Topics in Mathematical Physics: A Journey Through Differential Equations

Yisong Yang
Polytechnic Institute of New York University

Nonlinear equations play important roles in fundamental science and present great challenges and opportunities to research mathematicians. In theoretical physics, many profound concepts, predictions, and advances were pioneered through mathematical insights gained from the study of the equations governing basic physical laws. Why do some materials demonstrate zero electric resistance when cooled (superconductivity)? How do the two strands in an entangled DNA double helix become separated when heated (DNA denaturation)? Why does the universe have a finite past (big bang cosmology)? Why are all the electric charges integer multiples of a minimal unit (charge quantization)? Why is that the basic constituents of matter known as quarks can never be found in isolation (quark confinement)? These are some of the exemplary situations where mathematical investigation is essential. The purpose of this series of lectures, to be given at the Applied Mathematics Graduate Summer School of Beijing University in July 2011, is to provide a vista-type overview of a broad range of nonlinear equations arising in classical field theory. Emphasis will be given to the mathematical structure and physical descriptions of various basic problems and to the appreciation of the power of functional analysis, although close attention will also be given to the links with other areas of mathematics, including geometry, algebra, and topology.

Here is a table of contents of the notes.

Con n

on n y

In this section, we begin our study with the Hamiltonian or Lagrangian formalism of classical mechanics, which is the conceptual foundation of all later developments.

Mo on o p c

Consider the motion of a point particle of mass m and coordinates $(q^{\mathbf{i}}) = q$ in a potential field $V(q,t)$ described by Newtonian mechanics. The equations of motion are

$$m\ddot{q^{\mathbf{i}}} = -\frac{\partial V}{\partial q^{\mathbf{i}}}, \quad i = 1, 2, \cdots, n, \tag{1.1}$$

where $\cdot$ denotes time derivative. Since

$$-\nabla V = -\left(\frac{\partial V}{\partial q^{\mathbf{i}}}\right) \tag{1.2}$$

defines the direction along which the potential energy V decreases most rapidly, the equation (1.1) says that the particle is accelerated along the direction of the flow of the steepest descent of V.

With the *Lagrangian function*

$$L(q, \dot{q}, t) = \frac{1}{2}m\sum_{\mathbf{i}=1}^{\mathbf{n}}(\dot{q^{\mathbf{i}}})^2 - V(q,t), \tag{1.3}$$

which is simply the difference of the kinetic and potential energies, (1.1) are simply the *Euler–Lagrange equations* of the action

$$\int_{\mathbf{t}_1}^{\mathbf{t}_2} L(q(t), \dot{q}(t), t)\,\mathrm{d}t \tag{1.4}$$

over the admissible space of trajectories $\{q(t) \,|\, t_1 < t < t_2\}$ starting and terminating at fixed points at $t = t_1$ and $t = t_2$, respectively.

The *Hamiltonian* function or *energy* at any time t is the sum of kinetic and potential energies given by

$$H = \frac{1}{2}m\sum_{\mathbf{i}=1}^{\mathbf{n}}(\dot{q^{\mathbf{i}}})^2 + V(q,t) = m\sum_{\mathbf{i}=1}^{\mathbf{n}}(\dot{q^{\mathbf{i}}})^2 - L. \tag{1.5}$$

Introduce the *momentum vector* $p = (p_{\mathbf{i}})$,

$$p_{\mathbf{i}} = m\dot{q^{\mathbf{i}}} = \frac{\partial L}{\partial \dot{q^{\mathbf{i}}}}, \quad i = 1, 2, \cdots, n. \tag{1.6}$$

Then, in view of (1.5), H is defined by

$$H(q,p,t) = \sum_{\mathbf{i}=1}^{\mathbf{n}} p_{\mathbf{i}} \dot{q^{\mathbf{i}}} - L(q,\dot{q},t) \tag{1.7}$$

and the equations of motion, (1.1), are a *Hamiltonian system*,

$$\dot{q^{\mathbf{i}}} = \frac{\partial H}{\partial p_{\mathbf{i}}}, \quad \dot{p_{\mathbf{i}}} = -\frac{\partial H}{\partial q^{\mathbf{i}}}, \quad i = 1, 2, \cdots, n. \tag{1.8}$$

The above formulations may be extended to the case when L is an arbitrary function of $q, \dot{q}$, and t. The equations of motion are the Euler–Lagrange equations of (1.4),

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q^{\mathbf{i}}}}\right) = \frac{\partial L}{\partial q^{\mathbf{i}}}, \quad i = 1, 2, \cdots, n. \tag{1.9}$$

In order to make a similar Hamiltonian formulation, we are motivated from (1.6) to introduce the *generalized momentum vector* $p = (p_{\mathbf{i}})$ by

$$p_{\mathbf{i}} = \frac{\partial L}{\partial \dot{q^{\mathbf{i}}}}, \quad i = 1, 2, \cdots, n. \tag{1.10}$$

We still use (1.7) to define the corresponding Hamiltonian function H. A direct calculation shows that (1.9) are now equivalent to the Hamiltonian system (1.8).

We note that an important property of a Hamiltonian function is that it is independent of the variables $\dot{q^{\mathbf{i}}}$ $(i = 1, 2, \cdots, n)$. In fact, from the definition of the generalized momentum vector given by (1.10), we have

$$\frac{\partial H}{\partial \dot{q^{\mathbf{i}}}} = p_{\mathbf{i}} - \frac{\partial L}{\partial \dot{q^{\mathbf{i}}}} = 0, \quad i = 1, 2, \cdots, n. \tag{1.11}$$

This fact justifies our notation of $H(p,q,t)$ in (1.7) instead of $H(p,q,\dot{q},t)$.

Let F be an arbitrary function depending on $p_{\mathbf{i}}, q^{\mathbf{i}}$ $(i = 1, 2, \cdots, n)$ and time t. We see that F varies its value along a trajectory of the equations of motion, (1.8), according to

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}t} &= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial q^{\mathbf{i}}}\dot{q^{\mathbf{i}}} + \frac{\partial F}{\partial p_{\mathbf{i}}}\dot{p_{\mathbf{i}}} \\ &= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial q^{\mathbf{i}}}\frac{\partial H}{\partial p_{\mathbf{i}}} - \frac{\partial F}{\partial p_{\mathbf{i}}}\frac{\partial H}{\partial q^{\mathbf{i}}}, \end{aligned} \tag{1.12}$$

where and in the sequel we observe the *summation convention* over repeated indices, although occasionally we also spell out the summation explicitly. Thus, we are motivated to use the *Poisson bracket* $\{\cdot,\cdot\}$,

$$\{f,g\} = \frac{\partial f}{\partial q^{\mathbf{i}}}\frac{\partial g}{\partial p_{\mathbf{i}}} - \frac{\partial f}{\partial p_{\mathbf{i}}}\frac{\partial g}{\partial q^{\mathbf{i}}}, \tag{1.13}$$

to rewrite the rate of change of F with respect to time t as

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + \{F,H\}. \tag{1.14}$$

In particular, when the Hamiltonian H does not depend on time t explicitly, $H = H(p,q)$, then (1.14) implies that

$$\frac{\mathrm{d}H}{\mathrm{d}t} = 0, \tag{1.15}$$

which gives the fact that energy is conserved and the mechanical system is thus called *conservative*.

It will be useful to '*complexify*' our formulation of classical mechanics. We introduce the complex variables

$$u_{\mathbf{i}} = \frac{1}{\sqrt{2}}(q^{\mathbf{i}} + \mathrm{i}p_{\mathbf{i}}), \quad i = 1, 2, \cdots, n, \quad \mathrm{i} = \sqrt{-1}. \tag{1.16}$$

Then the Hamiltonian function H depends only on $u = (u_{\mathbf{i}})$ and its complex conjugate $\overline{u} = (\overline{u}_{\mathbf{i}})$,

$$H = H(u, \overline{u}, t). \tag{1.17}$$

Hence, in terms of the differential operators,

$$\frac{\partial}{\partial u_{\mathbf{i}}} = \frac{\sqrt{2}}{2}\left(\frac{\partial}{\partial q^{\mathbf{i}}} - \mathrm{i}\frac{\partial}{\partial p_{\mathbf{i}}}\right), \quad \frac{\partial}{\partial \overline{u}_{\mathbf{i}}} = \frac{\sqrt{2}}{2}\left(\frac{\partial}{\partial q^{\mathbf{i}}} + \mathrm{i}\frac{\partial}{\partial p_{\mathbf{i}}}\right), \tag{1.18}$$

the Hamiltonian system (1.8) takes the concise form

$$\mathrm{i}\dot{u}_{\mathbf{i}} = \frac{\partial H}{\partial \overline{u}_{\mathbf{i}}}, \quad i = 1, 2, \cdots, n. \tag{1.19}$$

Again, let F be a function depending on $u, \overline{u}$, and t. Then (1.19) gives us

$$\begin{aligned} \frac{\mathrm{d}F}{\mathrm{d}t} &= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial u_{\mathbf{i}}}\dot{u}_{\mathbf{i}} + \frac{\partial F}{\partial \overline{u}_{\mathbf{i}}}\dot{\overline{u}}_{\mathbf{i}} \\ &= \frac{\partial F}{\partial t} - \mathrm{i}\frac{\partial F}{\partial u_{\mathbf{i}}}\frac{\partial H}{\partial \overline{u}_{\mathbf{i}}} + \mathrm{i}\frac{\partial F}{\partial \overline{u}_{\mathbf{i}}}\frac{\partial H}{\partial u_{\mathbf{i}}}. \end{aligned}$$

With the notation

$$\{f, g\} = \frac{\partial f}{\partial u_{\mathbf{i}}}\frac{\partial g}{\partial \overline{u}_{\mathbf{i}}} - \frac{\partial f}{\partial \overline{u}_{\mathbf{i}}}\frac{\partial g}{\partial u_{\mathbf{i}}} \tag{1.20}$$

for the Poisson bracket, we have

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + \frac{1}{\mathrm{i}}\{F, H\}. \tag{1.21}$$

In particular, the complexified Hamiltonian system (1.19) becomes

$$\dot{u}_{\mathbf{i}} = \frac{1}{\mathrm{i}}\{u_{\mathbf{i}}, H\}, \quad i = 1, 2, \cdots, n, \tag{1.22}$$

which is in a close resemblance of the *Schrödinger equation*, in the *Heisenberg representation*, in quantum mechanics, which will be detailed later.

In the rest of this section, we present a few important examples.

[illegible]

It will be instructive to start from a general discussion. Let $\mathbf{v}$ be the velocity field of a fluid. Then

$$\boldsymbol{\omega} = \nabla \times \mathbf{v} \tag{1.23}$$

describes the tendency that the fluid swirls itself, which is commonly called the *vorticity* field. Imagine that we form a vortex tube by vortex lines, similar to streamlines induced from the

velocity field. Then cut off two cross-sections, say S_1 and S_2, to form a cylindrically shaped finite vortex tube, say T. Then the divergence theorem says that

$$\int_{\mathbf{T}} \nabla \cdot \, \mathrm{d}x = 0, \tag{1.24}$$

which then implies

$$\int_{\mathbf{S}_1} \cdot \mathrm{d}\mathbf{S} = \int_{\mathbf{S}_2} \cdot \mathrm{d}\mathbf{S}, \tag{1.25}$$

where the orientations on S_1 and S_2 are chosen in an obviously compatible way. In other words, the flux of vortex lines across the vortex tube is constant along the tube. This common flux is called the strength or tension of the vortex tube. On the other hand, the circulation of a vector field $\mathbf{v}$ along a closed curve C is defined to be

$$\oint_{\mathbf{C}} \mathbf{v} \cdot \mathrm{d}\mathbf{s}. \tag{1.26}$$

Thus, if C is the boundary curve of a cross-section of a vortex tube of the fluid with velocity field $\mathbf{v}$, the above discussion indicates that the strength of the vortex tube may be expressed as the circulation of the fluid around the vortex tube.

Now consider the motion confined in a horizontal plane so that $\mathbf{v} = (v_1, v_2, 0)$. Then the vorticity field is always along the vertical direction so that we may express it as a scalar field given by

$$w = \partial_1 v_1 - \partial_2 v_1. \tag{1.27}$$

Of course, vortex lines are all vertical to the plane.

A Kirchhoff vortex centered at the origin of $\mathbb{R}^2$ is an idealized situation where the velocity field is centrally generated from a specified scalar potential function according to the relations

$$v_{\mathbf{j}} = \epsilon_{\mathbf{jk}} \partial_{\mathbf{k}} U, \quad j, k = 1, 2, \quad U(x) = -\frac{\gamma}{2\pi} \ln |x|, \quad x \in \mathbb{R}^2, \tag{1.28}$$

where $\epsilon_{\mathbf{jk}}$ is the standard skew-symmetric Kronecker symbol with $\epsilon_{12} = 1$ and $\gamma > 0$ is a parameter. It is clear that the flow-lines are concentric circles around the origin. Let $C_{\mathbf{r}}$ be any one of such circles of radius $r > 0$. Then, we have

$$\oint_{\mathbf{C}_r} \mathbf{v} \cdot \mathrm{d}\mathbf{s} = \gamma, \tag{1.29}$$

which says the circulation along any flow line or the strength of any vortex tube containing the center of the vortex takes the constant value γ. In other words, the quantity γ gives the circulation or strength of the vortex centered at the origin. Furthermore, we can also compute the vorticity field directly,

$$w = -\Delta U = -(\partial_1^2 + \partial_2^2) U = \gamma \delta(x), \tag{1.30}$$

which clearly reveals a *point vortex* at the origin given by the Dirac function and justifies again the quantity γ as the *strength of the point vortex*.

Following the model of Kirchhoff, the dynamical interaction of N point vortices located at $\mathbf{i} = \mathbf{i}(t) \in \mathbb{R}^2$ of respective strengths $\gamma_{\mathbf{i}}$'s $(i = 1, \cdots, N)$ at time t is governed by the interaction potential

$$U(\ 1, \cdots, \ \mathbf{N}) = -\frac{1}{2\pi} \sum_{1 \le \mathbf{i} < \mathbf{i}' \le \mathbf{N}} \gamma_{\mathbf{i}} \gamma_{\mathbf{i}'} \ln |\ \mathbf{i} - \ \mathbf{i}'|, \tag{1.31}$$

and the equations of motion

$$\gamma_{\mathbf{i}}\dot{\ }_{\mathbf{i}} = J\nabla_{\mathbf{x}_i}U, \quad i = 1, \cdots, N, \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{1.32}$$

Rewriting in the coordinate form with $= (x^{\mathbf{j}}) \in \mathbb{R}^2$ $(j = 1, 2)$ and setting

$$x^1_{\mathbf{i}} = q_{\mathbf{i}}, \quad \gamma_{\mathbf{i}} x^2_{\mathbf{i}} = p_{\mathbf{i}}, \quad i = 1, \cdots, N, \tag{1.33}$$

we arrive at

$$\dot{q}_{\mathbf{i}} = \frac{\partial U}{\partial p_{\mathbf{i}}}, \quad \dot{p}_{\mathbf{i}} = -\frac{\partial U}{\partial q_{\mathbf{i}}}, \quad i = 1, \cdots, N, \tag{1.34}$$

which is a Hamiltonian system. Note that the 'momenta' $p_{\mathbf{i}}$'s appear in the Hamiltonian function U in a 'non-standard' way. The reason for this odd appearance is that the $p_{\mathbf{i}}$'s actually do not have a mechanical meaning as momenta and are artificially identified as the momentum variables. However, it is interesting to note how the circulations $\gamma_{\mathbf{i}}$'s are being absorbed into these momenta so consistently.

N, ody p o

The *N-body problem* is motivated from *celestial mechanics* which treats celestial bodies as point particles interacting through Newton's law of gravitation.

Consider N point particles, each of mass $m_{\mathbf{i}}$, located at $_{\mathbf{i}} \in \mathbb{R}^3$, $i = 1, \cdots, N$. Then Newton's law gives us the equations of motion

$$m_{\mathbf{i}}\ddot{}_{\mathbf{i}} = -G\sum_{\mathbf{i}' \neq \mathbf{i}}^{\mathbf{N}} \frac{m_{\mathbf{i}} m_{\mathbf{i}'}(\ _{\mathbf{i}} - \ _{\mathbf{i}'})}{|\ _{\mathbf{i}} - \ _{\mathbf{i}'}|^3} = -\nabla_{\mathbf{x}_i} U, \quad i = 1, \cdots, N, \tag{1.35}$$

where $G > 0$ is the universal gravitational constant and

$$U(\ _1, \cdots, \ _{\mathbf{n}}) = -G \sum_{1 \leq \mathbf{i} < \mathbf{i}' \leq \mathbf{N}}^{\mathbf{N}} \frac{m_{\mathbf{i}} m_{\mathbf{i}'}}{|\ _{\mathbf{i}} - \ _{\mathbf{i}'}|} \tag{1.36}$$

is the gravitational potential. Thus, the motion is governed by the principle that the particles are accelerated along the directions of the fastest descendents that would lower the potential energy.

In order to recast the system into a Hamiltonian system, we relabel the coordinate variables and masses according to

$$\begin{aligned} (\ _1, \cdots, \ _{\mathbf{N}}) &\mapsto (q_1, q_2, \cdots, q_{3\mathbf{N}}), \\ (m_1, m_1, m_1, \cdots, m_{\mathbf{N}}, m_{\mathbf{N}}, m_{\mathbf{N}}) &\mapsto (m_1, m_2, \cdots, m_{3\mathbf{N}}), \end{aligned}$$

which allows us to introduce the momentum variables

$$p_{\mathbf{i}} = m_{\mathbf{i}} \dot{q}_{\mathbf{i}}, \quad i = 1, 2, \cdots, 3N. \tag{1.37}$$

It can be seen that the equations of motion now take the form of a Hamiltonian system

$$\dot{q}_{\mathbf{i}} = \frac{\partial H}{\partial p_{\mathbf{i}}}, \quad \dot{p}_{\mathbf{i}} = -\frac{\partial H}{\partial q_{\mathbf{i}}}, \quad i = 1, 2, \cdots, 3N, \tag{1.38}$$

where the Hamiltonian function H is defined by

$$H = \sum_{\mathbf{i}=1}^{3\mathbf{N}} \frac{p_{\mathbf{i}}^2}{2m_{\mathbf{i}}} + U = \sum_{\mathbf{i}=1}^{3\mathbf{N}} \frac{1}{2}\dot{q}_{\mathbf{i}} p_{\mathbf{i}} + U. \tag{1.39}$$

Since the system (1.38) consists of $6N$ first order equations, its complete integration (solution) requires obtaining $6N$ independent integrals. By exploring mechanical properties of the system, we have the following immediate integrals (or conserved quantities), namely, the center of masses $_0$ determined by

$$_0 \sum_{\mathbf{i}=1}^{\mathbf{N}} m_{\mathbf{i}} = \sum_{\mathbf{i}=1}^{\mathbf{N}} m_{\mathbf{i}} \ _{\mathbf{i}}; \tag{1.40}$$

the total (linear) momentum $\mathbf{L}_0$ given by

$$\mathbf{L}_0 = \sum_{\mathbf{i}=1}^{\mathbf{N}} m_{\mathbf{i}} \dot{} _{\mathbf{i}} = \sum_{\mathbf{i}=1}^{\mathbf{N}} \mathbf{p_i}; \tag{1.41}$$

the total angular momentum $_0$ expressed as

$$_0 = \sum_{\mathbf{i}=1}^{\mathbf{N}} m_{\mathbf{i}} \ _{\mathbf{i}} \times \dot{} _{\mathbf{i}} = \sum_{\mathbf{i}=1}^{\mathbf{N}} \ _{\mathbf{i}} \times \mathbf{p_i}; \tag{1.42}$$

and the conserved total energy H stated in (1.39). Thus, we have a total of 10 obvious first integrals. This number count indicates that the N-body problem quickly becomes highly non-trivial when N increases. Indeed, the $N = 3$ is already notoriously hard and is known as the 3-body problem. In general, it is believed that the N-body problem is not integrable. The best understood situation is the 2-body problem [91, 108]. Important applications of the 2-body problem include derivation of Kepler's laws. When the masses are replaced by charges so that Newton's gravitation is placed by Coulomb's law of electrostatics, we can study the N-body problem of charged particles. The quantum mechanical version of this is called the quantum N-body problem [64] which has important applications in theory of atoms and molecules and is of contemporary research interest [75].

on n nc on nd odyn c

To study the thermodynamical properties of a Hamiltonian system, we need the notation of *partition function* in statistical mechanics. For simplicity, consider a closed system which can occupy a countable set of states indexed by $s \in \mathbb{N}$ (the set of non-negative integers) and of distinct energies $E_{\mathbf{s}}$ ($s \in \mathbb{N}$). Then the partition function of the system is defined by

$$Z = \sum_{\mathbf{s}=0}^{\infty} \mathrm{e}^{-\boldsymbol{\beta}\mathbf{E}_s}, \tag{1.43}$$

where $\beta = 1/k_{\mathrm{B}}T$ is called the *inverse temperature* for which k_{B} is the *Boltzmann constant* and T is the *absolute temperature*. Thus, in order that (1.43) makes sense, the sequence $\{E_{\mathbf{s}}\}$ cannot have a limiting point and has to diverge sufficiently rapidly as $s \to \infty$. Assuming all conditions are valid so that $Z < \infty$, we see that

$$P_{\mathbf{s}} = \frac{1}{Z}\mathrm{e}^{-\boldsymbol{\beta}\mathbf{E}_s}, \quad s \in \mathbb{N}, \tag{1.44}$$

may naturally be interpreted as the probability that the system occupies the state s so that its energy is $E = E_{\mathbf{s}}$ ($s \in \mathbb{N}$). The quantity $\mathrm{e}^{-\boldsymbol{\beta}\mathbf{E}_s}$ is also called the *Boltzmann factor*. With such an understanding, the partition function Z may be regarded as the normalization factor of the sequence of the Boltzmann factors which give rise to the probability distribution of the random energy, E, of the system.

We now illustrate how to use Z to obtain statistical information of the system. First, the expected value of the energy (the thermodynamic value of the energy) is

$$\langle E\rangle = \sum_{\mathbf{s}=0}^{\infty} E_{\mathbf{s}} P_{\mathbf{s}} = \frac{1}{Z}\sum_{\mathbf{s}=0}^{\infty} E_{\mathbf{s}}\mathrm{e}^{-\boldsymbol{\beta}\mathbf{E}_s} = -\frac{\partial \ln Z}{\partial \beta} = k_{\mathrm{B}}T^2\frac{\partial \ln Z}{\partial T}, \tag{1.45}$$

which is also commonly denoted as U. Next, the variance is

$$\sigma_{\mathbf{E}}^2 = \langle (E - \langle E\rangle)^2\rangle = \frac{\partial^2 \ln Z}{\partial \beta^2}, \tag{1.46}$$

which gives rise to the heat capacity

$$C_{\mathbf{v}} = \frac{\partial\langle E\rangle}{\partial T} = \frac{1}{k_{\mathrm{B}}T^2}\sigma_{\mathbf{E}}^2 = \frac{1}{k_{\mathrm{B}}T^2}\frac{\partial^2 \ln Z}{\partial \beta^2}. \tag{1.47}$$

Besides, the *entropy* of the system, S, which measures the *disorder* or uncertainty of the system, is given by

$$S = -k_{\mathrm{B}}\sum_{\mathbf{s}=0}^{\infty} P_{\mathbf{s}}\ln P_{\mathbf{s}} = k_{\mathrm{B}}(\ln Z + \beta\langle E\rangle) = \frac{\partial}{\partial T}(k_{\mathrm{B}}T\ln Z) = -\frac{\partial A}{\partial T}, \tag{1.48}$$

where

$$A = -k_{\mathrm{B}}T\ln Z = U - TS \quad (U = \langle E\rangle) \tag{1.49}$$

is the *Helmholtz free energy*. These examples have shown the usefulness of the partition function.

For a classical Hamiltonian system with generalized coordinates $q = (q_1, \cdots, q_{\mathbf{n}})$ and momenta $p = (p_1, \cdots, p_{\mathbf{n}})$, governed by the Hamiltonian function $H(q, p)$, the partition function is expressed by

$$Z = \int \mathrm{e}^{-\boldsymbol{\beta}\mathbf{H}(\mathbf{q},\mathbf{p})}\,\mathrm{d}q\mathrm{d}p, \tag{1.50}$$

where q, p take over the role of the state index s and integral replaces the summation in our earlier discussion. Therefore, a similar collection of knowledge can be gathered as before. For example, if $F(q, p)$ is a mechanical quantity of interest, then its expected or thermodynamic value is given by

$$\langle F\rangle = \int F(q,p)\mathrm{e}^{-\boldsymbol{\beta}\mathbf{H}(\mathbf{q},\mathbf{p})}\,\mathrm{d}q\mathrm{d}p. \tag{1.51}$$

A fairly thorough treatment of statistical mechanics may be found in [61, 71]. In the next subsection, we will apply the ideas here to study a thermodynamic property of DNA.

1.5 Motion of DNA

DNA, the short name for *deoxyribonucleic acid*, is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms. Chemically, a DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups. These two strands run in parallel and form a double helix. Attached to each sugar is one of four types of nucleotide molecules, also called *bases*, named by letters A (adenine), C (cytosine), G (guanine), T (thymine), so that only A and T, C and G, from opposite strands may bind to form pairs.[1] During the last three decades, biologists and physicists have carried out heated research on dynamics of DNA, using mathematical modeling and computer simulation, and obtained profound knowledge about DNA and its function.

Mathematical modeling of dynamics of DNA was initiated in 1980 by Englander *et al* [39] who presented a discrete sine–Gordon soliton interpretation of the DNA of n pairs of bases and used the solitary wave in the continuous limit as an approximation in the limit $n \to \infty$ to obtain some qualitative behavior of DNA. In 1989, Peyrard and Bishop [107] published their pioneering work on DNA dynamical modeling in which the base pairing due to *hydrogen bonding* is recognized, the discreteness of the model is maintained, and a statistical mechanics study is fully carried out which describes the inter-strand separation in the double helix as a function of temperature, leading to a mathematical formulation of DNA denaturation. Following [107], here we initially allow two degrees of freedom for each pair of bases and use $u_{\mathbf{i}}$ and $v_{\mathbf{i}}$ to denote the displacements of the bases from their equilibrium positions along the direction of the hydrogen bonds that connect the two bases in a pair. The governing Hamiltonian for the *double helix model* containing a *harmonic coupling* between neighboring bases due to *stacking* and n pairs of bases is given as

$$H = \sum_{\mathbf{i}=1}^{\mathbf{n}} \left\{ \frac{1}{2} m(\dot{u}_{\mathbf{i}}^2 + \dot{v}_{\mathbf{i}}^2) + \frac{1}{2}\kappa([u_{\mathbf{i}} - u_{\mathbf{i}-1}]^2 + [v_{\mathbf{i}} - v_{\mathbf{i}-1}]^2) + V(u_{\mathbf{i}} - v_{\mathbf{i}}) \right\}, \tag{1.52}$$

where a common mass m is taken for all bases, a uniform 'elastic' (*stacking force*) constant κ is assumed for simplicity, and the potential energy V is defined by

$$V(u) = D(\mathrm{e}^{-\mathbf{a}u} - 1)^2, \tag{1.53}$$

which accounts for the *hydrogen bonding*, with a, D some positive constants, and is of the *Morse type* [94]. In terms of the new variables $x_{\mathbf{i}}, y_{\mathbf{i}}$ and the associated momenta $p_{\mathbf{i}}, P_{\mathbf{i}}$, defined by

$$x_{\mathbf{i}} = \frac{(u_{\mathbf{i}} + v_{\mathbf{i}})}{\sqrt{2}}, \quad y_{\mathbf{i}} = \frac{(u_{\mathbf{i}} - v_{\mathbf{i}})}{\sqrt{2}}, \quad p_{\mathbf{i}} = m\dot{x}_{\mathbf{i}}, \quad P_{\mathbf{i}} = m\dot{y}_{\mathbf{i}}, \tag{1.54}$$

the Hamiltonian (1.52) is normalized into the form

$$H = \sum_{\mathbf{i}=1}^{\mathbf{n}} \left\{ \frac{p_{\mathbf{i}}^2}{2m} + \frac{1}{2}\kappa(x_{\mathbf{i}} - x_{\mathbf{i}-1})^2 \right\} + \sum_{\mathbf{i}=1}^{\mathbf{n}} \left\{ \frac{P_{\mathbf{i}}^2}{2m} + \frac{1}{2}\kappa(y_{\mathbf{i}} - y_{\mathbf{i}-1})^2 + D(\mathrm{e}^{-\mathbf{a}\sqrt{2}\mathbf{y}_i} - 1)^2 \right\}, \tag{1.55}$$

where the definitions of x_0 and y_0 depend on the specific boundary condition to be considered. It is important to realize that the variable $y_{\mathbf{i}}$ measures the stretching distance between the bases

[1] See Wikipedia for more information.

in a pair of bases. To understand the thermal dynamics of stretching, note that the partition function Z is seen to be factored as [2]

$$Z = \int \mathrm{e}^{-\boldsymbol{\beta}\mathbf{H}(\mathbf{p},\mathbf{x},\mathbf{P},\mathbf{y})}\,\mathrm{d}x\mathrm{d}y\mathrm{d}p\mathrm{d}P = Z_{\mathbf{p}}Z_{\mathbf{x}}Z_{\mathbf{P}}\,Z_{\mathbf{y}}, \tag{1.56}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$, with T the absolute temperature and k_{B} the Boltzmann constant. From this, Peyrard and Bishop [107] recognized that the mean stretching $\langle y_{\ell}\rangle$ of the bases at the position $\ell = 1, \cdots, n$, due to the hydrogen bonding, is given by

$$\langle y_{\ell}\rangle = \frac{1}{Z}\int y_{\ell}\mathrm{e}^{-\boldsymbol{\beta}\mathbf{H}}\,\mathrm{d}x\mathrm{d}y\mathrm{d}p\mathrm{d}P = \frac{1}{Z_{\mathbf{y}}}\int y_{\ell}\mathrm{e}^{-\boldsymbol{\beta}\sum_{i=1}^{n}\mathbf{f}(\mathbf{y}_i,\mathbf{y}_{i-1})}\,\mathrm{d}y, \tag{1.57}$$

where the factors involving x, p, P are dropped as a consequence of the decomposed Hamiltonian (1.55) and $f(y, y')$ is the reduced potential given by the y-dependent terms in (1.55) as

$$f(y, y') = \frac{1}{2}\kappa(y - y')^2 + D(\mathrm{e}^{-\mathbf{a}\sqrt{2}\mathbf{y}} - 1)^2. \tag{1.58}$$

It is still rather difficult to analyze the quantity (1.57) as a function of the temperature T without further simplification. In [107], Peyrard and Bishop take $n \to \infty$ in (1.57) to arrive at

$$\langle y_{\ell}\rangle = \langle\varphi_0|y_{\ell}|\varphi_0\rangle = \int \varphi_0^2(y)y\,\mathrm{d}y, \tag{1.59}$$

where

$$\varphi_0(y) = \frac{(\sqrt{2}a)^{1/2}(2d)^{\mathbf{d}-1/2}}{\Gamma(2d-1)^{1/2}}\exp(-d\mathrm{e}^{-\sqrt{2}\mathbf{a}\mathbf{y}})\mathrm{e}^{-(\mathbf{d}-\frac{1}{2})\sqrt{2}\mathbf{a}\mathbf{y}}, \quad d = \frac{1}{a}\beta(\kappa D)^{1/2} > \frac{1}{2}. \tag{1.60}$$

Based on the above formalism, Peyrard and Bishop [107] succeeded in finding a thermodynamical description of the DNA denaturation phenomenon. Using (1.60) and numerical evaluation, it is shown [107] that the base mean stretching $\langle y_{\ell}\rangle$ increases significantly as the temperature climbs to a particular level which is an unambiguous indication of DNA denaturation. Another interesting by-product of such a calculation is that, since the dependence of the ground state on the absolute temperature $T = (k_{\mathrm{B}}\beta)^{-1}$ is through the parameter d given earlier, a greater value of the elastic constant κ leads to a higher DNA denaturation temperature, which is what observed [46, 107] in laboratory.

In particular, we have seen that the dynamics of the DNA molecule is effectively described by the reduced Hamiltonian that contains the 'out-of-phase' motion of the bases only given in terms of the y-variables as

$$H = \sum_{\mathbf{i}=1}^{\mathbf{n}}\left\{\frac{1}{2}m\dot{y}_{\mathbf{i}}^2 + \frac{1}{2}\kappa(y_{\mathbf{i}} - y_{\mathbf{i}-1})^2 + D(\mathrm{e}^{-\mathbf{a}\sqrt{2}\mathbf{y}_i} - 1)^2\right\}. \tag{1.61}$$

See [106] for a review of related topics and directions. This example shows how a simple system of ordinary differential equations may be used to investigate a fundamental problem in biophysics.

[2] To save space, we use $\mathrm{d}x$ (say) to denote $\mathrm{d}x^1\cdots\mathrm{d}x^n$ and use x to denote the vector coordinates (x^i) or a single variable interchangeably if there is no risk of confusion in the context.

Introduction and Background of Quantum Mechanics

Quantum mechanics (QM) was developed at the beginning of the last century aimed at explaining physical phenomena at microscopic scales (small mass and small distance) and based upon several celebrated experiments which could not be explained within the conceptual framework of classical physics. Even today, QM remains a challenge to human intuition and continues to stun away students who study it. See [84, 96] for some extra reading of the mathematical formulation and history of QM and [35, 55, 90, 115] for textbook introductions to QM oriented towards people interested in its physical origins, mathematical structure, and computational details. Referring to QM, in Preface of his book [55], Griffiths states "there is no general consensus as to what its fundamental principles are, how it should be taught, or what it really means" and quotes Richard Feynman's words "I think I can say safely that nobody understands quantum mechanics." All these sound rather pessimistic about QM. Nevertheless, QM is one of the greatest successes of modern physics. In this section, we will focus on the Schrödinger equation which is the core of QM and attempt to achieve a reasonable level of understanding of some basics of QM.

Some early-day discoveries

We start with a brief discussion of several milestone early-day discoveries that led to the formulation of QM. We will mainly follow the presentation in [55] in the first two subsections.

The photoelectric effect

Place a piece of metal in a vacuum tube and shoot a beam of light onto it. The electrons in the metal may become sufficiently energized to be emitted from the metal. This is the so-called *photoelectric effect* and finds wide range of applications in today's electronics. Now measure the energy carried by an emitted electron and denote it by $E_{\mathbf{e}}$. It is known that $E_{\mathbf{e}}$ may be written as the difference of two quantities, one is proportional to the frequency, ν, of the light beam so that the proportioality constant, h, is universal and independent of the metal, the other, ϕ, depends on the metal but is independent of the light frequency. Therefore, we have

$$E_{\mathbf{e}} = h\nu - \phi. \tag{2.1}$$

Einstein's postulate

Light, a special form of electromagnetic waves, is composed of particles called photons. Each photon carries an amount of energy equal to $h\nu$. That is,

$$E = h\nu. \tag{2.2}$$

When the photon hits an electron in a metal, the electron receives this amount of energy, consumes the amount of the binding energy of the metal to the electron to escape from the metal, and becomes an emitted electron of the energy given by (2.1).

Measurements

In physics, frequency ν is measured in hertz with unit second^{-1} (times per second), and angular frequency ω is related to ν by $\omega = 2\pi\nu$ (radians per second). Hence, in terms of ω, the Einstein formula becomes

$$E = \hbar\omega, \quad \hbar = \frac{h}{2\pi}. \tag{2.3}$$

Recall that energy is measured in unit of Joules and one Joule is equal to one Newton×meter. The constant $\hbar$ in (2.3), called the *Planck constant*, is a tiny number of the unit of Joules×second and accepted to be

$$\hbar = 1.05457 \times 10^{-34}. \tag{2.4}$$

(Historically, h is called the Planck constant, and $\hbar$ the *Dirac constant* or extended Planck constant.)

The Co[redacted] pton e ect

After the 1905 postulate of Einstein that light is composed of photons, physicists began to wonder whether a photon might exhibit its (kinetic) momentum in interaction (i.e., in collision with another particle). In 1922, Compton and Debye came with a very simple (and bold) mathematical description of this, which was then experimentally observed by Compton himself in 1923 and further proved by Y. H. Woo,[3] then a graduate student of Compton. In simple terms, when a photon hits an electron, it behaves like indeed like a particle collides with another particle so that one observes energy as well as momentum conservation relations, which is evidenced by a wavelength shift after the collision.

Mathematically, we write the energy of a photon by the *Einstein formula*, $E = mc^2$, where c is the speed of light in vacuum and m is the 'virtual rest mass' of photon (note that a photon in fact has no rest mass). In view of (2.3), we have

$$E = mc^2 = \hbar\omega. \tag{2.5}$$

Recall that the wavenumber (also called the *angular wavenumber*) k, wavelength λ, frequency ν, angular frequency ω, and speed c of a photon are related by

$$k = \frac{2\pi}{\lambda}, \quad c = \lambda\nu = \lambda\,\frac{\omega}{2\pi}. \tag{2.6}$$

Consequently, the momentum of the photon is given by

$$\begin{aligned} p &= mc = \frac{E}{c} = \hbar\,\frac{\omega}{c} \\ &= \hbar\, k. \end{aligned} \tag{2.7}$$

The de Brog ie ʾʌ ve p rtic e du ity hypothesis

In 1924, de Broglie formulated his celebrated *wave-particle duality* hypothesis in his Ph. D. thesis which equalizes waves and particles, takes the Einstein formula (2.3) and the *Compton–Debye formula* (2.7) as the two axioms, and reiterates the wave and particle characteristics of all interactions in nature:

$$E = \hbar\,\omega, \tag{2.8}$$

$$p = \hbar\, k. \tag{2.9}$$

[3]Here is what found in Wikipedia about Y. H. Woo: Wu graduated from the Department of Physics of Nanjing Higher Normal School (later renamed National Central University and Nanjing University), and was later associated with the Department of Physics at Tsinghua University. He was once the president of National Central University (later renamed Nanjing University and r

In other words, a particle of energy E and momentum p behaves like a wave of wavenumber k and a wave of wavenumber k behaves like a particle of energy E and momentum p such that E, p, and k are related through (2.8) and (2.9).

[illegible]od n [illegible]on

With the above preparation, we are now ready to derive the *Schrödinger equation* which was first published by Schrödinger in 1926.

Consider a stationary wave distributed over the x-axis of wavenumber k (the wave has k repe[illegible]

(2n.9sw

For a particle moving in a potential field V, the energy-momentum relation (2.17) becomes

$$E = \frac{p^2}{2m} + V. \tag{2.19}$$

Therefore the Schrödinger equation (2.18) for a free particle is modified into form

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\phi}{\partial x^2} + V\phi. \tag{2.20}$$

This is called the *Schrödinger wave equation* whose solution, ϕ, is called a *wave function*.

St tistic interpret tion of the ve function y Born

Consider the 1D Schrödinger equation (2.20) describing a particle of mass m and assume that ϕ is a 'normalized' solution of (2.20) which satisfies normalization condition

$$\int |\phi(x,t)|^2\,\mathrm{d}x = 1 \tag{2.21}$$

and characterizes the 'state' of the particle. According to Born, the mathematical meaning of such a wave function is that $\rho(x,t) = |\phi(x,t)|^2$ gives the *probability density* of the location of the particle at time t. In other words, the probability of finding the particle in an interval (a,b) at time t is

$$P(\{a < x(t) < b\}) = \int_{\mathbf{a}}^{\mathbf{b}} |\phi(x,t)|^2\,\mathrm{d}x. \tag{2.22}$$

With this interpretation, we see that the expected location of the particle at time t is

$$\langle x\rangle(t) = \int x|\phi(x,t)|^2\,\mathrm{d}x = \int \overline{\phi}(x,t)\,x\,\phi(x,t)\,\mathrm{d}x. \tag{2.23}$$

Oper tor represent tions of physic qu ntities

Naturally, the expected value of the momentum of the particle should be equal to the product of the particle mass and the expected value of the particle velocity. Therefore, in view of (2.20), we have

$$\begin{aligned}
\langle\hat{p}\rangle(t) &= m\frac{\mathrm{d}\langle x\rangle(t)}{\mathrm{d}t} \\
&= m\int (\overline{\phi}_{\mathbf{t}}(x,t)\,x\,\phi(x,t) + \overline{\phi}(x,t)\,x\,\phi_{\mathbf{t}}(x,t))\,\mathrm{d}x \\
&= \mathrm{i}\,\frac{\hbar}{2}\int (x\overline{\phi}\phi_{\mathbf{xx}} - \overline{\phi}_{\mathbf{xx}}\phi)\,\mathrm{d}x \\
&= \int \overline{\phi}(x,t)\Big(-\mathrm{i}\hbar\frac{\partial}{\partial x}\Big)\phi(x,t)\,\mathrm{d}x.
\end{aligned} \tag{2.24}$$

Hence, formally, the expected value of the momentum is the 'expected value' of the operator

$$\hat{p} = -\mathrm{i}\hbar\frac{\partial}{\partial x}. \tag{2.25}$$

In other words, within the framework of Born's statistical interpretation of the wave function, momentum has its elegant operator representation (2.25).

