Gauge Theory

Dong-Sheng Wang

April 21, 2021

Definition. *A gauge theory is a model of quantum systems with gauge redundancy, described by a group, and there are a finite number of gauge fields which can be coupled to external matter fields following the minimal-coupling rules.*

Keywords

Maxwell equations; Gauge redundancy; Higgs mechanism; Yang-Mills theory; Lattice gauge models; Wegner models; Gauging; Standard model.

Abstract

A brief review of gauge theory in the contexts of standard model of particle physics and lattice gauge theory is given. We introduce the idea of gauge from electromagnetism, and then explain the Higgs mechanism of gauge-matter coupling for the U(1) case. We then give a brief account of general Yang-Mills theory and the difficulties. Furthermore, we use concrete examples to explain lattice gauge theory and the idea of gauging. Finally, we survey the standard model and some frontiers of gauge theory.

1 Minimal version of Theory

1.1 Opening

Although being centuries' old, electromagnetism has profound impact of our world, which cannot be matched yet by quantum theory. It is the first example of a gauge (field) theory, which, together with Relativity, opens up the study of the structure of spacetime of the universe. Gauge theory plays central roles in electrodynamics, the standard model of particle physics, and also string theory, etc.

What is a gauge? It is usually called a gauge symmetry, but many people strongly disagree and prefer to called it "gauge freedom" or "gauge redundancy". As we will see, it does not relate to conserved quantities of the system itself, instead it indicates some missing information of a more complete description of the system.

The basic framework to construct a gauge theory is the Yang-Mills theory, which could be defined for many Lie groups. To solve these systems prove to be very difficult. Besides, the proper coupling with matter fields is also sophisticated. The Yang-Mills gauge fields are gapless, and when applied to physics, in particular, particle physics, the Higgs mechanism needs to be involved for gauge bosons to obtain mass. Particle physics or high-energy physics aims to explain the behavior of a big family of particles. Right now there are still lots of mystery. The lattice gauge formalism, together with quantum simulators for them, are expected to resolve many puzzles.

1.2 Basics: U(1) gauge theory

The simplest gauge theory is the electromagnetism described by Maxwell equations, which has U(1) gauge group. Without source terms, the two dynamic equations are

$$-\dot{\vec{B}} = \nabla \times \vec{E},\ \dot{\vec{E}} = \nabla \times \vec{B}, \tag{1}$$

and the two static constraints, as Gauss's law, are

$$\nabla \cdot \vec{E} = 0,\ \nabla \cdot \vec{B} = 0. \tag{2}$$

Note we ignored constants such as speed of light c and Plank constant h for simplicity. The first interesting thing is the dynamics can be written as

Schrödinger equation $i|\dot{\psi}\rangle = H|\psi\rangle$ with

$$|\psi\rangle = \vec{E} + i\vec{B},\ H = \nabla \times . \tag{3}$$

We could use $\nabla = i\vec{p}$, and then it is obvious the Hamiltonian is a kind of spin-orbit coupling of photons, similar with Dirac equation for electrons. The solution is easy to obtain $|\psi(t)\rangle = e^{itH}|\psi(0)\rangle$, but there is a lot of freedom to fix the initial state $|\psi(0)\rangle$.

The Gauss's law now comes into play: it provides constraints at each point of the space. In all, the four equations describes the gauge fields, i.e. wavefunctions of photons, in free space except for some extrinsic charge or current defects. Now, where does the gauge freedom comes from? First, we introduce the scalar ϕ and vector potential $\vec{A}$, and usually

$$-\vec{E} = \nabla\phi + \dot{\vec{A}},\ \vec{B} = \nabla \times \vec{A}. \tag{4}$$

Let the 4-potential $A^\mu = (\phi, \vec{A})$, then the gauge freedom means that $A^\mu \to A^\mu + \partial^\mu f$ would not change the physics, for any scalar function with $\partial_\mu \partial^\mu f = 0$.

