [[arXiv:2205.05086](#)].
- [94] C. Cordova and K. Ohmori, *Non-Invertible Chiral Symmetry and Exponential Hierarchies*, [[arXiv:2205.06243](#)].
- [95] A. Antinucci, G. Galati, and G. Rizzi, *On Continuous 2-Category Symmetries and Yang-Mills Theory*, [[arXiv:2206.05646](#)].
- [96] V. Bashmakov, M. Del Zotto, and A. Hasan, *On the 6d Origin of Non-invertible Symmetries in 4d*, [[arXiv:2206.07073](#)].
- [97] J. A. Damia, R. Argurio, and L. Tizzano, *Continuous Generalized Symmetries in Three Dimensions*, [[arXiv:2206.14093](#)].
- [98] J. A. Damia, R. Argurio, and E. Garcia-Valdecasas, *Non-Invertible Defects in 5d, Boundaries and Holography*, [[arXiv:2207.02831](#)].
- [99] Y. Choi, H. T. Lam, and S.-H. Shao, *Non-invertible Time-reversal Symmetry*, [[arXiv:2208.04331](#)].
- [100] L. Bhardwaj, S. Schafer-Nameki, and J. Wu, *Universal Non-Invertible Symmetries*, [[arXiv:2208.05973](#)].
- [101] T. Bartsch, M. Bullimore, A. E. V. Ferrari, and J. Pearson, *Non-invertible Symmetries and Higher Representation Theory I*, [[arXiv:2208.05993](#)].
- [102] L. Lin, D. G. Robbins, and E. Sharpe, *Decomposition, condensation defects, and fusion*, *Fortsch. Phys.* **70** (2022) 2200130, [[arXiv:2208.05982](#)].

- [103] I. García Etxebarria, *Branes and Non-Invertible Symmetries*, [arXiv:2208.07508].
- [104] F. Apruzzi, I. Bah, F. Bonetti, and S. Schafer-Nameki, *Non-Invertible Symmetries from Holography and Branes*, [arXiv:2208.07373].
- [105] J. J. Heckman, M. Hübner, E. Torres, and H. Y. Zhang, *The Branes Behind Generalized Symmetry Operators*, [arXiv:2209.03343].
- [106] P. Niro, K. Roumpedakis, and O. Sela, *Exploring Non-Invertible Symmetries in Free Theories*, [arXiv:2209.11166].
- [107] N. Mekareeya and M. Sacchi, *Mixed Anomalies, Two-groups, Non-Invertible Symmetries, and 3d Superconformal Indices*, [arXiv:2210.02466].
- [108] A. Antinucci, F. Benini, C. Copetti, G. Galati, and G. Rizi, *The holography of non-invertible self-duality symmetries*, [arXiv:2210.09146].
- [109] S. Chen and Y. Tanizaki, *Solitonic symmetry beyond homotopy: invertibility from bordism and non-invertibility from TQFT*, [arXiv:2210.13780].
- [110] V. Bashmakov, M. Del Zotto, A. Hasan, and J. Kaidi, *Non-invertible Symmetries of Class $\mathcal{S}$ Theories*, [arXiv:2211.05138].
- [111] A. Karasik, *On anomalies and gauging of $U(1)$ non-invertible symmetries in 4d QED*, [arXiv:2211.05802].
- [112] C. Córdova, S. Hong, S. Koren, and K. Ohmori, *Neutrino Masses from Generalized Symmetry Breaking*, [arXiv:2211.07639].
- [113] T. D. Décoppet and M. Yu, *Gauging Noninvertible Defects: A 2-Categorical Perspective*, [arXiv:2211.08436].
- [114] I. García Etxebarria and N. Iqbal, *A Goldstone theorem for continuous non-invertible symmetries*, [arXiv:2211.09570].

- [115] Y. Choi, H. T. Lam, and S.-H. Shao, *Non-invertible Gauss Law and Axions*, [arXiv:2212.04499].
- [116] R. Yokokura, *Non-invertible symmetries in axion electrodynamics*, [arXiv:2212.05001].
- [117] L. Bhardwaj, S. Schafer-Nameki, and A. Tiwari, *Unifying Constructions of Non-Invertible Symmetries*, [arXiv:2212.06159].
- [118] L. Bhardwaj, L. E. Bottini, S. Schafer-Nameki, and A. Tiwari, *Non-Invertible Symmetry Webs*, [arXiv:2212.06842].
- [119] T. Bartsch, M. Bullimore, A. E. V. Ferrari, and J. Pearson, *Non-invertible Symmetries and Higher Representation Theory II*, [arXiv:2212.07393].
- [120] P.-S. Hsin, *Non-Invertible Defects in Nonlinear Sigma Models and Coupling to Topological Orders*, [arXiv:2212.08608].
- [121] A. Das, N. Iqbal, and N. Poovuttikul, *Towards an effective action for chiral magnetohydrodynamics*, [arXiv:2212.09787].
- [122] J. J. Heckman, M. Hubner, E. Torres, X. Yu, and H. Y. Zhang, *Top Down Approach to Topological Duality Defects*, [arXiv:2212.09743].
- [123] A. Antinucci, C. Copetti, G. Galati, and G. Rizi, *"Zoology" of non-invertible duality defects: the view from class $\mathcal{S}$* , [arXiv:2212.09549].
- [124] C. Córdova and S. Koren, *Higher Flavor Symmetries in the Standard Model*, [arXiv:2212.13193].
- [125] E. P. Verlinde, *Fusion Rules and Modular Transformations in 2D Conformal Field Theory*, *Nucl. Phys. B* **300** (1988) 360–376.