In this manner, we have the trivial operator representations

$$\hat{x} = x, \quad \hat{f}(x) = f(x), \tag{2.26}$$

for the particle coordinate x and its functions. Besides, (2.19) gives us the energy representation

$$\hat{E} = \frac{1}{2m}\hat{p}^2 + V \quad (= \text{the Hamiltonian}). \tag{2.27}$$

Thus,

$$\langle \hat{E} \rangle = \int \overline{\phi}(x,t)\bigg(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V\bigg)\phi(x,t)\,\mathrm{d}x. \tag{2.28}$$

Using (2.20) in (2.28), we have

$$\langle \hat{E} \rangle = \int \overline{\phi}(x,t)\bigg(\mathrm{i}\hbar\frac{\partial}{\partial t}\bigg)\phi(x,t)\,\mathrm{d}x. \tag{2.29}$$

In other words, energy should be represented by the operator

$$\hat{E} = \mathrm{i}\hbar\frac{\partial}{\partial t}. \tag{2.30}$$

The afore-discussed fundamental representations of various physical quantities form the foundation of QM.

Conserv tion ʌ nd pro i ity current

It is easily checked that the normalization condition is well posed because

$$\frac{\mathrm{d}}{\mathrm{d}t}\int |\phi(x,t)|^2\,\mathrm{d}x = 0, \tag{2.31}$$

by virtue of the equation (2.20) so that it suffices to require the condition

$$\int |\phi(x,0)|^2\,\mathrm{d}x = 1 \tag{2.32}$$

initially. Here, we look for some additional consequences from the *global conservation law* (2.31). For this purpose, we differentiate the probability density ρ to get

$$\begin{aligned} \rho_{\mathbf{t}} &= \overline{\phi}_{\mathbf{t}}\phi + \overline{\phi}\phi_{\mathbf{t}} \\ &= -\mathrm{i}\frac{\hbar}{2m}(\phi\overline{\phi}_{\mathbf{x}} - \overline{\phi}\phi_{\mathbf{x}})_{\mathbf{x}}, \end{aligned} \tag{2.33}$$

where we have used (2.20) again. It is interesting to view ρ as a '*charge*' density and rewrite (2.33) in the form of a *conservation law*,

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial x}j = 0, \tag{2.34}$$

where j may be viewed as a '*current*' density which is identified as

$$j = \mathrm{i}\,\frac{\hbar}{2m}(\phi\overline{\phi}_{\mathbf{x}} - \overline{\phi}\phi_{\mathbf{x}}). \tag{2.35}$$

We note that it is crucial that ϕ is complex-valued: if it is real-valued, the current density will be zero identically and ρ will be time-independent.

Furthermore, differentiating (2.22) and using (2.35), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P(\{a<x(t)<b\}) &= \frac{\mathrm{d}}{\mathrm{d}t}Q(a,b)(t)\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbf{a}}^{\mathbf{b}}\rho(x,t)\,\mathrm{d}x\\ &= j(a,t)-j(b,t),\end{aligned}\tag{2.36}$$

where $Q(a,b)$ may be interpreted as the charge contained in the interval (a,b) at time t so that its rate of change is equal to the net current following into such an interval. Or more correctly, we may call Q the '*probability charge*' and j the '*probability current.*'

The Ehrenfest theore

Differentiating (2.24) and using (2.20), we have

$$\begin{aligned}\frac{\mathrm{d}\langle\hat{p}\rangle}{\mathrm{d}t} &= \int\left(\overline{\phi}_{\mathbf{t}}\Big[-\mathrm{i}\hbar\frac{\partial}{\partial x}\Big]\phi+\overline{\phi}\Big[-\mathrm{i}\hbar\frac{\partial}{\partial x}\Big]\phi_{\mathbf{t}}\right)\mathrm{d}x\\ &= -\mathrm{i}\hbar\int\left(\Big[-\mathrm{i}\,\frac{\hbar}{2m}\overline{\phi}_{\mathbf{xx}}+\mathrm{i}\,\frac{V}{\hbar}\overline{\phi}\Big]\phi_{\mathbf{x}}+\overline{\phi}\frac{\partial}{\partial x}\Big[\mathrm{i}\,\frac{\hbar}{2m}\phi_{\mathbf{xx}}-\mathrm{i}\,\frac{V}{\hbar}\phi\Big]\right)\mathrm{d}x\\ &= -\frac{\hbar^2}{2m}\int([\overline{\phi}_{\mathbf{x}}\phi_{\mathbf{x}}]_{\mathbf{x}}-[\overline{\phi}\phi_{\mathbf{xx}}]_{\mathbf{x}})\,\mathrm{d}x-\int V_{\mathbf{x}}|\phi|^2\,\mathrm{d}x\\ &= -\langle V_{\mathbf{x}}\rangle,\end{aligned}\tag{2.37}$$

which may be compared with the equation of motion in the classical Newtonian mechanics,

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=\frac{\mathrm{d}p}{\mathrm{d}t}=-V_{\mathbf{x}}.\tag{2.38}$$

In other words, in QM, in sense of expected value, quantum operators obey the equation of motion of Newtonian mechanics. This statement is known as the *Ehrenfest theorem.*

nst e p rtic es

The profound meaning of the conservation law (2.21) is that a particle can never disappear once it is present. Here we show that a small modification may be made so that we are able to describe unstable particles which may disappear after some time elapse. We will not justify whether such a modification is physically correct but will only be content to know that there is room in QM to accommodate theoretical explorations. To this end, we assume that the potential energy V in (2.20) is perturbed by an imaginary quantity,

$$V=V_1+\mathrm{i}V_2,\quad V_1 \text{ and } V_2 \text{ are both real-valued.}\tag{2.39}$$

Hence, (2.20) becomes

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t}=-\frac{\hbar^2}{2m}\frac{\partial^2\phi}{\partial x^2}+(V_1+\mathrm{i}V_2)\phi.\tag{2.40}$$

In view of (2.40), we see that the probability that there is a particle present at time t, i.e.,

$$P(t)=\int|\phi(x,t)|^2\,\mathrm{d}x,\tag{2.41}$$

satisfies the equation

$$P'(t) = \frac{1}{\hbar}\int V_2(x)|\phi(x,t)|^2\,\mathrm{d}x. \tag{2.42}$$

For simplicity, we further assume that there is a constant $\Gamma > 0$ such that

$$V_2(x) \le -\Gamma, \quad \forall x. \tag{2.43}$$

Then (2.42) and (2.43) lead us to

$$P'(t) \le -\frac{\Gamma}{\hbar}P(t). \tag{2.44}$$

If a particle is present initially, $P(0) = 1$, then (2.44) implies that

$$P(t) \le \mathrm{e}^{-\frac{\Gamma}{\hbar}\mathbf{t}}, \quad t > 0. \tag{2.45}$$

In other words, in a bulk situation, we will observe loss of particles, which suggests that we encounter unstable particles.

Higher dimensions

It is immediate that our discussion about the 1D Schrödinger equations can be extended to arbitrarily high dimensions. For this purpose, we consider the Minkowski space of dimension $(n+1)$ with coordinates $t = x^0$, $= (x^1, \cdots, x^{\mathbf{n}})$, for time and space, respectively. We use the Greek letter μ, ν, etc, to denote the spacetime indices, $\mu, \nu = 0, 1, \cdots, n$, the Latin letter i, j, k, etc, to denote the space indices, $i, j, k = 1, \cdots, n$, and ∇ to denote the gradient operator on functions depending on $x^1, \cdots, x^{\mathbf{n}}$.

The Schrödinger equation that quantum-mechanically governs a particle of mass m in $\mathbb{R}^{\mathbf{n}}$ is given by

$$\mathrm{i}\hbar\frac{\partial\phi}{\partial t} = -\frac{\hbar^2}{2m}\Delta\phi + V\phi. \tag{2.46}$$

The energy and momentum operators are, respectively, given by

$$\hat{E} = \mathrm{i}\hbar\frac{\partial}{\partial t}, \quad \hat{\mathbf{p}} = -\mathrm{i}\hbar\nabla. \tag{2.47}$$

The total energy operator, or the Hamiltonian, is

$$\hat{H} = \frac{1}{2m}\hat{\mathbf{p}}^2 + V. \tag{2.48}$$

The probability current $j = (j^{\boldsymbol{\mu}}) = (j^0, \mathbf{j}) = (\rho, j^{\mathbf{i}})$ is defined by

$$j^{\mathbf{i}} = \mathrm{i}\,\frac{\hbar}{2m}(\phi\partial_{\mathbf{i}}\overline{\phi} - \overline{\phi}\partial_{\mathbf{i}}\phi), \quad i = 1, \cdots, n. \tag{2.49}$$

The conservation law relating probability density ρ (or 'charge') and probability current $\mathbf{j}$ (or 'current') reads

$$\partial_{\boldsymbol{\mu}} j^{\boldsymbol{\mu}} = 0, \quad \text{or} \quad \frac{\partial}{\partial t}\rho + \nabla\cdot\mathbf{j} = 0. \tag{2.50}$$

Moreover, the Ehrenfest theorem says

$$\frac{\mathrm{d}\langle\hat{\mathbf{p}}\rangle}{\mathrm{d}t} = -\langle\nabla V\rangle. \tag{2.51}$$

Note also that, in applications, the potential function V may be self-induced by the wave function ϕ. For example, $V = |\phi|^2$.

Ste dy st te nd energy spectru

We return to the one-dimensional situation again and look for solution of (2.20) in the separable form

$$\phi(x,t) =$$

Therefore, in view of the normalization condition (2.21), we arrive at

$$\sum_{\mathbf{n}=1}^{\infty} |c_{\mathbf{n}}|^2 = 1. \tag{2.61}$$

This result is amazing since it reminds us of the total probability of a discrete random variable whose probability mass density function is given by the sequence $\{|c_{\mathbf{n}}|^2\}$. In QM, indeed, such a random variable is *postulated* as the measured value of energy for the particle that occupies the state given by (2.60). In other words, if we use $\mathcal{E}$ to denote the random reading of the energy of the particle occupying the state (2.60), then $\mathcal{E}$ may only take $E_1, E_2, \cdots, E_{\mathbf{n}}, \cdots$, as possible values. Furthermore, if these values are distinct, then

$$P(\{\mathcal{E} = E_{\mathbf{n}}\}) = |c_{\mathbf{n}}|^2, \quad n = 1, 2, \cdots. \tag{2.62}$$

In QM, the above statement appears as a major postulate which is also referred to as the '*generalized statistical interpretation*' of eigenstate representation. In particular, when $\phi(x,t)$ itself is separable as given in (2.52), since E is an eigenvalue itself, we see that $\mathcal{E}$ takes the single value E with probability one. Thus, we recover the earlier observation made on separable state.

Finally, using (2.60), we can compute the expected value of the energy operator $\hat{H}$ immediately:

$$\begin{aligned} \langle \hat{H} \rangle &= \int \overline{\phi} \hat{H} \phi \, \mathrm{d}x \\ &= \sum_{\mathbf{n}=1}^{\infty} E_{\mathbf{n}} |c_{\mathbf{n}}|^2. \end{aligned} \tag{2.63}$$

It is interesting to note that (2.63) is consistent with the postulate (2.62). In fact, we may also understand that the interpretation (2.62) is motivated (or supported) by (2.63).

Note also that the right-hand side of (2.63) is independent of time t. Thus, (2.63) may be viewed as a QM version of the energy conservation law.

n N, ody p o

Consider the QM description of N particles of respective masses $m_{\mathbf{i}}$ and electric charges $Q_{\mathbf{i}}$ $(i = 1, \cdots, N)$ interacting solely through the Coulomb force. The locations of these N particles are at $\mathbf{x}_{\mathbf{i}} \in \mathbb{R}^3$. Thus we can express their respective momenta as

$$\hat{\mathbf{p}}_{\mathbf{i}} = -\mathrm{i}\hbar \nabla_{\mathbf{x}_i}, \quad i = 1, \cdots, N. \tag{2.64}$$

The potential function is given by

$$V(\mathbf{x}_1, \cdots, \mathbf{x}_{\mathbf{N}}) = \sum_{1 \le \mathbf{i} < \mathbf{i}' \le \mathbf{N}}^{\mathbf{N}} \frac{Q_{\mathbf{i}} Q_{\mathbf{i}'}}{|\mathbf{x}_{\mathbf{i}} - \mathbf{x}_{\mathbf{i}'}|}. \tag{2.65}$$

Thus, the N-particle system Hamiltonian reads

$$\hat{H} = \sum_{\mathbf{i}=1}^{\mathbf{N}} \frac{1}{2m_{\mathbf{i}}} \hat{\mathbf{p}}_{\mathbf{i}}^2 + V(\mathbf{x}_1, \cdots, \mathbf{x}_{\mathbf{N}}), \tag{2.66}$$

which forms the foundation of the *quantum N-body problem*.

An important special situation of the quantum N-body problem is the classical atom model in which Z electrons, each of electric charge e and mass m, orbit around a heavy nucleus of electric charge Ze resting at the origin. In this case, the Hamiltonian becomes

$$\hat{H} = -\sum_{\mathbf{i}=1}^{\mathbf{Z}} \frac{\hbar^2}{2m} \nabla_{\mathbf{x}_i}^2 - \sum_{\mathbf{i}=1}^{\mathbf{Z}} \frac{Ze^2}{|\ \mathbf{i}|} + \sum_{1 \le \mathbf{i} < \mathbf{i}' \le \mathbf{Z}} \frac{e^2}{|\ \mathbf{i} - \ \mathbf{i}'|}, \tag{2.67}$$

where the second term describes the Coulomb interaction of the electrons with the nucleus and the third term that between the electrons. In the case of a hydrogen atom, $Z = 1$, and the third term disappears. Thus, we arrive at the simplest possible Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla_{\mathbf{x}}^2 - \frac{e^2}{|\ |}. \tag{2.68}$$

The spectrum of (2.68), say $\{E\}$, consists of two different portions: $E > 0$ (which happens to be continuous) and $E < 0$ (which happens to be discrete). Since the Coulomb potential vanishes at infinity, the state with $E < 0$ indicates that the electron is in a state which lies inside the 'potential well' of the nucleus, the proton, which is called a *bound state*, and describes a situation when the electron and proton 'bind' to form a composite particle, the hydrogen. Likewise, the state with $E > 0$ indicates that the electron is in a state which lies outside the potential well of the proton, which is called a *scattering state*, and describes a situation when the electron and proton interact as two charged 'free' particles which do not appear to have the characteristics of a composite particle, namely, a hydrogen atom. Hence we will be interested in the bound state situation only.

Restricting to spherically symmetric configurations, it can be shown [55] that the bound-state energy spectrum of (2.68) is given by

$$E_{\mathbf{n}} = -\frac{me^4}{2\hbar^2 n^2} = \frac{E_1}{n^2}, \quad n = 1, 2, \cdots, \tag{2.69}$$

known as the *Bohr formula*. The ground-state energy, E_1, is about -13.6 eV which is what is needed to ionize a hydrogen atom.

Suppose that the hydrogen atom absorbs or emits an amount of energy, $E_{\boldsymbol{\delta}}$, so that the initial and final energies are E_{i} and E_{f}, respectively. Then

$$E_{\boldsymbol{\delta}} = E_{\mathrm{i}} - E_{\mathrm{f}} = E_1 \left(\frac{1}{n_{\mathrm{i}}^2} - \frac{1}{n_{\mathrm{f}}^2} \right). \tag{2.70}$$

It will be instructive to examine in some detail that the hydrogen atom is made to emit energy through the form of light. The Plank formula states that the frequency ν of the light obeys

$$E_{\boldsymbol{\delta}} = h\nu = h\frac{c}{\lambda}, \tag{2.71}$$

where c is the speed and λ is the wavelength of light. Substituting (2.71) into (2.70), we obtain the celebrated *Rydberg formula*

$$\frac{1}{\lambda} = \frac{me^4}{4\pi c\hbar^3} \left(\frac{1}{n_{\mathrm{f}}^2} - \frac{1}{n_{\mathrm{i}}^2} \right), \quad n_{\mathrm{i}} > n_{\mathrm{f}}. \tag{2.72}$$

Specifically, transitions to the ground state $n_{\rm f} = 1$ give rise to ultraviolet (higher-frequency) lights with

$$\frac{1}{\lambda} = \frac{me^4}{4\pi c\hbar^3}\left(1 - \frac{1}{n^2}\right), \quad n = 2, 3, \cdots, \tag{2.73}$$

called the Lyman series; transitions to the first excited state $n_{\rm f} = 2$ lead to visible (medium-frequency) lights with $n = 3, 4, \cdots$, called the Balmer series; transitions to the second excited state $n_{\rm f} = 3$ correspond to infrared (lower-frequency) lights with $n = 4, 5, \cdots$, called the Paschen series. The series with $n_{\rm f} = 4, 5, 6$ are named under Brackett, Pfund, and Humphreys, respectively. The Rydberg formula was presented by Johannes Robert Rydberg (a Swedish experimental physicist at Lund University) in 1888,[4] many years before the formulation of the Schrödinger equation and QM.

The model for helium, with $Z = 2$, immediately becomes more difficult because the Hamiltonian takes the form

$$\hat{H} = -\left(\frac{\hbar^2}{2m}\nabla^2_{\mathbf{x}_1} + \frac{2e^2}{|\ _1|}\right) - \left(\frac{\hbar^2}{2m}\nabla^2_{\mathbf{x}_2} + \frac{2e^2}{|\ _2|}\right) + \frac{e^2}{|\ _1 -\ _2|}, \tag{2.74}$$

in which the last term renders the problem non-separable.[5] The model for lithium, with $Z = 3$, shares the same difficulty. Thus, we see that the quantum N-body problem is important for particle physics and quantum chemistry but difficult to deal with when $N \geq 2$. A way out of this is to develop approximation methods. Along this direction, two well-known approaches are the Hartree–Fock method and the Thomas–Fermi model, both based on variational techniques. We will briefly discuss these approaches below.

We note that, while the classical N-body problem is nonlinear, its QM version, which asks about the spectrum of the N-body Hamiltonian, becomes linear.

- oc od

To motivate our discussion, we use $\{E_{\mathbf{n}}\}$ to denote the complete sequence of eigenvalues of the Hamiltonian $\hat{H}$ so that

$$E_1 \leq E_2 \leq \cdots \leq E_{\mathbf{n}} \leq \cdots, \tag{2.75}$$

and $\{u_{\mathbf{n}}\}$ the corresponding eigenstates which form an orthonormal basis. Let ψ be any normalized function, satisfying

$$\langle\psi|\psi\rangle = \int |\psi|^2 \,\mathrm{d}x = 1. \tag{2.76}$$

[4]The original Rydberg formula reads

$$\frac{1}{\lambda} = R\left(\frac{1}{n_{\rm f}^2} - \frac{1}{n_{\rm i}^2}\right), \quad n_{\rm i} > n_{\rm f},$$

where R is the Rydberg constant with the observed value $R = 1.097 \times 10^7\ \mathrm{m}^{-1}$, which amazingly agrees with that given in (2.72) in terms of the fundamental constants $c, e, m, \hbar$.

[5]One might be tempted to ignore the inter-electron interaction spelled out by the last term. In this situation, the Hamiltonian is the sum of the two hydrogen Hamiltonians, which is separable and renders the ground-state energy $E_1 = 8(-13.6) = -108.8$ eV. This value is much lower than the experimentally determined value -79 eV for helium. Hence, we see that the inter-electron potential, which indeed adds a positive contribution, should not be neglected.

Thus, the expansion

$$\psi = \sum_{\mathbf{n}=1}^{\infty} c_{\mathbf{n}} u_{\mathbf{n}}, \tag{2.77}$$

gives us

$$\sum_{\mathbf{n}=1}^{\infty} |c_{\mathbf{n}}|^2 = 1. \tag{2.78}$$

Consequently, we have

$$\langle \psi | \hat{H} | \psi \rangle = \int \overline{\psi} \hat{H} \psi \, \mathrm{d}x = \sum_{\mathbf{n}=1}^{\infty} E_{\mathbf{n}} |c_{\mathbf{n}}|^2 \geq E_1 \sum_{\mathbf{n}=1}^{\infty} |c_{\mathbf{n}}|^2 = E_1. \tag{2.79}$$

In other words, the lowest eigen-pair (E_1, u_1) may be obtained from solving the minimization problem

$$\min \left\{ \langle \psi | \hat{H} | \psi \rangle \mid \langle \psi | \psi \rangle = 1 \right\}. \tag{2.80}$$

In practice, it is often hard to approach (2.80) directly due to lack of compactness. Instead, one may come up with a reasonably good wave-function configuration, a trial approximation, depending on finitely many parameters, say $\alpha_1, \cdots, \alpha_{\mathbf{m}}$, of the form

$$\psi(x) = \psi(\alpha_1, \cdots, \alpha_{\mathbf{m}})(x). \tag{2.81}$$

Then one solves the minimization problem

$$\min \left\{ \langle \psi(\alpha_1, \cdots, \alpha_{\mathbf{m}}) | \hat{H} | \psi(\alpha_1, \cdots, \alpha_{\mathbf{m}}) \rangle \mid \langle \psi(\alpha_1, \cdots, \alpha_{\mathbf{m}}) | \psi(\alpha_1, \cdots, \alpha_{\mathbf{m}}) \rangle = 1 \right\}, \tag{2.82}$$

involving multivariable functions of $\alpha_1, \cdots, \alpha_{\mathbf{m}}$ only.

We now rewrite the Hamiltonian (2.67) as

$$\hat{H} = \sum_{\mathbf{i}=1}^{\mathbf{Z}} \hat{H}_{\mathbf{i}} + \frac{1}{2} \sum_{\mathbf{i} \neq \mathbf{j}} V_{\mathbf{ij}}, \tag{2.83}$$

where

$$\hat{H}_{\mathbf{i}} = -\frac{\hbar^2}{2m} \nabla^2_{\mathbf{x}_i} - \frac{Ze^2}{|\ \mathbf{i}|}, \quad V_{\mathbf{ij}} = \frac{e^2}{|\ \mathbf{i} - \ \mathbf{j}|}, \quad i \neq j, \quad i, j = 1, \cdots, Z, \tag{2.84}$$

are the i-th electron Hamiltonian, without inter-electron interaction, and the inter-electron Coulomb potential between the i- and j-th electrons, respectively.

Since the non-interacting Hamiltonian $\sum_{\mathbf{i}=1}^{\mathbf{Z}} \hat{H}_{\mathbf{i}}$ allows separation of variables, we are prompted to use the trial configuration

$$\psi(\ \ _1, \cdots, \ \ \mathbf{z}) = \phi_1(\ \ _1) \cdots \phi_{\mathbf{Z}}(\ \ \mathbf{z}), \qquad _1, \cdots, \ \ \mathbf{z} \in \mathbb{R}^3, \tag{2.85}$$

known as the *Hartree product*, where $\phi_1, \cdots, \phi_{\mathbf{Z}}$ are unknowns. In order to implement the normalization condition $\langle \psi | \psi \rangle = 1$, we impose

$$\langle \phi_{\mathbf{i}} | \phi_{\mathbf{i}} \rangle = \int |\phi_{\mathbf{i}}|^2(\ \ \mathbf{i}) \, \mathrm{d}\ \ \mathbf{i} = \int |\phi_{\mathbf{i}}|^2(\ \ ) \, \mathrm{d}\ \ = 1, \quad i = 1, \cdots, Z. \tag{2.86}$$

Inserting (2.85) and using (2.86), we arrive at

$$\begin{aligned} I(\phi_1,\cdots,\phi_{\mathbf{Z}}) &= \int \overline{\psi}\hat{H}\psi \,\mathrm{d}\ _1\cdots\mathrm{d}\ _{\mathbf{Z}} \\ &= \sum_{\mathbf{i}=1}^{\mathbf{Z}} \int \overline{\phi}_{\mathbf{i}}\hat{H}_{\mathbf{i}}\phi_{\mathbf{i}}\,\mathrm{d}\ _{\mathbf{i}} + \frac{1}{2}\sum_{\mathbf{i}\neq\mathbf{j}} \int \overline{\phi}_{\mathbf{i}}\overline{\phi}_{\mathbf{j}}V_{\mathbf{ij}}\phi_{\mathbf{i}}\phi_{\mathbf{j}}\,\mathrm{d}\ _{\mathbf{i}}\,\mathrm{d}\ _{\mathbf{j}} \\ &= \sum_{\mathbf{i}=1}^{\mathbf{Z}} \int \left(\frac{\hbar}{2m}|\nabla\phi_{\mathbf{i}}|^2 - \frac{Ze^2}{|\ |}|\phi_{\mathbf{i}}|^2\right)\mathrm{d}\ \ + \frac{e^2}{2}\sum_{\mathbf{i}\neq\mathbf{j}} \int \frac{|\phi_{\mathbf{i}}(\ \ )|^2|\phi_{\mathbf{j}}(\mathbf{y})|^2}{|\ \ -\mathbf{y}|}\,\mathrm{d}\ \,\mathrm{d}\mathbf{y}, \end{aligned} \tag{2.87}$$

where we have renamed the dummy variables with $\ ,\mathbf{y}\in\mathbb{R}^3$. Consequently, we are led to considering the reduced constrained minimization problem

$$\min\left\{I(\phi_1,\cdots,\phi_{\mathbf{Z}})\,|\,\langle\phi_{\mathbf{i}}|\phi_{\mathbf{i}}\rangle = 1,\, i=1,\cdots,Z\right\}, \tag{2.88}$$

whose solutions may be obtained by solving the following system of nonlinear integro-differential equations

$$\frac{\hbar}{2m}\Delta\phi_{\mathbf{i}} + \frac{Ze^2}{|\ |}\phi_{\mathbf{i}} + \lambda_{\mathbf{i}}\phi_{\mathbf{i}} = \frac{e^2}{2}\left(\sum_{\mathbf{j}\neq\mathbf{i}}^{\mathbf{Z}}\int\frac{|\phi_{\mathbf{j}}(\mathbf{y})|^2}{|\ \ -\mathbf{y}|}\,\mathrm{d}\mathbf{y}\right)\phi_{\mathbf{i}}, \quad i=1,\cdots,Z, \tag{2.89}$$

with the Lagrange multipliers $\lambda_1,\cdots,\lambda_{\mathbf{Z}}$ appearing as eigenvalues. Thus, in particular, we see that, in order to solve a linear problem with interacting potential, we are offered a highly nontrivial nonlinear problem to tackle instead.

Note that the above discussion of the Hartree–Fock method is over-simplified. Since electrons are *fermions* which obey the *Pauli exclusion principle*, the wave function to be considered should have been taken to be skew-symmetric instead of being symmetric, with respect to permutations of $_1,\cdots,\ _{\mathbf{Z}}$ and $\phi_1,\cdots,\phi_{\mathbf{Z}}$, which give rise to their joint wave function $\psi(\ _1,\cdots,\ _{\mathbf{Z}})$. Thus, practically, we need to consider the problem with the redesigned skew-symmetric wave function

$$\psi(\ _1,\cdots,\ _{\mathbf{Z}}) = \frac{1}{\sqrt{Z!}}\begin{vmatrix} \phi_1(\ _1) & \phi_1(\ _2) & \cdots & \phi_1(\ _{\mathbf{Z}}) \\ \phi_2(\ _1) & \phi_2(\ _2) & \cdots & \phi_2(\ _{\mathbf{Z}}) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_{\mathbf{Z}}(\ _1) & \phi_{\mathbf{Z}}(\ _2) & \cdots & \phi_{\mathbf{Z}}(\ _{\mathbf{Z}}) \end{vmatrix}, \tag{2.90}$$

known as the *Slater determinant* [131], which makes the problem more complicated.

The Hartree–Fock method is effective when the atom number Z is small. When Z is large, the problem quickly becomes difficult and alternative methods are to be developed. The Thomas [140] and Fermi [45] approach is such a method which treats electrons as a *static electron gas cloud* surrounding a nucleus and subject to a continuously distributed electrostatic potential. The electron at assumes the maximum energy, say $-eA$, where A is a constant otherwise the electrons will not be staying in the static state. Let the electrostatic potential be $\phi(\ )$. Then

$-e\phi(\ \ )$ will be the potential energy carried by the electron. Thus, if we use $p(\ \ )$ to denote the maximum momentum of the electron, we have the relation

$$-eA = \frac{p^2(\ \ )}{2m} - e\phi(\ \ ). \tag{2.91}$$

On the other hand, let $n(\ \ )$ be the number of electrons per unit volume of the space, which is taken to be a tiny domain, say $\delta\Omega$, centered around $\ \ $. Then $p(\ \ )$ is approximately a constant over $\delta\Omega$. We assume that all states in the momentum space are occupied by the electrons which take up a volume

$$\frac{4\pi}{3}p^3(\ \ ) \tag{2.92}$$

in the momentum space. Since each state can be occupied by exactly one electron, due to Pauli's exclusion principle, we arrive at the electron number count (in $\delta\Omega$)

$$n(\ \ ) = 2\frac{\frac{4\pi}{3}p^3(\ \ )}{h^3}, \tag{2.93}$$

where h is the Planck constant and the factor 2 takes account of the two possible spins of the electrons. Inserting (2.91) into (2.93), we have

$$n(\ \ ) = \frac{8\pi}{3h^3}\left(2me[\phi(\ \ ) - A]\right)^{\frac{3}{2}}. \tag{2.94}$$

On the other hand, we know that the electrostatic potential function ϕ and the electron number density n are related through the Poisson equation

$$\Delta\phi = 4\pi en, \tag{2.95}$$

where $-ne = \rho$ is the charge density (cf. §3). In view of (2.94) and (2.95), we obtain the *self-consistency equation*

$$\Delta\phi = \alpha(\phi - A)^{\frac{3}{2}}, \quad \alpha = \frac{32\pi^2 e}{3h^3}(2me)^{\frac{3}{2}}, \tag{2.96}$$

which serves as the governing equation of the Thomas–Fermi method, also called the *Thomas–Fermi equation*. Of course, a meaningful solution must satisfy $\phi \geq A$.

Since the electron cloud surrounds a nucleus of charge Ze, we see that ϕ should behave like a central Coulomb potential, $Ze/|\ \ |$, near the origin. Hence, we have the singular boundary condition

$$\lim_{|\mathbf{x}|\to 0} |\ \ |\phi(\ \ ) = Ze. \tag{2.97}$$

Besides, if we assume the electron cloud is concentrated in a bounded domain, say Ω, then $n = 0$ on $\partial\Omega$. Thus, (2.94) leads to the boundary condition

$$\phi(\ \ ) = A, \qquad \in \partial\Omega. \tag{2.98}$$

It is interesting to note that the Thomas–Fermi semi-classical treatment of the quantum N-body which is linear turns the problem back into a nonlinear problem. See [74, 75, 76] for the mathematical work on the Thomas–Fermi model and many important extensions.

M on c onopo c

In this section, we start with the electromagnetic duality in the Maxwell equations. We then present the Dirac monopole and Dirac strings. Next, we consider the motion of a charged particle in an electromagnetic field and introduce the notion of gauge fields. Finally, we derive Dirac's charge quantization formula.

M on nd c o n c d y

Let the vector fields $\mathbf{E}$ and $\mathbf{B}$ denote the electric and magnetic fields, respectively, which are induced from the presence of an electric charge density distribution, ρ, and a current density, $\mathbf{j}$. Then these fields are governed in the Heaviside–Lorentz rationalized units by the Maxwell equations

$$\nabla \cdot \mathbf{E} = \rho, \tag{3.1}$$

$$\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{j}, \tag{3.2}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{3.3}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \quad . \tag{3.4}$$

In vacuum where $\rho = 0$, $\mathbf{j} =$, these equations are invariant under the dual correspondence

$$\mathbf{E} \mapsto \mathbf{B}, \quad \mathbf{B} \mapsto -\mathbf{E}. \tag{3.5}$$

In other words, in another world 'dual' to the original one, electricity and magnetism are seen e

c[illegible] onopo nd c n

In order to motivate our study, consider the classical situation that electromagnetism is generated from an ideal point electric charge q lying at the origin,

$$\rho_{\mathbf{e}} = 4\pi q\delta(\ ), \quad \mathbf{j_e} = \ , \quad \rho_{\mathbf{m}} = 0, \quad \mathbf{j_m} = \ . \tag{3.11}$$

It is clear that, inserting (3.11), the system (3.6)–(3.9) can be solved to yield $\mathbf{B} =$ and

$$\mathbf{E} = \frac{q}{|\ |^3}\ , \tag{3.12}$$

which is the well-known Coulomb law in static electricity.

We now consider the case of a point magnetic charge g, or a monopole, at the origin,

$$\rho_{\mathbf{e}} = 0, \quad \mathbf{j_e} = \ , \quad \rho_{\mathbf{m}} = 4\pi g\,\delta(\ ), \quad \mathbf{j_m} = \ . \tag{3.13}$$

Hence $\mathbf{E} = 0$ and

$$\mathbf{B} = \frac{g}{|\ |^3}\ = -g\nabla\left(\frac{1}{|\ |}\right). \tag{3.14}$$

Consequently, the magnetic flux through a sphere centered at the origin and of radius $r > 0$ is

$$\Phi = \int_{|\mathbf{x}|=\mathbf{r}} \mathbf{B}\cdot\,\mathrm{d}\mathbf{S} = 4\pi g, \tag{3.15}$$

which is independent of r and is identical to the Gauss law for static electricity. Nevertheless, we show below through quantum mechanics that the introduction of a magnetic charge yields drastically new physics because electric and magnetic fields are induced differently from a gauge vector potential.

We now evaluate the energy of a monopole. Recall that the total energy of an electromagnetic field with electric component $\mathbf{E}$ and magnetic component $\mathbf{B}$ is given by

$$E = \frac{1}{2}\int_{\mathbb{R}^3} (\mathbf{E}^2 + \mathbf{B}^2)\,\mathrm{d}x. \tag{3.16}$$

Inserting (3.14) into (3.16) and using $r = |\ |$, we have

$$E = 2\pi g^2 \int_0^\infty \frac{1}{r^2}\,\mathrm{d}r = \infty. \tag{3.17}$$

This energy blow-up seems to suggest that the idea of a magnetic monopole encounters an unacceptable obstacle. However, since the Coulomb law expressed in (3.12) for a point electric charge also leads to a divergent energy of the same form, (3.17), the infinite energy problem for a monopole is not a more serious one than that for a point electric charge which has been used effectively as good approximation for various particle models.

We now study the magnetic field generated from a monopole more closely.

Recall that for the electric field generated from a point electric charge q, the Coulomb law (3.12) gives us a scalar potential function $\phi = -q/|\ |$ such that $\mathbf{E} = \nabla\phi$ holds everywhere away from the point electric charge.

Similarly, we consider the magnetic field $\mathbf{B}$ generated from a monopole of charge g, given in (3.14). Based on the classical knowledge on magnetic field, we know that $\mathbf{B}$ should be solenoidal. That is, there should exist a vector field $\mathbf{A}$ such that

$$\mathbf{B} = \nabla \times \mathbf{A}, \tag{3.18}$$

except at the origin where the point monopole is placed. Unfortunately, using (3.15) and the Stokes theorem, it is easy to see that (3.18) cannot hold everywhere on any closed sphere centered at the origin. In other words, any such sphere would contain a singular point at which (3.18) fails. Shrinking a sphere to the origin would give us a continuous locus of singular points which is a string that links the origin to infinity. Such a string is called a *Dirac string*.

It may be checked directly that

$$\mathbf{A}^+ = (A_1^+, A_2^+, A_3^+), \quad A_1^+ = \frac{-x^2}{|\ |(|\ |+x^3)}g, \quad A_2^+ = \frac{x^1}{|\ |(|\ |+x^3)}g, \quad A_3^+ = 0, \tag{3.19}$$

satisfies (3.18) everywhere except on the negative x^3-axis, $x^1 = 0, x^2 = 0, x^3 \leq 0$. That is, with $\mathbf{A}^+$, the Dirac string S^- is the negative x^3-axis. Similarly,

$$\mathbf{A}^- = (A_1^-, A_2^-, A_3^-), \quad A_1^- = \frac{x^2}{|\ |(|\ |-x^3)}g, \quad A_2^- = \frac{-x^1}{|\ |(|\ |-x^3)}g, \quad A_3^- = 0, \tag{3.20}$$

satisfies (3.18) everywhere except on the positive x^3-axis, $x^1 = 0, x^2 = 0, x^3 \geq 0$. That is, with $\mathbf{A}^-$, the Dirac string S^+ is the positive x^3-axis.

As a consequence, $\mathbf{A}^+$ and $\mathbf{A}^-$ do not agree away from $S^+ \cup S^-$ because there holds

$$\mathbf{A}^+ - \mathbf{A}^- = \quad = (a_1, a_2, a_3), \qquad \in \mathbb{R}^3 \setminus (S^+ \cup S^-), \tag{3.21}$$

where

$$a_1 = \frac{-2gx^2}{(x^1)^2 + (x^2)^2}, \quad a_2 = \frac{2gx^1}{(x^1)^2 + (x^2)^2}, \quad a_3 = 0. \tag{3.22}$$

In order to understand the physical meaning associated with the appearance of the Dirac strings which give rise to the puzzling ambiguity (3.21), we need to consider the motion of an electric charge in an electromagnetic field.

C d p c n n c o n c d

Consider a point particle of mass m and electric charge Q moving in an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$, in addition to a potential field V, in the Euclidean space $\mathbb{R}^3$ so that $= (x^1, x^2, x^3)$ gives the location of the particle. The equation of motion is

$$m\ddot{} = Q(\mathbf{E} + \dot{} \times \mathbf{B}) - \nabla V, \tag{3.23}$$

where $Q\mathbf{E}$ is the electric force and $Q\dot{} \times \mathbf{B}$ is the *Lorentz force* of the magnetic field $\mathbf{B}$ exerted on the particle of velocity $\dot{}$.

Let $\mathbf{B}$ and $\mathbf{E}$ be represented by a vector potential $\mathbf{A}$ and a scalar potential Ψ as follows

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A}, \\ \mathbf{E} &= \nabla\Psi - \frac{\partial \mathbf{A}}{\partial t}. \end{aligned}$$

At classical level, Ψ and $\mathbf{A}$ do not contribute to the underlying physics because they do not make appearance in the governing equations (3.1)–(3.4) and (3.23). However, at quantum-mechanical level, they do make observable contributions. Such a phenomenon was predicted by Aharonov and Bohm [5, 6] and is known as the *Aharonov–Bohm effect*. Thus, in order to explore the meaning of the 'ambiguity' of the vector potential associated with (3.21), we need to consider the QM description of the motion of the charged particle.