The question to ask is that: if $\vec{E}$ and $\vec{B}$ are wavefunctions, then what are the potentials? Are they metric tensor? It turns out they are not. The potential A^μ is usually treated as a connection of a U(1) vector bundle, and the fields are components of the curvature $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. The curvature is invariant under the gauge transformation, and it does not depend on any metric tensor. However, this geometric interpretation does not explain the physics clear. It is still not clear why A^μ exists, and we really do not know! Perhaps it relates to the inherent nature of the space(-time) itself. For instance, the dependence of the field $\vec{E}(x^\mu)$ on the coordinate x^μ may come from an ignorance of the quantum nature of the space; namely, let operator $\hat{x}^\mu$ act as $\hat{x}^\mu|x^\mu\rangle = x^\mu|x^\mu\rangle$, and

$$|\vec{E}\rangle|x^\mu\rangle \to |\vec{E}(x^\mu)\rangle \tag{5}$$

ignoring the space. Such a tracing out process is described by *completely positive* maps, which are common in the study of quantum open-system dynamics. The gauge freedom can now be easily understood: it is due to the ignorance of some underlying features of the space and the interaction between the space and fields. However, we would not pursue this point too far.

The U(1) gauge freedom can also be seen if we put an electron in the field. Classically, it will be acted upon by the Lorentz force $\vec{F} = q_e(\vec{E}+\vec{v}\times\vec{B})$, which however cannot see the U(1) gauge. We have to consider the wavefunction of the electron and it will pick up a phase

$$e^{ie\oint A^\mu dl}, \tag{6}$$

which is the observable effect of A^μ. The U(1) freedom will make a global phase shift which clearly does not change the physics. Such phases can be described as *Berry phase* which is of geometric nature, and it can even be topological if the field configuration is topologically nontrivial. This is revealed by the famous Aharonov-Bohm effect.

We still notice an asymmetry for the relation between fields and potentials above. It turns out this is only apparent: if there are magnetic monopoles, then there will be a full duality between the electric and the magnetic fields. One way is to introduce in a new 4-potential $C^\mu = (\varphi, \vec{C})$, and

$$-\vec{E} = \nabla\phi + \dot{\vec{A}} + \nabla\times\vec{C}, \tag{7}$$

$$\vec{B} = -\nabla\varphi - \dot{\vec{C}} + \nabla\times\vec{A}. \tag{8}$$

The Lorentz force becomes

$$\vec{F} = q_e(\vec{E}+\vec{v}\times\vec{B}) + q_m(\vec{B}-\vec{v}\times\vec{E}). \tag{9}$$

However, monopoles are still a missing piece of the electromagnetic world.

1.3 Higgs mechanism and superconductors

Now let's see how gauge fields couple to matter fields, which could be bosonic or fermionic. We know photons are massless. Let's see how photons can obtain an effective mass. This is the Higgs mechanism and can be easily illustrated by superconductors!

We are very familiar with the phenomenology of superconductors: it needs a phase transition at lower temperatures, it shows zero resistance, Meissner effect, surface current, and it is gapped. To understand the deeper mechanism, we need a U(1) gauge theory coupled to a scalar matter field. Here the gauge field is photons, while the scalar field ϕ is the wavefunction of Cooper pairs. This is a U(1) Higgs model

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + |D_\mu\phi|^2 - V \tag{10}$$

for a potential $V = m^2|\phi|^2 + \lambda|\phi|^4$, similar with the phi-4 theory. The covariant derivative is $D_\mu = \partial_\mu + iq_e A_\mu$. This model has two phases depending on m^2 and λ: a phase with definite value of $|\phi|$, and a phase with $|\phi| = 0$. The first phase is the superconducting phase, which is usually understood as a spontaneous breaking of the global U(1) symmetry.

Now we can see what the Higgs mechanism is. We can set ϕ to its fixed value, ϕ_{sc}, then the effective model for the superconducting phase is

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}q_e^2\phi_{sc}^2 A^\mu A_\mu. \tag{11}$$

This means that photons have a mass $m = q_e\phi_{sc}$, for q_e as the charge of a Cooper pair. It is usually said that the Goldstone boson ϕ is "eaten" by the gauge boson, which is the photon, and this provides the origin of mass. Now because of the photon mass, the electromagnetic field cannot propagate in the bulk of a superconductor, leading to the repulsion of magnetic flux of Meissner effect. The Higgs mechanism plays central roles to explain the origin of mass in particle physics.