- [126] V. B. Petkova and J. B. Zuber, *Generalized twisted partition functions*, *Phys. Lett. B* **504** (2001) 157–164, [[hep-th/0011021](#)].
- [127] J. Fuchs, I. Runkel, and C. Schweigert, *TFT construction of RCFT correlators 1. Partition functions*, *Nucl. Phys. B* **646** (2002) 353–497, [[hep-th/0204148](#)].
- [128] J. Frohlich, J. Fuchs, I. Runkel, and C. Schweigert, *Kramers-Wannier duality from conformal defects*, *Phys. Rev. Lett.* **93** (2004) 070601, [[cond-mat/0404051](#)].
- [129] J. Frohlich, J. Fuchs, I. Runkel, and C. Schweigert, *Duality and defects in rational conformal field theory*, *Nucl. Phys.* **B763** (2007) 354–430, [[hep-th/0607247](#)].
- [130] A. Feiguin, S. Trebst, A. W. W. Ludwig, M. Troyer, A. Kitaev, Z. Wang, and M. H. Freedman, *Interacting anyons in topological quantum liquids: The golden chain*, *Phys. Rev. Lett.* **98** (2007), no. 16 160409, [[cond-mat/0612341](#)].
- [131] J. Frohlich, J. Fuchs, I. Runkel, and C. Schweigert, *Defect lines, dualities, and generalised orbifolds*, in *Proceedings, 16th International Congress on Mathematical Physics (ICMP09): Prague, Czech Republic, August 3-8, 2009*, 2009. [arXiv:0909.5013](#).
- [132] N. Carqueville and I. Runkel, *Orbifold completion of defect bicategories*, *Quantum Topol.* **7** (2016) 203, [[arXiv:1210.6363](#)].
- [133] D. Aasen, R. S. K. Mong, and P. Fendley, *Topological Defects on the Lattice I: The Ising model*, *J. Phys. A* **49** (2016), no. 35 354001, [[arXiv:1601.07185](#)].
- [134] L. Bhardwaj and Y. Tachikawa, *On finite symmetries and their gauging in two dimensions*, *JHEP* **03** (2018) 189, [[arXiv:1704.02330](#)].
- [135] W. Ji, S.-H. Shao, and X.-G. Wen, *Topological Transition on the Conformal Manifold*, *Phys. Rev. Res.* **2** (2020), no. 3 033317, [[arXiv:1909.01425](#)].

- [136] Y.-H. Lin and S.-H. Shao, *Duality Defect of the Monster CFT*, *J. Phys. A* **54** (2021), no. 6 065201, [[arXiv:1911.00042](#)].
- [137] R. Thorngren and Y. Wang, *Fusion Category Symmetry I: Anomaly In-Flow and Gapped Phases*, [[arXiv:1912.02817](#)].
- [138] D. Gaiotto and J. Kulp, *Orbifold groupoids*, *JHEP* **02** (2021) 132, [[arXiv:2008.05960](#)].
- [139] Z. Komargodski, K. Ohmori, K. Roumpedakis, and S. Seifnashri, *Symmetries and strings of adjoint QCD_2* , *JHEP* **03** (2021) 103, [[arXiv:2008.07567](#)].
- [140] D. Aasen, P. Fendley, and R. S. K. Mong, *Topological Defects on the Lattice: Dualities and Degeneracies*, [[arXiv:2008.08598](#)].
- [141] C.-M. Chang and Y.-H. Lin, *Lorentzian dynamics and factorization beyond rationality*, *JHEP* **10** (2021) 125, [[arXiv:2012.01429](#)].
- [142] M. Nguyen, Y. Tanizaki, and M. Ünsal, *Noninvertible 1-form symmetry and Casimir scaling in 2D Yang-Mills theory*, *Phys. Rev. D* **104** (2021), no. 6 065003, [[arXiv:2104.01824](#)].
- [143] R. Thorngren and Y. Wang, *Fusion Category Symmetry II: Categoriosities at $c = 1$ and Beyond*, [[arXiv:2106.12577](#)].
- [144] E. Sharpe, *Topological operators, noninvertible symmetries and decomposition*, [[arXiv:2108.13423](#)].
- [145] T.-C. Huang, Y.-H. Lin, and S. Seifnashri, *Construction of two-dimensional topological field theories with non-invertible symmetries*, *JHEP* **12** (2021) 028, [[arXiv:2110.02958](#)].

- [146] T.-C. Huang, Y.-H. Lin, K. Ohmori, Y. Tachikawa, and M. Tezuka, *Numerical Evidence for a Haagerup Conformal Field Theory*, *Phys. Rev. Lett.* **128** (2022), no. 23 231603, [[arXiv:2110.03008](#)].
- [147] R. Vanhove, L. Lootens, M. Van Damme, R. Wolf, T. J. Osborne, J. Haegeman, and F. Verstraete, *Critical Lattice Model for a Haagerup Conformal Field Theory*, *Phys. Rev. Lett.* **128** (2022), no. 23 231602, [[arXiv:2110.03532](#)].
- [148] I. M. Burbano, J. Kulp, and J. Neuser, *Duality defects in E_8* , *JHEP* **10** (2022) 186, [[arXiv:2112.14323](#)].
- [149] K. Inamura, *Fermionization of fusion category symmetries in 1+1 dimensions*, [[arXiv:2206.13159](#)].
- [150] C.-M. Chang, J. Chen, and F. Xu, *Topological Defect Lines in Two Dimensional Fermionic CFTs*, [[arXiv:2208.02757](#)].
- [151] Y.-H. Lin, M. Okada, S. Seifnashri, and Y. Tachikawa, *Asymptotic density of states in 2d CFTs with non-invertible symmetries*, [[arXiv:2208.05495](#)].
- [152] D. Robbins, E. Sharpe, and T. Vandermeulen, *Decomposition, Trivially-Acting Symmetries, and Topological Operators*, [[arXiv:2211.14332](#)].
- [153] C. Wang and T. Senthil, *Interacting fermionic topological insulators/superconductors in three dimensions*, *Phys. Rev. B* **89** (2014), no. 19 195124, [[arXiv:1401.1142](#)].
[Erratum: *Phys.Rev.B* 91, 239902 (2015)].
- [154] C. Wang and T. Senthil, *Composite fermi liquids in the lowest Landau level*, *Phys. Rev. B* **94** (2016), no. 24 245107, [[arXiv:1604.06807](#)].