Using $\mathbf{y} = (y_{\mathbf{i}}) = m\dot{} = m(\dot{x^{\mathbf{i}}})$ to denote the mechanical momentum vector, the equation (3.23) becomes

$$\begin{aligned} \dot{y_{\mathbf{i}}} &= Q\left(\frac{\partial\Psi}{\partial x^{\mathbf{i}}} - \frac{\partial A_{\mathbf{i}}}{\partial t}\right) + Q\dot{x^{\mathbf{j}}}\left(\frac{\partial A_{\mathbf{j}}}{\partial x^{\mathbf{i}}} - \frac{\partial A_{\mathbf{i}}}{\partial x^{\mathbf{j}}}\right) - \frac{\partial V}{\partial x^{\mathbf{i}}} \\ &= -Q\frac{\mathrm{d}A_{\mathbf{i}}}{\mathrm{d}t} + Q\frac{\partial\Psi}{\partial x^{\mathbf{i}}} + Q\dot{x^{\mathbf{j}}}\frac{\partial A_{\mathbf{j}}}{\partial x^{\mathbf{i}}} - \frac{\partial V}{\partial x^{\mathbf{i}}}, \end{aligned}$$

which may be recast into the form

$$\frac{\mathrm{d}}{\mathrm{d}t}(y_{\mathbf{i}} + QA_{\mathbf{i}}) = \frac{\partial}{\partial x^{\mathbf{i}}}(Q\Psi + Q\dot{x^{\mathbf{j}}}A_{\mathbf{j}} - V),$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x^{\mathbf{i}}}}\right) = \frac{\partial L}{\partial x^{\mathbf{i}}}, \quad i = 1, 2, 3, \tag{3.24}$$

if we define the function L to be

$$\begin{aligned} L(\ \ ,\dot{\ \ },t) &= \frac{1}{2}m(\dot{x^{\mathbf{i}}})^2 + Q\Psi + Q\dot{x^{\mathbf{i}}}A_{\mathbf{i}} - V \\ &= \frac{1}{2}m\dot{}^{\,2} + Q\Psi(\ \ ,t) + Q\dot{} \cdot \mathbf{A}(\ \ ,t) - V(\ \ ,t). \end{aligned} \tag{3.25}$$

In other words, the formula (3.25) gives us the Lagrangian function of the problem. It is interesting to note that the momentum vector has a correction due to the presence of the electromagnetic field through the vector potential $\mathbf{A}$,

$$p_{\mathbf{i}} = \frac{\partial L}{\partial \dot{x^{\mathbf{i}}}} = \dot{y_{\mathbf{i}}} + QA_{\mathbf{i}}, \quad i = 1, 2, 3. \tag{3.26}$$

Hence the Hamiltonian function becomes

$$\begin{aligned} H &= p_{\mathbf{i}}\dot{x^{\mathbf{i}}} - L = \frac{1}{2m}\dot{y_{\mathbf{i}}}^2 - Q\Psi + V \\ &= \frac{1}{2m}(p_{\mathbf{i}} - QA_{\mathbf{i}})^2 - Q\Psi + V. \end{aligned} \tag{3.27}$$

Finally, if we use $A = (A_{\boldsymbol{\mu}})$ $(\mu = 0, 1, 2, 3)$ to denote a vector with four components, $A = (\Psi, \mathbf{A})$, the Hamiltonian function (3.27) takes the form

$$H = \frac{1}{2m}(p_{\mathbf{i}} - QA_{\mathbf{i}})^2 - QA_0 + V. \tag{3.28}$$

From the Hamiltonian (3.28) and the correspondence (2.47), we have

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = -\frac{1}{2m}(\hbar\partial_{\mathbf{i}} - \mathrm{i}QA_{\mathbf{i}})^2\psi - QA_0\psi + V\psi. \tag{3.29}$$

Thus, if we introduce the *gauge-covariant derivatives*

$$D_{\boldsymbol{\mu}}\psi = \partial_{\boldsymbol{\mu}}\psi - \mathrm{i}\frac{Q}{\hbar}A_{\boldsymbol{\mu}}\psi, \quad \mu = 0, 1, 2, 3, \tag{3.30}$$

then the *gauged Schrödinger equation* (3.29) assumes an elegant form,

$$\mathrm{i}\hbar D_0\psi = -\frac{\hbar^2}{2m}D_{\mathbf{i}}^2\psi + V\psi. \tag{3.31}$$

Note that (3.29) or (3.31) is semi-quantum mechanical in the sense that the point particle of mass m is treated quantum mechanically by the Schrödinger equation but the electromagnetic field is a classical field, through the coupling of the vector potential $A_{\boldsymbol{\mu}}$, also called the *gauge field*, which will be made more specific in the next section.

It can be examined that (3.31) in invariant under the transformation

$$\psi \mapsto \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}}\psi, \quad A_{\boldsymbol{\mu}} \mapsto A_{\boldsymbol{\mu}} + \frac{\hbar}{Q}\partial_{\boldsymbol{\mu}}\omega, \tag{3.32}$$

which is also called the *gauge transformation*, *gauge equivalence*, or *gauge symmetry*. Two gauge equivalent field configurations, $(\psi, A_{\boldsymbol{\mu}})$ and $(\psi', A'_{\boldsymbol{\mu}})$, describe identical physics.

In view of differential geometry, the above structure defines a *complex line bundle* ξ over the Minkowski spacetime $\mathbb{R}^{3,1}$ where the symmetry group is $U(1) = \{\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}} \,|\, \omega \in \mathbb{R}\}$ so that ψ is a *cross-section* and $A_{\boldsymbol{\mu}}$ a *connection* which obey the transformation property

$$\psi' = \Omega\psi, \quad A'_{\boldsymbol{\mu}} = A_{\boldsymbol{\mu}} - \mathrm{i}\frac{\hbar}{Q}\Omega^{-1}\partial_{\boldsymbol{\mu}}\Omega, \quad \Omega \in C^2(\mathbb{R}^{3,1}, U(1)). \tag{3.33}$$

It will be convenient to consider the problem in the context of such a global transformation property.

[illegible] o o c n nc o [illegible] onopo nd c [illegible] n z on

We are now prepared to study the relation between the vector potentials $\mathbf{A}^+$ and $\mathbf{A}^-$ induced from a point magnetic charge g placed at the origin.

From (3.22), we see that if we use (r, θ, φ) to denote the spherical coordinates where θ is the azimuth angle and φ the inclination angle, then

$$a_1 = 2g\partial_1\theta, \quad a_2 = 2g\partial_2\theta, \quad \theta = \tan^{-1}\left(\frac{x^2}{x^1}\right). \tag{3.34}$$

Thus, inserting $\Omega = \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}}$ into

$$A^+_{\boldsymbol{\mu}} - A^-_{\boldsymbol{\mu}} = -\mathrm{i}\frac{\hbar}{Q}\Omega^{-1}\partial_{\boldsymbol{\mu}}\Omega, \tag{3.35}$$

and assuming ω depends on θ only on $\mathbb{R}^3 \setminus (S^+ \cup S^-)$, we have, in view of (3.34), the relation

$$\frac{\hbar}{Q}\frac{\partial\omega}{\partial\theta} = 2g. \tag{3.36}$$

Integrating the above equation and using the requirement of the single-valuedness of Ω, we arrive at

$$\frac{\hbar}{Q}(2\pi n) = 2g(2\pi), \quad n \in \mathbb{N}, \tag{3.37}$$

which leads to the *Dirac charge quantization formula*

$$gQ = \frac{\hbar}{2}n, \quad n \in \mathbb{N}. \tag{3.38}$$

Consequently, when the condition (3.38) holds, the magnetic field away from a point magnetic charge g is well defined everywhere and is generated piecewise from suitable gauge potentials defined on their corresponding domains. In particular, the Dirac strings are seen to be artifacts and are removed. Consequently, like point electric charges, magnetic monopoles are also truly point magnetic charges, which will simply be referred to as monopoles from now on.

An immediate popular-science implication of the formula (3.38) is that the existence of a single monopole in the universe would predict that all electric charges are integer multiples of a basic unit charge. Indeed, this is what observed in nature since all electric charges are measured to be the multiples of the charge of the electron.

Although a monopole has never been found in nature, there are some recent heated activities among experimental physicists leading to the discovery of monopole-like structures in condensed-matter systems [15, 27, 49, 93].

The charge quantization .5591(t36(76(h))-1.55762]TJ4458(t)5605(t)-7dl2(l)-8.-o)-6(n)-214.441((()-7.11134(55

Thus, the dual basis transformation follows a reversed direction. For this reason, we say that the bases in $\mathcal{V}$ are transformed in a *covariant* way, while the bases in $\mathcal{V}'$ are transformed in a *contravariant* way.

If we express A explicitly in terms of a matrix $(A_{\mathbf{ij}})$ by

$$Au_{\mathbf{j}} = \sum_{\mathbf{i}=1}^{\mathbf{n}} A_{\mathbf{ij}} u_{\mathbf{i}}, \quad i = 1, \cdots, n, \tag{4.4}$$

then for any vector expanded with respect to the bases $\mathcal{B}$ i.e., $u = \sum_{\mathbf{i}=1}^{\mathbf{n}} x_{\mathbf{i}} u_{\mathbf{i}}$, with coordinate vector $(x_1, \cdots, x_{\mathbf{n}})$, we have

$$v = Au = \sum_{\mathbf{i}=1}^{\mathbf{n}} x_{\mathbf{i}} \left(\sum_{\mathbf{j}=1}^{\mathbf{n}} A_{\mathbf{ji}} u_{\mathbf{j}} \right) = \sum_{\mathbf{i}=1}^{\mathbf{n}} \left(\sum_{\mathbf{j}=1}^{\mathbf{n}} A_{\mathbf{ij}} x_{\mathbf{j}} \right) u_{\mathbf{i}}. \tag{4.5}$$

In other words, if we write the coordinates of v under $\mathcal{B}$ as $(y_1, \cdots, y_{\mathbf{n}})$, we have the same 'forward' relationship

$$y_{\mathbf{i}} = \sum_{\mathbf{j}=1}^{\mathbf{n}} A_{\mathbf{ij}} x_{\mathbf{j}}, \quad i = 1, \cdots, n. \tag{4.6}$$

Thus, we may call the coordinate vectors for vectors in $\mathcal{V}$ the *covariant vectors*. Likewise, for a vector $u' = \sum_{\mathbf{i}=1}^{\mathbf{n}} x'_{\mathbf{i}} u'_{\mathbf{i}}$, correspondingly, we see that $v' = (A')^{-1} u' = \sum_{\mathbf{i}=1}^{\mathbf{n}} y'_{\mathbf{i}} u'_{\mathbf{i}}$ satisfies

$$y'_{\mathbf{i}} = \sum_{\mathbf{j}=1}^{\mathbf{n}} A^{-1}_{\mathbf{ji}} x'_{\mathbf{j}} \quad \text{or} \quad x'_{\mathbf{i}} = \sum_{\mathbf{i}=1}^{\mathbf{n}} A'_{\mathbf{ij}} y'_{\mathbf{j}} = \sum_{\mathbf{i}=1}^{\mathbf{n}} A_{\mathbf{ji}} y'_{\mathbf{j}}. \tag{4.7}$$

Hence, we see that the corresponding coordinate vectors are transformed in a 'backward' manner with the transposed matrix. For this reason, the coordinate vectors for vectors in $\mathcal{V}'$ are called the *contravariant vectors*.

Simply speaking, covariant and contravariant vectors transform themselves with respect to the column and row indices of a transformation matrix, respectively. Analogously, we have covariant, contravariant, and mixed tensors, as well.

Note that, since $\mathcal{V}'' = \mathcal{V}$, the terms 'covariant' and 'contravariant' are relative and interchangeable, since the column and row indices of a matrix are interchangeable.

Now we come back to field theory. The Minkowski spacetime $\mathbb{R}^{3,1}$ is chosen to have the 4×4 metric matrix given as

$$g = (g_{\boldsymbol{\mu\nu}}) = \operatorname{diag}\{1, -1, -1, -1\}, \quad \mu, \nu = 0, 1, 2, 3. \tag{4.8}$$

The Lorentz transformation are 4×4 invertible matrices, of the form $(L^{\boldsymbol{\mu}}_{\boldsymbol{\nu}})$ (μ, ν are the row and column indices, respectively), which leave g invariant, i.e.,

$$g = L^{\mathbf{t}} g L, \quad \text{or} \quad g_{\boldsymbol{\alpha\beta}} = L^{\boldsymbol{\mu}}_{\boldsymbol{\alpha}} g_{\boldsymbol{\mu\nu}} L^{\boldsymbol{\nu}}_{\boldsymbol{\beta}} \quad (\alpha, \beta, \mu, \nu = 0, 1, 2, 3), \tag{4.9}$$

where summation convention is assumed over repeated indices. Thus $g_{\boldsymbol{\mu\nu}}$ in particular is a contravariant 2-tensor. From (4.9), we have

$$g^{-1} = (L^{-1}) g^{-1} (L^{-1})^{\mathbf{t}}. \tag{4.10}$$

Since matrix transposition switches the indices of columns and rows of a matrix, we see that g^{-1} is a covariant 2-tensor, written as $g^{\mu\nu}$, which happens to be identical to g,

$$g^{-1} = (g^{\mu\nu}) = (g_{\mu\nu}) = g. \tag{4.11}$$

Now the coordinates of the Minkowski space $\mathbb{R}^{3,1}$ is written as x^μ, which is a covariant vector, of course, which is transformed according to

$$y^\mu = L^\mu_\nu x^\nu. \tag{4.12}$$

We may use $g^{\mu\nu}$ and $g_{\mu\nu}$ to raise and lower indices, which transform contravariant and covariant quantities to their counterparts (i.e., covariant and contravariant quantities). For example,

$$x_\mu = g_{\mu\nu}x^\nu, \quad (x_\mu) = (x_0, x_1, x_2, x_3) = (x^0, -x^1, -x^2, -x^3). \tag{4.13}$$

As a consequence of (4.9), (4.12), and (4.13), we have the invariance

$$y_\mu y^\mu = g_{\mu\nu}L^\nu_\alpha x^\alpha L^\mu_\beta x^\beta = x^\alpha g_{\alpha\beta}x^\beta = x_\mu x^\mu, \tag{4.14}$$

for the contraction of any coordinate vector under the Lorentz transformations, which is essential for relativity.

The set of all Lorentz transformations, or the *Lorentz group*, is generated by the *Lorentz boosts* [6] and spatial rotations. For example, the Lorentz boost along the x^1-axis connecting two inertial coordinate frames, (x^μ) and y^μ, moving apart with relative speed v ($|v| < c$), is given by

$$y^0 = \frac{x^0 + \frac{vx^1}{c^2}}{(1 - \frac{v^2}{c^2})^{\frac{1}{2}}}, \quad y^1 = \frac{x^1 + vx^0}{(1 - \frac{v^2}{c^2})^{\frac{1}{2}}}, \quad y^2 = x^2, \quad y^3 = x^3. \tag{4.15}$$

Obviously the Lorentz group may be parametrized by six parameters, three for spatial rotations and three for the relative speeds of inertial frames with respect to the three spatial directions. Since the three speeds assume values in the open interval $(-c, c)$, the Lorentz group is *non-compact*. The transformation group consisting of the Lorentz group and all spacetime translations is called the *Poincaré group*, which is ten-dimensional, of course.

In the same fashion, if $T^{\mu\nu}$ is a covariant 2-tensor which transforms according to

$$S^{\mu\nu} = L^\mu_\alpha L^\nu_\beta T^{\alpha\beta}, \tag{4.16}$$

then we have

$$S_{\mu\nu}S^{\mu\nu} = g_{\mu\mu'}g_{\nu\nu'}S^{\mu'\nu'}S^{\mu\nu} = g_{\mu\mu'}g_{\nu\nu'}L^{\mu'}_{\alpha'}L^{\nu'}_{\beta'}T^{\alpha'\beta'}L^\mu_\alpha L^\nu_\beta T^{\alpha\beta} = T_{\mu\nu}T^{\mu\nu}, \tag{4.17}$$

which is seen again as a Lorentz invariant quantity.

Since physical laws are independent of the coordinate system of an observer and the Lorentz transformations link all allowed coordinate systems, these laws may only be expressed by those quantities which are Lorentz invariant.

It may be checked that the standard operator $\partial_\mu = \partial/\partial x^\mu$ behaves contravariantly and the differential $\mathrm{d}x^\mu$ behaves covariantly. Thus, the spacetime line element

$$\mathrm{d}s^2 = \mathrm{d}x_\mu \mathrm{d}x^\mu = (\mathrm{d}x^0)^2 - (\mathrm{d}x^1)^2 - (\mathrm{d}x^2)^2 - (\mathrm{d}x^3)^2 \tag{4.18}$$

is an invariant, which implies in particular that the speed ($c = 1$) of light remains the same in all coordinate frames related by the Lorentz transformations. If the gauge field A_μ introduced in the last section is contravariant, then the gauge-covariant derivative D_μ defined in (3.30) is also contravariant.

[6] Sometimes the Lorentz transformations are specifically referred to as the Lorentz boosts.

c d on

Recall that the Schrödinger equation for the motion of a free particle of mass m is derived from the Newtonian energy-momentum relation

$$E = \frac{\mathbf{p}^2}{2m}, \tag{4.19}$$

which is non-relativistic. In order to extend the Schrödinger equation to the relativistic realm, a direct approach is to replace the Newtonian relation (4.19) by its relativistic extension, which is quite simple to do.

In fact, recall that, according to relativity, the energy contained in a resting particle of mass m is given by the popular-science formula

$$E = mc^2, \tag{4.20}$$

where c is the speed of light in vacuum. When the particle is considered with motion, the relation (4.20), after setting $c = 1$, becomes the following energy-momentum relation

$$E^2 = \mathbf{p}^2 + m^2, \tag{4.21}$$

due to Einstein. Thus, using the quantization procedure stated in (2.47), we arrive at the equation

$$-\frac{\partial^2\psi}{\partial t^2} = -\Delta\psi + m^2\psi \quad \text{or} \quad -\Box\psi = m^2\psi, \tag{4.22}$$

governing a complex scalar field ψ, which is relativistic and commonly called the Klein–Gordon equation. Here and in the sequel, we set $\hbar = 1$ for convenience. Alternatively, we may also rewrite (4.22) as

$$-\partial_\mu\partial^\mu\psi = m^2\psi, \tag{4.23}$$

which is clearly Lorentz invariant and is the Euler–Lagrange equation of the invariant action

$$L = \int \left(\frac{1}{2}\partial_\mu\psi\overline{\partial^\mu\psi} - \frac{1}{2}m^2|\psi|^2\right) \mathrm{d}x. \tag{4.24}$$

G o nd oc sy etries

There is no harm to consider a slightly generalized Lagrange action density

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\overline{\partial^\mu\phi} - V(|\phi|^2), \tag{4.25}$$

where ϕ is a complex-valued scalar field as before. It is clear that the above Lagrangian is invariant under the *phase change* for the field ϕ,

$$\phi(x) \mapsto \mathrm{e}^{\mathrm{i}\omega}\phi(x), \tag{4.26}$$

where ω is a real constant. Such a symmetry is called a *global* symmetry because it simply says that an everywhere identical phase shift for the field ϕ does not change anything. However, when this global symmetry is *enlarged* to a *local* one for which ω becomes a function of the spacetime coordinates, $\omega = \omega(x)$, the invariance is no longer valid. A way out of this is to replace the ordinary derivatives by covariant derivatives and modify the Lagrangian into

$$\mathcal{L} = \frac{1}{2}D_\mu\phi\overline{D^\mu\phi} - V(|\phi|^2), \quad D_\mu\phi = \partial_\mu\phi - \mathrm{i}eA_\mu\phi, \tag{4.27}$$

where $e > 0$ is a coupling constant resembling an electric charge as before, and require that the vector field $A_{\boldsymbol{\mu}}$ obey the transformation rule

$$A_{\boldsymbol{\mu}} \mapsto A_{\boldsymbol{\mu}} + \frac{1}{e}\partial_{\boldsymbol{\mu}}\omega \tag{4.28}$$

and behave like a contravariant vector field under the Lorentz transformations. It can be directly checked that, under a local (x-dependent) phase shift

$$\phi(x) \mapsto \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}(\mathbf{x})}\phi(x), \tag{4.29}$$

$D_{\boldsymbol{\mu}}\phi$ changes itself covariantly according to

$$D_{\boldsymbol{\mu}}\phi \mapsto \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}(\mathbf{x})}D_{\boldsymbol{\mu}}\phi \tag{4.30}$$

so that the modified Lagrangian indeed becomes invariant under the *gauge transformation* consisting of (4.28) and (4.29).

Unfortunately, the Lagrangian (4.27) is incomplete because it cannot give rise to an equation of motion for the newly introduced gauge field $A_{\boldsymbol{\mu}}$ which is an additional dynamical variable. In order to derive a suitable dynamic law for the motion of $A_{\boldsymbol{\mu}}$, we compare $A_{\boldsymbol{\mu}}$ with ϕ and demand that the qualified Lagrangian contain quadratic terms of the first-order derivatives of $A_{\boldsymbol{\mu}}$. Since these terms should be invariant under (4.28), we see that a minimal way to do so is to introduce the contravariant 2-tensor

$$F_{\boldsymbol{\mu\nu}} = \partial_{\boldsymbol{\mu}}A_{\boldsymbol{\nu}} - \partial_{\boldsymbol{\nu}}A_{\boldsymbol{\mu}}, \quad \mu, \nu = 0, 1, 2, 3. \tag{4.31}$$

Using contraction to observe Lorentz invariance, we arrive at the minimally modified complete Lagrange action density

$$\mathcal{L} = -\frac{1}{4}F_{\boldsymbol{\mu\nu}}F^{\boldsymbol{\mu\nu}} + \frac{1}{2}D_{\boldsymbol{\mu}}\phi\overline{D^{\boldsymbol{\mu}}\phi} - V(|\phi|^2), \tag{4.32}$$

where factor $\frac{1}{4}$ is for convenience, which is invariant under the gauge and Lorentz transformations and governs the evolution of the fields ϕ and $A_{\boldsymbol{\mu}}$. The Euler–Lagrange equations of the updated Lagrangian (4.32) are

$$D_{\boldsymbol{\mu}}D^{\boldsymbol{\mu}}\phi = -2V'(|\phi|^2)\phi, \tag{4.33}$$

$$\partial_{\boldsymbol{\nu}}F^{\boldsymbol{\mu\nu}} = -J^{\boldsymbol{\mu}}, \tag{4.34}$$

where

$$J^{\boldsymbol{\mu}} = \frac{\mathrm{i}}{2}e(\overline{\phi}D^{\boldsymbol{\mu}}\phi - \phi\overline{D^{\boldsymbol{\mu}}\phi}). \tag{4.35}$$

The equation (4.33) is a gauged wave equation which extends (4.22) or (4.23). It will be instructive to recognize what is contained in the equation (4.34). To do so, we can first use (4.33) to check that $J^{\boldsymbol{\mu}}$ is a conserved '4-current' satisfying

$$\partial_{\boldsymbol{\mu}}J^{\boldsymbol{\mu}} = 0. \tag{4.36}$$

Thus, if we introduce the charge density ρ and current density $\mathbf{j}$ by setting

$$\rho = J^0 = \frac{\mathrm{i}}{2}e(\overline{\phi}D^0\phi - \phi\overline{D^0\phi}), \quad \mathbf{j} = (J^{\mathbf{i}}), \quad J^{\mathbf{i}} = \frac{\mathrm{i}}{2}e(\overline{\phi}D^{\mathbf{i}}\phi - \phi\overline{D^{\mathbf{i}}\phi}), \quad i = 1, 2, 3, \tag{4.37}$$

we have the conservation law

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0, \tag{4.38}$$

which indicates that (4.34) is to be identified with the Maxwell equations. For this purpose, we introduce the electric and magnetic fields $\mathbf{E} = (E^{\mathbf{i}})$ and $\mathbf{B} = (B^{\mathbf{i}})$ by setting

$$(F^{\boldsymbol{\mu\nu}}) = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{4.39}$$

We can check that the $\mu = 0$ component of (4.34) is simply $\nabla \cdot \mathbf{E} = \rho$, which is the Gauss law (3.1). The spatial components of (4.34), with $\mu = i = 1, 2, 3$, render the equation

$$\frac{\partial \mathbf{E}}{\partial t} + \mathbf{j} = \nabla \times \mathbf{B}, \tag{4.40}$$

which is (3.2). Hence we have partially recover the Maxwell equations (the part with charge and current sources). In order to recover the source-free part, i.e., the equations (3.3) and (3.4), we note that (4.31) implies the *Bianchi identity*

$$\partial^{\boldsymbol{\gamma}} F^{\boldsymbol{\mu\nu}} + \partial^{\boldsymbol{\mu}} F^{\boldsymbol{\nu\gamma}} + \partial^{\boldsymbol{\nu}} F^{\boldsymbol{\gamma\mu}} = 0, \tag{4.41}$$

which lead to (3.3) and (3.4). In other words, (3.3) and (3.4) automatically hold as a consequence of the definition (4.31). In summary, (4.34) is indeed the Maxwell equations.

Therefore, we have seen that *the Maxwell equations can be derived as a consequence of imposing gauge symmetry*, which is a great lesson to have.

At this spot, it is worthwhile to pause and make some comments about what we have just learned. The spacetime where the fields are defined on is called the *external space*. The scalar field may be viewed as a cross-section of a principal (complex line) bundle over the spacetime. The bundle into which the scalar field takes values in is call the *internal space*. The external space has the Lorentz group as the symmetry group which leads to relativity. The internal space has the $U(1)$ group as the symmetry group, whose local invariance or gauge invariance leads to the introduction of gauge field which generates electromagnetism. The external space symmetry is called *external symmetry* and the internal space symmetry is called *internal symmetry*. Local $U(1)$ internal symmetry requires the presence of electromagnetism. The Lorentz symmetry discussed is a global external symmetry which gives rise to special relativity. When such a global external symmetry is elevated into a local one, new fields arise. Indeed, in this situation, we are led to general relativity and presence of gravitation, as shown by Einstein. Furthermore, it is foreseeable that, when the internal symmetry is modified to be given by larger gauge groups such as $SU(N)$ ($N \geq 2$), other physical forces may be generated. In fact, this is the case and the forces generated can be weak and strong forces for nuclear interactions. This procedure exhausts all four known forces in nature: gravitational, electromagnetic, weak, and strong forces. In other words, we can draw the conclusion that *external symmetry leads to the presence of gravity and internal symmetry leads to the presence of electromagnetic, weak, and strong interactions.*

We also remark that, if the complex scalar field ϕ becomes real-valued, the charge density ρ given in (4.37) vanishes identically. In other words, a real-valued scalar field leads to an *electrically neutral* situation and a complex-valued scalar field may be used to generate electricity.

Moreover, we remark that the charge (4.37) is generated by a scalar field but the charge e in the definition of the gauge-covariant derivative is switched on 'by hand'.

Note that

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} = -\frac{1}{2}\sum_{0\le\mu<\nu\le 3} F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}\left(\sum_{i=1}^{3} F_{0i}^2 - \sum_{1\le i<j\le 3} F_{ij}^2\right) = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2). \tag{4.42}$$

We see that the terms

$$\frac{1}{2}\sum_{i=1}^{3} F_{0i}^2 = \frac{1}{2}\mathbf{E}^2, \quad \frac{1}{2}|D_0\phi|^2, \tag{4.43}$$

and

$$\frac{1}{2}\sum_{1\le i<j\le 3} F_{ij}^2 = \frac{1}{2}\mathbf{B}^2, \quad \frac{1}{2}|D_i\phi|^2, \tag{4.44}$$

in the Lagrangian (4.32) play rather different roles: The former are 'kinetic' terms which describe the dynamics of the fields and the latter are 'potential' or 'elastic' terms because they do not contain time derivatives. In particular, electric energy density is of kinetic type but magnetic energy density is of elastic or potential type.

Co p d non n yp o c nd p c on

It may be enlightening to view the equations of motion (4.33) and (4.34) as two coupled wave equations. For this purpose, we use the notation

$$\Box_D = D_\mu D^\mu, \tag{4.45}$$

to denote the gauged D'Lambertian operator with respect to the gauge-covariant derivatives. Hence $\Box_\partial = \Box$ is the standard one and we see that (4.33) and (4.34) become

$$\Box_D \phi = -2V'(|\phi|^2)\phi, \tag{4.46}$$

$$\Box_\partial A^\mu = J^\mu + \partial^\mu(\partial_\nu A^\nu). \tag{4.47}$$

We see that, if we may impose the so-called *Lorentz gauge* condition

$$\partial_\mu A^\mu = 0, \tag{4.48}$$

the equation (4.47) gives rise to non-homogeneous wave equations governing A^μ ($\mu = 0, 1, 2, 3$).

It is also interesting to consider the static situation when the fields are independent of the time variable $x^0 = t$. In this situation, it is consistent to impose the so-called *temporal gauge* condition

$$A_0 = 0. \tag{4.49}$$

With $\mathbf{A} = (A_1, A_2, A_3)$ and

$$D_\mathbf{A}\phi = \nabla\phi - ie\mathbf{A}\phi, \quad D_i D^i\phi = -D_\mathbf{A}^2\phi = -\Delta_\mathbf{A}\phi, \tag{4.50}$$

we have

$$\Delta_\mathbf{A}\phi = \Delta\phi - 2ie\mathbf{A}\cdot\nabla\phi - ie(\nabla\cdot\mathbf{A})\phi - e^2|\mathbf{A}|^2\phi. \tag{4.51}$$

Thus, (4.46) and (4.47) become a system of nonlinear equations

$$\Delta\phi = 2ie\mathbf{A}\cdot\nabla\phi + ie(\nabla\cdot\mathbf{A})\phi + e^2|\mathbf{A}|^2\phi + 2V'(|\phi|^2)\phi, \tag{4.52}$$

$$\Delta\mathbf{A} = \nabla(\nabla\cdot\mathbf{A}) + \frac{i}{2}(\phi\overline{D_{\mathbf{A}}\phi} - \overline{\phi}D_{\mathbf{A}}\phi), \tag{4.53}$$

which are strictly elliptic and considerably simplified when $\mathbf{A}$ satisfies the condition

$$\nabla\cdot\mathbf{A} = 0, \tag{4.54}$$

known as the *Coulomb gauge* condition. It may be useful to note that the equations (4.52) and (4.53) are the Euler–Lagrange equations of the energy functional

$$E(\phi,\mathbf{A}) = \int\left\{\frac{1}{2}|\nabla\times\mathbf{A}|^2 + \frac{1}{2}|D_{\mathbf{A}}\phi|^2 + V(|\phi|^2)\right\}\,dx. \tag{4.55}$$

The equations (4.52) and (4.53) and the energy (4.55) are actually known as the *Ginzburg–Landau equations* and the *Ginzburg–Landau energy* arising in the *theory of superconductivity*, which will be discussed in more detail later in the notes.

A n r od

Although the Abelian Higgs model originated from particle physics, it has broad applications in solid-state or condensed-matter physics and supplies as a profound source of important ideas. We begin by considering the static solutions of the model with temperature dependence.

Te per ture dependent potenti function

The simplest situation is that the potential density V is of the form

$$V(|\phi|^2) = \frac{1}{2}m^2(T)|\phi|^2 + \frac{1}{8}\lambda|\phi|^4, \tag{4.56}$$

where m^2 is a function of the temperature T, which is typically of the form

$$m^2(T) = a\left(\left[\frac{T}{T_{\mathbf{c}}}\right]^2 - 1\right), \tag{4.57}$$

$\lambda, a > 0$ are suitable parameters, and $T_{\mathbf{c}} > 0$ is a critical temperature.

Vacuum solutions, or ground states, are the lowest energy static solutions. In high temperature, $T > T_{\mathbf{c}}$, we have $m^2(T) > 0$ and, in view of the Klein–Gordon equation, the quantity $m(T) > 0$ is the mass of two real scalar particles represented by ϕ_1 and ϕ_2 where $\phi = \phi_1 + i\phi_2$ and higher order terms of ϕ_1 and ϕ_2 describe self-interactions. The only minimum of the Hamiltonian density

$$\mathcal{H} = \frac{1}{2}|\nabla\phi|^2 + \frac{1}{2}m^2(T)|\phi|^2 + \frac{1}{8}\lambda|\phi|^4 \tag{4.58}$$

is

$$\phi_{\mathrm{V}} = 0, \tag{4.59}$$

which is the unique vacuum state of the problem. This vacuum state is of course invariant under the $U(1)$-symmetry group $\phi \mapsto e^{i\omega}\phi$.

Sy etry nd sy etry re ing

In general, given a symmetry group, the Lagrangian density should be invariant if the vacuum state is already invariant, based on some consideration from quantum field theory. Such a statement is known as the *Coleman theorem* (the invariance of the vacuum state implies the invariance of the universe). If both the vacuum state and the Lagrangian density are invariant, we say that there is *exact symmetry*. If the vacuum state is non-invariant, the Lagrangian density may be non-invariant or invariant. In both cases, we say that the symmetry as a whole is *broken*. The former case is referred to as *explicit symmetry-breaking* and the latter case is referred to as *spontaneous symmetry-breaking*, which is one of the fundamental phenomena in low-temperature physics.

To explain some of the above concepts, we assume that $T < T_{\mathbf{c}}$. We have $m^2(T) < 0$ and we see that there is a phase transition: although $\phi = 0$ is still a solution but it is no longer stable. In fact the minimum of $\mathcal{H}$ is attained instead at any of the configurations

$$\phi_{\mathrm{v},\boldsymbol{\theta}} = \phi_0 \mathrm{e}^{\mathrm{i}\boldsymbol{\theta}}, \quad \phi_0 = \sqrt{\left(1 - \left[\frac{T}{T_{\mathbf{c}}}\right]^2\right)\frac{2a}{\lambda}} > 0, \quad \theta \in \mathbb{R}, \tag{4.60}$$

which give us a continuous family of distinct vacuum states (a circle) labeled by θ. Since for $\omega \neq 2k\pi$

$$\phi \mapsto \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}}\phi \tag{4.61}$$

transforms any given vacuum state, $\phi_{\mathrm{v},\boldsymbol{\theta}}$, to a different one, $\phi_{\mathrm{v},\boldsymbol{\theta}+\boldsymbol{\omega}}$, we observe the non-invariance of vacuum states although the Lagrangian density is still invariant. Consequently, the symmetry is spontaneously broken. The quantity ϕ_0 measures the *scale of the broken symmetry*.

Illustration. Consider what happens when exerting pressure on an upright stick.

Go dstone p rtic es nd Higgs [redacted] ech nis[redacted]

We continue to consider the system at low temperature, $T < T_{\mathbf{c}}$. Since $m^2(T) < 0$, it seems that we would have particles of imaginary mass. However, this is not the case as will be seen below.

In fact, the fore-discussed Lagrangian density governs fluctuations around vacuum state. For $T > T_{\mathbf{c}}$, the vacuum state is the zero state and $m(T)$ clearly defines mass. For $T < T_{\mathbf{c}}$, we need to consider fluctuations around a given nonzero vacuum state, say ϕ_0, represented by two real scalar functions ϕ_1 and ϕ_2,

$$\phi(x) = \phi_0 + \phi_1(x) + \mathrm{i}\phi_2(x). \tag{4.62}$$

In this case, the minimum of V is strictly negative,

$$V(\phi_0^2) = \frac{1}{2}m^2(T)\phi_0^2 + \frac{1}{8}\lambda\phi_0^4 = -\frac{1}{8}\lambda\phi_0^4. \tag{4.63}$$

Thus, in order to maintain finite energy in an unbounded space, we need to shift the potential energy density by the quantity given above,

$$V \mapsto V + \frac{1}{8}\lambda\phi_0^4 = \frac{\lambda}{8}(|\phi|^2 - \phi_0^2)^2, \tag{4.64}$$

and the new minimum energy level is zero. Using the updated potential function, we have the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial_{\boldsymbol{\mu}}\phi_1\partial^{\boldsymbol{\mu}}\phi_1 + \frac{1}{2}\partial_{\boldsymbol{\mu}}\phi_2\partial^{\boldsymbol{\mu}}\phi_2 - \frac{\lambda}{2}\phi_0^2\phi_1^2 - \frac{\lambda}{8}(\phi_1^4 + \phi_2^4 + 2\phi_1^2\phi_2^2 + 4\phi_0\phi_1^3 + 4\phi_0\phi_1\phi_2^2), \tag{4.65}$$

which governs the scalar fields ϕ_1 and ϕ_2 fluctuating around the vacuum state, $\phi_1 = 0, \phi_2 = 0$. The coefficient of ϕ_1^2 defines the mass of the ϕ_1-particles,

$$m_1 = \sqrt{\lambda}\phi_0 = \sqrt{2}|m(T)| > 0. \tag{4.66}$$

However, since the ϕ_2^2 term is absent (the higher=order terms describe interactions), these ϕ_2-particles are massless and are called the *Goldstone particles*. Hence, we see that spontaneous symmetry-breaking leads to the presence of the Goldstone particles, namely, particles of zero mass instead of particles of imaginary mass. This statement is known as the *Goldstone theorem*.

The Goldstone particles are massless, and hence, are curious. We see in the following that these particles may be removed from the system when gauge fields are switched on. For this purpose, we return to the locally invariant Lagrangian density to get

$$\mathcal{L} = -\frac{1}{4}F_{\boldsymbol{\mu\nu}}F^{\boldsymbol{\mu\nu}} + \frac{1}{2}D_{\boldsymbol{\mu}}\phi\overline{D^{\boldsymbol{\mu}}\phi} - \frac{\lambda}{8}(|\phi|^2 - \phi_0^2)^2, \tag{4.67}$$

and we consider fluctuations around the vacuum state

$$\phi_{\rm V} = \phi_0, \quad (A_{\boldsymbol{\mu}})_{\rm V} = 0. \tag{4.68}$$

Using the earlier decomposition of ϕ into ϕ_0, ϕ_1, ϕ_2, we obtain the Lagrangian density for the interaction of the fields ϕ_1, ϕ_2, and $A_{\boldsymbol{\mu}}$ as follows,

$$\mathcal{L} = -\frac{1}{4}F_{\boldsymbol{\mu\nu}}F^{\boldsymbol{\mu\nu}} + \frac{1}{2}\partial_{\boldsymbol{\mu}}\phi_1\partial^{\boldsymbol{\mu}}\phi_1 + \frac{1}{2}\partial_{\boldsymbol{\mu}}\phi_2\partial^{\boldsymbol{\mu}}\phi_2 + \frac{1}{2}e^2\phi_0^2A_{\boldsymbol{\mu}}A^{\boldsymbol{\mu}} - \frac{\lambda}{2}\phi_0^2\phi_1^2 + \mathcal{L}_{\rm inter}, \tag{4.69}$$

where $\mathcal{L}_{\rm inter}$ contains all off-diagonal interaction terms involving the mixed products of the fields $\phi_1, \phi_2, A_{\boldsymbol{\mu}}$, and their derivatives. Recall that the definition of the gauge transformation. Hence, $\phi_1, \phi_2, A_{\boldsymbol{\mu}}$ transform themselves according to the rule

$$\begin{aligned}
\phi_0 + \phi_1 + \mathrm{i}\phi_2 &\mapsto \phi_0 + \phi_1' + \mathrm{i}\phi_2', \quad A_{\boldsymbol{\mu}} \mapsto A_{\boldsymbol{\mu}}', \\
\phi_1' &= \phi_1\cos\omega - \phi_2\sin\omega + \phi_0(\cos\omega - 1), \\
\phi_2' &= \phi_1\sin\omega + \phi_2\cos\omega + \phi_0\sin\omega, \\
A_{\boldsymbol{\mu}}' &= A_{\boldsymbol{\mu}} + \frac{1}{e}\partial_{\boldsymbol{\mu}}\omega.
\end{aligned}$$

From the Lagrangian density, we see that ϕ_2 remains massless. Besides, the gauge field $A_{\boldsymbol{\mu}}$ becomes massive (a mass of $e\phi_0$). However, using the above expression, we can find a suitable gauge transformation so that $\phi_2' = 0$. For example, we may choose

$$\omega = -\arctan\frac{\phi_2}{\phi_0 + \phi_1}. \tag{4.70}$$

If we use the phase function ω determined above in the transformation and the new field variables $\phi_1', A_{\boldsymbol{\mu}}'$, and suppress the prime sign $'$, we see that the Lagrangian becomes

$$\mathcal{L} = -\frac{1}{4}F_{\boldsymbol{\mu\nu}}F^{\boldsymbol{\mu\nu}} + \frac{1}{2}\partial_{\boldsymbol{\mu}}\phi_1\partial^{\boldsymbol{\mu}}\phi_1 + \frac{1}{2}e^2\phi_0^2A_{\boldsymbol{\mu}}A^{\boldsymbol{\mu}} - \frac{\lambda}{2}\phi_0^2\phi_1^2 + \mathcal{L}_{\rm inter}, \tag{4.71}$$

where $\mathcal{L}_{\rm inter}$ contains all off-diagonal interaction terms involving the mixed products of the fields $\phi_1, A_{\boldsymbol{\mu}}$, and their derivatives. Thus we see that, in such a fixed gauge, we are left with a massive

real scalar field and a massive gauge field and the massless Goldstone particle is eliminated. In other words, spontaneous breaking of a continuous symmetry does not lead to the appearance of a massless Goldstone particle but to the disappearance of a scalar field and the appearance of a massive gauge field. This statement is known as the *Higgs mechanism*[7] and the massive scalar particles are called the *Higgs particles*. In particular, the Lagrangian density we have seen is commonly referred to as the *Abelian Higgs model* and the complex scalar field ϕ is called the *Higgs field*.

nz - L nd on o p cond c y

In this section, we consider the Ginzburg–Landau equations for superconductivity and Abrikosov's vortices and their topological characterizations. We will end with an excursion to the monopole and quark confinement problem.