1.4 Main: Yang-Mills theory

We have hinted two points of views of Maxwell theory above: symmetry and geometry. The geometric one is developed for unification theory with general relativity, which eventually leads to string theory. The symmetrical one leads to Yang-Mills theory, which forms the foundation of the standard model of particle physics.

Let's discuss the SU(n) Yang-Mills theory. Denote the generators of SU(n) as T^a, and there are $n-1$ of them. The potential A_μ will be generalized to a set of them A_μ^a, same for the tensor $F_{\mu\nu}^a$. The curvature and connection will take the form

$$F_{\mu\nu} = T^a F_{\mu\nu}^a,\ A_\mu = T^a A_\mu^a, \tag{12}$$

with summation of repeated index. Different from the abelian case, there is a commutator in the curvature form

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]. \tag{13}$$

The geometry of the SU(n) manifold defines a natural covariant derivative $D_\mu = \partial_\mu - iA_\mu$. Now geometric identity leads to the analog of Maxwell equations

$$D^\mu F_{\mu\nu} = 0,\ D_\mu G^{\mu\nu} = 0, \tag{14}$$

for the dual tensor $G^{\mu\nu} = \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$.

We notice a key difference from the U(1) case: here the covariant derivative D_μ is used instead of the ∂_μ. This prevents from breaking the equations down to the level of fields, and there are self-interaction of fields leading to nonlinear dynamics which is apparently inconsistent with Schrödinger equation! Well, we could interpret this as a nonlinear modification of it, just like Gross–Pitaevskii equation describing the BEC condensate. Also because of the nonlinearity, the Yang-Mills equations are extremely hard to solve. There are only a few of ansatz despite more that half a century's work.

How the gauge group acts? This is also a key difference from the abelian case: the curvature F is not invariant, instead it is only covariant $F_{\mu\nu} \to g(x)F_{\mu\nu}g(x)^\dagger$, for $g \in \mathrm{SU}(n)$. However, the Hamiltonian and action is invariant since they are of the form $\mathrm{tr}F^{\mu\nu}F_{\mu\nu}$, for which the effect of gauge transformation cancels. Besides, another key difference from the abelian case is there is no duality between the electric and magnetic fields anymore. However, such a duality can be enforced by hand as a type of solutions.

The Yang-Mills equations are difficult to solve in general due to its nonlinearity. Ways to find solutions are usually from analog with nonlinear optics, wave mechanics, or from ansatz. The famous Wu-Yang monopole describes a pure single monopole with $\frac{1}{r}$-form potential; as we know no monopole of any kind has been found in nature yet. Another type of solutions are instantons, which have short lifetime, and they can be used to explain tunneling and the structure of vacuum.

1.5 More: lattice gauge theory

To consider the "quanta" of gauge fields, we have to do a second quantization of the gauge fields. This could be done using path integral, or more abstract methods. However, like QED, there are lots of problems relating to divergence. So there is a different approach: instead of using continuous variables, we use discrete space and more general field variables. This is Wilson's approach of lattice gauge theory.

Putting a gauge theory on a lattice brings many benefits: it is suitable for numerical simulation, it leads to a lot of lattice models, and etc. But there may also be non-universal features depending on the lattice, which we should keep in mind. It is very useful to study non-Abelian gauge theory, especially for strong interaction of quarks.

To explain lattice gauge theory, we shall start from the simplest one: Z_2 gauge, which can be easily generalized. It applies to spatial dimension $D \geq 2$. Consider the 3D cubic lattice, and put a spin-1/2 on each edge. Now define the Wegner model

$$H = J_1 \sum_r X_r + J_4 \sum_\square Z_\square \tag{15}$$

with Pauli X on each spin and 4-body product of Pauli Z on each unit square. The model looks simple, but it is not commuting. First, we can tell that when J_1 terms dominate, the system is a paramagnet (PM); when J_4 terms dominate, it is another phase. The system has a gauge symmetry: $\bar{X}$ with 6 Pauli X around a site, for all local sites.

The construction can be generalized. For U(1) case, we can put a phase $e^{i\theta}$ on each link, and for unitary group, we can put a group element g on each link. What is subtle for the non-Abelian case is that the gauge symmetry, $Q(r)$, do not commute with each other, although they commute with the Hamiltonian.