- [155] I. Sodemann, I. Kimchi, C. Wang, and T. Senthil, *Composite fermion duality for half-filled multicomponent Landau Levels*, *Phys. Rev. B* **95** (2017), no. 8 085135, [[arXiv:1609.08616](#)].
- [156] C. Wang, A. Nahum, M. A. Metlitski, C. Xu, and T. Senthil, *Deconfined quantum critical points: symmetries and dualities*, *Phys. Rev. X* **7** (2017), no. 3 031051, [[arXiv:1703.02426](#)].
- [157] R. Kobayashi, K. Shiozaki, Y. Kikuchi, and S. Ryu, *Lieb-Schultz-Mattis type theorem with higher-form symmetry and the quantum dimer models*, *Phys. Rev. B* **99** (2019), no. 1 014402, [[arXiv:1805.05367](#)].
- [158] J. Wang, X.-G. Wen, and E. Witten, *Symmetric Gapped Interfaces of SPT and SET States: Systematic Constructions*, *Phys. Rev. X* **8** (2018), no. 3 031048, [[arXiv:1705.06728](#)].
- [159] Z. Wan and J. Wang, *Adjoint QCD₄, Deconfined Critical Phenomena, Symmetry-Enriched Topological Quantum Field Theory, and Higher Symmetry-Extension*, *Phys. Rev. D* **99** (2019), no. 6 065013, [[arXiv:1812.11955](#)].
- [160] M. Guo, K. Ohmori, P. Putrov, Z. Wan, and J. Wang, *Fermionic Finite-Group Gauge Theories and Interacting Symmetric/Crystalline Orders via Cobordisms*, *Commun. Math. Phys.* **376** (2020), no. 2 1073–1154, [[arXiv:1812.11959](#)].
- [161] R. Kobayashi, K. Ohmori, and Y. Tachikawa, *On gapped boundaries for SPT phases beyond group cohomology*, *JHEP* **11** (2019) 131, [[arXiv:1905.05391](#)].
- [162] C. Córdova and K. Ohmori, *Anomaly Obstructions to Symmetry Preserving Gapped Phases*, [[arXiv:1910.04962](#)].
- [163] C. Córdova and K. Ohmori, *Anomaly Constraints on Gapped Phases with Discrete Chiral Symmetry*, *Phys. Rev. D* **102** (2020), no. 2 025011, [[arXiv:1912.13069](#)].

- [164] R. Thorngren, *Topological quantum field theory, symmetry breaking, and finite gauge theory in 3+1D*, *Phys. Rev. B* **101** (2020), no. 24 245160, [[arXiv:2001.11938](#)].
- [165] P. Putrov, J. Wang, and S.-T. Yau, *Braiding Statistics and Link Invariants of Bosonic/Fermionic Topological Quantum Matter in 2+1 and 3+1 dimensions*, *Annals Phys.* **384** (2017) 254–287, [[arXiv:1612.09298](#)].
- [166] P.-S. Hsin and A. Turzillo, *Symmetry-enriched quantum spin liquids in $(3 + 1)d$* , *JHEP* **09** (2020) 022, [[arXiv:1904.11550](#)].
- [167] P.-S. Hsin, H. T. Lam, and N. Seiberg, *Comments on one-form global symmetries and their gauging in 3d and 4d*, *SciPost Physics* **6** (Mar., 2019).
- [168] D. Delmastro and J. Gomis, *Symmetries of abelian chern-simons theories and arithmetic*, [[arXiv:1904.12884](#)].
- [169] A. Kapustin, R. Thorngren, A. Turzillo, and Z. Wang, *Fermionic Symmetry Protected Topological Phases and Cobordisms*, *JHEP* **12** (2015) 052, [[arXiv:1406.7329](#)].
- [170] I. Hason, Z. Komargodski, and R. Thorngren, *Anomaly Matching in the Symmetry Broken Phase: Domain Walls, CPT, and the Smith Isomorphism*, *SciPost Phys.* **8** (2020) 62.
- [171] C. Córdova, K. Ohmori, S.-H. Shao, and F. Yan, *Decorated z_2 symmetry defects and their time-reversal anomalies*, *Physical Review D* **102** (Aug., 2020).
- [172] A. Kapustin and R. Thorngren, *Higher symmetry and gapped phases of gauge theories*, [[arXiv:1309.4721](#)].
- [173] R. Thorngren and Y. Wang, *Fusion category symmetry i: Anomaly in-flow and gapped phases*, .

- [174] J. Villain, *Theory of one-dimensional and two-dimensional magnets with an easy magnetization plane. 2. The Planar, classical, two-dimensional magnet*, *J. Phys. (France)* **36** (1975) 581–590.
- [175] F. J. Wegner, *Duality in Generalized Ising Models and Phase Transitions Without Local Order Parameters*, *J. Math. Phys.* **12** (1971) 2259–2272.
- [176] E. Fradkin and L. Susskind, *Order and disorder in gauge systems and magnets*, *Phys. Rev. D* **17** (May, 1978) 2637–2658.
- [177] D. Horn, M. Weinstein, and S. Yankielowicz, *Hamiltonian approach to $z(n)$ lattice gauge theories*, *Phys. Rev. D* **19** (Jun, 1979) 3715–3731.
- [178] S. Elitzur, R. B. Pearson, and J. Shigemitsu, *Phase structure of discrete abelian spin and gauge systems*, *Phys. Rev. D* **19** (Jun, 1979) 3698–3714.
- [179] A. Ukawa, P. Windey, and A. H. Guth, *Dual Variables for Lattice Gauge Theories and the Phase Structure of $Z(N)$ Systems*, *Phys. Rev. D* **21** (1980) 1013.
- [180] R. Savit, *Duality in Field Theory and Statistical Systems*, *Rev. Mod. Phys.* **52** (1980) 453.
- [181] M. Creutz, L. Jacobs, and C. Rebbi, *Monte carlo study of abelian lattice gauge theories*, *Phys. Rev. D* **20** (Oct, 1979) 1915–1922.
- [182] J. L. Cardy and E. Rabinovici, *Phase structure of zp models in the presence of a θ parameter*, *Nuclear Physics B* **205** (Apr., 1982) 1–16.
- [183] J. L. Cardy, *Duality and the Theta Parameter in Abelian Lattice Models*, *Nucl. Phys. B* **205** (1982) 17–26.