In the static situation,

$$\partial_0 = 0, A_0 = 0, \tag{5.1}$$

the Abelian Higgs model is the well-known *Ginzburg–Landau theory* for *superconductivity*. The complex scalar field ϕ gives rise to density distribution of superconducting electron pairs known as the *Cooper pairs* and the fact that the electromagnetic field behaves like a massive field due to the Higgs mechanism is simply a consequence of the *Meissner effect* which is equivalent to saying that the magnetic field becomes massive and cannot penetrate a superconductor.

For simplicity, we focus here on the two-dimensional situation. This is the most interesting situation because it allows us to consider the Abrikosov [1] or Nielsen–Olesen [99] vortices.

c p oo o M n c

We start with a 'proof' of the Meissner effect which states that a superconductor screens an external magnetic field when this magnetic field is not strong enough to destroy its superconducting phase and that the superconducting phase in the superconductor may be switched to the normal-conducting phase permitting the full penetration of the external magnetic field when this magnetic field is made strong enough. For this purpose, we recall that the energy density or Hamiltonian of the two-dimensional Ginzburg–Landau theory in the presence of a constant external magnetic field, $H_{\rm ext} > 0$ (say), is given by

$$\mathcal{H}(\phi, A_{\mathbf{j}}) = \frac{1}{2}F_{12}^2 + \frac{1}{2}|D_1\phi|^2 + \frac{1}{2}|D_2\phi|^2 + \frac{\lambda}{8}(|\phi|^2 - \phi_0^2)^2 - F_{12}H_{\rm ext}. \tag{5.2}$$

The normal and completely superconducting phases are represented by $(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n})$ and $(\phi^{\rm s}, A_{\mathbf{j}}^{\rm s})$, respectively, so that

$$\phi^{\rm n} = 0, \quad F_{12}^{\rm n} = H_{\rm ext}; \quad \phi^{\rm s} = \phi_0, \quad A_{\mathbf{j}}^{\rm s} = 0, \quad j = 1, 2. \tag{5.3}$$

Therefore, we have

$$\mathcal{H}(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n}) = \frac{\lambda}{8}\phi_0^4 - \frac{1}{2}H_{\rm ext}^2, \quad \mathcal{H}(\phi^{\rm s}, A_{\mathbf{j}}^{\rm s}) = 0. \tag{5.4}$$

[7]This fundamental mechanism in fact was due to the independent work by three different groups of people: Englert and Brout [40], Higgs [58], and Guralnik, Hagen, and Kibble [57], published in the same journal and at the same time.

Consequently, when $H_{\rm ext}$ satisfies

$$H_{\rm ext} < \frac{1}{2}\phi_0^2\sqrt{\lambda}, \tag{5.5}$$

we have $\mathcal{H}(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n}) > \mathcal{H}(\phi^{\rm S}, A_{\mathbf{j}}^{\rm S})$ and $(\phi^{\rm S}, A_{\mathbf{j}}^{\rm S})$ is *energetically favored* over $(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n})$. Thus the superconductor is in the superconducting phase described by $(\phi^{\rm S}, A_{\mathbf{j}}^{\rm S})$ and the magnetic field is completely expelled from the superconductor, $F_{12}^{\rm S} = 0$. On the other hand, when

$$H_{\rm ext} > \frac{1}{2}\phi_0^2\sqrt{\lambda}, \tag{5.6}$$

we have $\mathcal{H}(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n}) < \mathcal{H}(\phi^{\rm S}, A_{\mathbf{j}}^{\rm S})$ and $(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n})$ is energetically favored over $(\phi^{\rm S}, A_{\mathbf{j}}^{\rm S})$. Thus the superconductor is in the normal phase described by $(\phi^{\rm n}, A_{\mathbf{j}}^{\rm n})$ and the externally applied magnetic field penetrates the superconductor completely, $F_{12}^{\rm n} = H_{\rm ext}$. Such a picture has fully depicted the Meissner effect stated earlier.

n y p on n z on nd opo o c p op

Below, we will neglect the external magnetic field and set $\phi_0 = 1, e = 1$ for convenience. The Euler–Lagrange equations, or the static Abelian Higgs equations (also known as the Ginzburg–Landau equations) are

$$D_{\mathbf{k}}D_{\mathbf{k}}\phi = \frac{\lambda}{2}(|\phi|^2 - 1)\phi, \tag{5.7}$$

$$\partial_{\mathbf{k}}F_{\mathbf{jk}} = \frac{\rm i}{2}(\phi\overline{D_{\mathbf{j}}\phi} - \overline{\phi}D_{\mathbf{j}}\phi), \tag{5.8}$$

where $j, k = 1, 2$. Since the total energy is

$$E(\phi, A_{\mathbf{j}}) = \int_{\mathbb{R}^2}\left(\frac{1}{2}F_{12}^2 + \frac{1}{2}|D_1\phi|^2 + \frac{1}{2}|D_2\phi|^2 + \frac{\lambda}{8}(|\phi|^2 - 1)^2\right), \tag{5.9}$$

the finite-energy condition implies that a solution (ϕ, A) must satisfy the boundary condition

$$F_{12} \to 0, \quad |D_{\mathbf{j}}\phi| \to 0, \quad |\phi| \to 1 \quad \text{as } |x| \to \infty. \tag{5.10}$$

In fact, it can be shown that the above decay may be achieved exponentially fast.

Note that the magnetic field F_{12} is also viewed as the vorticity field if the vector field $A = (A_1, A_2)$ is viewed as the velocity field in a two-dimensional fluid. Therefore, wherever $F_{12} \neq 0$, nontrivial vorticity is present. For this reason, nontrivial solutions of two-dimensional gauge field equations are also called 'vortices' or 'vortex-lines'.

P rtition identity nd consequence

Exploring the conformal structure of the problem, we arrive at

$$4\int_{\mathbb{R}^2} F_{12}^2\,{\rm d}x = \lambda\int_{\mathbb{R}^2}(|\phi|^2 - 1)^2\,{\rm d}x. \tag{5.11}$$

This identity says that there holds an exact partition between the magnetic energy and the potential energy of the Higgs particle. In particular, the absence of the magnetic field, $F_{12} \equiv 0$ implies the triviality of the Higgs scalar field, $|\phi| \equiv 1$, and vice versa. In particular, the equations do not allow reduction to

$$A_1 = A_2 \equiv 0, \quad \Delta\phi = \frac{\lambda}{2}(|\phi|^2 - 1)\phi, \quad x \in \mathbb{R}^2, \tag{5.12}$$

among finite-energy solutions. Such a property may be regarded as a demonstration of the Meissner effect from a different angle: A finite-energy solution has a nontrivial scalar field sector if and only of the solution has a nontrivial gauge field sector. In the following, we will present such solutions.

F u qu ntiz tion nd topo ogic inv ri nts

Since $|\phi(x)| \to 1$ as $|x| \to \infty$, we see that

$$\Gamma = \frac{\phi}{|\phi|} : \quad S^1_{\mathbf{R}} \to S^1 \tag{5.13}$$

is well defined when $R > 0$ is large enough, where $S^1_{\mathbf{R}}$ denotes the circle in $\mathbb{R}^2$ centered at the origin and of radius R. Therefore Γ may be viewed as an element in the fundamental group

$$\pi_1(S^1) = \mathbb{Z} \tag{5.14}$$

and represented by an integer N. In fact, this integer N is the winding number of ϕ around $S^1_{\mathbf{R}}$ and may be expressed by the integral

$$N = \frac{1}{2\pi \mathrm{i}} \int_{\mathbf{S}^1_R} \mathrm{d} \ln \phi. \tag{5.15}$$

It is interesting to note that the continuous dependence of the right-hand side of the above relation with respect to R implies that it is actually independent of R. An important consequence of such an observation is the famous *flux quantization condition*

$$\Phi = \int_{\mathbb{R}^2} F_{12} \, \mathrm{d}x = 2\pi N \tag{5.16}$$

which follows from

$$\begin{aligned} \left| \int_{|\mathbf{x}| \le \mathbf{R}} F_{12} \, \mathrm{d}x + \mathrm{i} \int_{|\mathbf{x}| = \mathbf{R}} \mathrm{d} \ln \phi \right| &= \left| \int_{|\mathbf{x}| = \mathbf{R}} A_{\mathbf{j}} \, \mathrm{d}x_{\mathbf{j}} + \mathrm{i} \int_{|\mathbf{x}| = \mathbf{R}} \phi^{-1} \partial_{\mathbf{j}} \phi \, \mathrm{d}x_{\mathbf{j}} \right| \\ &\le \int_{|\mathbf{x}| = \mathbf{R}} |\phi^{-1}| |D_{\mathbf{A}} \phi| \, \mathrm{d}s \\ &\le C \mathrm{e}^{-\boldsymbol{\delta} \mathbf{R}} \int_{|\mathbf{x}| = \mathbf{R}} \mathrm{d}s \\ &= 2\pi R C \mathrm{e}^{-\boldsymbol{\delta} \mathbf{R}} \to 0 \quad \text{as } R \to \infty. \end{aligned} \tag{5.17}$$

Note that, when the theory is formulated in the language of a complex line bundle say ξ, so that ϕ is a cross section, A is a connection 1-form, $F = \mathrm{d}A$ is the curvature, and $D_{\mathbf{A}}$ is the bundle connection, then the integer N is nothing but the first Chern class $c_1(\xi)$, which completely classifies the line bundle up to an isomorphism. That is,

$$\frac{\Phi}{2\pi} = N = c_1(\xi). \tag{5.18}$$

An important open question is whether for any given $N \in \mathbb{Z}$ there is a solution to the constrained minimization problem

$$E_{\mathbf{N}} \equiv \inf \left\{ E(\phi, A) \,\middle|\, \int_{\mathbb{R}^2} F_{12} \, \mathrm{d}x = 2\pi N \right\}. \tag{5.19}$$

The problem is known to be solvable only when $\lambda = 1$ due to Taubes [66, 138, 139].

o , n o on nd p c

In the theory of superconductivity, the complex scalar field is an *'order' parameter* which characterizes the two phases, superconducting and normal states, of a solid. Mathematically, $|\phi|^2$ gives rise to the density of superconducting electron pairs, also called the Cooper pairs, so that $\phi \neq 0$ indicates the presence of electron pairs and onset of superconductivity and $\phi = 0$ indicates the absence of electron pairs and the dominance of normal state. As a consequence, when the order parameter ϕ is such that it is nonvanishing somewhere but vanishing elsewhere, we are then having a '*mixed state*'. Recall that, according to the Meissner effect, a superconductor screens the magnetic field. In other words, the presence of superconductivity prevents the penetration of a magnetic field. Therefore, in a mixed state, the magnetic field, F_{12}, always has its maximum penetration at the spots where $\phi = 0$. Or equivalently, $|F_{12}|$ assumes its local maximum values at the zeros of ϕ. Since F_{12} may be interpreted as a vorticity field, the zeros of ϕ give rise to centers of vortices or locations of vortex-lines distributed over $\mathbb{R}^2$. In fact, although it is not obvious, it may be checked that $|D_1\phi|^2 + |D_2\phi|^2$ also attains its local maxima at the zeros of ϕ. Hence, we have seen that the energy density

$$\mathcal{H}(\phi, A)(x) = \left(\frac{1}{2}F_{12}^2 + \frac{1}{2}|D_1\phi|^2 + \frac{1}{2}|D_2\phi|^2 + \frac{\lambda}{8}(|\phi|^2 - 1)^2\right)(x) \tag{5.20}$$

attains its local maxima at the zeros of ϕ as well. However, since energy and mass are equivalent, we have observed *mass concentration* centered at the zeros of ϕ. In other words, we have produced a distribution of *solitons* which may be identified with *'particles'* in quantum field theory.

We now show that the presence of zeros of ϕ is essential for a solution to be nontrivial: *If (ϕ, A) is a finite-energy solution of the Abelian Higgs equations so that ϕ never vanishes, then (ϕ, A) is gauge-equivalent to the trivial solution $A \equiv 0, \phi \equiv 1$.*

Here is a quick proof.

Since ϕ never vanishes, we may rewrite ϕ as

$$\phi = \varphi e^{i\omega} \tag{5.21}$$

for globally defined real-valued smooth functions φ and ω over $\mathbb{R}^2$. In fact, we may assume $\varphi > 0$. Using the gauge transformation

$$\phi \mapsto \phi e^{-i\omega}, \quad A_j \mapsto A_j - \partial_j \omega, \tag{5.22}$$

we see that (ϕ, A) becomes 'unitary' in the sense that ϕ is real valued,

$$\phi = \varphi, \tag{5.23}$$

and we say that we have chosen a 'unitary gauge.' In unitary gauge, the equations of motion decompose significantly to take the form

$$\begin{aligned}
\Delta\varphi &= |A|^2\varphi + \frac{\lambda}{2}(\varphi^2 - 1)\varphi, \\
2A_k\partial_k\varphi + (\partial_k A_k)\varphi &= 0, \\
\partial_1 F_{12} &= \varphi^2 A_2, \\
\partial_2 F_{12} &= -\varphi^2 A_1
\end{aligned}$$

From the last two equations, we have

$$\int_{\mathbb{R}^2} (A_2\partial_1 F_{12} - A_1\partial_2 F_{12})\,\mathrm{d}x = \int_{\mathbb{R}^2} |A|^2\varphi^2\,\mathrm{d}x. \tag{5.24}$$

Integrating further by parts and dropping the boundary terms, we obtain

$$\int_{\mathbb{R}^2} (F_{12}^2 + |A|^2\varphi^2)\,\mathrm{d}x = 0. \tag{5.25}$$

Since φ never vanishes, we have $A \equiv 0$. Returning to the governing equations, we arrive at the single remaining equation

$$\Delta\varphi = \frac{\lambda}{2}\varphi(\varphi^2 - 1), \tag{5.26}$$

which may be rewritten as

$$\Delta(\varphi - 1) = \frac{\lambda}{2}\varphi(\varphi + 1)(\varphi - 1). \tag{5.27}$$

Using the boundary condition $\varphi - 1 \to 0$ as $|x| \to \infty$ and the maximum principle, we deduce $\varphi \equiv 1$ as claimed.

Estimate of energy from the topological lower bound

A useful identity involving gauge-covariant derivatives is

$$|D_{\mathbf{j}}\phi|^2 = |D_1\phi \pm \mathrm{i}D_2\phi|^2 \pm F_{12}|\phi|^2 \pm \mathrm{i}(\partial_1[\phi\overline{D_2\phi}] - \partial_2[\phi\overline{D_1\phi}]). \tag{5.28}$$

Therefore the total energy satisfies

$$\begin{aligned} E(\phi, A) &\geq \min\{\lambda, 1\} \int_{\mathbb{R}^2} \left(\frac{1}{2}F_{12}^2 + \frac{1}{2}|D_{\mathbf{j}}\phi|^2 + \frac{1}{8}(|\phi|^2 - 1)^2\right)\mathrm{d}x \\ &= \min\{\lambda, 1\} \int_{\mathbb{R}^2} \left(\frac{1}{2}\left|F_{12} \pm \frac{1}{2}(|\phi|^2 - 1)\right| + \frac{1}{2}|D_1\phi \pm \mathrm{i}D_2\phi|^2 \pm \frac{1}{2}F_{12}\right)\mathrm{d}x \\ &\geq \min\{\lambda, 1\}\pi|N|, \end{aligned} \tag{5.29}$$

where $\Phi = 2\pi N$ and the signs $\pm$ follow $N = \pm|N|$. In the sequel, we focus on $N \geq 0$ for convenience. When $\lambda = 1$, we have

$$E(\phi, A) \geq \pi N \tag{5.30}$$

and such an energy lower bound is saturated if and only if (ϕ, A) satisfies the *self-dual* system of equations

$$D_1\phi + \mathrm{i}D_2\phi = 0, \tag{5.31}$$

$$F_{12} = \frac{1}{2}(1 - |\phi|^2), \tag{5.32}$$

which is a reduction of the original equations of motion and is also often called a *BPS system* after Bogomol'nyi [10] and Prasad and Sommerfield [113] who first derived these equations.

Structure of BPS system

It will be convenient to complexify our variables and use

$$A = A_1 + \mathrm{i}A_2, \quad A_1 = \frac{1}{2}(A + \overline{A}), \quad A_2 = \frac{1}{2\mathrm{i}}(A - \overline{A}), \quad z = x^1 + \mathrm{i}x^2, \tag{5.33}$$

$$\partial = \frac{1}{2}(\partial_1 - \mathrm{i}\partial_2), \quad \overline{\partial} = \frac{1}{2}(\partial_1 + \mathrm{i}\partial_2), \quad \partial_1 = \partial + \overline{\partial}, \quad \partial_2 = \mathrm{i}(\partial - \overline{\partial}), \tag{5.34}$$

$$\Delta = \partial_1^2 + \partial_2^2 = 4\partial\overline{\partial} = 4\overline{\partial}\partial. \tag{5.35}$$

Hence the first equation in the BPS system assumes the form

$$\overline{\partial}\phi = \frac{\mathrm{i}}{2} A\phi. \tag{5.36}$$

To see what this relation means, we recall the $\overline{\partial}$-Poincaré lemma which states that the equation

$$\overline{\partial}\omega(z) = \mathrm{i}\alpha(z) \tag{5.37}$$

over a disk $B \subset \mathbb{C}$ always has a solution. In fact, this solution may be represented by [66]

$$\omega(z) = \frac{1}{2\pi} \int_{\mathbf{B}} \frac{\alpha(z')}{z' - z}\, \mathrm{d}z' \wedge \mathrm{d}\overline{z'}. \tag{5.38}$$

Now let ψ solve $\overline{\partial}\psi = \frac{\mathrm{i}}{2} A$ locally. Then we see that the complex-valued function $f = \phi \mathrm{e}^{-\boldsymbol{\Psi}}$ satisfies the Cauchy–Riemann equation

$$\overline{\partial} f = \overline{\partial}(\phi \mathrm{e}^{-\boldsymbol{\Psi}}) = \mathrm{e}^{-\boldsymbol{\Psi}}(\overline{\partial}\phi - \phi\overline{\partial}\psi) = 0. \tag{5.39}$$

Therefore $f(z)$ is analytic. In particular, f (and hence ϕ) may only have isolated zeros with integer multiplicities. In other words, if z_0 is a zero of ϕ, then

$$\phi(z) = (z - z_0)^{\mathbf{n}} h(z) \tag{5.40}$$

for z near z_0 and the function $h(z)$ never vanishes, where n is a positive integer which is also the local winding number of ϕ around z_0. From the second equation of the BPS system, we see clearly that the vorticity field F_{12} achieves it maximum value at z_0 as well,

$$\max\{F_{12}\} = F_{12}(z_0) = \frac{1}{2} \tag{5.41}$$

so that the point z_0 defines the center of a (magnetic) vortex. The integer n is also called the local vortex charge. Besides, since $|\phi| \to 1$ as $|x| \to \infty$, we see that ϕ can only have a finite number of zeros over $\mathbb{C}$. Assume that the zeros of ϕ and their respective multiplicities are

$$z_1, n_1, \quad z_2, n_2, \quad \cdots, \quad z_{\mathbf{k}}, n_{\mathbf{k}}. \tag{5.42}$$

Counting multiplicities of these zeros (i.e., a zero of multiplicity m is counted as m zeros), the total vortex charge is the total number of zeros of ϕ, say $N(\phi)$,

$$N(\phi) = \sum_{\mathbf{s}=1}^{\mathbf{k}} n_{\mathbf{s}}. \tag{5.43}$$

On the other hand, away from the zeros of ϕ, the first equation in the BPS system is

$$A = -\mathrm{i}2\overline{\partial} \ln \phi. \tag{5.44}$$

Therefore, there, we can represent F_{12} as

$$F_{12} = \partial_1 A_2 - \partial_2 A_1 = -\mathrm{i}(\partial A - \overline{\partial A}) = -2\partial\overline{\partial} \ln |\phi|^2 = -\frac{1}{2}\Delta \ln |\phi|^2. \tag{5.45}$$

Inserting the above into the second equation in the BPS system, we have

$$\Delta \ln |\phi|^2 + 1 - |\phi|^2 = 0 \quad \text{(away from the zeros of } \phi\text{)}. \tag{5.46}$$

Now define

$$u = \ln |\phi|^2. \tag{5.47}$$

Then, near $z_{\mathbf{s}}$ $(s = 1, 2, \cdots, k)$, we have

$$u(z) = 2n_{\mathbf{s}} \ln |z - z_{\mathbf{s}}| + \text{ a regular term}. \tag{5.48}$$

Consequently, we arrive at the Liouville type equation

$$\Delta u = \mathrm{e}^{\mathbf{u}} - 1 + 4\pi \sum_{\mathbf{s}=1}^{\mathbf{k}} n_{\mathbf{s}} \delta(z - z_{\mathbf{s}}) \quad \text{ in } \mathbb{C} = \mathbb{R}^2 \tag{5.49}$$

subject to the boundary condition

$$u \to 0 \quad \text{as } |x| = |z| \to \infty \tag{5.50}$$

(since $|\phi| \to 1$ as $|x| \to \infty$), which may be solved by various techniques.

Conversely, for any given data $\{(z_{\mathbf{s}}, n_{\mathbf{s}})\}$, the solution of the above elliptic equation gives rise to a solution pair (ϕ, A) which represents multiply distributed vortices at $\{z_{\mathbf{s}}\}$ with the corresponding local vortex charges $\{n_{\mathbf{s}}\}$.

Formally, it may be easily convinced that

$$\int_{\mathbb{R}^2} \Delta u \, \mathrm{d}x = 0. \tag{5.51}$$

Hence, from the equation, we obtain

$$\int_{\mathbb{R}^2} (1 - |\phi|^2) \, \mathrm{d}x = \int_{\mathbb{R}^2} (1 - \mathrm{e}^{\mathbf{u}}) \, \mathrm{d}x = 4\pi \sum_{\mathbf{s}=1}^{\mathbf{k}} n_{\mathbf{s}} = 4\pi N(\phi). \tag{5.52}$$

Integrating the second equation in the BPS system and inserting the above relation, we get the beautiful result

$$N = c_1(\xi) = \frac{1}{2\pi} \int_{\mathbb{R}^2} F_{12} \, \mathrm{d}x = \frac{1}{4\pi} \int_{\mathbb{R}^2} (1 - |\phi|^2) \, \mathrm{d}x = N(\phi). \tag{5.53}$$

In other words, the total vortex number is nothing but the first Chern class of the solution we have seen before, which also determines the total magnetic flux, $\Phi = 2\pi N$.

Let us record the important conclusion that our solution carries the minimum energy,

$$E_{\mathbf{N}} = \pi N. \tag{5.54}$$

Note that such an exact result is only known for $\lambda = 1$ but unknown for $\lambda \neq 1$. It is worth mentioning that λ classifies superconductivity so that $\lambda < 1$ corresponds to type I and $\lambda > 1$ corresponds to type II superconductivity, respectively.

Esti[redacted]te of energy fro[redacted] ove topo ogic upper ound

Using (ϕ, A) to denote a solution of the BPS system with $N = \pm|N|$ as a trial field configuration pair, we have

$$\begin{aligned} E(\phi, A) &\leq \max\{\lambda, 1\} \int_{\mathbb{R}^2} \left(\frac{1}{2} F_{12}^2 + \frac{1}{2} |D_{\mathbf{j}} \phi|^2 + \frac{1}{8} (|\phi|^2 - 1)^2 \right) \mathrm{d}x \\ &= \max\{\lambda, 1\} \int_{\mathbb{R}^2} \left(\frac{1}{2} \left| F_{12} \pm \frac{1}{2} (|\phi|^2 - 1) \right| + \frac{1}{2} |D_1 \phi \pm \mathrm{i} D_2 \phi|^2 \pm \frac{1}{2} F_{12} \right) \mathrm{d}x \\ &\leq \max\{\lambda, 1\} \pi |N|. \end{aligned} \tag{5.55}$$

Therefore, we have obtained the following lower and upper bounds

$$\min\{\lambda, 1\} \pi |N| \leq E_{\mathbf{N}} \leq \max\{\lambda, 1\} \pi |N|, \tag{5.56}$$

which implies that the energy E grows in proportion to the total vortex number N and suggests that these vortices may indeed be viewed as particles. It is also seen that a nonvanishing N is essential for the existence of a nontrivial solution.

Recently, it is demonstrated in [21] that $E_{\mathbf{N}}$ is asymptotically like $\frac{\pi}{2} N^2 \ln \lambda$ for λ large.

Energy g p

An interesting fact contained in (5.56) is that the Ginzburg–Landau equations have no nontrivial energy-minimizing solution with an energy in the open interval

$$I = (0, \min\{\lambda, 1\} \pi). \tag{5.57}$$

Such a result may be viewed as an *energy* or *mass gap theorem* at classical level.

o onopo con n o con n

Consider two masses, m_1 and m_2, initially placed r distance away. The gravitational attractive force between the masses is then

$$F(r) = G \frac{m_1 m_2}{r^2}, \tag{5.58}$$

following Newton's law. Thus, the work or energy needed to completely split them so that they eventually stay away as non-interacting 'free masses' is

$$W = \int_{\mathbf{r}}^{\infty} F(\rho) \, \mathrm{d}\rho = G \frac{m_1 m_2}{r}. \tag{5.59}$$

Of course, we can do the same thing for a pair of monopole and anti-monopole because the magnetic force between them obeys the same inverse-square law.

However, when the monopole and anti-monopole are placed in a type-II superconductor, we will encounter an entirely different situation. Due to the presence of superconducting Cooper pairs (y)R1792(s)-1.55628(44324(w)18.2268(e)4.44334(h)5.55379(o)-6(l)-8.22315(l)-8.22JΩR74)5.55506(n)27.560

which is a linear function of r. In particular, the work needed to completely split the monopole pair so that the monopole and anti-monopole are seen as isolated free entities will be infinite. Thus, it will be practically impossible to separate the monopole and anti-monopole immersed in a type-II superconductor. In other words, the monopole and anti-monopole are confined. More precisely, we see that, due to the Meissner effect, the monopole and anti-monopole placed in a type-II superconductor interact with each other through narrowly formed vortex-lines which give rise to a constant inter-monopole binding force and the linear law of potential expressed in (5.60). As a result, it is impossible to separate such a pair of monopole and anti-monopole, and the monopole confinement phenomenon takes place. Such a confinement picture is also called *linear confinement*.

A fundamental puzzle in physics, known as the *quark confinement* [54], is that quarks, which make up elementary particles such as mesons and baryons, cannot be observed in isolation. A well accepted confinement mechanism, exactly known as the linear confinement model, interestingly states that, when one tries to separate a pair of quarks, such as a quark and an anti-quark constituting a meson, the energy consumed would grow linearly with respect to the the separation distance between the quarks so that it would require an infinite amount of energy in order to split the pair. The quark and anti-quark may be regarded as a pair of source and sink of color-charged force fields. The source and sink interact through color-charged fluxes which are screened in the bulk of space but form thin tubes in the form of color-charged vortex-lines so that the strength of the force remains constant over arbitrary distance, resulting in a linear dependence relation for the potential energy with regard to the separation distance. Such a situation is similar to the above-described magnetic monopole and anti-monopole pair immersed in a type-II superconductor. We have seen that the magnetic fluxes mediating the interacting monopoles are not governed by the Maxwell equations, which would otherwise give rise to an inverse-square-power law type of decay of the forces and lead to non-confinement, but rather by the Ginzburg–Landau equations or the static Abelian Higgs model in the temporal gauge, which produce narrowly distributed vortex-lines, known as the Abrikosov vortices [1] or the Nielsen–Olesen strings [99], as presented, leading to a linear confinement result, (5.60).

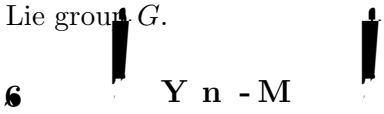
Inspired by the above-described monopole confinement in a type-II superconductor, Mandelstam [86, 87], Nambu [95], and 't Hooft [135, 137] proposed in the 1970s that the ground state of quantum-chromodynamics (QCD) is a condensate of chromomagnetic (color-charged) monopoles, causing the chromoelectric fluxes between quarks to be squeezed into narrowly formed tubes or vortex-lines, similar to the electron condensation in the bulk of a superconductor, in the form of the Cooper pairs, resulting in the formation of color-charged flux-tubes or vortex-lines which mediate the interaction between quarks, following a non-Abelian version of the Meissner effect, called the '*dual Meissner effect*', which is responsible for the screening of chromoelectric fluxes [124, 125].

Thus, we have seen that the notion of magnetic monopoles is not only a theoretical construction as a result of the electromagnetic duality but also supplies as a useful '*theoretical phenomenon*' that provides a crucial hint to a hopeful solution of one of the greatest puzzles of modern physics – quark confinement.

6 Non-Abelian Gauge Fields

In the last two sections, we discussed a gauge field theory with the Abelian group $U(1)$. The *Yang–Mills theory* is now a generic name for the gauge field theory with an arbitrary non-Abelian

Lie group G.



For any $A \in \mathcal{G}$, since $A^\dagger = -A$, we see that

$$|A|^2 = (A, A) = -\text{Tr } (A^2). \tag{6.10}$$

In complete analogy with the electromagnetic field in the Abelian case, we can examine the non-commutativity of the gauge-covariant derivatives to get

$$D_\mu D_\nu \phi - D_\nu D_\mu \phi = (\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])\phi, \tag{6.11}$$

where $[\cdot, \cdot]$ is the Lie bracket (or commutator) of $\mathcal{G}$. Hence we are motivated to define the skew-symmetric Yang–Mills field (curvature) 2-tensor $F_{\mu\nu}$ as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]. \tag{6.12}$$

It is clear to see that $F_{\mu\nu}$ transforms itself according to

$$F_{\mu\nu} \mapsto F'_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu + [A'_\mu, A'_\nu] = \Omega F_{\mu\nu} \Omega^{-1}. \tag{6.13}$$

Hence we obtain the analogous invariant term,

$$\frac{1}{4}\text{Tr } (F_{\mu\nu} F^{\mu\nu}). \tag{6.14}$$

Hence, we arrive at the final form of our locally gauge-invariant Lagrangian action density

$$\mathcal{L} = \frac{1}{4}\text{Tr } (F_{\mu\nu} F^{\mu\nu}) + \frac{1}{2}(D_\mu \phi)(D^\mu \phi)^\dagger - V(|\phi|^2). \tag{6.15}$$

which defines a non-Abelian gauge field theory called the Yang–Mills theory. The Euler–Lagrange equations of the above action are called the Yang–Mills equations. In the situation where the potential density V introduces a spontaneously broken symmetry, the theory is called the Yang–Mills–Higgs theory and the equations are the Yang–Mills–Higgs equations.

In the case where the matter component (containing ϕ) is neglected, the action density becomes

$$\mathcal{L} = \frac{1}{4}\text{Tr } (F_{\mu\nu} F^{\mu\nu}), \tag{6.16}$$

which is simply called the (pure) Yang–Mills theory. The Euler–Lagrange equations of such a Lagrangian are called the (pure) Yang–Mills equations, which are non-Abelian extension of the Maxwell equations in vacuum.

Like the Maxwell electromagnetic field, the Yang–Mills fields are also *mediating* (force) fields. In the non-Abelian case the commutators introduce nonlinearity and new physics appears: these non-Abelian gauge fields are in fact *nuclear force* fields which become significant only in short distances. More precisely, like the $U(1)$ group giving rise to electromagnetic interactions, the $SU(2)$ group gives rise to weak, $SU(3)$ strong, and $SU(5)$ grand unified interactions.

6 [illegible] n dyon

In order to motivate our presentation of some well-known *particle-like solutions* of the Yang–Mills–Higgs equations, we briefly discuss the *Schwinger dyons* which are hypothetical point particles carrying both electric and magnetic charges and were introduced by Schwinger [123] to model quarks.

Consider the motion of a dyon with mass m and electric and magnetic charges q_1 and g_1 in the electromagnetic field $(\mathbf{E}, \mathbf{B})$ of another dyon with electric and magnetic charges q_2 and g_2 placed at the origin. We assume that the second dyon is so heavy that it stays stationary throughout our study.

Solving (3.6)–(3.9) for the second dyon, we obtain the generated electric and magnetic fields

$$\mathbf{E} = q_2 \frac{}{|\ |^3}, \quad \mathbf{B} = g_2 \frac{}{|\ |^3}. \tag{6.17}$$

On the other hand, the non-relativistic motion of the first dyon is governed by the Lorentz force so that

$$m\ddot{} = q_1(\mathbf{E} + \dot{} \times \mathbf{B}) + g_1(\mathbf{B} - \dot{} \times \mathbf{E}). \tag{6.18}$$

Inserting (6.17) into (6.18), we arrive at

$$m\ddot{} = (q_1 q_2 + g_1 g_2) \frac{}{|\ |^3} + (q_1 g_2 - q_2 g_1)\dot{} \times \frac{}{|\ |^3}. \tag{6.19}$$

Using (6.19), we arrive at the conserved total angular momentum defined by

$$= \times m\dot{} - (q_1 g_2 - q_2 g_1) \frac{}{|\ |}. \tag{6.20}$$

It is important to notice that, in the special case of the motion of an electrically charged particle in the magnetic field generated by a monopole, we have $g_1 = 0, q_2 = 0$ and (6.20) becomes

$$= \times m\dot{} - q_1 g_2 \frac{}{|\ |}. \tag{6.21}$$

We recall that, in this case, the Dirac charge quantization formula reads

$$g_2 q_1 = \frac{n}{2}, \quad n = 0, \pm 1, \pm 2, \cdots. \tag{6.22}$$

As an immediate consequence of the similarity between (6.20) and (6.21), we obtain the celebrated *Schwinger charge quantization formula for dyons*

$$g_2 q_1 - g_1 q_2 = \frac{n}{2}, \quad n = 0, \pm 1, \pm 2, \cdots. \tag{6.23}$$

Like that of a Dirac monopole, a Schwinger dyon also carries infinite energy.

6 oo - o y o onopo nd - dyon

In the above study, we discussed monopoles and dyons in terms of the Maxwell equations for electromagnetism which is a theory of Abelian gauge fields. In fact it is more natural for monopoles and dyons to exist in non-Abelian gauge field-theoretical models because nonvanishing commutators themselves are now present as electric and magnetic source terms. In other words, non-Abelian monopoles and dyons are self-induced and inevitable. In this section, we shall present the simplest non-Abelian dyons known as the Julia–Zee dyons [67], which contain as special solutions the 't Hooft–Polyakov monopoles.

Consider the simplest non-Abelian Lie group $SO(3)$, which has a set of generators $\{t_{\mathbf{a}}\}$ $(a = 1, 2, 3)$ satisfying $[t_{\mathbf{a}}, t_{\mathbf{b}}] = \epsilon^{\mathbf{abc}} t_{\mathbf{c}}$. Consequently, the $so(3)$-valued quantities $A = A^{\mathbf{a}} t_{\mathbf{a}}$ and $B = B^{\mathbf{a}} t_{\mathbf{a}}$ give rise to a commutator,

$$[A, B] = \epsilon_{\mathbf{abc}} A^{\mathbf{a}} B^{\mathbf{b}} t_{\mathbf{c}}. \tag{6.24}$$

For convenience, we may view A and B as two 3-vectors, $\mathbf{A} = (A^{\mathbf{a}})$ and $\mathbf{B} = (B^{\mathbf{a}})$. Then, by (6.24), $[A, B]$ corresponds to the vector cross-product, $\mathbf{A} \times \mathbf{B}$. With these in mind, we make the following introduction to the $SO(3)$ (or $SU(2)$ since $SO(3)$ and $SU(2)$ have identical Lie algebras) Yang–Mills–Higgs theory.

Let $\mathbf{A_\mu} = (A^{\mathbf{a}}_{\boldsymbol{\mu}})$ $(\mu = 0, 1, 2, 3)$ and $\phi = (\phi^{\mathbf{a}})$ $(a = 1, 2, 3)$ be a gauge and matter Higgs fields, respectively, interacting through the action density

$$\mathcal{L} = -\frac{1}{4}\mathbf{F^{\mu\nu}} \cdot \mathbf{F_{\mu\nu}} + \frac{1}{2} D^{\boldsymbol{\mu}}\phi \cdot D_{\boldsymbol{\mu}}\phi - \frac{\lambda}{4}(|\phi|^2 - 1)^2, \tag{6.25}$$

where the field strength tensor $\mathbf{F_{\mu\nu}}$ is defined by

$$\mathbf{F_{\mu\nu}} = \partial_{\boldsymbol{\mu}}\mathbf{A_\nu} - \partial_{\boldsymbol{\nu}}\mathbf{A_\mu} - e\mathbf{A_\mu} \times \mathbf{A_\nu}, \tag{6.26}$$

and the gauge-covariant derivative $D_{\boldsymbol{\mu}}$ is defined by

$$D_{\boldsymbol{\mu}}\phi = \partial_{\boldsymbol{\mu}}\phi - e\mathbf{A_\mu} \times \phi. \tag{6.27}$$

Based on consideration on interactions [67, 134, 136], it is recognized that the electromagnetic field $F_{\boldsymbol{\mu\nu}}$ is defined by the formula

$$F_{\boldsymbol{\mu\nu}} = \frac{1}{|\phi|}\phi \cdot \mathbf{F_{\mu\nu}} - \frac{1}{e|\phi|^3}\phi \cdot (D_{\boldsymbol{\mu}}\phi \times D_{\boldsymbol{\nu}}\phi). \tag{6.28}$$

The equations of motion of (6.25) can be derived as

$$\begin{aligned} D^{\boldsymbol{\nu}}\mathbf{F_{\mu\nu}} &= -e\phi \times D_{\boldsymbol{\mu}}\phi, \\ D^{\boldsymbol{\mu}} D_{\boldsymbol{\mu}}\phi &= -\lambda\phi(|\phi|^2 - 1). \end{aligned} \tag{6.29}$$

We are interested in static solutions of (6.29). In general, this is a difficult problem. Here we can only consider radially symmetric solutions.

Set $r = |\ \ |$. Following Julia and Zee [67], the most general radially symmetric solutions of (6.29) are of the form

$$\begin{aligned} A^{\mathbf{a}}_0 &= \frac{x^{\mathbf{a}}}{er^2} J(r), \\ A^{\mathbf{a}}_{\mathbf{i}} &= \epsilon_{\mathbf{abi}} \frac{x^{\mathbf{b}}}{er^2}(K(r) - 1), \\ \phi^{\mathbf{a}} &= \frac{x^{\mathbf{a}}}{er^2} H(r), \quad a, b, c = 1, 2, 3. \end{aligned} \tag{6.30}$$

Inserting (6.30) into (6.29) and using prime to denote differentiation with respect to the radial variable r, we have

$$\begin{aligned} r^2 J'' &= 2JK^2, \\ r^2 H'' &= 2HK^2 - \lambda r^2 H\left(1 - \frac{1}{e^2 r^2} H^2\right), \\ r^2 K'' &= K(K^2 - J^2 + H^2 - 1). \end{aligned} \tag{6.31}$$

We need to specify boundary conditions for these equations. First, we see from (6.30) and regularity requirement that H, J, K must satisfy

$$\lim_{\mathbf{r}\to 0}\left(\frac{H(r)}{er}, \frac{J(r)}{er}, K(r)\right) = (0,0,1). \tag{6.32}$$

Secondly, since the Hamiltonian density of (6.25) takes the form

$$\begin{aligned}\mathcal{H} &= \mathbf{F_{0i}}\cdot\mathbf{F_{0i}} + D_0\phi\cdot D_0\phi - \mathcal{L} \\ &= \frac{1}{e^2r^2}(K')^2 + \frac{1}{2}(u')^2 + \frac{1}{2}(v')^2 + \frac{1}{2e^2r^4}(K^2-1)^2 + \frac{1}{r^2}K^2(u^2+v^2) + \frac{\lambda}{4}(u^2-1)^2,\end{aligned} \tag{6.33}$$

where $u = H/er$ and $v = J/er$. Finite energy condition,

$$E = \int_{\mathbb{R}^3}\mathcal{H}\,\mathrm{d}x < \infty, \tag{6.34}$$

and the formula (6.33) imply that $u(r)\to 1$ and $K(r)\to 0$ as $r\to\infty$. Besides, it is seen from (6.33) that $v(r)\to$ some constant C_0 as $r\to\infty$. However, C_0 cannot be determined completely. We record these results as follows,

$$\lim_{\mathbf{r}\to\infty}\left(\frac{H(r)}{er}, \frac{J(r)}{er}, K(r)\right) = (1, C_0, 0). \tag{6.35}$$

So utions in the BPS [redacted] it

Inserting (6.30) into (6.28), we find the electric and magnetic fields, $\mathbf{E} = (E^{\mathbf{i}})$ and $\mathbf{B} = (B^{\mathbf{i}})$, as follows,

$$E^{\mathbf{i}} = -F^{0\mathbf{i}} = \frac{x^{\mathbf{i}}}{er}\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{J}{r}\right), \tag{6.36}$$

$$B^{\mathbf{i}} = -\frac{1}{2}\epsilon_{\mathbf{ijk}}F^{\mathbf{jk}} = \frac{x^{\mathbf{i}}}{er^3}. \tag{6.37}$$

It is interesting to note that both $E^{\mathbf{i}}$ and $B^{\mathbf{i}}$ obey the inverse-square law.