The gauge-invariant observable are Wilson loop operators, W. It turns out there are 2 phases: one is a confined phase with string tensions for Wilson loops, and one is deconfined with no string tension. The confined phase is the PM, and string tensions (i.e., extensive energy cost) come from each local term X_r. The thermal average $\langle W \rangle \sim e^{-A\beta}$, as an "area law", while for the deconfined phase $\langle W \rangle \sim e^{-L\beta}$, as a "perimeter law". Here A is the area enclosed by a loop and L is the length of the loop. This phase transition is similar with that for 2D classical Ising model which has global Z_2 symmetry. We know that it has a thermal phase transition between the PM and FM. The driving force for the disordered PM phase is the "proliferation" of domains, while in the FM phase the energy of a domain scales with the length of its boundary.

Furthermore, spatial dimensionality also plays central roles here, just like that for systems with global symmetry. The 2D version of Wegner model on a square lattice does not have a thermal phase transition. This can be seen if it is mapped to a 1D Ising model. What is interesting is that the model leads to topological order. If we include the gauge terms in it, the model looks like

$$H = J_1 \sum_r X_r + J_z \sum_\square Z_\square + J_x \sum_+ X_+. \tag{16}$$

The phase for vanishing J_1 is the *toric code*, which has Z_2 topological order. On a torus, the ground state degeneracy is 4. This can be seen from the algebra of four Wilson loop operators, X_l^x, X_l^y, Z_l^x, Z_l^y, for x, y as the two directions of the torus. There are 2 species of excitations each created by a local X or a local Z. Excitations appear in pair and are deconfined so that they can be separated arbitrarily. The 3D case is the 3D toric code, for which one type of excitation becomes string or loop-like, just like magnetic flux. It turns out the Z_2 topological order can also describe superconductors, which we would not explain in details here.

2 Advanced topics: gauging

Now we ask the following question: how to convert a symmetry of being global and local? This is a process called "gauging" if we convert from global to local symmetry. The origin of it is still the U(1) case: replace the derivative ∂_μ by D_μ is a gauging. In general, gauging follows a *minimal-coupling* rules: use covariant derivative, a proper metric tensor, and avoid curvature terms. It plays important roles in string theory. Here, instead we survey its usage in many-body physics, as lattice gauge theory has very close connection with it.

We have discussed Ising model and toric code above. So the question is: is there any relation between them, i.e., the global and local Z_2 symmetry? It turns out there is. Consider the 2D square lattice Ising model

$$H = J_x \sum_r X_r + J_z \sum_r Z_r Z_{r+1} \tag{17}$$

for onsite kinetic terms X_r and 2-body potential terms $Z_r Z_{r+1}$. The global symmetry is $\prod_r X_r$. It has the symmetry-breaking FM phase and non-breaking disordered PM phase. Now to gauge it, we need to add gauge variables, and they shall be on each link: here they are also spins. There should be 2 types of excitations just like Maxwell case: electric and magnetic fields. The metric is flat so no worry about covariant derivative. The symmetry becomes local, but shall it involves both matter and gauge fields? It seems it is to ensure a coupling between them. Denote gauge spins as τ^i. It is not hard to find that the term $Z_r Z_{r+1}$ can be replaced by $Z_r \tau^x Z_{r+1}$, the symmetry term is $X_r \prod_{\langle r\rangle} \tau^z$, for $\langle r\rangle$ as the four edge neighbors of r. The term X_r shall remain since it commutes with the symmetry. We also need a

Gauss law 4-body term $\prod_\square \tau^x$ for each unit cell. With these terms, we arrive at the gauged model

$$\begin{aligned} H &= J_x \sum_r X_r + J_z \sum_r Z_r \tau^x Z_{r+1} + \\ & J_s \sum_r X_r \prod_{\langle r \rangle} \tau^z + J_g \sum_\square \prod_\square \tau^x. \end{aligned} \tag{18}$$

Now when the matter field is in the PM phase, X_r has a classical value so we can delete it from the symmetry term, and J_z terms vanish. The matter field and gauge field actually decouples. So we just obtain the toric code! From the FM phase instead, the J_x terms vanish, and the J_g terms become redundant since the J_z terms will reproduce them. So we only have the J_z and J_s terms, and this model now defines the so-called cluster-state phase. Recall that a cluster state is obtained by replacing local X operators by $X \prod_i Z_i$ for some patterns of $\prod_i Z_i$. For a 1D cluster state, these so-called "stabilizers" are $Z_{r-1} X_r Z_{r+1}$. The gauged model from the PM phase is a system of coupled 1D cluster wires. This phase is often called a Higgs phase with massive gauge fields. This can be seen by noting that the J_z terms are nothing but the mass terms of gauge fields.