- [184] 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](#)].
- [185] 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](#)].
- [186] M. A. Metlitski, *S -duality of $u(1)$ gauge theory with $\theta = \pi$ on non-orientable manifolds: Applications to topological insulators and superconductors*, [[arXiv:1510.05663](#)].
- [187] M. Honda and Y. Tanizaki, *Topological aspects of 4d abelian lattice gauge theories with the θ parameter*, *Journal of High Energy Physics* **2020** (Dec., 2020).
- [188] Y. Hayashi and Y. Tanizaki, *Non-invertible self-duality defects of cardy-rabinovici model and mixed gravitational anomaly*, *Journal of High Energy Physics* **2022** (Aug., 2022).
- [189] B. Moy, H. Goldman, R. Sohal, and E. Fradkin, *Theory of oblique topological insulators*, [[arXiv:2206.07725](#)].
- [190] S. Alexandru, *The Wild World of 4-manifolds*. American Mathematical Society, Providence, RI.
- [191] T. Nagell, *Introduction to number theory*. American Mathematical Society, Providence, RI.
- [192] B. Durhuus and T. Jonsson, *Classification and construction of unitary topological field theories in two dimensions*, *Journal of Mathematical Physics* **35** (Oct., 1994) 5306–5313.

- [193] S. Sawin, *Direct sum decompositions and indecomposable TQFTs*, *Journal of Mathematical Physics* **36** (Dec., 1995) 6673–6680.
- [194] K. G. Wilson, *Confinement of Quarks*, *Phys. Rev. D* **10** (1974) 2445–2459.
- [195] J. B. Kogut and L. Susskind, *Hamiltonian Formulation of Wilson’s Lattice Gauge Theories*, *Phys. Rev. D* **11** (1975) 395–408.
- [196] F. Wilczek, *QCD made simple*, *Physics Today* **53** (2000), no. 8 22–28.
- [197] W. Greiner, S. Schramm, and S. Eckart, *Quantum Chromodynamics*. Springer, 3 ed., 2006.
- [198] J. Greensite, *An Introduction to the Confinement Problem*. Springer Berlin Heidelberg, 2011.
- [199] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1.*, *Phys. Rev.* **122** (1961) 345–358.
- [200] Y. Nambu and G. Jona-Lasinio, *Dynamical model of elementary particles based on an analogy with superconductivity. II.*, *Phys. Rev.* **124** (1961) 246–254.
- [201] C. Rebbi, *Lattice Gauge Theories and Monte Carlo Simulations*. World Scientific, 1983.
- [202] Z. Fodor and S. D. Katz, *Lattice determination of the critical point of QCD at finite T and μ* , *JHEP* **03** (2002) 014, [[hep-lat/0106002](#)].
- [203] BMW Collaboration, S. Durr *et al.*, *Ab-Initio Determination of Light Hadron Masses*, *Science* **322** (2008) 1224–1227, [[arXiv:0906.3599](#)].
- [204] Y. Zhang, T. Grover, and A. Vishwanath, *Topological entanglement entropy of z_2 spin liquids and lattice Laughlin states*, *Phys. Rev. B* **84** (2011) 075128, [[arXiv:1106.0015](#)].

- [205] J. McMinis and N. M. Tubman, *Renyi entropy of the interacting Fermi liquid*, *Phys. Rev. B* **87** (2013) 081108, [[arXiv:1207.4188](#)].
- [206] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, *Nature* **521** (2015), no. 7553 436–444.
- [207] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, 4 ed., 2013.
- [208] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, *Automatic differentiation in machine learning: a survey*, *J. Mach. Learn. Res.* **18** (2018), no. 153 1–43, [[arXiv:1502.05767](#)].
- [209] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, *Nature* **323** (1986), no. 6088 533–536.
- [210] M. B. Hastings, *An area law for one-dimensional quantum systems*, *J. Stat. Mech.: Theory Exp.* **2007** (2007), no. 08 P08024–P08024.
- [211] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, *Math. Control Signals Syst.* **2** (1989), no. 4 303–314.
- [212] Z. Denis, A. Sinibaldi, and G. Carleo, *Comment on ‘Can Neural Quantum States Learn Volume-Law Ground States?’*, [[arXiv:2309.11534](#)].
- [213] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Equivalence of restricted boltzmann machines and tensor network states*, *Phys. Rev. B* **97** (2018), no. 8 085104, [[arXiv:1701.04831](#)].
- [214] O. Sharir, A. Shashua, and G. Carleo, *Neural tensor contractions and the expressive power of deep neural quantum states*, *Phys. Rev. B* **106** (2022), no. 20 205136, [[arXiv:2103.10293](#)].

- [215] S. Weinzierl, *Introduction to Monte Carlo methods*, [hep-ph/0006269].
- [216] Y. Nomura, N. Yoshioka, and F. Nori, *Purifying deep Boltzmann machines for thermal quantum states*, *Phys. Rev. Lett.* **127** (2021) 060601, [arXiv:2103.04791].
- [217] H. Lange, A. Van de Walle, A. Abedinnia, and A. Bohrdt, *From Architectures to Applications: A Review of Neural Quantum States*, [arXiv:2402.09402].
- [218] M. Medvidović and J. R. Moreno, *Neural-network quantum states for many-body physics*, [arXiv:2402.11014].
- [219] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Lattice Gauge Symmetry in Neural Networks*, *PoS LATTICE2021* (2022) 185, [arXiv:2111.04389].
- [220] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Preserving gauge invariance in neural networks*, *EPJ Web Conf.* **258** (2022) 09004, [arXiv:2112.11239].
- [221] M. Favoni, A. Ipp, and D. I. Müller, *Applications of Lattice Gauge Equivariant Neural Networks*, *EPJ Web Conf.* **274** (2022) 09001, [arXiv:2212.00832].
- [222] J. Aronsson, D. I. Müller, and D. Schuh, *Geometrical aspects of lattice gauge equivariant convolutional neural networks*, [arXiv:2303.11448].
- [223] U. Wenger, K. Holland, A. Ipp, and D. I. Müller, *Fixed point actions from convolutional neural networks*, *PoS LATTICE2023* (2024) 038, [arXiv:2311.17816].
- [224] K. Holland, A. Ipp, D. I. Müller, and U. Wenger, *Machine learning a fixed point action for $SU(3)$ gauge theory with a gauge equivariant convolutional neural network*, [arXiv:2401.06481].