From (6.36) we see that, if $J = 0$, $\mathbf{E} =$ and there is no electric field. The magnetic charge g may be obtained through integrating (6.37),

$$g = \frac{1}{4\pi}\lim_{\mathbf{r}\to\infty}\oint_{|\mathbf{x}|=\mathbf{r}}\mathbf{B}\cdot\mathrm{d}\mathbf{S} = \frac{1}{e}, \tag{6.38}$$

which is similar to the Dirac quantization formula except that electric charge is doubled. Moreover, the equations of motion (6.31) are simplified into the form

$$\begin{aligned}r^2H'' &= 2HK^2 - \lambda r^2 H\left(1 - \frac{1}{e^2r^2}H^2\right), \\ r^2K'' &= K(K^2 + H^2 - 1).\end{aligned} \tag{6.39}$$

These equations cannot be solved explicitly for general $\lambda > 0$ but an existence theorem has been established by using functional analysis [117]. Here we present a family of explicit solutions at the BPS limit $\lambda = 0$ due to Bogomol'nyi [10] and Prasad–Sommerfield [113].

When $\lambda = 0$, the system (6.39) becomes

$$\begin{aligned} r^2 H'' &= 2HK^2, \\ r^2 K'' &= K(K^2 + H^2 - 1), \end{aligned} \tag{6.40}$$

with the associated energy

$$E = \int_{\mathbb{R}^3} \mathcal{H}\, \mathrm{d}x = \frac{4\pi}{e^2} \int_0^\infty \left\{ (K')^2 + \frac{1}{2r^2}(rH' - H)^2 + \frac{1}{2r^2}(K^2 - 1)^2 + \frac{1}{r^2} K^2 H^2 \right\} \mathrm{d}r. \tag{6.41}$$

It is clear that the system (6.40) is the Euler–Lagrange equations of (6.41). Besides, using the boundary conditions (6.32) and (6.35), we have

$$\begin{aligned} E &= \frac{4\pi}{e^2} \int_0^\infty \left\{ \left(K' + \frac{1}{r} KH \right)^2 + \frac{1}{2r^2} \left(rH' - H + (K^2 - 1) \right)^2 + \frac{\mathrm{d}}{\mathrm{d}r} \left(\frac{H}{r} - \frac{K^2 H}{r} \right) \right\} \mathrm{d}r \\ &\geq \frac{4\pi}{e}. \end{aligned} \tag{6.42}$$

Hence, we have the energy lower bound, $E \geq 4\pi/e$, which is attained when (H, K) satisfies

$$\begin{aligned} rK' &= -KH, \\ rH' &= H - (K^2 - 1). \end{aligned} \tag{6.43}$$

Of course, any solution of (6.43) also satisfies (6.40). It was Maison [85] who first showed that (6.40) and (6.43) are actually equivalent, which is a topic we will not get into here.

We now obtain a family of explicit solutions of (6.43) (hence (6.40)).

Introduce a change of variables from (H, K) to (U, V),

$$-H = 1 + rU, \quad K = rV. \tag{6.44}$$

Then (6.43) becomes $U' = V^2, V' = UV$, which implies that $U^2 - V^2 =$ constant. Thus, by virtue of (6.35) and (6.44), we have

$$U^2 - V^2 = e^2, \quad r > 0. \tag{6.45}$$

Inserting (6.45) into $U' = V^2$ and using (6.32), i.e., $U(r) \sim -1/r$ for $r > 0$ small, we obtain an explicit solution of (6.43),

$$\begin{aligned} H(r) &= er \coth(er) - 1, \\ K(r) &= \frac{er}{\sinh(er)}, \end{aligned} \tag{6.46}$$

which gives rise to a smooth, minimum energy (mass) $E = 4\pi/e$, monopole ($J = 0$) solution of the equations of motion (6.29) at the BPS limit $\lambda = 0$ through the radially symmetric prescription (6.30). When $\lambda > 0$, one may use variational methods to obtain an existence theory for solutions. These solutions are collectively known as the *'t Hooft–Polyakov monopoles* [109, 133] which are *smooth* and of *finite energy*.

We next present a continuous family of explicit dyon solutions. At the BPS limit, $\lambda = 0$, the system (6.31) is

$$\begin{aligned} r^2 J'' &= 2JK^2, \\ r^2 H'' &= 2HK^2, \\ r^2 K'' &= K(K^2 - J^2 + H^2 - 1), \end{aligned} \tag{6.47}$$

which becomes (6.40) when we compress H, J through

$$H \mapsto (\cosh\alpha)H, \quad J \mapsto (\sinh\alpha)H,$$

where α is a constant. Therefore, using (6.46), we have

$$\begin{aligned} H(r) &= \cosh\alpha(er\coth(er)-1), \\ J(r) &= \sinh\alpha(er\coth(er)-1), \\ K(r) &= \frac{er}{\sinh(er)}. \end{aligned} \tag{6.48}$$

Consequently, in view of (6.30), we have obtained a family of explicit dyon solutions of (6.29). Note that all the boundary conditions stated in (6.32) and (6.35) are fulfilled except one, namely,

$$\lim_{\mathbf{r}\to\infty}\frac{H(r)}{er} = \cosh\alpha \neq 1. \tag{6.49}$$

However, since $\lambda = 0$, (6.49) is of no harm to the finite energy condition (6.34).

To compute the total electric charge, we use (6.36). We have

$$\begin{aligned} q &= \frac{1}{4\pi}\lim_{\mathbf{r}\to\infty}\oint_{|\mathbf{x}|=\mathbf{r}}\mathbf{E}\cdot\mathrm{d}\mathbf{S} \\ &= \frac{1}{e}\lim_{\mathbf{r}\to\infty}(rJ'(r)-J(r)) = \frac{\sinh\alpha}{e}. \end{aligned} \tag{6.50}$$

It is a comfort to see that the solution becomes electrically neutral, $q = 0$, when $\alpha = 0$ and (6.48) reduces to the monopole solution (6.46).

Tot e ectric ch rge y e non qu ntized

It should be noted that, since α in (6.50) is arbitrary, the electric charge q given in (6.50) is *not quantized* and may assume value in a continuum. The main reason for the discrepancy with what expressed by the Dirac charge quantization formula is that the electric charge q here is the total charge induced from a continuously distributed electric field but not a pure point charge.

The Higgs e d y gener te ch rges

The calculation carried out here also shows that, like mass, both electric and magnetic charges may be *generated* from the Higgs fields.

6 o - n - c o od

The bosonic Lagrangian action density of the *Glashow–Weinberg–Salam electroweak model* may be written as

$$\mathcal{L} = -\frac{1}{4}\mathbf{F}_{\boldsymbol{\mu\nu}}\cdot\mathbf{F}^{\boldsymbol{\mu\nu}} - \frac{1}{4}H_{\boldsymbol{\mu\nu}}H^{\boldsymbol{\mu\nu}} + (\hat{D}_{\boldsymbol{\mu}}\phi)\cdot(\hat{D}^{\boldsymbol{\mu}}\phi)^{\dagger} - \frac{\lambda}{2}(|\phi|^2-\phi_0^2)^2, \tag{6.51}$$

where ϕ now is a Higgs complex doublet lying in the fundamental representation space of $SU(2)$, $\mathbf{F}_{\boldsymbol{\mu\nu}}$ and $H_{\boldsymbol{\mu\nu}}$ are the gauge fields of $SU(2)$ and $U(1)$ with the potential $\mathbf{A}_{\boldsymbol{\mu}}$ and $B_{\boldsymbol{\mu}}$ and the corresponding coupling constants g and g', respectively, for the generators of $SU(2)$ we use the conventional Pauli spin matrices $\tau^{\mathbf{a}}$ $(a = 1, 2, 3)$,

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau^2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \tau = (\tau^{\mathbf{a}}), \tag{6.52}$$

the coordinate vectors $\mathbf{A}_{\boldsymbol{\mu}} = (A^{\mathbf{a}}_{\boldsymbol{\mu}})$, and the gauge-covariant derivatives are defined by the expressions

$$\hat{D}_{\boldsymbol{\mu}}\phi = \left(\partial_{\boldsymbol{\mu}} - \mathrm{i}g\tau^{\mathbf{a}}A^{\mathbf{a}}_{\boldsymbol{\mu}} - \mathrm{i}g'B_{\boldsymbol{\mu}}\right)\phi = D_{\boldsymbol{\mu}}\phi - \mathrm{i}g'B_{\boldsymbol{\mu}}\phi. \tag{6.53}$$

Therefore, within the above framework, the Euler–Lagrangian equations of (6.51) are

$$\hat{D}_{\boldsymbol{\mu}}\hat{D}^{\boldsymbol{\mu}}\phi = \lambda(|\phi|^2 - \phi_0^2)\phi, \tag{6.54}$$

$$D^{\boldsymbol{\mu}}\mathbf{F}_{\boldsymbol{\mu\nu}} = \mathrm{i}g(\phi^\dagger\tau[\hat{D}_{\boldsymbol{\nu}}\phi] - [\hat{D}_{\boldsymbol{\nu}}\phi]^\dagger\tau\phi), \tag{6.55}$$

$$\partial^{\boldsymbol{\mu}}H_{\boldsymbol{\mu\nu}} = \mathrm{i}g'(\phi^\dagger[\hat{D}_{\boldsymbol{\nu}}\phi] - [\hat{D}_{\boldsymbol{\nu}}\phi]^\dagger\phi). \tag{6.56}$$

Both dyon and vortex solutions of these equations have been obtained (cf. [152] and references therein).[8]

n n on nd d op c

We start from a quick introduction to Riemannian geometry and a derivation of the metric energy-momentum tensor. We next derive the Einstein equations for gravitation. We then discuss some direct cosmological consequences from the Einstein equations, the origin of the cosmological constant and its interpretation as vacuum mass-energy density, and the Schwarzschild blackhole solution and its extensions. We end with an excursion to the ADM mass and related topics such as the positive mass theorem and the Penrose inequality.

n n d on

Let $(g_{\boldsymbol{\mu\nu}})$ be the metric tensor of spacetime. The spacetime line element or the first fundamental form is defined by

$$\mathrm{d}s^2 = g_{\boldsymbol{\mu\nu}}\mathrm{d}x^{\boldsymbol{\mu}}\mathrm{d}x^{\boldsymbol{\nu}}, \tag{7.1}$$

which is also a measurement of the *proper time* (see (4.18) for its flat-spacetime version). A freely moving particle in spacetime follows a curve that stationarizes the action

$$\int \mathrm{d}s. \tag{7.2}$$

We now derive the equations of motion from the above action principle.

[8] The following paragraph posted at Wikipedia about the Glashow–Weinberg–Salam model, under the subject title 'Unified Field Theory,' is worth reading in the context of our study here: In 1963 American physicist Sheldon Glashow proposed that the weak nuclear force and electricity and magnetism could arise from a partially unified electroweak theory. In 1967, Pakistani Abdus Salam and American Steven Weinberg independently revised Glashow's theory by having the masses for the W particle and Z particle arise through spontaneous symmetry breaking with the Higgs mechanism. This unified theory was governed by the exchange of four particles: the photon for electromagnetic interactions, a neutral Z particle and two charged W particles for weak interaction. As a result of the spontaneous symmetry breaking, the weak force becomes short range and the Z and W bosons acquire masses of 80.4 and 91.2 GeV, respectively. Their theory was first given experimental support by the discovery of weak neutral currents in 1973. In 1983, the Z and W bosons were first produced at CERN by Carlo Rubbia's team. For their insights, Salam, Glashow and Weinberg were awarded the Nobel Prize in Physics in 1979. Carlo Rubbia and Simon van der Meer received the Prize in 1984.

We use the notation $x^{\mu}(s)$ to denote the desired curve (trajectory of the particle) and $\delta x^{\mu}(s)$ a small variation, both parametrized by s. Then, to the first order of variation, we have

$$\begin{aligned}\delta(\mathrm{d}s^2) &= 2\mathrm{d}s\delta(\mathrm{d}s) = (\delta g_{\mu\nu})\mathrm{d}x^{\mu}\mathrm{d}x^{\nu} + 2g_{\mu\nu}\mathrm{d}x^{\mu}\delta(\mathrm{d}x^{\nu}) \\ &= (g_{\mu\nu,\alpha}\delta x^{\alpha})\mathrm{d}x^{\mu}\mathrm{d}x^{\nu} + 2g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}(\delta x^{\nu}),\end{aligned} \tag{7.3}$$

where and in the sequel we use the notation

$$f_{,\alpha}, \quad A_{\mu,\alpha}, \quad F_{\mu\nu,\alpha}, \quad T^{\mu\nu}{}_{,\alpha}, \tag{7.4}$$

etc., to denote the conventional partial derivative with respect to the variable x^{α} of various quantities. Using v^{μ} to denote the components of the *4-velocity*,

$$v^{\mu}(s) = \frac{\mathrm{d}x^{\mu}(s)}{\mathrm{d}s}, \tag{7.5}$$

we then obtain

$$\begin{aligned}\delta(\mathrm{d}s) &= \left(\frac{1}{2}g_{\mu\nu,\alpha}v^{\mu}v^{\nu}\delta x^{\alpha} + g_{\mu\nu}v^{\mu}\frac{\mathrm{d}}{\mathrm{d}s}(\delta x^{\nu})\right)\mathrm{d}s \\ &= (\delta x^{\alpha})\left(\frac{1}{2}g_{\mu\nu,\alpha}v^{\mu}v^{\nu} - \frac{\mathrm{d}}{\mathrm{d}s}(g_{\mu\alpha}v^{\mu})\right)\mathrm{d}s + \mathrm{d}(g_{\mu\nu}v^{\mu}\delta x^{\nu}).\end{aligned} \tag{7.6}$$

Inserting the above into the stationary condition

$$\delta\int \mathrm{d}s = 0 \tag{7.7}$$

and using the fact that δx^{μ} vanishes at the two end points of the curve, we arrive at the equations of motion

$$\frac{\mathrm{d}}{\mathrm{d}s}(g_{\mu\alpha}v^{\mu}) - \frac{1}{2}g_{\mu\nu,\alpha}v^{\mu}v^{\nu} = 0. \tag{7.8}$$

Again, since $g_{\mu\nu}$ is symmetric, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}(g_{\mu\alpha}v^{\mu}) &= g_{\mu\alpha}\frac{\mathrm{d}v^{\mu}}{\mathrm{d}s} + g_{\mu\alpha,\nu}v^{\mu}v^{\nu} \\ &= g_{\mu\alpha}\frac{\mathrm{d}v^{\mu}}{\mathrm{d}s} + \frac{1}{2}(g_{\alpha\mu,\nu} + g_{\alpha\nu,\mu})v^{\mu}v^{\nu}.\end{aligned} \tag{7.9}$$

Consequently the equations of motion become

$$g_{\mu\alpha}\frac{\mathrm{d}v^{\mu}}{\mathrm{d}s} + \Gamma_{\alpha\mu\nu}v^{\mu}v^{\nu} = 0 \quad \text{or} \quad \frac{\mathrm{d}v^{\alpha}}{\mathrm{d}s} + \Gamma^{\alpha}_{\mu\nu}v^{\mu}v^{\nu} = 0, \tag{7.10}$$

where $\Gamma_{\mu\nu\alpha}$ and $\Gamma^{\alpha}_{\mu\nu}$ are called the *Christoffel symbols*, which are defined by

$$\Gamma_{\mu\nu\alpha} = \frac{1}{2}(g_{\mu\nu,\alpha} + g_{\mu\alpha,\nu} - g_{\nu\alpha,\mu}), \quad \Gamma^{\alpha}_{\mu\nu} = g^{\alpha\beta}\Gamma_{\mu\nu\beta}. \tag{7.11}$$

The curves that are solutions of (7.10) are called *geodesics*.

We see immediately that $\Gamma_{\mu\nu\alpha} = \Gamma_{\mu\alpha\nu}$. Besides, it is also useful to note that the definition of $\Gamma_{\mu\nu\alpha}$ gives us the identity

$$\Gamma_{\mu\nu\alpha} + \Gamma_{\nu\mu\alpha} = g_{\mu\nu,\alpha}. \tag{7.12}$$

One of the most important applications of the Christoffel symbols is their role in the definition of covariant derivatives for covariant and contravariant quantities,

$$\begin{aligned} A_{\mu;\alpha} &= A_{\mu,\alpha} - \Gamma^{\beta}_{\mu\alpha} A_{\beta}, \\ T_{\mu\nu;\alpha} &= T_{\mu\nu,\alpha} - \Gamma^{\beta}_{\mu\alpha} T_{\beta\nu} - \Gamma^{\beta}_{\nu\alpha} T_{\mu\beta}, \\ A^{\mu}{}_{;\alpha} &= A^{\mu}{}_{,\alpha} + \Gamma^{\mu}_{\beta\alpha} A^{\beta}, \\ T^{\mu\nu}{}_{;\alpha} &= T^{\mu\nu}{}_{,\alpha} + \Gamma^{\mu}_{\beta\alpha} T^{\beta\nu} + \Gamma^{\nu}_{\beta\alpha} T^{\mu\beta}. \end{aligned} \tag{7.13}$$

We will sometimes use ∇_{α} to denote covariant derivative. A direct consequence of the above definition and the identity (7.12) is that

$$\begin{aligned} g_{\mu\nu;\alpha} &= g_{\mu\nu,\alpha} - \Gamma^{\beta}_{\mu\alpha} g_{\beta\nu} - \Gamma^{\beta}_{\nu\alpha} g_{\mu\beta} \\ &= g_{\mu\nu,\alpha} - \Gamma_{\mu\alpha\nu} - \Gamma_{\nu\alpha\mu} = 0. \end{aligned} \tag{7.14}$$

Similarly, $g^{\mu\nu}{}_{;\alpha} = 0$. Therefore we have seen that the covariant and contravariant metric tensors, $g_{\mu\nu}$ and $g^{\mu\nu}$, behave like constants under covariant differentiation.

Let A_{μ} be a test covariant vector. Following (7.13), we obtain through an easy calculation the commutator

$$A_{\mu;\alpha;\beta} - A_{\mu;\beta;\alpha} = [\nabla_{\alpha}, \nabla_{\beta}] A_{\mu} = R^{\nu}_{\mu\alpha\beta} A_{\nu}, \tag{7.15}$$

where

$$R^{\nu}_{\mu\alpha\beta} = \Gamma^{\nu}_{\mu\beta,\alpha} - \Gamma^{\nu}_{\mu\alpha,\beta} + \Gamma^{\gamma}_{\mu\beta}\Gamma^{\nu}_{\gamma\alpha} - \Gamma^{\gamma}_{\mu\alpha}\Gamma^{\nu}_{\gamma\beta} \tag{7.16}$$

is a mixed 4-tensor called the *Riemann* curvature tensor. There hold the simple properties

$$R^{\nu}_{\mu\alpha\beta} = -R^{\nu}_{\mu\beta\alpha}, \tag{7.17}$$

$$R^{\nu}_{\mu\alpha\beta} + R^{\nu}_{\alpha\beta\mu} + R^{\nu}_{\beta\mu\alpha} = 0. \tag{7.18}$$

Furthermore, similar to (7.15), for covariant 2-tensors, we have

$$T_{\mu\nu;\alpha;\beta} - T_{\mu\nu;\beta;\alpha} = R^{\gamma}_{\mu\alpha\beta} T_{\gamma\nu} + R^{\gamma}_{\nu\alpha\beta} T_{\mu\gamma}. \tag{7.19}$$

Therefore, in particular, for a covariant vector field A_{μ}, we have

$$A_{\mu;\nu;\alpha;\beta} - A_{\mu;\nu;\beta;\alpha} = R^{\gamma}_{\mu\alpha\beta} A_{\gamma;\nu} + R^{\gamma}_{\nu\alpha\beta} A_{\mu;\gamma}. \tag{7.20}$$

We now make permutations of the indices ν, α, β and add the three resulting equations. In view of (7.15), the left-hand side is

$$\begin{aligned} & (A_{\mu;\alpha;\beta;\nu} - A_{\mu;\beta;\alpha;\nu}) + \text{permutations} \\ = \; & (R^{\gamma}_{\mu\alpha\beta} A_{\gamma})_{;\nu} + \text{permutations} \\ = \; & (R^{\gamma}_{\mu\alpha\beta} A_{\gamma;\nu} + R^{\gamma}_{\mu\alpha\beta;\nu} A_{\gamma}) + \text{permutations}. \end{aligned} \tag{7.21}$$

In view of (7.18), the right-hand side is

$$R^{\gamma}_{\mu\alpha\beta} A_{\gamma;\nu} + \text{permutations}. \tag{7.22}$$

Equating (7.21) and (7.22), we arrive at

$$R^{\gamma}_{\mu\alpha\beta;\nu} A_{\gamma} + \text{permutations} = 0. \tag{7.23}$$

Since A_{μ} is arbitrary, we find that

$$R^{\gamma}_{\mu\nu\alpha;\beta} + R^{\gamma}_{\mu\alpha\beta;\nu} + R^{\gamma}_{\mu\beta\nu;\alpha} = 0. \tag{7.24}$$

This result is also known as the *Bianchi identity*.

The *Ricci tensor* $R_{\mu\nu}$ is defined from $R^{\nu}_{\mu\alpha\beta}$ through contraction,

$$R_{\mu\nu} = R^{\alpha}_{\mu\nu\alpha}. \tag{7.25}$$

It is clear that $R_{\mu\nu}$ is symmetric. The *scalar curvature* R is then defined by

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{7.26}$$

In the Bianchi identity (7.24), set $\gamma = \nu$ and multiply by $g^{\mu\alpha}$. We obtain

$$(g^{\mu\alpha} R^{\nu}_{\mu\nu\alpha})_{;\beta} + (g^{\mu\alpha} R^{\nu}_{\mu\alpha\beta})_{;\nu} + (g^{\mu\alpha} R^{\nu}_{\mu\beta\nu})_{;\alpha} = 0, \tag{7.27}$$

which is simply

$$2R^{\beta}_{\alpha;\beta} - R_{;\alpha} = 0. \tag{7.28}$$

Multiplying the above by $g^{\mu\alpha}$, we have the following very important result,

$$G^{\mu\nu}{}_{;\nu} = 0, \tag{7.29}$$

where

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R, \tag{7.30}$$

or its covariant partner, $G_{\mu\nu}$, is called the *Einstein tensor*.

We next consider physics over the curved spacetime of metric $(g_{\mu\nu})$ governed by a matter field u which is either a scalar field or a vector field and governed by the action

$$S = \int \mathcal{L}(x, Du, g) \sqrt{|g|}\, \mathrm{d}x, \tag{7.31}$$

where we have emphasized the influence of the metric tensor $g = (g_{\mu\nu})$ and used the *canonical volume element* $\sqrt{|g|}\, \mathrm{d}x$. Here $|g|$ is the absolute value of the determinant of the metric g. Since physics is independent of coordinates, $\mathcal{L}$ must be a scalar. For example, the real Klein–Gordon action density now reads

$$\mathcal{L}(x, u, Du, g) = \frac{1}{2} g^{\mu\nu} \partial_{\mu} u \partial_{\nu} u - V(u), \tag{7.32}$$

which is g-dependent. In other words, physics can no longer be purely material.

It is easily seen that the Euler–Lagrange equations, or the equations of motion, of (7.31) are now

$$\frac{1}{\sqrt{|g|}} \partial_{\mu} \left(\sqrt{|g|} \frac{\partial \mathcal{L}}{\partial(\partial_{\mu} u^{a})} \right) = \frac{\partial \mathcal{L}}{\partial u^{a}}, \quad a = 1, 2, \cdots, m; \quad u = (u^{a}). \tag{7.33}$$

Using the translation invariance of the action and (7.33), we can derive the *energy-momentum tensor*,[9] also called the *stress tensor*, $T^{\mu\nu}$, given as

$$T^{\mu\nu} = -2 \frac{\partial \mathcal{L}}{\partial g_{\mu\nu}} - g^{\mu\nu} \mathcal{L} = 2 g^{\alpha\mu} g^{\beta\nu} \frac{\partial \mathcal{L}}{\partial g^{\alpha\beta}} - g^{\mu\nu} \mathcal{L}, \tag{7.34}$$

[9]The foundational framework that allows one to derive conserved quantities as a result of symmetry properties of the action is called the *Noether theorem* which will not be elaborated here.

which obeys the conservation law

$$T^{\mu\nu}{}_{;\nu} = 0. \tag{7.35}$$

The basic principle which led Einstein to write down his fundamental equations for gravitation states that the geometry of a spacetime is determined by the matter it contains. Mathematically, Einstein's idea was to consider the equation

$$Q^{\mu\nu} = -\kappa T^{\mu\nu}, \tag{7.36}$$

where $Q^{\mu\nu}$ is a 2-tensor generated from the spacetime metric ($g_{\mu\nu}$) which is purely geometric, $T^{\mu\nu}$ is the energy-momentum tensor which is purely material, κ is a constant called the Einstein gravitational constant, and the negative sign in front of κ is inserted for convenience. This equation imposes severe restriction to the possible form of the 2-tensor $Q^{\mu\nu}$. For example, $Q^{\mu\nu}$ should also satisfy the same conservation law (or the divergence-free condition),

$$Q^{\mu\nu}{}_{;\nu} = 0, \tag{7.37}$$

as $T^{\mu\nu}$ (see (7.35)). The simplest candidate for $Q^{\mu\nu}$ could be $g^{\mu\nu}$. However, since $g^{\mu\nu}$ is non-degenerate, this choice is seen to be incorrect because it makes $T^{\mu\nu}$ non-degenerate, which is absurd in general. The next candidate could be the Ricci curvature $R^{\mu\nu}$. Since $R^{\mu\nu}$ does not satisfy the required identity (7.37), it has to be abandoned. Consequently, based on both the compatibility condition (7.37) and simplicity consideration, we are naturally led to the choice of the Einstein tensor, $G^{\mu\nu}$, defined in (7.30). Therefore we obtain the *Einstein equations*,

$$G^{\mu\nu} = -\kappa T^{\mu\nu} \quad \text{or} \quad G_{\mu\nu} = -\kappa T_{\mu\nu}. \tag{7.38}$$

It can be shown that the equation (7.38) recovers Newton's law of gravitation,

$$F = -G\frac{m_1 m_2}{r^2}, \tag{7.39}$$

which gives the magnitude of an attractive force between two point particles of masses m_1 and m_2 with a distance r apart, in the static spacetime and slow motion limit, if and only if $\kappa = 8\pi G$. Recall that the constant G is called the Newton universal gravitational constant, which is extremely small compared to other quantities.

In summary, we have just derived the Einstein gravitational field equations,

$$G_{\mu\nu} = -8\pi G\, T_{\mu\nu}. \tag{7.40}$$

Co[illegible] o o c con [illegible] nc

In modern cosmology, the universe is believed to be *homogeneous* (the number of stars per unit volume is uniform throughout large regions of space) and *isotropic* (the number of stars per unit solid angle is the same in all directions). This basic property is known as the *Cosmological Principle* and has been evidenced by astronomical observations. A direct implication of such a principle is that synchronized clocks may be placed throughout the universe to give a uniform measurement of time (*Cosmic time*). Another is that the space curvature, K, is constant at any fixed cosmic time t. Hence we have the following simple mathematical descriptions for the space.

(a) If $K = K(t) > 0$, the space is closed and may be defined as a 3-sphere embedded in the flat Euclidean space of the form

$$x^2 + y^2 + z^2 + w^2 = a^2, \quad a = a(t) = \frac{1}{\sqrt{K(t)}}. \tag{7.41}$$

(b) If $K = K(t) < 0$, the space is open and may be defined similarly by the equation

$$-x^2 - y^2 - z^2 + w^2 = a^2, \quad a = a(t) = \frac{1}{\sqrt{|K(t)|}}, \tag{7.42}$$

which is embedded in the flat Minkowski space with the line element

$$\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - \mathrm{d}w^2, \tag{7.43}$$

known as the *anti-de Sitter space* (or the 'adS space' in short).[10]

(c) If $K = K(t) = 0$, the space is the Euclidean space $\mathbb{R}^3$. In particular the space is open and its line element is given by

$$\mathrm{d}\ell^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{7.44}$$

Use the conventional spherical coordinates (r, θ, χ) to replace the Cartesian coordinates (x, y, z). We have

$$x = r\cos\theta\sin\chi, \quad y = r\sin\theta\sin\chi, \quad z = r\cos\chi. \tag{7.45}$$

Thus, in the cases (a) and (b), we have

$$\pm r^2 + w^2 = a^2, \quad \pm r\mathrm{d}r = w\mathrm{d}w. \tag{7.46}$$

Substituting (7.45) and (7.46) into the line elements of the Euclidean space and of the Minkowski space given by (7.43), respectively, we obtain the induced line element $\mathrm{d}\ell^2$ of the space,

$$\mathrm{d}\ell^2 = \frac{a^2}{(a^2 \mp r^2)}\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\chi^2. \tag{7.47}$$

Finally, inserting (7.47) into the spacetime line element and making the rescaling $r \mapsto ar$, we have

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left(\frac{1}{(1-kr^2)}\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\chi^2\right), \tag{7.48}$$

where $k = \pm 1$ or $k = 0$ according to $K > 0, K < 0$ or $K = 0$. This is the most general line element of a homogeneous and isotropic spacetime and is known as the *Robertson–Walker metric.*

In cosmology, the large-scale viewpoint allows us to treat stars or galaxies as particles of a perfect 'gas' that fills the universe and is characterized by its mass-energy density ρ, counting both rest mass and kinetic energy per unit volume, and pressure p, so that the associated energy-momentum tensor $T_{\mu\nu}$ is given by

$$T_{\mu\nu} = (\rho + p)v_\mu v_\nu + pg_{\mu\nu}, \tag{7.49}$$

[10] In the Minkowski spacetime $\mathbb{R}^{n,1}$ with the metric $\mathrm{d}s^2 = (\mathrm{d}x^0)^2 - \sum_{i=1}^n (\mathrm{d}x^i)^2$, the de Sitter or anti-de Sitter space is the hyperbolic submanifold defined by $(x^0)^2 - \sum_{i=1}^n (x^i)^2 = \mp a^2$ $(a > 0)$.

where v_{μ} is the 4-velocity of the gas particles and $g_{\mu\nu}$ is the spacetime metric. The cosmological principle requires that ρ and p depend on time t only.

We now consider some possible consequences of a homogeneous and isotropic universe in view of the Einstein theory. From (7.16) and (7.25), we can represent the Ricci tensor in terms of the Christoffel symbols by

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\alpha,\nu} - \Gamma^{\alpha}_{\mu\nu,\alpha} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\alpha\nu} - \Gamma^{\alpha}_{\mu\nu}\Gamma^{\beta}_{\alpha\beta}. \tag{7.50}$$

Naturally, we label our coordinates according to $x^0 = t, x^1 = r, x^2 = \theta, x^3 = \chi$. Then the nonzero Christoffel symbols induced from the *Robertson–Walker line element* are [11]

$$\begin{aligned}
\Gamma^0_{11} &= \frac{a(t)a'(t)}{(1-kr^2)}, \quad \Gamma^0_{22} = a(t)a'(t)r^2, \quad \Gamma^0_{33} = a(t)a'(t)r^2\sin^2\theta,\\
\Gamma^1_{01} &= \frac{a'(t)}{a(t)}, \quad \Gamma^1_{11} = \frac{kr}{(1-kr^2)},\\
\Gamma^1_{22} &= -r(1-kr^2), \quad \Gamma^1_{33} = -r(1-kr^2)\sin^2\theta,\\
\Gamma^2_{02} &= \frac{a'(t)}{a(t)}, \quad \Gamma^2_{12} = \frac{1}{r}, \quad \Gamma^2_{33} = -\sin\theta\cos\theta,\\
\Gamma^3_{03} &= \frac{a'(t)}{a(t)}, \quad \Gamma^3_{13} = \frac{1}{r}, \quad \Gamma^3_{23} = \cot\theta,
\end{aligned} \tag{7.51}$$

where $a'(t) = \mathrm{d}a(t)/\mathrm{d}t$. Inserting (7.51) into (7.50), we see that the Ricci tensor $R_{\mu\nu}$ becomes diagonal with

$$\begin{aligned}
R_{00} &= \frac{3a''}{a}, \quad R_{11} = -\frac{aa'' + 2(a')^2 + 2k}{1-kr^2},\\
R_{22} &= -(aa'' + 2(a')^2 + 2k)r^2,\\
R_{33} &= -(aa'' + 2(a')^2 + 2k)r^2\sin^2\theta.
\end{aligned} \tag{7.52}$$

Hence, in view of (7.26), the scalar curvature (7.26) becomes

$$R = \frac{6}{a^2}(aa'' + (a')^2 + k). \tag{7.53}$$

On the other hand, from (7.10) and (7.51), we see that the geodesics of the metric (7.48), which are the trajectories of moving stars and galaxies when net local interactions are neglected, are given by $r, \theta, \chi =$ constant. Thus in (7.49) we have $v_0 = 1$ and $v_i = 0$, $i = 1, 2, 3$. Therefore $T_{\mu\nu}$ is also diagonal with

$$T_{00} = \rho, \quad T_{11} = \frac{pa^2}{1-kr^2}, \quad T_{22} = pa^2r^2, \quad T_{33} = pa^2r^2\sin^2\theta. \tag{7.54}$$

Substituting (7.52), (7.53), and (7.54) into the Einstein equations (7.40), we arrive at the following two equations,

$$\begin{aligned}
\frac{3a''}{a} &= -4\pi G(\rho + 3p),\\
aa'' + 2(a')^2 + 2k &= 4\pi G(\rho - p)a^2.
\end{aligned} \tag{7.55}$$

[11] It should be noted that, nowadays, the computation of Riemannian tensors has been facilitated enormously by available symbolic packages. See [19, 88, 112] and references therein.

Eliminating a'' from these equations, we obtain the well-known *Friedmann equation*

$$(a')^2 + k = \frac{8\pi}{3} G\rho a^2. \tag{7.56}$$

We can show that, in the category of time-dependent solutions, the Einstein cosmological equations, (7.55), are in fact equivalent to the single Friedmann equation (7.56). To this end, recall that both systems are to be subject to the conservation law for the energy-momentum tensor, namely, $T^{\mu\nu}{}_{;\nu} = 0$ or

$$\rho' + 3(\rho + p)\frac{a'}{a} = 0. \tag{7.57}$$

Differentiating (7.56) and using (7.57), we get the first equation in (7.55). Inserting (7.56) into the first equation in (7.55), we get the second equation in (7.55).

The relative rate of change of the radius of the universe is recognized as *Hubble's 'constant'*, $H(t)$, which is given by

$$H(t) = \frac{a'(t)}{a(t)}. \tag{7.58}$$

Recent estimates for Hubble's constant put it at about $(18 \times 10^9 \text{ years})^{-1}$. In particular, $a' > 0$ at present. However, since the first equation in (7.55) indicates that $a'' < 0$ everywhere, we can conclude that $a' > 0$ for all time in the past. In other words, *the universe has undergone a process of expansion in the past.*

We now investigate whether the universe has a beginning time. For this purpose, let t_0 denote the present time and t denote any past time, $t < t_0$. The property $a'' < 0$ again gives us $a'(t) > a'(t_0)$, which implies that

$$a(t_0) - a(t) > a'(t_0)(t_0 - t).$$

Thus there must be a finite time t in the past, $t < t_0$, when a vanishes. Such a time may be defined as the time when the universe begins. For convenience, we may assume that the universe begins at $t = 0$, namely, $a(0) = 0$. Hence we arrive at the general picture of the *Big Bang cosmology* that the universe started at a finite time in the past from a singular point and has been expanding in all its history of evolution.

It is easy to see that the equations (7.55) do not allow *static* (time-independent) solutions. When Einstein applied his gravitational equations to cosmology, he hoped to obtain a homogeneous, isotropic, static, and compact universe. Therefore he was led to his modified equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -8\pi G\, T_{\mu\nu}, \tag{7.59}$$

where Λ is a constant called the *cosmological constant*. Of course the added cosmological term, $\Lambda g_{\mu\nu}$, does not violate the required divergence-free condition. Although static models of the universe have long been discarded since Hubble's discovery in 1929 that the universe is expanding, a nonvanishing cosmological constant gives important implications in the theoretical studies of the early-universe cosmology. In fact, the equations (7.59) may also be rewritten

$$\begin{aligned} G_{\mu\nu} &= -8\pi G \left(T_{\mu\nu} + T^{(\text{vac})}_{\mu\nu} \right), \\ T^{(\text{vac})}_{\mu\nu} &= \frac{\Lambda}{8\pi G} g_{\mu\nu}, \end{aligned} \tag{7.60}$$

where $T^{(\mathrm{vac})}_{\mu\nu}$ is interpreted as the energy-momentum tensor associated with the vacuum: the vacuum polarization of quantum field theory endows the vacuum with a nonzero energy-momentum tensor, which is completely unobservable except by its gravitational effects. In particular,

$$\rho^{(\mathrm{vac})} = T^{(\mathrm{vac})}_{00} = \frac{\Lambda}{8\pi G} \tag{7.61}$$

may be viewed as the mass-energy density of the vacuum. This viewpoint imposes a natural restriction on the sign of the cosmological constant, $\Lambda \geq 0$.

Multiplying (7.59) by the metric $g^{\mu\nu}$ and summing over repeated indices, we find

$$R = 8\pi G\, T + 4\Lambda, \quad T = g^{\mu\nu} T_{\mu\nu}. \tag{7.62}$$

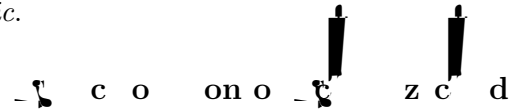
Inserting (7.62) into (7.59), we obtain the more elegant equations

$$R_{\mu\nu} - \Lambda g_{\mu\nu} = -8\pi G \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right). \tag{7.63}$$

In particular, in the absence of matter, we have the vacuum Einstein equations

$$R_{\mu\nu} = \Lambda g_{\mu\nu}. \tag{7.64}$$

Any spacetime satisfying (7.64) is called an *Einstein space* and its metric $g_{\mu\nu}$ is called an *Einstein metric*.