There are also interesting thing about symmetry or Wilson loops: on a torus, the toric code breaks loop symmetry X_l^x and X_l^y, resulting the 4-fold degeneracy; however, the cluster phase preserves these symmetry, hence the ground state is unique. It is said that the toric code has topological order while the cluster state has 1-form symmetric (symmetry-protected) order.

One may wonder: how about 1D case? Can we gauge 1D Ising model? Sure, we did it and we find the PM phase is gauged to a FM phase of the gauge field, and the matter field decouples, and the FM phase is gauged to the 1D cluster phase, which is gapped.

So, what we learned from gauging? Probably it is not about the symmetry itself, instead, it is the emerging symmetry (Wilson loops) from it. They are a so-called high-form symmetry, the understanding of which is still not complete yet.

3 Relevant theory: standard model

We turn to the major impact of gauge theory: the standard model of particle physics which is based on Yang-Mills theory and Higgs mechanism.

Currently, it contains the electroweak unification theory, and QCD for the strong interaction. Here we survey briefly how the electroweak theory works.

The electroweak theory starts from gauge theory with SU(2) $\times$ U(1), and via Higgs mechanism breaks down to U(1). Initially, there are 3 massless gauge fields W_μ^a (a=1,2,3; μ=x,y,z) from SU(2) and 1 massless field B_μ from U(1). They will give rise to 1 new massless gauge field, which is photon, from the combination of U(1) and the U(1) subgroup of SU(2), and 3 massive gauge fields for the weak interaction. In addition, there has to be a Higgs field, H, which will generate mass.

The Higgs field has 2 components since it is acted upon by the gauge group. It will obtain a non-zero value on ground state (vacuum), and this symmetry breaking will reduce the symmetry to U(1). This is just like superconductors, with the Cooper pair field operator ϕ as the Higgs field. The model takes this form

$$\mathcal{L} = -W^a_{\mu\nu}W^{a\mu\nu} - B_{\mu\nu}B^{\mu\nu} + D_\mu H^\dagger D^\mu H - V \tag{19}$$

for a Higgs potential V. The first 3 terms are taken as free parts, and the V term will replace H by its vacuum value, H_0. Usually, it is taken as $H_0 = [0, h]^t$ for a real number h, which can be achieved by tuning the gauge redundancy. The consequence is profound. First, H_0 is invariant under diagonal phase operators in SU(2), i.e., U(1) rotations, and this must be from a combination of the original U(1) and the U(1) subgroup of SU(2). This predicts the existence of photon. The term with $D_\mu H$ give mass to other gauge fields, with

$$D_\mu H_\alpha = \partial_\mu H_\alpha + ig_2\sigma^a_{\alpha\beta}W^a_\mu H_\beta + ig_1 B_\mu H_\alpha \tag{20}$$

as a covariant derivative. Here σ^a are Pauli matrices, g_1, g_2 are coupling parameters. The Weinberg mixing angle, $\theta_W = \arctan\frac{g_1}{g_2}$, has been experimentally confirmed to high precision with $\sin^2\theta \approx 0.23$. As this is too technical, we would not explain the details.

The SU(3) Yang-Mills shows different features. Different from SU(2), the fundamental irrep 3 is not real, and it has the conjugate $\bar{3}$. This leads to 3 quarks and 3 anti-quarks. As the adjoint irrep 8, there are 8 gluons. The two central features of QCD are *asymptotic freedom* and *quark confinement*. The asymptotic freedom says at high energies quarks interact weakly, and the quark confinement says at low energies quarks interact strongly. These phenomena are due to gluons which interact with themselves and quarks.

Finally, we survey the elementary particles in the standard model. As messenger bosons, there are photons γ for electromagnetism, Z and W for weak interaction, and the Higgs boson H, and gluons for strong interaction. As matter fermions, there are electrons e, muon μ, tau τ which are charged, and electron neutrino ν_e, muon neutrino ν_μ, and tau neutrino ν_τ which are neutral. The standard model is well established, but there are still lots of puzzles, see below.