- [225] D. Luo, G. Carleo, B. K. Clark, and J. Stokes, *Gauge Equivariant Neural Networks for Quantum Lattice Gauge Theories*, *Phys. Rev. Lett.* **127** (2021), no. 27 276402, [arXiv:2012.05232].
- [226] D. Luo, S. Yuan, J. Stokes, and B. K. Clark, *Gauge Equivariant Neural Networks for 2+1D $U(1)$ Gauge Theory Simulations in Hamiltonian Formulation*, [arXiv:2211.03198].
- [227] F. Vicentini *et al.*, *NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems*, *SciPost Phys. Codeb.* **2022** (2022) 7, [arXiv:2112.10526].
- [228] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, “JAX: composable transformations of Python + NumPy programs.” <https://github.com/google/jax>.
- [229] D. Boyda *et al.*, *Applications of Machine Learning to Lattice Quantum Field Theory*, in *Snowmass 2021*, 2022. arXiv:2202.05838.
- [230] S. Bulusu, M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, *Generalization capabilities of translationally equivariant neural networks*, *Phys. Rev. D* **104** (2021), no. 7 074504, [arXiv:2103.14686].
- [231] D. Luo, Z. Chen, K. Hu, Z. Zhao, V. M. Hur, and B. K. Clark, *Gauge Invariant and Anyonic Symmetric Autoregressive Neural Networks for Quantum Lattice Models*, *Phys. Rev. Res.* **5** (2023) 013216, [arXiv:2101.07243].
- [232] R. Bulirsch and J. Stoer, *Numerical treatment of ordinary differential equations by extrapolation methods*, *Numer. Math.* **8** (1966), no. 1 1–13.
- [233] M. Henkel and A. Patkos, *Conformal structure in the spectrum of an altered quantum Ising chain*, *J. Phys. A: Math. Gen.* **20** (1987), no. 8 2199.

- [234] M. Henkel and G. Schutz, *Finite-lattice extrapolation algorithms*, *J. Phys. A: Math. Gen* **21** (1988), no. 11 2617.
- [235] J. Choquette, W. Gandhi, O. Giroux, N. Stam, and R. Krashinsky, *NVIDIA A100 Tensor Core GPU: Performance and Innovation*, *IEEE Micro* **41** (2021), no. 2 29–35.
- [236] A. Athenodorou and M. Teper, *On the spectrum and string tension of $U(1)$ lattice gauge theory in $2 + 1$ dimensions*, *JHEP* **01** (2019) 063, [arXiv:1811.06280].
- [237] C. Luo, A. Athenodorou, S. Dubovsky, and M. Teper, *Confining strings and glueballs in $\mathbb{Z}_N$ gauge theories*, *PoS LATTICE2023* (2024) 375, [arXiv:2312.03855].
- [238] K. Beach, L. Wang, and A. W. Sandvik, *Data collapse in the critical region using finite-size scaling with subleading corrections*, [cond-mat/0505194].
- [239] D.-L. Deng, X. Li, and S. Das Sarma, *Quantum entanglement in neural network states*, *Phys. Rev. X* **7** (2017), no. 2 021021, [arXiv:1701.04844].
- [240] R. Medina, R. Vasseur, and M. Serbyn, *Entanglement transitions from restricted boltzmann machines*, *Phys. Rev. B* **104** (2021), no. 10 104205.
- [241] A. Bulgarelli and M. Panero, *Entanglement entropy from non-equilibrium Monte Carlo simulations*, *JHEP* **06** (2023) 030, [arXiv:2304.03311].
- [242] A. Bulgarelli and M. Panero, *Duality transformations and the entanglement entropy of gauge theories*, *JHEP* **06** (2024) 041, [arXiv:2404.01987].
- [243] D. Kochkov and B. K. Clark, *Variational optimization in the ai era: Computational graph states and supervised wave-function optimization*, 2018.
- [244] A. Szabó and C. Castelnovo, *Neural network wave functions and the sign problem*, *Phys. Rev. Research* **2** **2** (2020), no. 3 033075, [arXiv:2002.04613].

- [245] A. Cuzzocrea, A. Scemama, W. J. Briels, S. Moroni, and C. Filippi, *Variational principles in quantum Monte Carlo: The troubled story of variance minimization*, *J. Chem. Theory Comput.* **16** (2020), no. 7 4203–4212, [arXiv:2002.05951].
- [246] D.-B. Zhang, B.-L. Chen, Z.-H. Yuan, and T. Yin, *Variational quantum eigensolvers by variance minimization*, *Chin. Phys. B* **31** (2022), no. 12 120301, [arXiv:2006.15781].
- [247] T. S. Cohen and M. Welling, *Group Equivariant Convolutional Networks*, [arXiv:1602.07576].
- [248] C. Roth and A. H. MacDonald, *Group convolutional neural networks improve quantum state accuracy*, [arXiv:2104.05085].
- [249] P. Emonts, M. C. Bañuls, I. Cirac, and E. Zohar, *Variational Monte Carlo simulation with tensor networks of a pure $\mathbb{Z}_3$ gauge theory in $(2+1)d$* , *Phys. Rev. D* **102** (2020), no. 7 074501, [arXiv:2008.00882].
- [250] D. Horn, M. Weinstein, and S. Yankielowicz, *Hamiltonian Approach to $\mathbb{Z}_N$ Lattice Gauge Theories*, *Phys. Rev. D* **19** (1979) 3715.
- [251] P. Emonts, A. Kelman, U. Borla, S. Moroz, S. Gazit, and E. Zohar, *Finding the ground state of a lattice gauge theory with fermionic tensor networks: A $2+1D$ $\mathbb{Z}_2$ demonstration*, *Phys. Rev. D* **107** (2023), no. 1 014505, [arXiv:2211.00023].
- [252] S. Elitzur, *Impossibility of Spontaneously Breaking Local Symmetries*, *Phys. Rev. D* **12** (1975) 3978–3982.
- [253] E. H. Fradkin and L. Susskind, *Order and Disorder in Gauge Systems and Magnets*, *Phys. Rev. D* **17** (1978) 2637.