In the situation when both the cosmological constant Λ and the matter energy-momentum tensor vanish, the Einstein equations (7.63) become

$$R_{\mu\nu} = 0, \tag{7.65}$$

which says that the vacuum spacetime is characterized by its Ricci tensor being trivial. In the context of the static spherically symmetric limit with the standard coordinates (t, r, θ, ϕ) used to count for (x^{μ}), the metric element takes the form

$$\mathrm{d}s^2 = f(r)\,\mathrm{d}t^2 - \{h(r)\,\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\}. \tag{7.66}$$

With (7.66), the Ricci tensor becomes diagonal and the diagonal components are given by [144]

$$R_{00} = \frac{1}{2}(fh)^{-\frac{1}{2}}([fh]^{-\frac{1}{2}} f')' + (rfh)^{-1} f', \tag{7.67}$$

$$R_{11} = -\frac{1}{2}(fh)^{-\frac{1}{2}}([fh]^{-\frac{1}{2}} f')' + (rh^2)^{-1} f', \tag{7.68}$$

$$R_{22} = R_{33} = -\frac{1}{2}(rfh)^{-1} f' + (rh^2)^{-1} h' + r^{-2}(1 - h^{-1}). \tag{7.69}$$

Inserting these into (7.65), we have

$$\frac{\mathrm{d}}{\mathrm{d}r} \ln(fh) = 0, \tag{7.70}$$

which leads to $f = K/h$ where $K > 0$ is a constant. Since f appears as the coefficient of $\mathrm{d}t^2$ in the metric element (7.66), we may rescale the time coordinate with $t \mapsto K^{\frac{1}{2}}t$ to normalize K to unity. Thus, substituting $fh = 1$ into (7.69) and setting it to zero, we obtain [144]

$$f' = \frac{1}{r}(1 - f), \tag{7.71}$$

which may be integrated to give us

$$f(r) = 1 + \frac{C}{r}, \tag{7.72}$$

where C is an integrating constant. We can examine that the pair (f, h) where $h = 1/f$ and f given in (7.72) indeed makes (7.67)–(7.69) vanish identically. Therefore, the metric element (7.66) becomes

$$\mathrm{d}s^2 = \left(1 + \frac{C}{r}\right)\,\mathrm{d}t^2 - \left\{\left(1 + \frac{C}{r}\right)^{-1}\,\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right\}. \tag{7.73}$$

It is worth noting that, in the $r \to \infty$ limit, this metric assumes the form of that in the standard *flat* Minkowski spacetime, i.e.,

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \{\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\}. \tag{7.74}$$

Since regularity of the metric element requires $f > 0$, we may be tempted to take $C > 0$ in (7.72). Unfortunately, or fortunately, the situation we are facing here is not so simple and a more elaborate consideration needs to be carried out so that the solution is physically meaningful. Indeed, one may argue [144] that, since the solution should recover that given by the Newton law of gravitation generated from a centralized localized mass, say M, in the region where r is sufficiently large, one is led to the inevitable conclusion

$$C = -2GM, \tag{7.75}$$

where G is the Newton constant.

It may be instructive to take a pause and find how to work (7.75) out. For this purpose, consider the motion of a test particle far away from local region. Relativistically, the trajectory of the particle is parameterized by the proper time, $\mathrm{d}s$, in terms of the spacetime coordinates $x^{\boldsymbol{\mu}}(s)$, and following the geodesic equations, (7.10). We may assume that the speed of the motion is negligible compared with the speed of light. Thus we are able to take the approximation

$$\mathrm{d}s \approx \mathrm{d}t, \quad v^{\boldsymbol{\mu}} \approx (1, 0, 0, 0). \tag{7.76}$$

In view of (7.76), the equations (7.10) become

$$\frac{\mathrm{d}^2 x^{\boldsymbol{\alpha}}}{\mathrm{d}t^2} + \Gamma^{\boldsymbol{\alpha}}_{00} = 0, \quad \alpha = 0, 1, 2, 3. \tag{7.77}$$

However, by virtue of (7.66), we see that the only nontrivial component of (7.77) is at $\alpha = 1$ ($x^1 = r$) which is the single equation

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{f'(r)}{2h} = \frac{1}{2}ff' = -\frac{C}{2r^2}\left(1 + \frac{C}{r}\right), \tag{7.78}$$

which agrees with Newton's law

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{GM}{r^2} \tag{7.79}$$

for r sufficiently large only if C is taken to satisfy (7.75), as stated.

In summary, we have arrived at the solution represented by

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{r}\right)\,\mathrm{d}t^2 - \left\{\left(1 - \frac{2GM}{r}\right)^{-1}\,\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right\}, \tag{7.80}$$

which is the celebrated *Schwarzschild metric* or *Schwarzschild solution* of the Einstein equations. Schwarzschild obtained this solution in 1915, the same year when Einstein published his work of general relativity.[12] There are coordinate singularities at the radius

$$r = r_{\mathrm{S}} = 2GM, \tag{7.81}$$

referred to as the *Schwarzschild radius*. With r_{S}, we rewrite (7.80) as

$$\mathrm{d}s^2 = \left(1 - \frac{r_{\mathrm{S}}}{r}\right)\,\mathrm{d}t^2 - \left\{\left(1 - \frac{r_{\mathrm{S}}}{r}\right)^{-1}\,\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right\}, \tag{7.82}$$

The singular sphere, $r = r_{\mathrm{S}}$, in space, also called the *Schwarzschild surface*, is an *event horizon*. The solution (7.80) represents an empty-space solution or *exterior solution* which is valid outside a spherically distributed massive body occupying the region given by $r \leq R$ (say). In other words, the solution (7.80) is valid for $r > R$. To get the *interior solution* of the Einstein equations in $r < R$, we need to consider the full equations (7.40), with suitably given energy-momentum or stress tensor (7.49) (say) which enables us to match the exterior solution. A detailed discussion of the interior solution may be found in [144].

We now discuss some of the simplest consequences of the Schwarzschild solution.

First, we note that, since the Newton constant G is a tiny quantity, the Schwarzschild radius r_{S} given by (7.81) is usually very small compared with the radius R of the gravitating body of mass M under normal circumstances. For example, $R = 6371$ km and $r_{\mathrm{S}} = 9$ mm for the Earth; $R = 696000$ km and $r_{\mathrm{S}} = 3$ km for the Sun. Thus, normally the Schwarzschild surface is well hidden in the bulk of the gravitating body, $r_{\mathrm{S}} < R$, and there is no singular gravitational effect because as the exterior solution the expression (7.82) is only valid for $r > R$.

[12]Here is some interesting reading provided by Wikipedia: Einstein himself was pleasantly surprised to learn that the field equations admitted exact solutions, because of their prima facie complexity, and because he himself had only produced an approximate solution. Einstein's approximate solution was given in his famous 1915 article on the advance of the perihelion of Mercury. There, Einstein used rectangular coordinates to approximate the gravitational field around a spherically symmetric, non-rotating, non-charged mass. Schwarzschild, in contrast, chose a more elegant 'polar-like' coordinate system and was able to produce an exact solution which he first set down in a letter to Einstein of 22 December 1915, written while Schwarzschild was serving in the war stationed on the Russian front. Schwarzschild concluded the letter by writing: "As you see, the war treated me kindly enough, in spite of the heavy gunfire, to allow me to get away from it all and take this walk in the land of your ideas." In 1916, Einstein wrote to Schwarzschild on this result: "I have read your paper with the utmost interest. I had not expected that one could formulate the exact solution of the problem in such a simple way. I liked very much your mathematical treatment of the subject. Next Thursday I shall present the work to the Academy with a few words of explanation. – Albert Einstein." Schwarzschild accomplished this triumph while serving in the German army during World War I. He died the following year from pemphigus, a painful autoimmune disease which he developed while at the Russian front.

However, if the mass density of the gravitating body is so high that $R < r_{\rm S}$, the singular surface $r = r_{\rm S}$ gives rise to rich and complicated gravitational properties of the spacetime. To illustrate, we may consider 'light' propagation along the radial direction, characterized by the *null proper time* condition, $\mathrm{d}s^2 = 0$, subject to $\mathrm{d}\theta = 0$ and $\mathrm{d}\phi = 0$. Thus, we obtain

$$\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 = \left(1 - \frac{r_{\rm S}}{r}\right)^2. \tag{7.83}$$

Outside the event horizon, $r > r_{\rm S}$, we see clearly that light is slowed down when it is near a gravitating body. Thus, a gravitating body bends light. Inside the event horizon, $r < r_{\rm S}$, we have [13]

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \begin{cases} \left(\frac{r_{\rm S}}{r} - 1\right), & \text{for outward light,} \\ -\left(\frac{r_{\rm S}}{r} - 1\right), & \text{for inward light.} \end{cases} \tag{7.86}$$

From the first line in (7.86), we see that it takes infinite time for the outward light to reach the event horizon; from the second line in (7.86), we have

$$r + r_{\rm S} \ln(r_{\rm S} - r) = t + C_0, \tag{7.87}$$

which says the light is accelerated and even reach the spacetime singularity $r = 0$ in finite time if $r > 0$ is empty space. Thus we have observed that, inside the event horizon, it is 'easier' to fall towards the center than deviate from it. Indeed, such a space region, enclosed by the event horizon and popularly called the *Schwarzschild black hole*, allows nothing, not even light, to escape from it but rather tends to 'collapse' everything towards its center.

It may also be of interest to present some of the well-known extensions of the Schwarzschild solution.

The Reissner Nordstro[redacted] so ution

First, consider the situation where an electrostatic field comes into the picture. Setting $\mathbf{B} =$ (vanishing magnetic field) and $\mathbf{E} = (Q/r^3)$ (electric field given by the Coulomb law) in the empty space and solving the Einstein equations (7.40) there, we see that the metric (7.80) becomes

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{r} + \frac{GQ^2}{r^2}\right)\mathrm{d}t^2 - \left\{\left(1 - \frac{2GM}{r} + \frac{GQ^2}{r^2}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right\}, \tag{7.88}$$

which is called the *Reissner–Nordstrom metric*. An event horizon occurs when

$$1 - \frac{2GM}{r} + \frac{GQ^2}{r^2} = 0, \tag{7.89}$$

[13] Inside the event horizon, $r < r_{\rm S}$, and the metric becomes

$$\mathrm{d}s^2 = \left(\frac{r_{\rm S}}{r} - 1\right)^{-1}\mathrm{d}r^2 - \left\{\left(\frac{r_{\rm S}}{r} - 1\right)\mathrm{d}t^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\right\}, \tag{7.84}$$

which suggests in view of the spacetime signature that the time and radial coordinates, t and r, may actually switch roles,

$$t \leftrightarrow r, \tag{7.85}$$

inside the black hole. Here, however, in order to keep our discussion at a minimum level of complexity, we avoid such an interpretation but maintain the original meanings of t and r.

which has no solution when Q is sufficiently large so that

$$Q^2 > GM^2. \tag{7.90}$$

In this case, the metric is regular everywhere in $r > 0$. On the other hand, when

$$Q^2 < GM^2, \tag{7.91}$$

the equation (7.89) gives rise to two concentric event horizons with the radii

$$r_\pm = GM \pm \sqrt{G^2M^2 - GQ^2}, \tag{7.92}$$

which merge into a single event horizon with the half of the Schwarzschild radius, $\frac{1}{2}r_{\rm S} = GM$, in the degenerate limit, $Q^2 = GM^2$.

The err so ution

Next, consider a rotating spherically symmetric gravitating body. Recall that, for a mass occupying a region V with mass density ρ and rotating about a certain axis ℓ, the moment of inertia may be calculated by

$$I = \int_{\mathbf{V}} \delta^2(\ \ )\rho(\ \ )\,{\rm d}\ \ , \tag{7.93}$$

where $\delta(\ \ )$ denotes the distance from the point $\ \ $ to the axis ℓ. If the angular velocity of the rotating motion of the mass about ℓ is ω, then the angular momentum is

$$J = I\omega. \tag{7.94}$$

In terms of J, the exterior gravitational metric generated from a spherically distributed mass rotating around the vertical axis with angular velocity ω is found to be

$$\begin{aligned} {\rm d}s^2 &= \rho^{-2}(\rho^2 - r_{\rm S}r){\rm d}t^2 - \rho^2\left(r^2 - r_{\rm S}r + \alpha^2\right)^{-1}{\rm d}r^2 - \rho^2{\rm d}\theta^2 \\ &\quad -(r^2 + \alpha^2 + r_{\rm S}r\alpha^2\rho^{-2}\sin^2\theta)\sin^2\theta\,{\rm d}\phi^2 + 2r_{\rm S}r\alpha\rho^{-2}\sin^2\theta\,{\rm d}t{\rm d}\phi, \end{aligned} \tag{7.95}$$

where $\alpha = J/M$ and $\rho^2 = r^2 + \alpha^2\cos^2\theta$. This metric is known as the *Kerr metric*. Considering the conditions $g_{\mathbf{tt}} = 0$ and $g_{\mathbf{rr}} = 0$ again, we see that the event horizons would occur at the radii

$$r^*_\pm = \frac{r_{\rm S} \pm \sqrt{r_{\rm S}^2 - 4\alpha^2\cos^2\theta}}{2}, \quad r^{**}_\pm = \frac{r_{\rm S} \pm \sqrt{r_{\rm S}^2 - 4\alpha^2}}{2}. \tag{7.96}$$

The above extensions allow us to come up with the concepts of *charged black holes* and *rotating black holes*. For details, see [92].

The above two examples have shown us, in view of general relativity, the important fact that, gravity, or rather spacetime curvature, can be generated from electricity or spinning motion, as well as mass. Besides, the presence of event horizons or black holes is seen to be *a universal phenomenon* dictated by the Einstein equations.

It is worth noting that the work of Schwarzschild may further be extended to incorporate the situation of the black holes generated from charged *and* rotating masses. In such a context, the solution is called the *Kerr–Newman solution* due to E. Newman *et al* [97, 98].

A M[illegible] nd d op c

At this point, it may be interesting to discuss briefly the *Positive Energy Theorem* [103, 119, 120, 121, 122, 148] and the *Penrose Conjecture* [105] in general relativity. In this context, we are interested in an asymptotically Euclidean space-like hypersurface $(\mathcal{M}, g)$ embedded in a four-dimensional Lorentzian spacetime whose metric tensor satisfies the Einstein equations (7.40) for which the energy-momentum tensor $T_{\mu\nu}$ is subject to a physically motivated condition, called the *dominant energy condition*, which implies that the pointwise energy density $\mathcal{H} = T_{00}$ is nonnegative and is in fact bounded from below by the magnitude of the pointwise angular momentum according to

$$\mathcal{H} = T_{00} \geq \sqrt{\sum_{j=1}^{3} T_{0j}^2}. \tag{7.97}$$

The hypersurface $(\mathcal{M}, g)$ is said to be *asymptotically Euclidean* if there is a compact subset $\mathcal{K}$ in $\mathcal{M}$ such that $\mathcal{M} \setminus \mathcal{K}$ is the union of finitely many '*ends*' of $\mathcal{M}$ so that each end is diffeomorphic to $\mathbb{R}^3 \setminus B_r$ where B_r denotes the ball of radius $r > 0$ centered at the origin. Under this diffeomorphism, the metric g_{jk} of $\mathcal{M}$ at each end can be represented near infinity by

$$\begin{aligned} g_{jk}(x) &= \delta_{jk} + a_{jk}(x), \quad x \in \mathbb{R}^3 \setminus B_r, \\ a_{jk}(x) &= \mathrm{O}(|x|^{-1}), \quad \partial_{\square} a_{jk}(x) = \mathrm{O}(|x|^{-2}), \quad \partial_{\square}\partial_m a_{jk}(x) = \mathrm{O}(|x|^{-3}), \end{aligned} \tag{7.98}$$

and the second fundamental form (h_{jk}) there satisfies similar asymptotic estimates

$$h_{jk}(x) = \mathrm{O}(|x|^{-2}), \quad \partial_{\square} h_{jk}(x) = \mathrm{O}(|x|^{-3}), \quad x \in \mathbb{R}^3 \setminus B_r. \tag{7.99}$$

Without loss of generality, we assume only one end for convenience. According to Arnowitt, Deser, and Misner [7], the total energy E and the momentum $P_{\square}$ can be defined as the limits of integral fluxes

$$E = \frac{1}{16\pi} \lim_{r\to\infty} \int_{\partial B_r} (\partial_j g_{jk} - \partial_k g_{jj})\, \nu^k \mathrm{d}\sigma_r, \tag{7.100}$$

$$P_{\square} = \frac{1}{8\pi} \lim_{r\to\infty} \int_{\partial B_r} (h_{\square k} - \delta_{\square k} h_{jj})\, \nu^k \mathrm{d}\sigma_r, \tag{7.101}$$

where $\mathrm{d}\sigma_r$ is the area element of ∂B_r and ν denotes the outnormal vector to ∂B_r. The Positive Energy Theorem [103, 119, 120, 121, 148] states that the total energy (7.100) is bounded from below by the total momentum (7.101) by

$$E \geq |P| \tag{7.102}$$

and that $E = 0$ if and only if $(\mathcal{M}, g)$ is the Euclidean space $(\mathbb{R}^3, \delta)$. In the special case when the second fundamental form (h_{jk}) vanishes identically, $P \equiv 0$, the energy E is called the total mass or the *ADM mass*, M_{ADM}, which is always nonnegative, $M_{\mathrm{ADM}} \geq 0$. The *Positive Mass Theorem* [122] states that

$$M_{\mathrm{ADM}} > 0, \tag{7.103}$$

unless the hypersurface $(\mathcal{M}, g)$ is the Euclidean space $(\mathbb{R}^3, \delta)$.

Simply put, the above theorems imply that no energy or mass means no geometry or no gravitation.

Note that, using the Einstein equations (7.40), one may relate the scalar curvature $R_{\mathbf{g}}$ of $(\mathcal{M}, g)$ to the energy density T_{00} by

$$R_{\mathbf{g}} + (h^{\mathbf{k}}_{\mathbf{k}})^2 - h^{\mathbf{k}}_{\mathbf{j}} h^{\mathbf{j}}_{\mathbf{k}} = 16\pi G T_{00}. \tag{7.104}$$

Thus, as a consequence of the dominant energy condition (see (7.97)), the vanishing of the second fundamental form naturally leads to the positivity condition for the scalar curvature,

$$R_{\mathbf{g}} \geq 0. \tag{7.105}$$

Naturally, one would hope to bound M_{ADM} away from zero by some physical information in a gravitational system. For example, one may start from considering an isolated blackhole of mass $M > 0$ whose spacetime metric is known to be given by the Schwarzschild line element (7.80). It can be checked that the spatial slice at any fixed t has the property that its second fundamental form vanishes and that its ADM mass is the same as the black hole mass M. In this case, the singular surface or the event horizon, Σ, of the black hole is a sphere of radius $r_{\mathbf{s}} = 2GM$ whose surface area has the value

$$\mathrm{Area}(\Sigma) = 4\pi r_{\mathbf{s}}^2 = 16\pi G^2 M^2. \tag{7.106}$$

The *Penrose Conjecture* [105] states that the total energy E of the spacetime defined in (7.100) is bounded from below by the total surface area of its apparent horizon Σ, which coincides with the event horizon in the case of a Schwarzschild blackhole, by

$$16\pi G^2 E^2 \geq \mathrm{Area}(\Sigma). \tag{7.107}$$

In the special case when the second fundamental form of the hypersurface M vanishes, (7.107) becomes

$$16\pi G^2 M_{\mathrm{ADM}}^2 \geq \mathrm{Area}(\Sigma), \tag{7.108}$$

which is referred to as the *Riemannian Penrose Inequality*, for which the lower bound may be saturated only in the Schwarzschild limit [16, 17, 18, 62, 63].

C d o c nd C n- on on

We have seen that gauge theory in three dimensions allows the existence of magnetically and electrically charged particle-like solutions called dyons. Naturally, it will be interesting to know whether there are dyons in two dimensions. That is, whether there are magnetically and electrically charged static vortices in gauge theory. To answer this question, Julia and Zee studied the Abelian Higgs gauge field theory model in their now classic 1975 paper [67]. Using a radially symmetric field configuration ansatz and assuming a sufficiently fast decay rate at spatial infinity, they were able to conclude that a finite-energy static solution of the equations of motion over the $(2 + 1)$-dimensional Minkowski spacetime must satisfy the temporal gauge condition, and thus, is necessarily electrically neutral. This result, referred here as the Julia–Zee theorem, leads to many interesting consequences. For example, it makes it transparent that the static Abelian Higgs model is exactly the Ginzburg–Landau theory [51] which is purely magnetic [66, 99]. Since the work of Julia and Zee [67], it has been accepted [38, 60, 65, 70, 104, 142, 143] that, in order to obtain both electrically and magnetically charged static vortices, one needs to introduce into the Lagrangian action density the Chern–Simons topological terms [30, 31], which is an essential construct in anyon physics [146, 147]. See also [47]. The pur

Recall that the classical Abelian Higgs theory over the $(2+1)$-dimensional spacetime is governed by the Lagrangian action density (4.32) and the associated equations of motion are (4.33)–(4.35). In the static situation, the operator $\partial_0 = 0$ nullifies everything. Hence the electric charge density ρ becomes

$$\rho = J^0 = \frac{\mathrm{i}}{2}(\overline{\phi} D^0\phi - \phi\overline{D^0\phi}) = -A_0|\phi|^2, \tag{8.1}$$

where $D_{\boldsymbol{\mu}} = \partial_{\boldsymbol{\mu}} + \mathrm{i}A_{\boldsymbol{\mu}}$ is the renormalized gauge-covariant derivative, and a nontrivial temporal component of the gauge field, A_0, is necessary for the presence of electric charge. On the other hand, the $\mu = 0$ component of the left-hand side of the Maxwell equation (4.34) is

$$\partial_{\boldsymbol{\nu}} F^{0\boldsymbol{\nu}} = \partial_{\mathbf{i}}(F_{\mathbf{i}0}) = \partial_{\mathbf{i}}^2 A_0 = \Delta A_0. \tag{8.2}$$

Consequently, the static version of the equations of motion (4.33)–(4.35) may be written as

$$\begin{aligned} D_{\mathbf{i}}^2\phi &= 2V'(|\phi|^2)\phi - A_0^2\phi, && (8.3)\\ \partial_{\mathbf{j}} F_{\mathbf{ij}} &= \frac{\mathrm{i}}{2}(\overline{\phi} D_{\mathbf{i}}\phi - \phi\overline{D_{\mathbf{i}}\phi}), && (8.4)\\ \Delta A_0 &= |\phi|^2 A_0, && (8.5) \end{aligned}$$

in which (8.5) is the Gauss law. Moreover, since the energy-momentum (stress) tensor has the form

$$T_{\boldsymbol{\mu\nu}} = -\eta^{\boldsymbol{\mu'\nu'}} F_{\boldsymbol{\mu\mu'}} F_{\boldsymbol{\nu\nu'}} + \frac{1}{2}(D_{\boldsymbol{\mu}}\phi\overline{D_{\boldsymbol{\nu}}\phi} + \overline{D_{\boldsymbol{\mu}}\phi} D_{\boldsymbol{\nu}}\phi) - \eta_{\boldsymbol{\mu\nu}}\mathcal{L}, \tag{8.6}$$

the Hamiltonian density is given by

$$\mathcal{H} = T_{00} = \frac{1}{2}|\partial_{\mathbf{i}} A_0|^2 + \frac{1}{2}A_0^2|\phi|^2 + \frac{1}{4}F_{\mathbf{ij}}^2 + \frac{1}{2}|D_{\mathbf{i}}\phi|^2 + V(|\phi|^2), \tag{8.7}$$

so that the finite-energy condition reads

$$\int_{\mathbb{R}^2} \mathcal{H}\,\mathrm{d}x < \infty. \tag{8.8}$$

With the above formulation, the Julia–Zee theorem [67] may be stated as: *Suppose that $(A_0, A_{\mathbf{i}}, \phi)$ is a finite-energy solution of the static Abelian Higgs equations (8.3)–(8.5) over $\mathbb{R}^2$. Then either $A_0 = 0$ everywhere if ϕ is not identically zero or $A_0 \equiv$ constant and the solution is necessarily electrically neutral. In other words, the static Abelian Higgs model is exactly the Ginzburg–Landau theory.*

Here is a proof [132] of the theorem which relies on a crucial choice of a test function so that the argument is valid exactly in two dimensions.

Let $0 \le \eta \le 1$ be of compact support and define for $M > 0$ fixed the truncated function

$$A_0^{\mathbf{M}} = \begin{cases} M & \text{if } A_0 > M, \\ A_0 & \text{if } |A_0| \le M, \\ -M & \text{if } A_0 < -M. \end{cases} \tag{8.9}$$

Then, multiplying (8.5) by $\eta A_0^{\mathbf{M}}$ and integrating, we have

$$\int_{\mathbb{R}^2} [\eta \nabla A_0 \cdot \nabla A_0^{\mathbf{M}} + A_0^{\mathbf{M}} \nabla A_0 \cdot \nabla\eta + \eta|\phi|^2 A_0^{\mathbf{M}} A_0]\,\mathrm{d}x = 0. \tag{8.10}$$

Using (8.9) in (8.10), we find

$$\begin{aligned}
&\int_{\{|\mathbf{A}_0|<\mathbf{M}\}\cap\,\mathrm{supp}(\boldsymbol{\eta})} \eta|\phi|^2 A_0^2\,\mathrm{d}x + M^2\int_{\{|\mathbf{A}_0|>\mathbf{M}\}\cap\,\mathrm{supp}(\boldsymbol{\eta})} \eta|\phi|^2\,\mathrm{d}x \\
&+\int_{\{|\mathbf{A}_0|<\mathbf{M}\}\cap\,\mathrm{supp}(\boldsymbol{\eta})} \eta|\nabla A_0|^2\,\mathrm{d}x \\
\leq\;& M\bigg(\int_{\mathbb{R}^2}|\nabla A_0|^2\,\mathrm{d}x\bigg)^{\frac{1}{2}}\bigg(\int_{\mathbb{R}^2}|\nabla\eta|^2\,\mathrm{d}x\bigg)^{\frac{1}{2}}. \qquad (8.11)
\end{aligned}$$

For $R>0$, we now choose η to be a logarithmic cutoff function given as

$$\eta=\begin{cases} 1 & \text{if } |x|<R,\\ 2-\frac{\ln|\mathbf{x}|}{\ln\mathbf{R}} & \text{if } R\leq|x|\leq R^2,\\ 0 & \text{if } |x|>R^2. \end{cases} \qquad (8.12)$$

Then

$$\int_{\mathbb{R}^2}|\nabla\eta|^2\,\mathrm{d}x=\frac{2\pi}{\ln R}. \qquad (8.13)$$

Using (8.13) in (8.11) gives

$$\begin{aligned}
&\int_{\{|\mathbf{A}_0|<\mathbf{M}\}\cap\mathbf{B}_R}|\phi|^2A_0^2\,\mathrm{d}x+\int_{\{|\mathbf{A}_0|<\mathbf{M}\}\cap\mathbf{B}_R}|\nabla A_0|^2\,\mathrm{d}x\\
\leq\;&\int_{\{|\mathbf{A}_0|<\mathbf{M}\}\cap\mathbf{B}_R}|\phi|^2A_0^2\,\mathrm{d}x+M^2\int_{\{|\mathbf{A}_0|>\mathbf{M}\}\cap\mathbf{B}_R}|\phi|^2\,\mathrm{d}x+\int_{\{|\mathbf{A}_0|<\mathbf{M}\}\cap\mathbf{B}_R}|\nabla A_0|^2\,\mathrm{d}x\\
\leq\;&M\frac{\Big(2\pi\int_{\mathbb{R}^2}|\nabla A_0|^2\,\mathrm{d}x\Big)^{\frac{1}{2}}}{(\ln R)^{\frac{1}{2}}}. \qquad (8.14)
\end{aligned}$$

The right hand side of (8.14) tends to zero as R tends to infinity. Letting M tend to infinity proves conclusion.

The Julia–Zee theorem is also valid for non-Abelian gauge field theory. See [132].

Chern–Simons term and gauge symmetry

For simplicity, we again concentrate on the Abelian situation in $(2+1)$ dimensions. With the gauge field $A_{\boldsymbol{\mu}}$ and the induced field tensor $F_{\boldsymbol{\mu\nu}}=\partial_{\boldsymbol{\mu}}A_{\boldsymbol{\nu}}-\partial_{\boldsymbol{\nu}}A_{\boldsymbol{\mu}}$, the *Chern–Simons term* reads

$$\frac{1}{4}\epsilon^{\boldsymbol{\alpha\mu\nu}}A_{\boldsymbol{\alpha}}F_{\boldsymbol{\mu\nu}}, \qquad (8.15)$$

where $\epsilon^{\boldsymbol{\alpha\mu\nu}}$ is the skew-symmetric Kronecker symbol with $\epsilon^{012}=1$. It is clear that the term (8.15) is *not invariant* under the gauge transformation

$$A_{\boldsymbol{\mu}}\mapsto A_{\boldsymbol{\mu}}-\mathrm{i}\partial_{\boldsymbol{\mu}}\omega. \qquad (8.16)$$

However, such a violation of gauge symmetry is only local but not global. In other words, the integral of (8.15), i.e.,

$$\int\frac{1}{4}\epsilon^{\boldsymbol{\alpha\mu\nu}}A_{\boldsymbol{\alpha}}F_{\boldsymbol{\mu\nu}}\,\mathrm{d}x, \qquad (8.17)$$

remains gauge invariant, as may be checked easily through an integration by parts that

$$\begin{aligned}\int \epsilon^{\alpha\mu\nu}(\partial_\alpha\omega)F_{\mu\nu}\,\mathrm{d}x &= \int \epsilon^{\alpha\mu\nu}([\partial_\nu\partial_\alpha\omega]A_\mu - [\partial_\mu\partial_\alpha\omega]A_\nu)\,\mathrm{d}x\\ &= \int \epsilon^{\alpha\mu\nu}(\partial_\nu\partial_\alpha\omega)A_\mu\,\mathrm{d}x - \int \epsilon^{\nu\alpha\mu}(\partial_\alpha\partial_\nu\omega)A_\mu\,\mathrm{d}x\\ &= \int \epsilon^{\alpha\mu\nu}(\partial_\nu\partial_\alpha\omega - \partial_\alpha\partial_\nu\omega)A_\mu\,\mathrm{d}x = 0. \end{aligned} \tag{8.18}$$

Thus, there is no violation of gauge symmetry when a Chern–Simons term is added into an action density.

With the above preparation, the Abelian Chern–Simons–Higgs Lagrangian density introduced in [104, 142], which minimally extends the classical Abelian Higgs model [66, 99], defined over the Minkowski spacetime $\mathbb{R}^{2,1}$, may be written in the form

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{\kappa}{4}\varepsilon^{\mu\nu\alpha}A_\mu F_{\nu\alpha} + \frac{1}{2}D_\mu\phi\overline{D^\mu\phi} - \frac{\lambda}{8}(|\phi|^2-1)^2, \tag{8.19}$$

where κ is a constant referred to as the *Chern–Simons coupling parameter*. In some literature, the quantity $k \equiv \pi\kappa$ is also called the *level of the Chern–Simons theory* so that the Chern–Simons term in the action density assumes the form

$$\frac{k}{4\pi}\epsilon^{\alpha\mu\nu}A_\alpha F_{\mu\nu}. \tag{8.20}$$

The extremals of the Lagrangian density (8.19) formally satisfy its Euler–Lagrange equations, or the Abelian *Chern–Simons–Higgs equations* [104],

$$D_\mu D^\mu\phi = \frac{\lambda}{2}\phi(1-|\phi|^2), \tag{8.21}$$

$$\partial_\nu F^{\mu\nu} - \frac{\kappa}{2}\varepsilon^{\mu\nu\alpha}F_{\nu\alpha} = -J^\mu. \tag{8.22}$$

in which (8.22) expresses the *modified Maxwell equations* so that the current density J^μ is given by

$$J^\mu = \frac{\mathrm{i}}{2}(\overline{\phi}D^\mu\phi - \phi\overline{D^\mu\phi}), \tag{8.23}$$

as before. Since we will consider static configurations only so that all the fields are independent of the temporal coordinate, $t = x^0$, we have

$$\rho = J^0 = -A_0|\phi|^2. \tag{8.24}$$

Besides, also recall that the electric field $\mathbf{E} = E^j$, in the spatial plane, and magnetic fields H, perpendicular to the spatial plane, induced from the gauge field A_μ are

$$E^j = \partial_j A_0, \quad j = 1,2; \quad H = F_{12}, \tag{8.25}$$

respectively. The *static version of the Chern–Simons–Higgs equations* (8.21) and (8.22) take the explicit form

$$D_j^2\phi = \frac{\lambda}{2}(|\phi|^2-1)\phi - A_0^2\phi, \tag{8.26}$$

$$\partial_k F_{jk} - \kappa\,\varepsilon_{jk}\partial_k A_0 = \frac{\mathrm{i}}{2}(\overline{\phi}D_j\phi - \phi\overline{D_j\phi}), \tag{8.27}$$

$$\Delta A_0 = \kappa F_{12} + |\phi|^2 A_0. \tag{8.28}$$

On the other hand, since the Chern–Simons term gives rise to a *topological invariant* which is independent of the spacetime metric, $\eta_{\mu\nu}$, it makes no contribution to the energy-momentum tensor $T_{\mu\nu}$ of the action density (8.19) which may be calculated as

$$T_{\mu\nu} = -\eta^{\alpha\beta} F_{\mu\alpha} F_{\nu\beta} + \frac{1}{2}([D_\mu\phi]\overline{[D_\nu\phi]} + \overline{[D_\mu\phi]}[D_\nu\phi]) - \eta_{\mu\nu}\mathcal{L}_0, \tag{8.29}$$

where $\mathcal{L}_0$ is obtained from the Lagrangian (8.19) by setting $\kappa = 0$. Hence, it follows that the Hamiltonian $\mathcal{H} = T_{00}$ or the energy density of the theory is given by

$$\mathcal{H} = \frac{1}{2}|\nabla A_0|^2 + \frac{1}{2}|\phi|^2 A_0^2 + \frac{1}{2}F_{12}^2 + \frac{1}{2}(|D_1\phi|^2 + |D_2\phi|^2) + \frac{\lambda}{8}(|\phi|^2 - 1)^2, \tag{8.30}$$

which is positive-definite and the terms in (8.30) not containing A_0 are exactly those appearing in the classical Abelian Higgs model [66, 99]. Thus, the finite-energy condition

$$E(\phi, A_0, A_{\mathbf{j}}) = \int_{\mathbb{R}^2} \mathcal{H}(A_0, A_{\mathbf{j}}, \phi)(x)\,\mathrm{d}x < \infty \tag{8.31}$$

leads us to arriving at the following natural asymptotic behavior of the fields A_0, $A_{\mathbf{j}}$, and ϕ,

$$A_0,\, \partial_{\mathbf{j}} A_0 \;\to\; 0, \tag{8.32}$$

$$F_{12} \;\to\; 0, \tag{8.33}$$

$$|\phi| \to 1,\; |D_1\phi|,\; |D_2\phi| \;\to\; 0, \tag{8.34}$$

as $|x| \to \infty$. In analogue to the Abelian Higgs model [66, 99], we see that a finite-energy solution of the Chern–Simons–Higgs equations (8.26)–(8.28) should be classified by the winding number, say $N \in \mathbb{Z}$, of the complex scalar field ϕ near infinity, which is expected to give rise to the total quantized magnetic charge (or magnetic flux).

The resolution of the aforementioned open problem for the existence of charged vortices in the full Chern–Simons–Higgs theory amounts to prove that, for any integer N, the coupled nonlinear elliptic equations (8.26)–(8.28) over $\mathbb{R}^2$ possess a smooth solution $(A_0, A_{\mathbf{j}}, \phi)$ satisfying the finite-energy condition (8.31) and natural boundary conditions (8.32)–(8.34) so that the winding number of ϕ near infinity is N.

Below is the main existence theorem [28], which solves the above problem.

For any given integer N, the Chern–Simons–Higgs equations (8.26)–(8.28) over $\mathbb{R}^2$ have a smooth finite-energy solution $(\phi, A_0, A_{\mathbf{j}})$ satisfying the asymptotic properties (8.32)–(8.34) as $|x| \to \infty$ such that the winding number of ϕ near infinity is N, which is also the algebraic multiplicity of zeros of ϕ in $\mathbb{R}^2$, and the total magnetic charge $Q_{\mathbf{m}}$ and electric charge $Q_{\mathbf{e}}$ are given by the quantization formulas

$$Q_{\mathbf{m}} \;=\; \frac{1}{2\pi}\int_{\mathbb{R}^2} F_{12}\,\mathrm{d}x = N, \tag{8.35}$$

$$Q_{\mathbf{e}} \;=\; \frac{1}{2\pi}\int_{\mathbb{R}^2} \rho\,\mathrm{d}x = \kappa N. \tag{8.36}$$

Such a solution represents an N-vortex soliton which is indeed both magnetically and electrically charged.

It should be noted that the problem we encounter here is that the static Chern–Simons–Higgs equations (8.26)–(8.28) are not the Euler–Lagrange equations of the energy functional (8.31)

defined by the Hamiltonian density (8.30) but the action functional defined by the Lagrangian density (8.19), which may be written explicitly as

$$\begin{aligned} I(\phi, A_0, A_{\mathbf{j}}) &= \frac{1}{2}\int_{\mathbb{R}^2}\left(F_{12}^2 + |D_1\phi|^2 + |D_2\phi|^2 + \frac{\lambda}{4}(|\phi|^2-1)^2\right)\,\mathrm{d}x \\ &\quad -\frac{1}{2}\int_{\mathbb{R}^2}\left(|\nabla A_0|^2 + A_0^2|\phi|^2 + 2\kappa A_0 F_{12}\right)\,\mathrm{d}x. \end{aligned} \tag{8.37}$$

Thus, we need to find a critical point of the indefinite action functional (8.37) under the finite-energy condition (8.31) and the topological constraint $Q_{\mathsf{m}} = N$ expressed in (8.35).

In order to tackle the difficulty arising from the negative part of the action functional, we introduce the constraint

$$\int_{\mathbb{R}^2}\left\{\nabla A_0\cdot\nabla\tilde{A}_0 + |\phi|^2 A_0\tilde{A}_0 + \kappa F_{12}\tilde{A}_0\right\}\,\mathrm{d}x = 0, \quad \forall\tilde{A}_0, \tag{8.38}$$

for each pair of fixed ϕ and $A_{\mathbf{j}}$, maintaining the finite-energy condition. In particular, when $\tilde{A}_0 = A_0$, we have

$$\int_{\mathbb{R}^2}\kappa A_0 F_{12}\,\mathrm{d}x = -\int_{\mathbb{R}^2}\{|\nabla A_0|^2 + A_0^2|\phi|^2\}\,\mathrm{d}x. \tag{8.39}$$

Substituting (8.39) into (8.37), we see that the action functional becomes positive definite, which resolves the issue of dealing with an indefinite action functional in the original (unconstrained) setting.

Note that what is interesting in the Chern–Simons–Higgs context is that the electric charge Q_{e} is also quantized topologically. Such a property is naturally expected since the finite-energy condition (8.31) for the equations (8.26)–(8.28) implies the vanishing property

$$\int_{\mathbb{R}^2}\Delta A_0\,\mathrm{d}x = 0. \tag{8.40}$$

Thus, integrating (8.28) and using (8.24), we arrive at $Q_{\mathsf{e}} = \kappa Q_{\mathsf{m}}$, which gives us (8.36).

Combining the above existence theorem with the Julia–Zee theorem, we arrive at the conclusion: *The Abelian static Chern–Simons–Higgs equations (8.26)–(8.28), which are the Euler–Lagrange equations of the minimally coupled action density (8.19), have a nontrivial finite-energy solution if and only if the Chern–Simons term is present, which is characterized by the condition* $\kappa \neq 0$. *In such a situation, electricity and magnetism must co-exist.*

For extensions of this study to the case of existence of *non-Abelian Chern–Simons–Higgs vortices*, see [28].

[illegible] dz p o [illegible]

Related to the Chern–Simons theory and relevant in modeling electroweak interaction, Rubakov and Tavkhelidze [116] introduced in 1985 the 'Abelian Higgs energy functional'

$$E(\mathbf{A}, u) = \int\left\{\frac{1}{2}|\nabla\times\mathbf{A}|^2 + |\nabla u|^2 + g^2|\mathbf{A}|^2u^2 + \lambda(u^2-v^2)^2\right\}\,\mathrm{d}x, \tag{8.41}$$

over $\mathbb{R}^3$, governing a vector field $\mathbf{A}$ and a real scalar field u, subject to the *prescribed Chern–Simons charge*

$$N_{\mathrm{CS}} = \frac{g^2}{16\pi^2}\int\mathbf{A}\cdot(\nabla\times\mathbf{A})\,\mathrm{d}x, \tag{8.42}$$

where $g, v, \lambda > 0$ are coupling constants. The Euler–Lagrange equations of the problem are

$$\begin{aligned} \Delta u - g^2|\mathbf{A}|^2 u - 2\lambda(u^2 - v^2)u &= 0, && (8.43)\\ \nabla \times (\nabla \times \mathbf{A}) - \xi(\nabla \times \mathbf{A}) + 2g^2u^2\mathbf{A} &= , && (8.44) \end{aligned}$$

where ξ is the Lagrangian multiplier associated with varying the functional

$$I(\mathbf{A}, u) = E(\mathbf{A}, u) - \xi\left(\frac{8\pi^2}{g^2}\right) N_{\mathrm{CS}}(\mathbf{A}, u). \tag{8.45}$$

See [118] for a recent study and [114] for some discussion in the context of a survey. So far, a rigorous mathematical study of such an *Euclidean three-dimensional Chern–Simons problem* has not been carried out yet.