4 Frontiers

In this brief survey, we did not cover a lot of important subjects, such as the second quantized form of gauge theory, duality, path integral, renormalization, and Feynman diagram. Many problems of particle physics are due to the difficulty to compute Feynman diagrams, which mainly works in the spirit of perturbation theory. A more powerful approach especially for strong couplings is highly expected.

There are lots of open problems of the standard model. We refer to Wikipedia for a list of them. These include the explanation of 3 generation of matter, matter-antimatter asymmetry, consistency with general relativity, etc. At the heart of the standard model, to prove the Yang–Mills theory is gapped for any compact Lie group is still an open problem. This is of great mathematical interest. This relates to a *proof* of the quark confinement and asymptotic freedom. Supersymmetry, i.e., duality between fermion and boson, is also a frontier relating to string theory.

Based on the framework of lattice gauge theory, quantum simulation is a promising endeavor. Small scale quantum simulators, with about 100 qubits, are within reach in a few years. These simulators can solve some problems faster than the best supercomputers in the world, although their precision may be limited, but good enough. Combining with classical algorithms, their performance can be enhanced. They can be used to study phase transition, dynamics, thermalization, and simulate particle scattering, and many others. This approach is expected one day to substitute high-energy large-scale colliders for some tasks.

5 History, people, and story

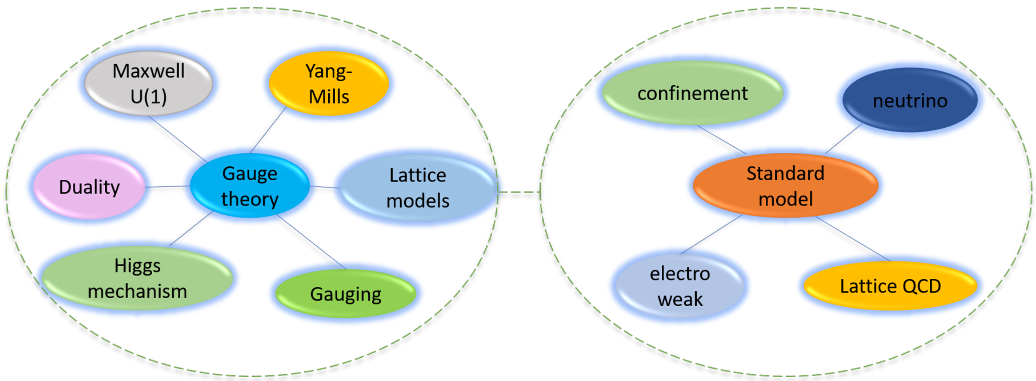
A story we had to mention is that Pauli, a geek, also formulated Yang-Mills independently but did not publish his result, since he found that the gauge bosons are massless. We also remember that he discovered the notion of spin, but did not publish the result since he found that a spinning particle might violate special relativity! What we learn from his story? Probably he is too critical about others and also himself; but we know these stories after all because he discussed with lots of people! It is important to let colleague know your results, whether with a publishable paper or not.

For the establishment of standard model, many conceptual breakthrough was made by individuals. Meanwhile, experiments are often large collaborations. This is because high-energy experiments are sophisticated and also generate huge amount of data. While deep physical insight is often grasped by a genius first, such as G. 't Hooft and E. Witten, then accepted by a small group of experts, and then gradually becomes popular in the research community. This is partly due to the toughness of path integral, renormalization, etc, but also the vagueness of the underlying physics, which is still unclear at present.

References

- J. Kogut, Lattice gauge theory and spin systems, Rev. Mod. Phys. 51, 1979.
- A. Actor, Classical solutions of SU(2) Yang—Mills theories, Rev. Mod. Phys. 51, 1979.
- A. Jaffe and E. Witten, Quantum Yang-Mills theory, 2000.
- Wikipedia, Physics beyond the Standard Model.

Concept map



Maxwell U(1)
Yang-Mills
Duality
Gauge theory
Lattice models
Higgs mechanism
Gauging
confinement
neutrino
Standard model
electro weak
Lattice QCD