- [254] A. Ritz-Zwilling, J.-N. Fuchs, and J. Vidal, *Wegner-Wilson loops in string nets*, *Phys. Rev. B* **103** (2021), no. 7 075128, [[arXiv:2011.12609](#)].
- [255] M. Creutz, *Asymptotic Freedom Scales*, *Phys. Rev. Lett.* **45** (1980) 313.
- [256] A. Gonzalez-Arroyo and M. Okawa, *The string tension from smeared Wilson loops at large N* , *Phys. Lett. B* **718** (2013) 1524–1528, [[arXiv:1206.0049](#)].
- [257] S. Sachdev, *Topological order, emergent gauge fields, and Fermi surface reconstruction*, *Rept. Prog. Phys.* **82** (2019), no. 1 014001, [[arXiv:1801.01125](#)].
- [258] B. C. Rayhaun and D. J. Williamson, *Higher-form subsystem symmetry breaking: Subdimensional criticality and fracton phase transitions*, *SciPost Phys.* **15** (2023), no. 1 017, [[arXiv:2112.12735](#)].
- [259] S. S. Bullock and G. K. Brennen, *Qudit surface codes and gauge theory with finite cyclic groups*, *J. Phys. A* **40** (2007) 3481–3505, [[quant-ph/0609070](#)].
- [260] N. Read and S. Sachdev, *Large- N expansion for frustrated quantum antiferromagnets*, *Phys. Rev. Lett.* **66** (1991) 1773–1776.
- [261] X. Wen, *Mean-field theory of spin-liquid states with finite energy gap and topological orders*, *Phys. Rev. B* **44** (1991), no. 6 2664–2672.
- [262] F. A. Bais, P. van Driel, and M. de Wild Propitius, *Quantum symmetries in discrete gauge theories*, *Phys. Lett. B* **280** (1992) 63–70, [[hep-th/9203046](#)].
- [263] J. M. Maldacena, G. W. Moore, and N. Seiberg, *D-brane charges in five-brane backgrounds*, *JHEP* **10** (2001) 005, [[hep-th/0108152](#)].
- [264] A. Y. Kitaev, *Fault tolerant quantum computation by anyons*, *Annals Phys.* **303** (2003) 2–30, [[quant-ph/9707021](#)].

- [265] M. Freedman, C. Nayak, K. Shtengel, K. Walker, and Z. Wang, *A class of P, T -invariant topological phases of interacting electrons*, *Annals Phys.* **310** (2004), no. 2 428–492, [[cond-mat/0307511](#)].
- [266] T. H. Hansson, V. Oganesyan, and S. L. Sondhi, *Superconductors are topologically ordered*, *Annals Phys.* **313** (2004), no. 2 497–538, [[cond-mat/0404327](#)].
- [267] S. Sachdev, *Quantum Phase Transitions*. Cambridge University Press, 2011.
- [268] J. Zinn-Justin, *Quantum field theory and critical phenomena*, *Int. Ser. Monogr. Phys.* **92** (1996) 1–1008.
- [269] X. Chen, Z. C. Gu, and X. G. Wen, *Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order*, *Phys. Rev. B* **82** (2010) 155138, [[arXiv:1004.3835](#)].
- [270] G. Bhanot and M. Creutz, *Phase diagram of $Z(N)$ and $U(1)$ gauge theories in three dimensions*, *Phys. Rev. D* **21** (1980), no. 10 2892–2902.
- [271] C. Omero, *Gauge-invariant variational study of the Hamiltonian $U(1)$ and Z_N model and critical space-time dimensionality*, *Phys. Lett. B* **110** (1982), no. 3–4 265–270.
- [272] O. Borisenko, V. Chelnokov, G. Cortese, M. Gravina, A. Papa, and I. Surzhikov, *Phase transitions in the three-dimensional $Z(N)$ models*, *PoS LATTICE2013* (2014) 347, [[arXiv:1311.0471](#)].
- [273] J. Vidal, S. Dusuel, and K. P. Schmidt, *Low-energy effective theory of the toric code model in a parallel field*, *Phys. Rev. B* **79** (2009) 033109, [[arXiv:0807.0487](#)].
- [274] M. D. Schulz, S. Dusuel, R. Orus, J. Vidal, and K. P. Schmidt, *Breakdown of a perturbed Z_N topological phase*, *New J. Phys.* **14** (2012) 025005, [[arXiv:1110.3632](#)].

- [275] H. Grosse, C. B. Lang, and H. Nicolai, *Equivalence of the $Z(4)$ and $Z(2) \times Z(2)$ Lattice Gauge Theories*, *Phys. Lett. B* **98** (1981) 69–73.
- [276] M. Campostrini, M. Hasenbusch, A. Pelissetto, P. Rossi, and E. Vicari, *Critical behavior of the three-dimensional XY universality class*, *Phys. Rev. B* **63** (2001) 214503, [[cond-mat/0010360](#)].
- [277] A. M. Polyakov, *Compact Gauge Fields and the Infrared Catastrophe*, *Phys. Lett. B* **59** (1975) 82–84.
- [278] A. M. Polyakov, *Quark Confinement and Topology of Gauge Groups*, *Nucl. Phys. B* **120** (1977) 429–458.
- [279] M. Creutz, L. Jacobs, and C. Rebbi, *Monte Carlo Study of Abelian Lattice Gauge Theories*, *Phys. Rev. D* **20** (1979) 1915.
- [280] H. W. J. Blöte and Y. Deng, *Cluster Monte Carlo simulation of the transverse Ising model*, *Phys. Rev. E* **66** (2002) 066110.
- [281] I. S. Tupitsyn, A. Kitaev, N. V. Prokof'ev, and P. C. E. Stamp, *Topological multicritical point in the Toric Code and 3D gauge Higgs Models*, *Phys. Rev. B* **82** (2010) 085114, [[arXiv:0804.3175](#)].
- [282] J. Tilly *et al.*, *The Variational Quantum Eigensolver: A review of methods and best practices*, *Phys. Rept.* **986** (2022) 1–128, [[arXiv:2111.05176](#)].
- [283] U. Wolff, *Critical slowing down*, *Nucl. Phys. B Proc. Suppl.* **17** (1990) 93–102.
- [284] I. Gühring, M. Raslan, and G. Kutyniok, *Expressivity of Deep Neural Networks*, p. 149–199. Cambridge University Press, 2022.