It should be noted that the model (8.41) lacks gauge invariance. In order to recover its gauge invariance, we may replace the real scalar field u by a complex scalar field ϕ and use the gauge-covariant derivative

$$D_{\mathbf{A}}\phi = \nabla\phi - g\mathrm{i}\mathbf{A}\phi \tag{8.46}$$

as before. Since

$$|D_{\mathbf{A}}\phi|^2 = |\nabla\phi|^2 + g^2|\mathbf{A}|^2|\phi|^2 + 2g\mathbf{A} \cdot \mathrm{Im}(\phi\nabla\overline{\phi}), \tag{8.47}$$

we see that the *Rubakov–Tavkhelidze energy* (8.41) is the real-scalar-field version of the *Ginzburg–Landau energy* [51]

$$E(\mathbf{A}, \phi) = \int \left\{ \frac{1}{2}|\nabla \times \mathbf{A}|^2 + |D_{\mathbf{A}}\phi|^2 + \lambda(|\phi|^2 - v^2)^2 \right\} \mathrm{d}x. \tag{8.48}$$

In other words, we may say that the model (8.41) is the Ginzburg–Landau model (8.48) stated in the *unitary gauge*. Hence the equations of motion (8.43)–(8.44) are modified into

$$\begin{aligned} D_{\mathbf{A}}^2\phi &= 2\lambda(|\phi|^2 - v^2)\phi, && (8.49)\\ \nabla \times \nabla \times \mathbf{A} &= \mathrm{i}g(\phi\overline{D_{\mathbf{A}}\phi} - \overline{\phi}D_{\mathbf{A}}\phi) + \xi(\nabla \times \mathbf{A}). && (8.50) \end{aligned}$$

In the *purely superconducting* limit when $|\phi|$ attains its maximum value v everywhere, the equations (8.49)–(8.50) are reduced into the single one governing the gauge potential $\mathbf{A}$:

$$\nabla \times \nabla \times \mathbf{A} = -2g^2v^2\mathbf{A} + \xi(\nabla \times \mathbf{A}). \tag{8.51}$$

In the special case when there is no Chern–Simons invariant present, $\xi = 0$, and we are left with

$$\nabla \times \nabla \times \mathbf{A} = -2g^2v^2\mathbf{A}. \tag{8.52}$$

Finally, recall that the induced magnetic field $\mathbf{B}$ may be expressed as $\mathbf{B} = \nabla \times \mathbf{A}$. Thus, in view of (8.52), we arrive at

$$\Delta\mathbf{B} = 2g^2v^2\mathbf{B}. \tag{8.53}$$

This equation is known as the *London equation* [83], discovered by the London brothers,[14] which clearly indicates that the induced magnetic field in a superconductor becomes '*massive*', and thus, fades out exponentially fast inside the superconductor. Such a statement gave the earliest mathematical proof of the Meissner effect.

[14] Fritz and Heinz London, German-born physicists. The full set of the London equations, although linear, actually are slightly more complicated than (8.53), and give rise to (8.53).

[redacted] od nd d op c

The idea that elementary particles may be described by continuously distributed fields with localized energy concentrations, also called solitons, has a long history. As a result, it will be interesting to know whether there exist static solitons describing particles at rest or in equilibrium.

[redacted] c o nd con nc

We start from the standard Klein–Gordon field theory whose Lagrangian has been defined earlier with an arbitrary potential density $V \geq 0$. In the static limit, the equations of motion become a semilinear elliptic equation

$$\Delta u = 2V'(|u|^2)u, \tag{9.1}$$

which is the Euler–Lagrange equation of the Hamilton energy

$$E(u) = \int_{\mathbb{R}^n} \left(\frac{1}{2}|\nabla u|^2 + V(|u|^2) \right) \mathrm{d}x. \tag{9.2}$$

Therefore, the solutions are simply the critical points of the energy functional.

Suppose that u is a critical point. Then $u_{\boldsymbol{\lambda}}(x) = u(\lambda x)$ is a critical point as well when $\lambda = 1$, which leads us to the assertion

$$\left\{ \frac{\mathrm{d}}{\mathrm{d}\lambda} E(u_{\boldsymbol{\lambda}}) \right\}\Big|_{\boldsymbol{\lambda}=1} = 0. \tag{9.3}$$

On the other hand, if we use $x_{\boldsymbol{\lambda}}$ to denote λx and $\nabla_{\boldsymbol{\lambda}}$ to denote the gradient operator with derivatives in terms of differentiation in $x_{\boldsymbol{\lambda}}$, then

$$\begin{aligned} E(u_{\boldsymbol{\lambda}}) &= \int_{\mathbb{R}^n} \left\{ \frac{1}{2}|\nabla u_{\boldsymbol{\lambda}}|^2 + V(u_{\boldsymbol{\lambda}}) \right\} \mathrm{d}x = \int_{\mathbb{R}^n} \left\{ \frac{1}{2}\lambda^2 |\nabla_{\boldsymbol{\lambda}} u_{\boldsymbol{\lambda}}|^2 + V(u_{\boldsymbol{\lambda}}) \right\} \mathrm{d}x \\ &= \int_{\mathbb{R}^n} \left\{ \frac{1}{2}\lambda^2 |\nabla_{\boldsymbol{\lambda}} u(x_{\boldsymbol{\lambda}})|^2 + V(u(x_{\boldsymbol{\lambda}})) \right\} \lambda^{-\mathbf{n}} \mathrm{d}x_{\boldsymbol{\lambda}} = \int_{\mathbb{R}^n} \left\{ \frac{1}{2}\lambda^{2-\mathbf{n}} |\nabla u|^2 + \lambda^{-\mathbf{n}} V(u) \right\} \mathrm{d}x. \end{aligned} \tag{9.4}$$

Combining (9.3) and (9.4), we obtain the identity

$$(2-n) \int_{\mathbb{R}^n} |\nabla u|^2 \, \mathrm{d}x = 2n \int_{\mathbb{R}^n} V(u) \, \mathrm{d}x. \tag{9.5}$$

Consequently, we see that there is no nontrivial solution if $n \geq 3$ which rules out the most physical dimension. This statement is known as the *Derrick theorem*. (Mathematicians also called the above integral identity the *Pohozaev identity* – see below.) Besides, the case $n = 2$ is interesting only in the absence of potential energy, $V = 0$. Only when $n = 1$, the potential density function V is not subject to any restriction and locally concentrated static solutions can indeed be constructed (which are often called kinks or domain walls).

The Pohoz ev identity n e cursion

This is a clever and sometimes very useful tool in analyzing semilinear partial differential equations. To see what it is, we consider the elliptic boundary value problem

$$-\Delta u = \lambda |u|^{\mathbf{p}-1} u \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \tag{9.6}$$

where Ω is a star-shaped bounded domain in $\mathbb{R}^{\mathbf{n}}$ with smooth boundary and $\lambda > 0$.

Multiplying Δu by $x \cdot \nabla u$, integrating over Ω, and using $u = 0$ on $\partial\Omega$, we have

$$\begin{aligned}
\int_\Omega x^{\mathbf{j}} \partial_{\mathbf{j}} u \partial_{\mathbf{k}} \partial_{\mathbf{k}} u \, \mathrm{d}x &= -n \int_\Omega u \Delta u \, \mathrm{d}x - \int_\Omega x^{\mathbf{j}} u \partial_{\mathbf{k}} \partial_{\mathbf{k}} \partial_{\mathbf{j}} u \, \mathrm{d}x \\
&= -n \int_\Omega u \Delta u \, \mathrm{d}x + \int_\Omega \delta^{\mathbf{j}}_{\mathbf{k}} u \partial_{\mathbf{k}} \partial_{\mathbf{j}} u \, \mathrm{d}x + \int_\Omega x^{\mathbf{j}} \partial_{\mathbf{k}} u \partial_{\mathbf{k}} \partial_{\mathbf{j}} u \, \mathrm{d}x \\
&= (1-n) \int_\Omega u \Delta u \, \mathrm{d}x + \frac{1}{2} \int_\Omega \partial_{\mathbf{j}} (x^{\mathbf{j}} |\nabla u|^2) \, \mathrm{d}x - \frac{n}{2} \int_\Omega |\nabla u|^2 \, \mathrm{d}x \\
&= (1-n) \int_\Omega u \Delta u \, \mathrm{d}x + \frac{1}{2} \int_{\partial\Omega} |\nabla u|^2 (\nu(x) \cdot x) \, \mathrm{d}S - \frac{n}{2} \int_\Omega |\nabla u|^2 \, \mathrm{d}x \\
&= \frac{(n-2)}{2} \int_\Omega |\nabla u|^2 \, \mathrm{d}x + \frac{1}{2} \int_{\partial\Omega} |\nabla u|^2 (\nu(x) \cdot x) \, \mathrm{d}S, \qquad (9.7)
\end{aligned}$$

where $\nu(x)$ denotes the unit outward normal at the boundary point $x \in \partial\Omega$. Besides, multiplying $|u|^{\mathbf{p}-1}u$ by $x \cdot \nabla u$ and integrating, we have

$$\int (x \cdot \nabla u) |u|^{\mathbf{p}-1} u \, \mathrm{d}x = \frac{1}{p+1} \int_\Omega \nabla \cdot (x|u|^{\mathbf{p}+1}) \, \mathrm{d}x - \frac{n}{p+1} \int_\Omega |u|^{\mathbf{p}+1} \, \mathrm{d}x = -\frac{n}{p+1} \int_\Omega |u|^{\mathbf{p}+1} \, \mathrm{d}x. \qquad (9.8)$$

Multiplying the differential equation in (9.6) by $x \cdot \nabla u$ and using (9.7) and (9.8), we obtain

$$\frac{(n-2)}{2} \int_\Omega |\nabla u|^2 \, \mathrm{d}x + \frac{1}{2} \int_{\partial\Omega} |\nabla u|^2 (\nu(x) \cdot x) \, \mathrm{d}S = \frac{\lambda n}{(p+1)} \int_\Omega |u|^{\mathbf{p}+1} \, \mathrm{d}x, \qquad (9.9)$$

which is called the *Pohozaev identity*. When Ω is star-shaped about the origin, we have $x \cdot \nu(x) \geq 0$. Therefore, we get

$$\frac{(n-2)}{2} \int_\Omega |\nabla u|^2 \, \mathrm{d}x \leq \frac{\lambda n}{(p+1)} \int_\Omega |u|^{\mathbf{p}+1} \, \mathrm{d}x. \qquad (9.10)$$

On the other hand, multiplying the differential equation in (9.6) simply by u and integrating, we have

$$\int_\Omega |\nabla u|^2 \, \mathrm{d}x = \lambda \int_\Omega |u|^{\mathbf{p}+1} \, \mathrm{d}x. \qquad (9.11)$$

Combining the above two results, we arrive at

$$\left(\frac{(n-2)}{2} - \frac{n}{(p+1)} \right) \int_\Omega |u|^{\mathbf{p}+1} \, \mathrm{d}x \leq 0, \qquad (9.12)$$

which establishes that the only solution is the trivial solution $u = 0$ when

$$\frac{(n-2)}{2} - \frac{n}{(p+1)} > 0 \qquad (9.13)$$

or

$$n \geq 3, \quad p > \frac{n+2}{n-2}. \qquad (9.14)$$

In other words, nontrivial solutions are possible only when $n = 1, 2$ or $n \geq 3$ and p satisfies

$$\text{the subcritical condition:} \qquad 1 < p < \frac{n+2}{n-2}, \qquad (9.15)$$

$$\text{or the critical condition:} \qquad p = \frac{n+2}{n-2}. \qquad (9.16)$$

The Derric, theore in presence of g uge e ds

We consider again the Derrick theorem over the Minkowski space $\mathbb{R}^{n,1}$. Of course, a finite-energy static solution (ϕ, A) in the temporal gauge $A_0 = 0$ is a critical point of the energy functional

$$E(\phi, A) = \int_{\mathbb{R}^n} \left\{ \frac{1}{4} F_{\mathbf{jk}}^2 + \frac{1}{2} |D_{\mathbf{j}}\phi|^2 + V(|\phi|^2) \right\} \mathrm{d}x.$$

We may use the same argument as before with $\phi^{\boldsymbol{\lambda}}(x) = \phi(\lambda x)$ and $A^{\boldsymbol{\lambda}}(x) = \lambda A(\lambda x)$ to arrive at the new identity

$$(4-n) \int_{\mathbb{R}^n} F_{\mathbf{jk}}^2 \, \mathrm{d}x + 2(2-n) \int_{\mathbb{R}^n} |D_{\mathbf{j}}\phi|^2 \, \mathrm{d}x = 4n \int_{\mathbb{R}^n} V(|\phi|^2) \, \mathrm{d}x.$$

Therefore, with a gauge field, the allowance of spatial dimensions is extended to $n \leq 4$.

The interesting individual cases are listed as follows.

(i) $n = 4$: The matter field sector must be trivial, $V = 0, D_{\mathbf{j}}\phi = 0$ $(j = 1, \cdots, n)$ and only gauge field is present (pure gauge situation). Solitons in this situations are called gauge instantons.

(ii) $n = 3, 2, 1$: All these are allowed with total freedom for choosing V.

Thus we have seen that the presence of gauge field component enhances dimensionality for the existence of static finite energy solutions of field equations.

y od

The Skyrme model [127, 128, 129, 130] is one of the most important particle models in physical three spatial dimensions without gauge fields. In the static case, its the energy functional reads

$$E(\phi) = \int_{\mathbb{R}^3} \Big(\sum_{1 \leq \mathbf{j} \leq 3} |\partial_{\mathbf{j}}\phi|^2 + \sum_{1 \leq \mathbf{j} < \mathbf{k} \leq 3} |\partial_{\mathbf{j}}\phi \wedge \partial_{\mathbf{k}}\phi|^2 \Big) \, \mathrm{d}x, \tag{9.17}$$

where the 'wave' function is a map from $\mathbb{R}^3$ into S^3. It is easy to see that energy has the scaling property

$$E(\phi^{\boldsymbol{\lambda}}) = \int_{\mathbb{R}^3} (\lambda^{-1} \sum_{1 \leq \mathbf{j} \leq 3} |\partial_{\mathbf{j}}\phi|^2 + \lambda \sum_{1 \leq \mathbf{j} < \mathbf{k} \leq 3} |\partial_{\mathbf{j}}\phi \wedge \partial_{\mathbf{k}}\phi|^2) \, \mathrm{d}x, \tag{9.18}$$

so that the corresponding Derrick identity reads

$$\int_{\mathbb{R}^3} \Big(\sum_{1 \leq \mathbf{j} \leq 3} |\partial_{\mathbf{j}}\phi|^2 \Big) \, \mathrm{d}x = \int_{\mathbb{R}^3} \Big(\sum_{1 \leq \mathbf{j} < \mathbf{k} \leq 3} |\partial_{\mathbf{j}}\phi \wedge \partial_{\mathbf{k}}\phi|^2 \Big) \, \mathrm{d}x, \tag{9.19}$$

which renders no objection to the existence of nontrivial critical points. In fact, we may assume that a finite energy map ϕ has a definitive limit at infinity so that ϕ can be viewed as a map from $\mathbb{R}^3 \cup \{\infty\} \approx S^3$ into S^3. In other words, ϕ can be represented by an integer (the homotopy class) in the group

$$\pi_3(S^3) = \mathbb{Z}. \tag{9.20}$$

In fact, such an integer, called the degree of ϕ and denoted as $\deg(\phi)$, has the following elegant integral representation,

$$\deg(\phi) = \frac{1}{2\pi^2} \int_{\mathbb{R}^3} \det(\phi, \partial_1\phi, \partial_2\phi, \partial_3\phi)(x) \, \mathrm{d}x. \tag{9.21}$$

From (9.21), we have

$$\begin{aligned} 6\pi^2|\deg(\phi)| &\leq \sum_{1\leq \mathbf{i}\leq 3}\sum_{1\leq \mathbf{a}\leq 4}\sum_{1\leq \mathbf{j}<\mathbf{k}\leq 3}^{\mathbf{j},\mathbf{k}\neq \mathbf{i}}\sum_{1\leq \mathbf{b}<\mathbf{c}\leq 4}^{\mathbf{b},\mathbf{c}\neq \mathbf{a}} \int_{\mathbb{R}^3} |\partial_{\mathbf{i}}\phi^{\mathbf{a}}|\left|\partial_{\mathbf{j}}\phi^{\mathbf{b}}\partial_{\mathbf{k}}\phi^{\mathbf{c}} - \partial_{\mathbf{j}}\phi^{\mathbf{c}}\partial_{\mathbf{k}}\phi^{\mathbf{b}}\right| \,\mathrm{d}x \\ &\leq \frac{1}{2}\int_{\mathbb{R}^3}\left(\sum_{1\leq \mathbf{i}\leq 3}|\partial_{\mathbf{i}}\phi|^2\,\mathrm{d}x + \sum_{1\leq \mathbf{i}<\mathbf{j}\leq 3}|\partial_{\mathbf{i}}\phi\wedge\partial_{\mathbf{j}}\phi|^2\right)\,\mathrm{d}x, \end{aligned} \tag{9.22}$$

in view of the Schwartz inequality, which leads to the classical *topological lower bound*

$$E(\phi) \geq 12\pi^2|\deg(\phi)|. \tag{9.23}$$

An important unsolved question asks: does the Skyrme energy have a minimizer among the prescribed homotopy (topological) class

$$\mathcal{C}_{\mathbf{N}} = \{\phi \,|\, E(\phi) < \infty,\, \deg(\phi) = N\} \tag{9.24}$$

defined by each integer N? This problem is only solved when $N = \pm 1$ [77]. A relaxed question asks: Does the Skyrme energy have a finite-energy critical point in each given homotopy class? This latter problem is solved for radially symmetric maps [41, 151]. Problems of this type are numerous in field theory physics.

no n dd od

In 1997, Faddeev and Niemi [44] published their seminal work on knotted solitons arising in a quantum field theory model, known as the Faddeev model [42] which may be regarded as a refined formalism of the Skyrme model. This work suggests that knots may indeed be used as candidates to model elementary particles, a proposal first put forth by Lord Kelvin in 1860s. Today, we know that the concept of knots has important applications in science. In the past 100 years, mathematicians have made great progress in topological and combinatorial classifications of knots. In turn, the development of knot theory has also facilitated the advancement of mathematics in several of its frontiers, especially low-dimensional topology. In knot theory, an interesting problem concerns the existence of "ideal knots," which promises to provide a natural link between the geometric and topological contents of knotted structures. This problem has its origin in theoretical physics in which one wants to prove the existence and predict the properties of knots "based on a first principle approach" [100]. In such an approach, one is interested in determining the detailed physical characteristics of a knot such as its energy (mass), geometric conformation, and topological identification, via conditions expressed in terms of temperature, viscosity, electromagnetic, nuclear, and possibly gravitational, interactions, which is also known as an Hamiltonian approach to knots as field-theoretical stable solitons. The Faddeev knots are such structures based a first-principle [8, 9, 43, 44, 100].

In normalized form, the action density of the Faddeev model over the standard Minkowski spacetime $\mathbb{R}^{3,1}$ reads

$$\mathcal{L} = \partial_{\boldsymbol{\mu}}\phi\cdot\partial^{\boldsymbol{\mu}}\phi - \frac{1}{2}F_{\boldsymbol{\mu\nu}}(\phi)F^{\boldsymbol{\mu\nu}}(\phi), \tag{9.25}$$

where the field ϕ assumes its values in the unit 2-sphere in $\mathbb{R}^3$ and

$$F_{\boldsymbol{\mu\nu}}(\phi) = \phi\cdot(\partial_{\boldsymbol{\mu}}\phi\wedge\partial_{\boldsymbol{\nu}}\phi). \tag{9.26}$$

Since ϕ is parallel to $\partial_{\mu}\phi \wedge \partial_{\nu}\phi$, it is seen that

$$F_{\mu\nu}(\phi)F^{\mu\nu}(\phi) = (\partial_{\mu}\phi \wedge \partial_{\nu}\phi) \cdot (\partial^{\mu}\phi \wedge \partial^{\nu}\phi), \tag{9.27}$$

which may be identified with the well-known Skyrme term [127, 128, 129, 130, 154] when one embeds S^2 into $S^3 \approx SU(2)$. Hence, the Faddeev model may be viewed as a refined Skyrme model. In what follows, we shall only be interested in static fields which make the Faddeev energy

$$E(\phi) = \int_{\mathbb{R}^3} \left\{ \sum_{j=1}^{3} |\partial_j \phi|^2 + \frac{1}{2} \sum_{j,k=1}^{3} |F_{jk}(\phi)|^2 \right\} \mathrm{d}x \tag{9.28}$$

finite. The finite-energy condition implies that ϕ approaches a constant vector ϕ_{∞} at spatial infinity (of $\mathbb{R}^3$). Hence we may compactify $\mathbb{R}^3$ into S^3 and view the fields as maps from S^3 to S^2. As a consequence, we see that each finite-energy field configuration ϕ is associated with an integer, $Q(\phi)$, in $\pi_3(S^2) = \mathbb{Z}$ (the set of all integers). In fact, such an integer $Q(\phi)$ is known as the *Hopf invariant* which has the following integral characterization: The differential form

$$F = F_{jk}(\phi)\, \mathrm{d}x^j \wedge \mathrm{d}x^k, \quad j,k = 1,2,3, \tag{9.29}$$

is closed in $\mathbb{R}^3$. Thus, there is a one form, $A = A_j \mathrm{d}x^j$, so that

$$F = \mathrm{d}A. \tag{9.30}$$

Then the Hopf charge $Q(\phi)$ of the map ϕ may be evaluated by the integral

$$Q(\phi) = \frac{1}{16\pi^2} \int_{\mathbb{R}^3} A \wedge F, \tag{9.31}$$

due to J. H. C. Whitehead [145], which is a special form of the *Chern–Simons invariant* [30, 31].

The existence of the Faddeev knotted solitons are realized as the solutions to the problem

$$E_N \equiv \inf\{E(\phi) \,|\, E(\phi) < \infty,\ Q(\phi) = N\}, \qquad N \in \mathbb{Z}, \tag{9.32}$$

referred to as the *Faddeev Knot Problem*.

Thus we encounter a direct minimization problem over the full space $\mathbb{R}^3$. In such a situation, a typical difficulty is that the minimizing sequence may fail to "concentrate" in a local region, which reminds us to look at what the *concentration-compactness principle* of P. L. Lions [81, 82] can offer. A careful examination of the Faddeev Knots Problem indicates that we cannot make direct use of this method due to the lack of several key ingredients in the Faddeev energy (9.28) and in the *Hopf–Whitehead topological integral* (9.31).

The Faddeev Knot Problem was partially solved in [77]. A key tool we used was called later by us as the "*Substantial Inequality*" [78] which may well be explained by what happens in a *nuclear fission process*: When a nucleus fissions, it splits into several smaller fragments. The sum of the masses of these fragments is less than the original mass. The "missing" mass has been converted into energy according to Einstein's equation.

On the other hand, in our general framework of minimization of a physical energy functional E subject to a topological constraint given by an integer invariant class $Q = N$, we may similarly expect an energy splitting of the configuration sequence into finitely many substantial constituents of topological charges $Q = N_s$ $(s = 1, 2, \cdots, k)$. We expect that the charge is

conserved and the energy of the "particle" of charge N splits into the sum of energies $E_{\mathbf{N}_s}$ ($s = 1, 2, \cdots, k$) of the "substantial particles" of respective charges $N_{\mathbf{s}}$ ($s = 1, 2, \cdots, k$). Therefore, we expect to have

$$\begin{aligned} N &= N_1 + N_2 + \cdots + N_{\mathbf{k}} \quad \text{(charge conservation equality)}, & (9.33) \\ E_{\mathbf{N}} &\geq E_{\mathbf{N}_1} + E_{\mathbf{N}_2} + \cdots + E_{\mathbf{N}_k} \quad \text{(energy conservation inequality)}. & (9.34) \end{aligned}$$

Note that (9.34) is read as an energy conservation relation since possible extra energy may be needed for the substances or constituents of energies $E_{\mathbf{N}_1}, E_{\mathbf{N}_2}, \cdots, E_{\mathbf{N}_k}$ to form a bound state or composite particle, of energy $E_{\mathbf{N}}$, and, as a result, the composite particle may carry more energy than the sum of the energies of its substances or constituents. Hence we collectively call the above two relations "the Substantial Inequality" which spells out a first kind of topological growth law describing how energy and topology split in a general minimization process. The importance of this inequality is that it characterizes the situation when concentration occurs for a minimizing sequence. In other words, the charge-energy splitting above is nontrivial ($k \geq 2$) for a certain charge N if and only if concentration fails there.

To see how (9.33) and (9.34) can be used to quickly deduce an existence theorem for the Faddeev minimization problem (9.32) in 3 dimensions, we recall the topological lower bound

$$E(\phi) \geq C|Q(\phi)|^{\frac{3}{4}} \tag{9.35}$$

established by Vakulenko and Kapitanski [141] where $C > 0$ is a universal constant. Hence $E_{\mathbf{N}} > 0$ for any $N \neq 0$.

Define

$$\mathbb{S} = \{N \in \mathbb{Z} \setminus \{0\} \,|\, \text{the Faddeev Problem (9.32) has a solution at } N\}. \tag{9.36}$$

The Faddeev Knot Problem asks whether or not there holds $\mathbb{S} = \mathbb{Z}$.

As a first step toward the above question, we have:

The set $\mathbb{S}$ is not empty.

The proof of this statement [77] amounts to establishing the Substantial Inequality for the Faddeev energy (9.28) and noting that if $\mathbb{S}$ is empty, then the splitting expressed in (9.33) and (9.34) will continue forever, which contradicts the finiteness and positiveness of $E_{\mathbf{N}}$ for any N.

With (9.33), we can learn more about the soluble set $\mathbb{S}$. For example, choose $N_0 \in \mathbb{Z} \setminus \{0\}$ so that

$$E_{\mathbf{N}_0} = \min\{E_{\mathbf{N}} \,|\, N \in \mathbb{Z} \setminus \{0\}\}. \tag{9.37}$$

Then we must have $N_0 \in \mathbb{S}$ because a nontrivial splitting given in the Substantial Inequality will be impossible by the definition of N_0. Thus we can state [77]:

The least energy point in the Faddeev energy spectrum $\{E_{\mathbf{N}} \,|\, N \in \mathbb{Z} \setminus \{0\}\}$ is attainable.

The set $\mathbb{S}$ *is an infinite subset of* $\mathbb{Z}$.
Here is a quick proof of this result.
Otherwise assume that $\mathbb{S}$ is finite. Set

$$N^0 = \max\{N \in \mathbb{S}\} \tag{9.39}$$

and let $N_0 \in \mathbb{S}$ be such that $E_{\mathbf{N}_0} = \min\{E_{\mathbf{N}} \,|\, N \in \mathbb{S}\}$, as defined earlier. Taking repeated decompositions if necessary, we may assume that all the integers $N_1, N_2, \cdots, N_{\mathbf{k}}$ in (9.33) and (9.34) are in $\mathbb{S}$ already. Hence $|N_1|, |N_2|, \cdots, |N_{\mathbf{k}}| \leq N^0$. Thus, in view of (9.33), we have $N \leq kN^0$; in view of (9.34), we have $E_{\mathbf{N}} \geq kE_{\mathbf{N}_0}$

which is a knot invariant. Naturally one expects the energy and the geometric complexity of the knot K to be closely related. Indeed, the combined results in [20, 26] lead to the relation

$$C_1 N(K)^{\mathbf{p}} \leq L(K) \leq C_2 N(K)^{\mathbf{p}}, \tag{9.43}$$

where $C_1, C_2 > 0$ are two universal constants and the exponent p satisfies $3/4 \leq p < 1$ so that in truly three-dimensional situations the preferred value of p is sharply at $p = 3/4$. This relation strikingly resembles the fractional-exponent growth law for the Faddeev knots just discussed and reminds us once more that a sublinear energy growth law with regard to the topological content involved is essential for knotted structures to occur.

The F ddeev[illegible] ode in gener Hopf di[illegible] ensions

In the Faddeev Knot Problem, it is the underlying property and structure of the homotopy group $\pi_3(S^2)$ and the Faddeev energy functional formula that guarantee the validity of the associated sublinear growth law. Generally, it seems that such a property may be related to the notion of quantitative homotopy introduced by Gromov [56]. For example, we may consider the Whitehead integral representation of the Hopf invariant and the "associated" knot energy ala Faddeev. More precisely, let $u : \mathbb{R}^{4\mathbf{n}-1} \to S^{2\mathbf{n}}$ $(n \geq 1)$ be a differentiable map which approaches a constant sufficiently fast at infinite. Denote by Ω the volume element of $S^{2\mathbf{n}}$ and $|S^{2\mathbf{n}}| = \int_{\mathbf{S}^{2n}} \Omega$. Then the integral representation of u in the homotopy group $\pi_{4\mathbf{n}-1}(S^{2\mathbf{n}})$, say $Q(u)$, which is the Hopf invariant of u, is given by

$$Q(u) = \frac{1}{|S^{2\mathbf{n}}|} \int_{\mathbb{R}^{4n-1}} v \wedge u^*(\Omega), \quad \mathrm{d}v = u^*(\Omega). \tag{9.44}$$

We can introduce a generalized Faddeev knot energy for such a map u as follows,

$$E(u) = \int_{\mathbb{R}^{4n-1}} \left(|\mathrm{d}u|^2 + \frac{1}{2} |u^*(\Omega)|^2 \right) \mathrm{d}x. \tag{9.45}$$

For this energy functional, we are able to establish the following generalized sublinear energy growth estimate [79, 80]

$$C_1 |N|^{(4\mathbf{n}-1)/4\mathbf{n}} \leq E_{\mathbf{N}} \leq C_2 |N|^{(4\mathbf{n}-1)/4\mathbf{n}}, \tag{9.46}$$

where

$$E_{\mathbf{N}} = \inf\{E(u) \,|\, E(u) < \infty,\, Q(u) = N\}, \tag{9.47}$$

and $C_1, C_2 > 0$ are universal constants. In particular, we are able to see that the fractional exponent in the generalized growth law is the ratio of the dimension of the domain space and twice of the dimension of the target space.

[illegible] n nd n

In this section, we present a brief introduction to relativistic strings and branes. We first discuss the relativistic motion of a point particle. We then generalize this discussion to consider the Nambu–Goto strings and branes. We next study the Polyakov strings and show that their quantization leads to the critical dimensionality counts of spacetime.

c f o on o p c

Let m be the mass of a free particle with coordinates $(x^{\mathbf{i}})$ in the space $\mathbb{R}^{\mathbf{n}}$, depending on time t. Recall that the Newtonian action and Lagrangian of the moving particle are

$$S = \int L \, \mathrm{d}t, \quad L = \frac{1}{2} m \sum_{\mathbf{i}=1}^{\mathbf{n}} (\dot{x}^{\mathbf{i}})^2, \quad \dot{f} = \frac{\mathrm{d}f}{\mathrm{d}t}. \tag{10.1}$$

On the other hand, relativistically, the motion of the particle follows the trajectory that extremizes the action

$$S = \kappa \int \mathrm{d}s, \tag{10.2}$$

where κ is an undetermined constant and $\mathrm{d}s^2$ is the metric element given by

$$\mathrm{d}s^2 = c^2 \mathrm{d}t^2 - \sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathrm{d}x^{\mathbf{i}})^2 = \eta_{\boldsymbol{\mu\nu}} \mathrm{d}x^{\boldsymbol{\mu}} \mathrm{d}x^{\boldsymbol{\nu}}, \tag{10.3}$$

with $c > 0$ the speed of light, $x^0 = ct$, and $\eta_{\boldsymbol{\mu\nu}} = \mathrm{diag}(1, -1, \cdots, -1)$ the standard Minkowskian metric tensor. In view of (10.2) and (10.3), if we consider the motion of the particle in terms of time t, we have

$$S = \kappa c \int \sqrt{1 - \sum_{\mathbf{i}=1}^{\mathbf{n}} \frac{(\dot{x}^{\mathbf{i}})^2}{c^2}} \, \mathrm{d}t. \tag{10.4}$$

It is clear that (10.2) gives the dynamics of (10.1) in low speed when $\kappa = -mc$. Thus, we arrived at the relativistic action and Lagrangian

$$S = \int L \, \mathrm{d}t, \quad L = -mc^2 \sqrt{1 - \sum_{\mathbf{i}=1}^{\mathbf{n}} \frac{(\dot{x}^{\mathbf{i}})^2}{c^2}}. \tag{10.5}$$

As a consequence, we can compute the associated momentum vector

$$p_{\mathbf{i}} = \frac{\partial L}{\partial \dot{x}^{\mathbf{i}}} = \frac{m \dot{x}^{\mathbf{i}}}{\sqrt{1 - \sum_{\mathbf{i}=1}^{\mathbf{n}} \frac{(\dot{\mathbf{x}}^i)^2}{\mathbf{c}^2}}}, \quad i = 1, \cdots, n, \tag{10.6}$$

which leads to the Hamiltonian

$$H = \sum_{\mathbf{i}=1}^{\mathbf{n}} p_{\mathbf{i}} \dot{x}^{\mathbf{i}} - L = \frac{mc^2}{\sqrt{1 - \sum_{\mathbf{i}=1}^{\mathbf{n}} \frac{(\dot{\mathbf{x}}^i)^2}{\mathbf{c}^2}}}, \tag{10.7}$$

which gives us the relation

$$H^2 = m^2 c^4 + c^2 \sum_{\mathbf{i}=1}^{\mathbf{n}} p_{\mathbf{i}}^2. \tag{10.8}$$

In particular, when the particle is at rest, we obtain the popular-science formula $E = H = mc^2$.

N - o o n

For simplicity, we set the speed of light to unity, $c = 1$, use the notation $^2 = \sum_{i=1}^{n}$, and $\mathbf{v}(t) = \dot{}$. Then the action of a free particle of mass m over a time span $[t_1, t_2]$ reads

$$S = -m \int_{t_1}^{t_2} \sqrt{1 - \mathbf{v}^2}\, \mathrm{d}t. \tag{10.9}$$

We now consider the motion of a free string of a uniform mass density, ρ_0, parametrized by a real parameter, s, with the spatial coordinates given as a parametrized curve,

$$= \quad (s, t), \quad s_1 \le s \le s_2, \tag{10.10}$$

at any fixed time t. Following (10.9), the action for the motion of the infinitesimal portion

$$\mathrm{d}\ell = \left|\frac{\partial}{\partial s}\right| \mathrm{d}s \equiv |\ '|\, \mathrm{d}s, \quad s_1 \le s \le s_2, \tag{10.11}$$

is given by

$$-\rho_0\, \mathrm{d}\ell \int_{t_1}^{t_2} \sqrt{1 - \mathbf{v}^2}\, \mathrm{d}t. \tag{10.12}$$

In the *Nambu–Goto theory* [52], the internal forces between neighboring points along a string do not contribute to the action so that the velocity vector $\mathbf{v}$ is perpendicular to the tangent of the string curve. Thus, we have

$$\mathbf{v} = \frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + a(s,t)\frac{\partial}{\partial s}, \quad \mathbf{v} \cdot \frac{\partial}{\partial s} = 0. \tag{10.13}$$

From (10.13), we can determine the scalar factor $a(s, t)$ and obtain $\mathbf{v}$ as follows,

$$\mathbf{v} = \frac{\partial}{\partial t} - \frac{\left(\frac{\partial \mathbf{x}}{\partial t} \cdot \frac{\partial \mathbf{x}}{\partial s}\right)}{\left(\frac{\partial \mathbf{x}}{\partial s}\right)^2} \frac{\partial}{\partial s}. \tag{10.14}$$

Integrating (10.12), we obtain the total action for a *Nambu–Goto string*,

$$S = -\rho_0 \int_{t_1}^{t_2} \int_{s_1}^{s_2} \sqrt{\left(\frac{\partial}{\partial s}\right)^2 (1 - \mathbf{v}^2)}\, \mathrm{d}s\mathrm{d}t. \tag{10.15}$$

To appreciate the geometric meaning of the action (10.15), we recall that, for two vectors, $x = (x^\mu) = (x^0, \ )$ and $y = (y^\mu) = (y^0, \mathbf{y})$, xy stands for the inner product

$$xy = x^\mu y_\mu = x^\mu \eta_{\mu\nu} y^\nu = x^0 y^0 - \sum_{i=1}^{n} x^i y^i = x^0 y^0 - \quad \cdot \mathbf{y}, \tag{10.16}$$

in the Minkowski spacetime $\mathbb{R}^{n,1}$. Thus, with the notation

$$\dot{f} = \frac{\partial f}{\partial t}, \quad f' = \frac{\partial f}{\partial s}, \tag{10.17}$$

we have

$$\dot{x}x' = \left(1, \frac{\partial}{\partial t}\right)\left(0, \frac{\partial}{\partial s}\right) = -\frac{\partial}{\partial t}\cdot\frac{\partial}{\partial s}, \tag{10.18}$$

$$\dot{x}^2 = 1 - \left(\frac{\partial}{\partial t}\right)^2, \quad x'^2 = -\left(\frac{\partial}{\partial s}\right)^2. \tag{10.19}$$

Therefore,

$$\begin{aligned}\left(\frac{\partial}{\partial s}\right)^2(1-\mathbf{v}^2) &= \left(\frac{\partial}{\partial s}\right)^2 - \left(\frac{\partial}{\partial t}\right)^2\left(\frac{\partial}{\partial s}\right)^2 + \left(\frac{\partial}{\partial t}\cdot\frac{\partial}{\partial s}\right)^2 \\ &= (\dot{x}x')^2 - \dot{x}^2x'^2. \end{aligned}\tag{10.20}$$

Hence the Nambu–Goto action becomes

$$\mathcal{A} = -\rho_0\int\sqrt{(\dot{x}x')^2 - \dot{x}^2x'^2}\,\mathrm{d}s\mathrm{d}t. \tag{10.21}$$

Besides, we may calculate the line element of the embedded 2-surface $x^{\boldsymbol{\mu}} = x^{\boldsymbol{\mu}}(s,t)$, in the flat Minkowski spacetime $\mathbb{R}^{\boldsymbol{n},1}$, parametrized by the parameters s and t by

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}x^{\boldsymbol{\mu}}\eta_{\boldsymbol{\mu\nu}}\mathrm{d}x^{\boldsymbol{\nu}} \\ &= (\dot{x}^{\boldsymbol{\mu}}\mathrm{d}t + x'^{\boldsymbol{\mu}}\mathrm{d}s)\eta_{\boldsymbol{\mu\nu}}(\dot{x}^{\boldsymbol{\nu}}\mathrm{d}t + x'^{\boldsymbol{\nu}}\mathrm{d}s) \\ &= \dot{x}^2\mathrm{d}t^2 + 2\dot{x}x'\mathrm{d}t\mathrm{d}s + x'^2\mathrm{d}s^2 \\ &= h_{\boldsymbol{ab}}\mathrm{d}u^{\boldsymbol{a}}\mathrm{d}u^{\boldsymbol{b}}, \quad a,b = 0,1, \end{aligned}\tag{10.22}$$

where $u^0 = t, u^1 = s$, and

$$(h_{\boldsymbol{ab}}) = \begin{pmatrix} \dot{x}^2 & \dot{x}x' \\ \dot{x}x' & x'^2 \end{pmatrix}. \tag{10.23}$$

From (10.21) and (10.23), we see that the Nambu–Goto string action is simply a surface integral

$$\mathcal{A} = -\rho_0\int_\Omega\sqrt{|h|}\,\mathrm{d}t\mathrm{d}s = -\rho_0\int_{\boldsymbol{S}}\mathrm{d}S, \tag{10.24}$$

where $|h|$ is the absolute value of the determinant of the matrix (10.23) and $\mathrm{d}S$ is the canonical area element of the embedded 2-surface, $(S, \{h_{\boldsymbol{ab}}\})$, in the Minkowski spacetime.