- [285] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. [arXiv:1512.03385](#).
- [286] J. Heek, A. Levskaya, A. Oliver, M. Ritter, B. Rondepierre, A. Steiner, and M. van Zee, “Flax: A neural network library and ecosystem for JAX.” <https://github.com/google/flax>.
- [287] R. Assaraf and M. Caffarel, *Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces*, *J. Chem. Phys.* **119** (2003), no. 20 10536–10552, [[physics/0310035](#)].
- [288] D. Mishkin and J. Matas, *All you need is a good init*, [[arXiv:1511.06422](#)].
- [289] S. Sorella, *Green function Monte Carlo with stochastic reconfiguration*, *Phys. Rev. Lett.* **80** (1998), no. 20 4558–4561, [[cond-mat/9803107](#)].
- [290] S. Sorella, *Generalized Lanczos algorithm for variational quantum Monte Carlo*, *Phys. Rev. B* **64** (2001), no. 2 024512, [[cond-mat/0009149](#)].
- [291] S.-i. Amari, *Natural Gradient Works Efficiently in Learning*, *Neural Comput.* **10** (1998), no. 2 251–276.
- [292] C.-Y. Park and M. J. Kastoryano, *Geometry of learning neural quantum states*, *Phys. Rev. Research* **2** (2020), no. 2 023232, [[arXiv:1910.11163](#)].
- [293] A. Chen and M. Heyl, *Efficient optimization of deep neural quantum states toward machine precision*, [[arXiv:2302.01941](#)].
- [294] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, *A simple linear algebra identity to optimize large-scale neural network quantum states*, [[arXiv:2310.05715](#)].

- [295] C. G. West, A. Garcia-Saez, and T.-C. Wei, *Efficient evaluation of high-order moments and cumulants in tensor network states*, *Phys. Rev. B* **92** (2015), no. 11 115103, [arXiv:1503.07045].
- [296] D. Wu *et al.*, *Variational Benchmarks for Quantum Many-Body Problems*, [arXiv:2302.04919].
- [297] D. H. Weinstein, *Modified ritz method*, *Proc. Natl. Acad. Sci.* **20** (1934), no. 9 529–532.
- [298] A. F. Stevenson, *On the lower bounds of Weinstein and Romberg in quantum mechanics*, *Phys. Rev.* **53** (1938) 199–199.
- [299] A. Kitaev, *Anyons in an exactly solved model and beyond*, *Annals Phys.* **321** (2006), no. 1 2–111, [cond-mat/0506438].
- [300] F. J. Wegner, *Duality in Generalized Ising Models and Phase Transitions Without Local Order Parameters*, *J. Math. Phys.* **12** (1971) 2259–2272.
- [301] S. Elitzur, R. B. Pearson, and J. Shigemitsu, *The Phase Structure of Discrete Abelian Spin and Gauge Systems*, *Phys. Rev. D* **19** (1979) 3698.
- [302] S. Trebst, P. Werner, M. Troyer, K. Shtengel, and C. Nayak, *Breakdown of a topological phase: Quantum phase transition in a loop gas model with tension*, *Phys. Rev. Lett.* **98** (2007) 070602, [cond-mat/0609048].
- [303] C.-N. Yang and T. D. Lee, *Statistical theory of equations of state and phase transitions. 1. Theory of condensation*, *Phys. Rev.* **87** (1952) 404–409.
- [304] T. D. Lee and C.-N. Yang, *Statistical theory of equations of state and phase transitions. 2. Lattice gas and Ising model*, *Phys. Rev.* **87** (1952) 410–419.

- [305] M. E. Fisher and M. N. Barber, *Scaling theory for finite-size effects in the critical region*, *Phys. Rev. Lett.* **28** (1972) 1516–1519.
- [306] S. Rychkov, *3D Ising Model: a view from the Conformal Bootstrap Island*, *Comptes Rendus Physique* **21** (2020), no. 2 185–198, [[arXiv:2007.14315](#)].
- [307] S. M. Bhattacharjee and F. Seno, *A measure of data collapse for scaling*, *J. Phys. A: Math. Gen.* **34** (2001), no. 33 6375–6380.
- [308] F. Kos, D. Poland, D. Simmons-Duffin, and A. Vichi, *Precision Islands in the Ising and $O(N)$ Models*, *JHEP* **08** (2016) 036, [[arXiv:1603.04436](#)].
- [309] D. Simmons-Duffin, *The Lightcone Bootstrap and the Spectrum of the 3d Ising CFT*, *JHEP* **03** (2017) 086, [[arXiv:1612.08471](#)].
- [310] J. Xu, A. M. Ferrenberg, and D. P. Landau, *92 Years of the Ising Model: A High Resolution Monte Carlo Study*, *J. Phys. Conf. Ser.* **1012** (2018), no. 1 012002.
- [311] M. Hasenbusch, *Finite size scaling study of lattice models in the three-dimensional Ising universality class*, *Phys. Rev. B* **82** (2010), no. 17 174433, [[arXiv:1004.4486](#)].
- [312] A. M. Ferrenberg, J. Xu, and D. P. Landau, *Pushing the limits of Monte Carlo simulations for the three-dimensional Ising model*, *Phys. Rev. E* **97** (2018), no. 4 043301, [[arXiv:1806.03558](#)].
- [313] R. Guida and J. Zinn-Justin, *Critical exponents of the N -vector model*, *J. Phys. A: Math. Gen.* **31** (1998), no. 40 8103–8121.
- [314] J. B. Kogut, *An Introduction to Lattice Gauge Theory and Spin Systems*, *Rev. Mod. Phys.* **51** (1979) 659.

- [315] D. Robaina, M. C. Bañuls, and J. I. Cirac, *Simulating $2 + 1D$ Z_3 Lattice Gauge Theory with an Infinite Projected Entangled-Pair State*, *Phys. Rev. Lett.* **126** (2021), no. 5 050401, [[arXiv:2007.11630](#)].
- [316] S. Iino, S. Morita, N. Kawashima, and A. W. Sandvik, *Detecting signals of weakly first-order phase transitions in two-dimensional Potts models*, *J. Phys. Soc. Jpn.* **88** (2019), no. 3 034006, [[arXiv:1801.02786](#)].
- [317] E. Buffenoir and S. Wallon, *The Correlation length of the Potts model at the first order transition point*, *J. Phys. A* **26** (1993) 3045–3062.
- [318] M. Fukugita and M. Okawa, *Correlation Length of the Three State Potts Model in Three-dimensions*, *Phys. Rev. Lett.* **63** (1989) 13.
- [319] J. Cotler and S. Rezchikov, *Renormalizing Diffusion Models*, [[arXiv:2308.12355](#)].
- [320] H. E. Stanley, *Scaling, universality, and renormalization: Three pillars of modern critical phenomena*, *Rev. Mod. Phys.* **71** (1999), no. 2 S358–S366.
- [321] B. Widom, *The critical point and scaling theory*, *Physica* **73** (1974), no. 1 107–118.
- [322] M. Laplace, *XIII. General method for determining the orbits of comets*, *The Philosophical Magazine* **39** (1812), no. 166 85–90.
- [323] B. Efron, *Bootstrap methods: Another look at the jackknife*, in *Breakthroughs in Statistics: Methodology and Distribution*, pp. 569–593. Springer New York, 1992.
- [324] H. Suwa, A. Sen, and A. W. Sandvik, *Level spectroscopy in a two-dimensional quantum magnet: Linearly dispersing spinons at the deconfined quantum critical point*, *Phys. Rev. B* **94** (2016), no. 14 144416, [[arXiv:1607.05110](#)].
- [325] D. Gaiotto and T. Johnson-Freyd, *Condensations in higher categories*, 2025.

- [326] T. D. Décoppet, P. Huston, T. Johnson-Freyd, D. Nikshych, D. Penneys, J. Plavnik, D. Reutter, and M. Yu, *The Classification of Fusion 2-Categories*, [arXiv:2411.05907].
- [327] L. Bhardwaj, T. Décoppet, S. Schafer-Nameki, and M. Yu, *Fusion 3-Categories for Duality Defects*, [arXiv:2408.13302].
- [328] T. D. Décoppet and M. Yu, *Fiber 2-Functors and Tambara–Yamagami Fusion 2-Categories*, *Commun. Math. Phys.* **406** (2025), no. 3 64, [arXiv:2306.08117].
- [329] C. Córdova, P.-S. Hsin, and C. Zhang, *Anomalies of non-invertible symmetries in $(3+1)d$* , *SciPost Physics* **17** (Nov., 2024).
- [330] J. Kaidi, E. Nardoni, G. Zafrir, and Y. Zheng, *Symmetry tfts and anomalies of non-invertible symmetries*, 2023.
- [331] A. Antinucci, F. Benini, C. Copetti, G. Galati, and G. Rizi, *Anomalies of non-invertible self-duality symmetries: fractionalization and gauging*, 2023.
- [332] A. Apte, C. Córdova, and H. T. Lam, *Obstructions to gapped phases from noninvertible symmetries*, *Physical Review B* **108** (July, 2023).
- [333] M. Creutz and M. Okawa, *Generalized actions in zp lattice gauge theory*, *Nuclear Physics* **220** (1983) 149–166.
- [334] A. Kan, L. Funcke, S. Kühn, L. Dellantonio, J. Zhang, J. F. Haase, C. A. Muschik, and K. Jansen, *$3+1D$ θ -Term on the Lattice from the Hamiltonian Perspective*, *PoS LATTICE2021* (2022) 112, [arXiv:2111.02238].
- [335] 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), no. 4 041008, [arXiv:2012.15845].

- [336] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, *Taming the sign problem of the finite density Hubbard model via Lefschetz thimbles*, [[arXiv:1906.02726](#)].
- [337] A. M. Czajka, Z.-B. Kang, H. Ma, and F. Zhao, *Quantum simulation of chiral phase transitions*, *JHEP* **08** (2022) 209, [[arXiv:2112.03944](#)].
- [338] **NuQS** Collaboration, T. D. Cohen, H. Lamm, S. Lawrence, and Y. Yamauchi, *Quantum algorithms for transport coefficients in gauge theories*, *Phys. Rev. D* **104** (2021), no. 9 094514, [[arXiv:2104.02024](#)].
- [339] M. G. Alford, A. Schmitt, K. Rajagopal, and T. Schäfer, *Color superconductivity in dense quark matter*, *Reviews of Modern Physics* **80** (Nov., 2008) 1455–1515.
- [340] Y. Y. Atas, J. Zhang, R. Lewis, A. Jahanpour, J. F. Haase, and C. A. Muschik, *$Su(2)$ hadrons on a quantum computer via a variational approach*, *Nature Communications* **12** (Nov., 2021).
- [341] K. Sakai and S. Sasaki, *Glueball spectroscopy in lattice QCD using gradient flow*, *Phys. Rev. D* **107** (2023), no. 3 034510, [[arXiv:2211.15176](#)].
- [342] S. Sachdev, *Quantum Phase Transitions*. Cambridge University Press, 2 ed., 2011.
- [343] I. L. Gutiérrez and C. B. Mendl, *Real time evolution with neural-network quantum states*, *Quantum* **6** (Jan., 2022) 627.
- [344] W. W. Ho, S. Choi, H. Pichler, and M. D. Lukin, *Periodic orbits, entanglement, and quantum many-body scars in constrained models: Matrix product state approach*, *Physical Review Letters* **122** (Jan., 2019).
- [345] D. Bluvstein, A. Omran, H. Levine, A. Keesling, G. Semeghini, S. Ebadi, T. T. Wang, A. A. Michailidis, N. Maskara, W. W. Ho, S. Choi, M. Serbyn, M. Greiner,

- V. Vuletić, and M. D. Lukin, *Controlling quantum many-body dynamics in driven rydberg atom arrays*, *Science* **371** (Mar., 2021) 1355–1359.
- [346] K. Donatella, Z. Denis, A. L. Boité, and C. Ciuti, *Dynamics with autoregressive neural quantum states: Application to critical quench dynamics*, *Phys. Rev. A* **108** (2023), no. 2 022210, [arXiv:2209.03241].
- [347] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, *Scaling laws for neural language models*, 2020.
- [348] I. Alabdulmohsin, B. Neyshabur, and X. Zhai, *Revisiting neural scaling laws in language and vision*, 2022.