As a comparison, the action (10.21) for a point particle is simply a path integral,

$$\mathcal{A} = -m\int_{\boldsymbol{\tau}_1}^{\boldsymbol{\tau}_2}\sqrt{\dot{x}^2}\,\mathrm{d}\tau = -m\int_{\boldsymbol{C}}\mathrm{d}C, \tag{10.25}$$

where $\mathrm{d}C$ is the line element of the path C parametrized by $x^{\boldsymbol{\mu}} = x^{\boldsymbol{\mu}}(\tau)$, $\tau_1 \le \tau \le \tau_2$, in the Minkowski spacetime.

Therefore the motion of a point particle follows an extremized path, the *world line*, and the motion of a Nambu–Goto string follows an extremized surface, the *world sheet*.

Of course, both (10.24) and (10.25) are parametrization invariant.

Returning to (10.21), using the generalized time and string coordinates, τ and σ, with

$$t = t(\tau,\sigma), \quad s = s(\tau,\sigma), \quad f_{\boldsymbol{\tau}} = \partial_{\boldsymbol{\tau}}f = \frac{\partial f}{\partial\tau}, \quad f_{\boldsymbol{\sigma}} = \partial_{\boldsymbol{\sigma}}f = \frac{\partial f}{\partial\sigma}, \tag{10.26}$$

and setting T_0 to unity, we see that the Nambu–Goto string action becomes

$$S = -\int \sqrt{(x_\tau x_\sigma)^2 - (x_\tau)^2 (x_\sigma)^2}\, \mathrm{d}\tau \mathrm{d}\sigma. \tag{10.27}$$

Using P^τ_μ and P^σ_μ to denote the generalized 'momenta' where

$$P^\tau_\mu = \frac{\partial L}{\partial x^\mu_\tau}, \quad P^\sigma_\mu = \frac{\partial L}{\partial x^\mu_\sigma}, \quad L = -\sqrt{(x_\tau x_\sigma)^2 - (x_\tau)^2 (x_\sigma)^2}. \tag{10.28}$$

Then the equations of motion of the Nambu–Goto string obtained from varying the action (10.27) may be written in the form of the conservation laws

$$\frac{\partial P^\tau_\mu}{\partial \tau} + \frac{\partial P^\sigma_\mu}{\partial \sigma} = 0, \quad \mu = 0, 1, \cdots, n, \tag{10.29}$$

or, more explicitly,

$$\partial_\tau \left(\frac{(x_\tau x_\sigma)\partial_\sigma x_\mu - (x_\sigma)^2 \partial_\tau x_\mu}{\sqrt{(x_\tau x_\sigma)^2 - (x_\tau)^2 (x_\sigma)^2}} \right) + \partial_\sigma \left(\frac{(x_\tau x_\sigma)\partial_\tau x_\mu - (x_\sigma)^2 \partial_\sigma x_\mu}{\sqrt{(x_\tau x_\sigma)^2 - (x_\tau)^2 (x_\sigma)^2}} \right) = 0. \tag{10.30}$$

More generally, consider an embedded $(p+1)$-dimensional hypersurface, or a membrane, parametrized by the coordinates (u^a) $(a = 0, 1, \cdots, p)$, so that the induced metric element is given by

$$\mathrm{d}s^2 = h_{ab} \mathrm{d}u^a \mathrm{d}u^b, \quad h_{ab} = \eta_{\mu\nu} \frac{\partial x^\mu}{\partial u^a} \frac{\partial x^\nu}{\partial u^b}. \tag{10.31}$$

The action of a *p-brane* is given by the volume integral of the hypersurface as follows,

$$S = -T_p \int \mathrm{d}V_{p+1} = -T_p \int \sqrt{|h|}\, \mathrm{d}u^0 \mathrm{d}u^1 \cdots \mathrm{d}u^p, \tag{10.32}$$

where T_p is a positive constant referred to as the *tension of the p-brane*. In particular, a point particle is a 0-brane, and a string is a 1-brane.

We now consider a '*time-independent* n-brane', M, in the Minkowski spacetime $\mathbb{R}^{n,1}$, which may be realized as a graph of a function depending on the spatial coordinates only given by

$$x^0 = f(x^1, \cdots, x^n). \tag{10.33}$$

Using (10.31), we see that the metric tensor (h_{ij}) of M is

$$h_{ij} = \eta_{\mu\nu} \partial_i x^\mu \partial_j x^\nu = \partial_i f \partial_j f - \delta_{ij}, \quad i, j = 1, \cdots, n. \tag{10.34}$$

It can be checked that

$$|h| = |\det(\partial_i f \partial_j f - \delta_{ij})| = 1 - |\nabla f|^2, \tag{10.35}$$

where we have assumed that M is spacelike, $|\nabla f| < 1$. Hence, ignoring the coupling constant, the action of an n-brane which happens to be a graph is

$$S = -\int \sqrt{1 - |\nabla f|^2}\, \mathrm{d}x, \tag{10.36}$$

whose Euler–Lagrange equation reads

$$\nabla \cdot \left(\frac{\nabla f}{\sqrt{1 - |\nabla f|^2}} \right) = 0. \tag{10.37}$$

A well-known theorem of Cheng and Yau [29] states that all solutions of (10.37) over the full space $\mathbb{R}^{\mathbf{n}}$ satisfying $|\nabla f| < 1$ must be affine linear, $f(x^1, \cdots, x^{\mathbf{n}}) = \sum_{\mathbf{i}=1}^{\mathbf{n}} a_{\mathbf{i}} x^{\mathbf{i}} + b$, where $a_{\mathbf{i}}$'s and b are constants.

Minimal hypersurfaces

It may be instructive to compare the above study of an n-brane with its Euclidean space counterpart where we replace the Minkowski spacetime $\mathbb{R}^{\mathbf{n},1}$ with the Euclidean space $\mathbb{R}^{\mathbf{n}+1}$ so that the inherited metric of the embedded n-hypersurface M defined by the graph of the function (10.33) is given by

$$h_{\mathbf{ij}} = \delta_{\boldsymbol{\mu\nu}} \partial_{\mathbf{i}} x^{\boldsymbol{\mu}} \partial_{\mathbf{j}} x^{\boldsymbol{\nu}} = \partial_{\mathbf{i}} f \partial_{\mathbf{j}} f + \delta_{\mathbf{ij}}, \quad i, j = 1, \cdots, n. \tag{10.38}$$

Consequently,

$$|h| = \det(h_{\mathbf{ij}}) = 1 + |\nabla f|^2, \tag{10.39}$$

such that the canonical volume of the hypersurface M reads

$$\mathcal{V}_{\mathbf{M}} = \int \sqrt{|h|}\, \mathrm{d}x = \int \sqrt{1 + |\nabla f|^2}\, \mathrm{d}x. \tag{10.40}$$

Minimizing (10.40) gives us the classical equation

$$\nabla \cdot \left(\frac{\nabla f}{\sqrt{1 + |\nabla f|^2}} \right) = 0, \quad x \in \mathbb{R}^{\mathbf{n}}, \tag{10.41}$$

known as the *minimal hypersurface equation* for *non-parametric minimal hypersurfaces* defined as the graph of a function. The Bernstein theorem for this equation states that all entire solutions are affine linear for $n \leq 7$ (cf. [101] and references therein).

It was Calabi who first observed that the equations (10.37) and (10.41) are equivalent [24] when $n = 2$. A proof of this fact is as follows.

Let u be a solution of (10.41) and

$$p = \partial_1 f, \quad q = \partial_2 f. \tag{10.42}$$

Set $w = \sqrt{1 + p^2 + q^2}$. Then (10.41) reads

$$\partial_1 \left(\frac{p}{w} \right) + \partial_2 \left(\frac{q}{w} \right) = 0. \tag{10.43}$$

Hence, there is a real-valued function U such that

$$\partial_1 U = P \equiv -\frac{q}{w}, \quad \partial_2 U = Q \equiv \frac{p}{w}. \tag{10.44}$$

Therefore, we have

$$1 - P^2 - Q^2 = \frac{1}{w^2} > 0 \tag{10.45}$$

and U is space-like (i.e., $|\nabla U|^2 < 1$). Inserting the relations $p = Qw$, $q = -Pw$, and $w = 1/W$ where $W = \sqrt{1 - P^2 - Q^2}$ into the identity $\partial_2 p = \partial_1 q$, we arrive at

$$\partial_1\left(\frac{P}{W}\right) + \partial_2\left(\frac{Q}{W}\right) = 0. \tag{10.46}$$

Thus, U solves (10.37). The inverse correspondence from (10.37) to (10.41) may be established similarly.

The above *equivalence theorem of Calabi* can be extended into arbitrary n-dimensional settings [152] which give rise to a rich range of open problems.

Conjecture. *Consider a map*

$$\phi : (M, \{h_{\mathbf{ab}}\}) \to (N, \{g_{\boldsymbol{\mu\nu}}\}), \quad \phi(u^0, u^1, \cdots, u^{\mathbf{m}}) = (x^0, x^1, \cdots, x^{\mathbf{n}}), \tag{10.47}$$

where $(M, \{h_{\mathbf{ab}}\})$ and $(N, \{g_{\boldsymbol{\mu\nu}}\})$ are $(m+1)$- and $(n+1)$-dimensional Minkowski manifolds parametrized with the coordinates $(u^{\mathbf{a}})$ and $(x^{\boldsymbol{\mu}})$, respectively. The *Polyakov action* is simply the 'harmonic map' functional defined as

$$S = -\int (D\phi)^2 \, \mathrm{d}V_{\mathbf{h}}, \quad (D\phi)^2 = g_{\boldsymbol{\mu\nu}} h^{\mathbf{ab}} \partial_{\mathbf{a}} x^{\boldsymbol{\mu}} \partial_{\mathbf{b}} x^{\boldsymbol{\nu}}, \quad a, b = 0, 1, \cdots, m, \tag{10.48}$$

where $\mathrm{d}V_{\mathbf{h}}$ is the canonical volume element of $(M, \{h_{\mathbf{ab}}\})$ given by $\mathrm{d}V_{\mathbf{h}} = \sqrt{|h|}\mathrm{d}u^0\mathrm{d}u^1 \cdots \mathrm{d}u^{\mathbf{m}}$, or customarily, for a p-brane,

$$S = \int \mathcal{L} \, \mathrm{d}V_{\mathbf{h}} = -\tau_{\mathbf{p}} \int \sqrt{|\det(h_{\mathbf{ab}})|} \, g_{\boldsymbol{\mu\nu}} h^{\mathbf{ab}} \partial_{\mathbf{a}} x^{\boldsymbol{\mu}} \partial_{\mathbf{b}} x^{\boldsymbol{\nu}} \, \mathrm{d}u^0 \mathrm{d}u^1 \cdots \mathrm{d}u^{\mathbf{p}}. \tag{10.49}$$

where we have attached the constant $\tau_{\mathbf{p}} > 0$ to account for the *Polyakov p-brane* tension.

When $p = 1$, the action (10.49) defines the *Polyakov string* [110] action which is conformally invariant (i.e., the action is invariant under the conformal transformation of the metric, $h_{\mathbf{ab}} \mapsto \Lambda h_{\mathbf{ab}}$). Such an invariance property is also called the *Weyl invariance*. The obvious advantage of the action (10.49) over (10.32) is that the former is quadratic in $x^{\boldsymbol{\mu}}$'s and gives rise to linear equations of motion for the branes. Thus, the theory is much easier to quantize.

In the special case when $(M, \{h_{\mathbf{ab}}\})$ is regarded as a submanifold of the spacetime $(\mathbb{R}^{\mathbf{n},1}, \{\eta_{\boldsymbol{\mu\nu}}\})$ so that the metric $h_{\mathbf{ab}}$ is induced from the map (10.47), we have

$$h_{\mathbf{ab}} = \eta_{\boldsymbol{\mu\nu}} \partial_{\mathbf{a}} x^{\boldsymbol{\mu}} \partial_{\mathbf{b}} x^{\boldsymbol{\nu}}, \tag{10.50}$$

which leads us to

$$h^{\mathbf{ab}} \eta_{\boldsymbol{\mu\nu}} \partial_{\mathbf{a}} x^{\boldsymbol{\mu}} \partial_{\mathbf{b}} x^{\boldsymbol{\nu}} = p + 1. \tag{10.51}$$

Consequently, we see that the Polyakov p-brane action (10.49) reduces into the Nambu–Goto p-brane action (10.32) when $T_{\mathbf{p}} = (p+1)\tau_{\mathbf{p}}$.

There is another point of view regarding the relationship between the Nambu–Goto strings and the Polyakov strings: Extremizing the metric tensor $h_{\mathbf{ab}}$ in the Polyakov p-brane action (10.49), we obtain after neglecting the constant factor $\tau_{\mathbf{p}}$ the *vanishing stress tensor condition*

$$T_{\mathbf{ab}} = 2 g_{\boldsymbol{\mu\nu}} \partial_{\mathbf{a}} x^{\boldsymbol{\mu}} \partial_{\mathbf{b}} x^{\boldsymbol{\nu}} - h_{\mathbf{ab}} \mathcal{L} = 0, \quad \forall a, b, \tag{10.52}$$

which gives us the solution for $h_{\mathbf{ab}}$ as follows,

$$h_{\mathbf{ab}} = \frac{2g_{\boldsymbol{\mu\nu}}\partial_{\mathbf{a}}x^{\boldsymbol{\mu}}\partial_{\mathbf{b}}x^{\boldsymbol{\nu}}}{\mathcal{L}} \equiv \frac{2}{\mathcal{L}}(\partial_{\mathbf{a}}x \cdot \partial_{\mathbf{b}}x). \tag{10.53}$$

Inserting (10.31) into (10.49), we have

$$S = -(p+1)\tau_{\mathbf{p}} \int \sqrt{|\det(\partial_{\mathbf{a}}x \cdot \partial_{\mathbf{b}}x)|} \left(\frac{2}{\mathcal{L}}\right)^{\frac{(p-1)}{2}} \mathrm{d}u, \tag{10.54}$$

which is clearly seen to become a pure volume integral exactly when $p = 1$. In other words, in such a context, the Nambu–Goto and Polyakov string actions are equivalent.

Quantization of the Polyakov string leads to the *string partition function* [110]

$$Z = \int D\varphi \exp\left(-\frac{1}{48\pi}(26 - D) \int \left\{\frac{1}{2}(\partial_{\mathbf{a}}\varphi)^2 + \kappa^2 \mathrm{e}^{\boldsymbol{\phi}}\right\} \mathrm{d}^2u\right), \tag{10.55}$$

where $D = n + 1 = 26$ is the spacetime dimension, the metric $h_{\mathbf{ab}}$ is Euclideanized into $\mathrm{e}^{\boldsymbol{\phi}}\delta_{\mathbf{ab}}$ through a *Wick rotation* $u^0 \mapsto \mathrm{i}u^0$, $\kappa > 0$ is constant, and $\int D\varphi$ denotes the *path integral* over the space of all possible conformal exponents. Note that, as already observed, the Polyakov string action is conformal invariant. However, the partition function (10.55) clearly spells out the fact that such a conformal invariance is no longer valid when $D \neq 26$. Such a phenomenon is called '*conformal anomaly*' and the vanishing of conformal anomaly gives us the unique condition

$$D = 26, \tag{10.56}$$

known as the *critical dimension* of *bosonic string theory*. Furthermore, when *fermions* are present, Polyakov's computation [111] of the partition function of quantized supersymmetric strings, or *superstrings*, gives us the unique condition

$$D = 10, \tag{10.57}$$

to avoid conformal anomaly, again. These results about the *critical dimensions of spacetime* are now standard facts in string theory [155]. Physicists [25, 68, 155] further conjectured that our 10-dimensional universe, $\mathcal{M}_{10}$, is a product of a 4-dimensional spacetime, $\mathcal{M}_4$, and a 6-dimensional compact manifold, $\mathcal{K}_6$, curled up in a tiny but highly sophisticated way, following a formalism called the '*string compactification*' [37, 53], so that the spacetime $\mathcal{M}_4$ is *maximally symmetric* (which implies that $\mathcal{M}_4$ can either be Minkowski, de Sitter, or anti-de Sitter)[15] and $\mathcal{K}_6$ is a *Calabi–Yau manifold* [22, 23, 153].[16]

Note also that, the action stemming out from (10.55), given by

$$L = \int \left\{\frac{1}{2}\partial_{\mathbf{a}}\varphi\partial^{\mathbf{a}}\varphi - \kappa^2 \mathrm{e}^{\boldsymbol{\phi}}\right\} \mathrm{d}^2u, \tag{10.58}$$

and the associated wave equation

$$\varphi_{\boldsymbol{\tau\tau}} - \varphi_{\boldsymbol{\sigma\sigma}} = -\kappa^2 \mathrm{e}^{\boldsymbol{\phi}}, \quad u^0 = \tau, \quad u^1 = \sigma, \tag{10.59}$$

as the equation of motion, are jointly known to define the *Liouville field theory* [34, 48], which is integrable [152] and of independent interest as a toy model.

[15] A manifold is maximally symmetric if it has the same number of symmetries as ordinary Euclidean space. More precisely, a Riemannian manifold is maximally symmetric if it has $\frac{1}{2}n(n+1)$ (n = dimension of the manifold) linearly independent Killing vector fields which generate isometric flows on the manifold.

[16] A Calabi–Yau manifold is a compact Kähler manifold with vanishing first Chern class.

Born–Infeld theory

As a natural development of the topics covered in the previous section, it will be interesting to present an introduction to the *Born–Infeld theory* [11, 12, 13, 14] formulated in the 1933–1934 and revived over the last 20 years due to its relevance in string theory.

Motivation

Recall that one of the major motivations for the introduction of the Born–Infeld electromagnetic field theory [11, 12, 13, 14] is to overcome the *infinity problem* associated with a *point charge* source in the original Maxwell theory. It is observed that, since the Einstein mechanics of special relativity may be obtained from the Newton mechanics by replacing the classical action function $\mathcal{L} = \frac{1}{2}mv^2$ by the relativistic expression

$$\mathcal{L} = mc^2\Big(1 - \sqrt{1 - \frac{v^2}{c^2}}\Big) = b^2\Big(1 - \sqrt{1 - \frac{1}{b^2}mv^2}\Big), \quad b^2 = mc^2, \tag{11.1}$$

so that no physical particle of a positive rest mass m can move at a speed v greater than the speed of light c, it will be acceptable to replace the action function of the Maxwell theory,

$$\mathcal{L} = \frac{1}{2}(\mathbf{E}^2 - \mathbf{B}^2), \tag{11.2}$$

where $\mathbf{E}$ and $\mathbf{B}$ are electric and magnetic fields, respectively, by a corresponding expression of the form

$$\mathcal{L} = b^2\Big(1 - \sqrt{1 - \frac{1}{b^2}(\mathbf{E}^2 - \mathbf{B}^2)}\Big), \tag{11.3}$$

where $b > 0$ is a suitable scaling parameter, often called the Born–Infeld parameter. It is clear that (11.3) defines a nonlinear theory of electromagnetism and the Maxwell theory, (11.2), may be recovered in the weak field limit $\mathbf{E}, \mathbf{B} \to$. Note that the choice of sign in front of the Lagrangian density (11.2) is the opposite of that of Born and Infeld [14] and is widely adopted in contemporary literature. This convention will be observed throughout the notes.

Intrinsically, if (11.2) is replaced by $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, then (11.3) takes the form

$$\mathcal{L} = b^2\Big(1 - \sqrt{1 + \frac{1}{2b^2}F_{\mu\nu}F^{\mu\nu}}\Big), \tag{11.4}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{11.5}$$

is the electromagnetic field strength curvature induced from a gauge vector potential A_μ. More precisely, if we use

$$\mathbf{E} = (E^1, E^2, E^3), \quad \mathbf{B} = (B^1, B^2, B^3) \tag{11.6}$$

to denote the electric and magnetic fields, respectively, as earlier, then there holds the standard identification

$$F^{0i} = -E^i, \quad F^{ij} = -\epsilon^{ijk}B^k, \quad i, j, k = 1, 2, 3, \tag{11.7}$$

which has the following matrix form,

$$(F^{\mu\nu}) = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix}. \tag{11.8}$$

The dual of $F^{\mu\nu}$ reads

$$*F^{\mu\nu} = \tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}. \tag{11.9}$$

From (11.5), there holds again the Bianchi identity

$$\partial_\mu \tilde{F}^{\mu\nu} = 0. \tag{11.10}$$

On the other hand, it is easy to find that the Euler–Lagrange equations of (11.4) are

$$\partial_\mu P^{\mu\nu} = 0, \tag{11.11}$$

$$P^{\mu\nu} = \frac{F^{\mu\nu}}{\sqrt{1+\frac{1}{2\mathfrak{b}^2}F_{\alpha\beta}F^{\alpha\beta}}}. \tag{11.12}$$

Corresponding to the *electric field* $\mathbf{E}$ and *magnetic field* $\mathbf{B}$, we introduce the *electric displacement* $\mathbf{D}$ and *magnetic intensity* $\mathbf{H}$,

$$\mathbf{D} = (D^1, D^2, D^3), \quad \mathbf{H} = (H^1, H^2, H^3), \tag{11.13}$$

and make the identification

$$P^{0i} = -D^i, \quad P^{ij} = -\epsilon^{ijk}H^k, \quad i,j,k = 1,2,3, \tag{11.14}$$

which has the following matrix form,

$$(P^{\mu\nu}) = \begin{pmatrix} 0 & -D^1 & -D^2 & -D^3 \\ D^1 & 0 & -H^3 & H^2 \\ D^2 & H^3 & 0 & -H^1 \\ D^3 & -H^2 & H^1 & 0 \end{pmatrix}. \tag{11.15}$$

Inserting (11.8) into (11.10) and (11.15) into (11.11), we obtain the fundamental *governing equations of the Born–Infeld electromagnetic theory*,

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = \ , \quad \nabla \cdot \mathbf{B} = 0, \tag{11.16}$$

$$-\frac{\partial \mathbf{D}}{\partial t} + \nabla \times \mathbf{H} = \ , \quad \nabla \cdot \mathbf{D} = 0, \tag{11.17}$$

which look exactly like the *vacuum Maxwell equations*, except that, in view of the relations (11.8), (11.12), and (11.15), the fields $\mathbf{E}, \mathbf{B}$ and $\mathbf{D}, \mathbf{H}$ are *related nonlinearly*,

$$\mathbf{D} = \frac{\mathbf{E}}{\sqrt{1+\frac{1}{\mathfrak{b}^2}(\mathbf{B}^2 - \mathbf{E}^2)}}, \tag{11.18}$$

$$\mathbf{H} = \frac{\mathbf{B}}{\sqrt{1+\frac{1}{\mathfrak{b}^2}(\mathbf{B}^2 - \mathbf{E}^2)}}. \tag{11.19}$$

Hence, the Born–Infeld electromagnetism introduces **E**- and **B**-dependent *dielectrics and permeability* 'coefficients',

$$\mathbf{D} = \varepsilon(\mathbf{E}, \mathbf{B})\mathbf{E}, \quad \mathbf{B} = \mu(\mathbf{E}, \mathbf{B})\mathbf{H}. \tag{11.20}$$

If there is an external current source, $(j^{\boldsymbol{\mu}}) = (\rho, \mathbf{j})$, the equation (11.11) will be replaced by

$$\partial_{\boldsymbol{\mu}} P^{\boldsymbol{\mu\nu}} = j^{\boldsymbol{\nu}} \tag{11.21}$$

and equivalently, the equations in (11.17) become

$$-\frac{\partial \mathbf{D}}{\partial t} + \nabla \times \mathbf{H} = \mathbf{j}, \quad \nabla \cdot \mathbf{D} = \rho, \tag{11.22}$$

We now examine the point charge problem.

Consider the electrostatic field generated from a point particle of electric charge q placed at the origin. Then $\mathbf{B} =$, $\mathbf{H} =$, and the Born–Infeld equations become a single one,

$$\nabla \cdot \mathbf{D} = 4\pi q \delta(\ ), \tag{11.23}$$

which can be solved to give us

$$\mathbf{D} = \frac{q}{|\ |^3}, \tag{11.24}$$

which is singular at the origin. However, from (11.18), we have

$$\mathbf{D} = \frac{\mathbf{E}}{\sqrt{1 - \frac{1}{\mathfrak{b}^2}\mathbf{E}^2}}, \tag{11.25}$$

which implies that

$$\begin{aligned} \mathbf{E} &= \frac{\mathbf{D}}{\sqrt{1 + \frac{1}{\mathfrak{b}^2}\mathbf{D}^2}} \\ &= \frac{q}{|\ |\sqrt{|\ |^4 + \left(\frac{\mathfrak{q}}{\mathfrak{b}}\right)^2}}. \end{aligned} \tag{11.26}$$

In particular, the electric field **E** is globally bounded. It is interesting to see that, when $|\ |$ is sufficiently large, **E** given in (11.26) approximates that given by the Coulomb law, a consequence of the Maxwell equations.

As for the energy, we obtain from the Lagrange density (11.4) the energy-momentum tensor

$$T^{\boldsymbol{\nu}}_{\boldsymbol{\mu}} = -\frac{F^{\boldsymbol{\gamma\nu}} F_{\boldsymbol{\gamma\mu}}}{\sqrt{1 + \frac{1}{2\mathfrak{b}^2} F_{\boldsymbol{\alpha\beta}} F^{\boldsymbol{\alpha\beta}}}} - \delta^{\boldsymbol{\nu}}_{\boldsymbol{\mu}} \mathcal{L}, \tag{11.27}$$

which gives us in the electrostatic case the Hamiltonian energy density

$$\begin{aligned} \mathcal{H} &= T^0_0 = b^2\left(\frac{1}{\sqrt{1 - \frac{1}{\mathfrak{b}^2}\mathbf{E}^2}} - 1\right) \\ &= b^2\left(\sqrt{1 + \frac{1}{b^2}\mathbf{D}^2} - 1\right) \\ &= b^2\left(\sqrt{1 + \left(\frac{q}{b}\right)^2 \frac{1}{|\ |^4}} - 1\right). \end{aligned} \tag{11.28}$$

From (11.28), it is seen that the total energy of a point electric charge is now finite,

$$E = \int_{\mathbb{R}^3} \mathcal{H}\,\mathrm{d} \quad < \infty. \tag{11.29}$$

Similarly, we can consider the magnetostatic field generated from a point magnetic charge g placed at the origin of $\mathbb{R}^3$. In this case, $\mathbf{D} =$, $\mathbf{E} =$, and the Born–Infeld equations become

$$\nabla \cdot \mathbf{B} = 4\pi g \delta(\ ). \tag{11.30}$$

From (11.30), we have as before,

$$\mathbf{B} = \frac{g}{|\ |^3}, \tag{11.31}$$

$$\mathbf{H} = \frac{g}{|\ |\sqrt{|\ |^4 + \left(\frac{g}{b}\right)^2}}. \tag{11.32}$$

Thus $\mathbf{H}$ is a bounded vector field. In view of (11.27), the Hamiltonian density of a magnetostatic field takes the form,

$$\mathcal{H} = b^2\left(\sqrt{1 + \frac{1}{b^2}\mathbf{B}^2} - 1\right). \tag{11.33}$$

Inserting (11.31) into (11.33), we see that the total energy of a point magnetic charge is also finite in the Born–Infeld theory.

We note that it is not hard to extend the above discussion to cover the situation of multiply distributed point electric charges or magnetic monopoles.

Bo n-[illegible] d o y nd n z d B n n p o

We now study sourceless static solutions. With $\mathbf{E} = \nabla\phi$ and $\mathbf{B} = \nabla \times \mathbf{A}$, the equations of motion of the Born–Infeld theory, (11.16) and (11.17), become

$$\nabla \cdot \left(\frac{\nabla\phi}{\sqrt{1 + \frac{1}{b^2}(|\nabla \times \mathbf{A}|^2 - |\nabla\phi|^2)}}\right) = 0, \tag{11.34}$$

$$\nabla \times \left(\frac{\nabla \times \mathbf{A}}{\sqrt{1 + \frac{1}{b^2}(|\nabla \times \mathbf{A}|^2 - |\nabla\phi|^2)}}\right) = \quad . \tag{11.35}$$

From (11.35), we see that there is a real scalar function ψ such that

$$\frac{\nabla \times \mathbf{A}}{\sqrt{1 + \frac{1}{b^2}(|\nabla \times \mathbf{A}|^2 - |\nabla\phi|^2)}} = \nabla\psi, \tag{11.36}$$

which leads us to the relation

$$|\nabla \times \mathbf{A}|^2 = \left(1 - \frac{1}{b^2}|\nabla\phi|^2\right)\frac{|\nabla\psi|^2}{\left(1 - \frac{1}{b^2}|\nabla\psi|^2\right)}. \tag{11.37}$$

Inserting (11.37) into (11.36), we obtain

$$\nabla \times \mathbf{A} = \nabla\psi \sqrt{\qquad\qquad}$$

As a 'warm-up question', we ask whether the solutions of (11.42)–(11.43) of finite action, $\mathcal{A}(f, g) < \infty$, are constant. See [126] for some further discussion.

Thus, (11.53) may also be recast into the form

$$\nabla \cdot \left(\frac{g(|\nabla \Psi|)}{|\nabla \Psi|} \nabla \Psi \right) = \rho. \tag{11.55}$$

Use to denote a point in $\mathbb{R}^3$. To study the electric confinement problem, we consider the situation where two point charges of opposite signs, Q and $-Q$, are placed at the points $_1$ and $_2$, so that the electric charge density is given by [73]

$$\rho(\ ) = Q(\delta(\ - \ _1) - \delta(\ - \ _2)). \tag{11.56}$$

The ultimate goal is to establish that the potential energy of the solution of (11.53) or (11.55) where ρ is given by (11.56) depends on the separation distance $|\ _1 - \ _2|$ linearly, at least asymptotically, provided that the function $\varepsilon(E)$ or $f(D)$ is suitably chosen.

In [2, 4], the dielectric function $\varepsilon(E)$ is taken to be

$$\varepsilon(E) = \begin{cases} 0, & \text{when } E \le E_0, \\ \varepsilon_0 \ln\left(\frac{\mathbf{E}}{\mathbf{E}_0}\right), & \text{when } E > E_0, \end{cases} \tag{11.57}$$

where $\varepsilon_0 > 0$ and $E_0 > 0$ are constants, and asymptotic analysis and numerical approximations are carried out to show electric confinement. In [73], the function $f(D)$ satisfies

$$\lim_{\mathbf{D}\to\infty} \frac{\ln f(D)}{\ln D} = 1. \tag{11.58}$$

Another well-known confinement model [32, 33, 50] is called the *MIT bag model*, which is purely based on *linear* electrostatics instead.

nc

A. A. A rikosov On the agnetic properties of superconductors of the second group o y **5** 2.

S. L. Adler and T. Piran Flu con ne ent in the leading logarith odel y Le B **113** 2.

S. L. Adler and T. Piran The heavy quark static potential in the leading log and the leading log log odels y Le B **117** 2.

S. L. Adler and T. Piran Rela ation odels for gauge eld equili riu equations e Mod y **56**.

Y. Aharonov and D. Boh Signi cance of electro agnetic potentials in quantu theory y e **115**.

Y. Aharonov and D. Boh Further considerations on electro agnetic potentials in the quantu theory y e **123** 2.

R. Arnowitt S. Deser and C. Misner Coordinate invariance and energy e pressions in general relativity y e **122**.

R. A. Battye and P. M. Sutcli e Knots as sta le solutions in a three di ensional classical eld theory y e Le **81**.

R. A. Battye and P. M. Sutcli e Solitons links and knots oc oy oc A **455**.

E. B. Bogo ol nyi The sta ility of classical solutions *o N a y* **24** .

M. Born Modi ed eld equations with a nite radius of the electron *N e* **132** .

M. Born On the quantu theory of the electro agnetic eld *oc oy oc A* **143** .

M. Born and L. Infeld Foundation of the new eld theory *N e* **132** .

M. Born and L. Infeld Foundation of the new eld theory *oc oy oc* A **144** .

S. T. Bra well S. R. Gi lin S. Calder R. Aldus D. Pra hakaran and T. Fennell Measure ent of the charge and current of agnetic onopoles in spin ice *N e* **461** .

H. L. Bray Proof of the Rie annian Penrose inequality using the positive ass theore *D i eo* **59** .

H. L. Bray Black holes geo etric ows and the Penrose inequality in general relativity *No ce A e M oc* **49** .

H. L. Bray and R. M. Schoen Recent proofs of the Rie annian Penrose conjecture *C i en De e op en n M e c* Int. Press So erville MA . pp. .

L. Brewin A rief introduction to Cada ra a tool for tensor co putations in General Relativity *Co p e y c Co n c on* **181** .

G. Buck Four thirds power law for knots and links *N e* **392** .

J. Burzla and F. Navarro Lerida Nielsen Olesen vortices for large Ginz urg Landau para eter *y e* D **82** .

E. Cala i The space of Khler etrics *oc n e n Con e M* A sterda pp. .

E. Cala i On Khler anifolds with vanishing canonical class in *A e c i eo e y nd opo o y A y po n ono o Le c e z* Princeton Mathe atical Series **12** Princeton University Press pp. .

E. Cala i E a ples of Bernstein Pro le s for so e nonlinear equations *i o An y* Proc. Sy pos. Pure Math. **15** Berkeley ed. S. S. Chern and S. S ale AMS pp. .

P. Candelas G. T. Horowitz A. Stro inger and E. itten Vacuu con gurations for superstrings *N a y* B **258** .

J. Cantarella R. B. Kusner and J. M. Sullivan Tight knot values deviate fro linear relations *N e* **392** .

C. Castelnovo R. Moessner and S. L. Sondhi Magnetic onopoles in spin ice *N e* **451** .

R. M. Chen Y. Guo D. Spirn and Y. Yang Electrically and agnetically charged vortices in the Chern Si ons Higgs theory *oc oy oc* A **465** .

S. Y. Cheng and S. T. Yau Ma i al spacelike hypersurfaces in the Lorentz Minkowski spaces *Ann o M* **104** .

S. S. Chern and J. Si ons So e coho ology classes in principal er undles and their application to Rie annian geo etry *oc N Ac d c A* **68** .

S. S. Chern and J. Si ons Characteristic for s and geo etric invariants *Ann o M* **99** .

A. Chodos R. L. Ja e K. Johnson and C. B. Thorn Baryon structure in the ag theory *y e* D **10**

A. Chodos R. L. Ja e K. Johnson C. B. Thorn and V. F. eisskopf New e tended odel of hadrons y e D **9** .

T. L. Curtright and C. B. Thorn Confor ally invariant quantization of the Liouville theory y e Le **48** .

A. S. Davydov n Mec n c Perga on Press O ford and New York .

P. A. M. Dirac Quantized singularities in the electro agnetic eld oc oy oc A **133** .

M. R. Douglas and S. Kachru Flu co pacti cation e Mod y **79** .

G. Dunne e D C e n on eo e Lecture Notes in Physics vol. **36** Springer Berlin .

S. . Englander N. R. Kallen ach A. J. Heeger J. A. Kru hansl and S. Litwin Nature of the open state in long polynucleotide dou le helices possi ility of soliton e citations oc N Ac d c A **77** .

F. Englert and R. Brout Broken sy etry and the ass of gauge vector esons y e Le **13** .

M. J. Este an E istence of sy etric solutions for the Skyr e s pro le Ann M App **147** .

L. Faddeev Einstein and several conte porary tendencies in the theory of ele entary particles e y n nd Co o o y vol. ed. M. Pantaleo and F. de Finis pp. .

L. Faddeev Knotted solitons Proc. ICM vol. Beijing August pp. .

L. Faddeev and A. J. Nie i Sta le knot like structures in classical eld theory N e **387** .

E. Fer i Eine statistische ethode zur esti ung einiger eigenschaften des ato s und ihre an wendung auf die theorie des periodischen syste s der ele ente Z y A **48** .

D. Freifelder Mo ec B o o y Jones and Bartlett Boston .

J. Frohlich and P.A. Marchetti Quantu eld theories of vortices and anyons Co n M y **121** .

J. L. Gervais and A. Neveu Dual string spectru in Polyakov s quantization parts I and II N a y B **199** d **209** .

S. R. Gi lin S. T. Bra well P. C. . Holdsworth D. Pra hakaran and I. Terry Creation and easure ent of long lived agnetic onopole currents in spin ice N e y c **7** .

R. Giles Force etween two point charges in a two di ensional ag y e D **18** .

V. L. Ginz urg and L. D. Landau On the theory of superconductivity in Co ec ed pe o L D L nd edited y D. Ter Haar pp. Perga on New York .

T. Goto Relativistic quantu echanics of one di ensional echanical continuu and su sidiary condition of dual resonance odel o eo e y **46** .

M. Grana Flu co pacti cations in string theory a co prehensive review y ep **423** .

J. Greensite An n od a on o e Con ne en o e Lecture Notes in Physics **821** Springer Verlag Berlin and New York .

D. J. Gri ths n od a on o n Mec n c Prentice Hall Englewood Cli s New Jersey .

M. Gro ov *n e o o opy eo y* Prospects in Math. Princeton U. Press Princeton New Jersey .

G. Guralnik C. R. Hagen and T. . B. Ki le Glo al conservation laws and assless particles *y e Le* **13** .

P. Higgs Broken sy etries and the asses of gauge osons *y e Le* **13** .

M. Hiraya a H. Ya akoshi and Ya ashita Esti ation of the Lin Yang ound of the least static energy of the Faddeev odel *o eo y* **116** .

J. Hong Y. Ki and P. Y. Pac Multivorte solutions of the A elian Chern Si ons Higgs theory *y e Le* **64** .

K. Huang *c Mec n c* John iley Sons New York .

G. Huisken and T. Il anen The Rie annian Penrose inequality *n M e No* **20** .

G. Huisken and T. Il anen The inverse ean curvature ow and the Rie annian Penrose inequality *D eo* **59** .

. Hunziker and I. M. Sigal The quantu *N* ody pro le *M y* **41** .

R. Jackiw and E. J. ein erg Self dual Chern Si ons vortices *y e Le* **64** .

A. Ja e and C. H. Tau es n Td . J .

F. Lin and Y. Yang Static knot energy Hopf charge and universal growth law *N* B **747**

F. Lin and Y. Yang Universal growth law for knot energy of Faddeev type in general di ensions *oc oy oc* A **646**

P. L. Lions The concentration co pactness principle in the calculus of variations. Part I and Part II *Ann n en y o nc e An y e Non L ne e* **1** **1**

P. L. Lions The concentration co pactness principle in the calculus of variations. Part I and Part II *e M e* **1** **2**

F. London and H. London The electro agnetic equations of the superconductor *oc oy oc* A **149**

G. Mackey *M e c o nd on o n Mec n c* . A. Benja in .

D. Maison Uniqueness of the Prasad So er eld onopole solution *N* B **182**

S. Mandelsta Vortices and quark con ne ent in non A elian gauge theories *y Le* B **53**

S. Mandelsta General introduction to con ne ent *y ep* C **67**

L. R. U. Manssur and R. Portugal The canon package a fast kernel for tensor anipulators *Co p e y c Co n c on* **157**

N. Manton and P. Sutcli e *opo o c o on* Ca ridge Monographs on Mathe atical Physics Ca ridge U. Press Ca ridge U. K.

A. Messiah *n Mec n c* Volu es I and II Dover New York

K. R. Meyer and G. R. Hall *n od on o on n Dyn c y e nd e N Body o e* Springer Verlag Applied Mathe atical Sciences **90** New York and Berlin

G. . Misner K. S. Thorne and J. A. heeler *on* H. . Free an New York

D. J. P. Morris D. A. Tennant S. A. Grigera B. Kle ke C. Castelnovo R. Moessner C. Czternasty M. Meissner K. C. Rule J. U. Ho ann K. Kiefer S. Gerischer D. Slo insky and R. S. Perry Dirac strings and agnetic onopoles in the spin ice $Dy_2Ti_2O_7$ *c ence* **326**

P. M. Morse Diato ic olecules according to the wave echanics. II. Vi rational levels *y e* **34**

Y. Na u Strings onopoles and gauge elds *y e* D **10**

J. von Neu ann *M e c o nd on o n Mec n c* Princeton U. Press

E. New an and A. Janis Note on the Kerr spinning particle etric *M y* **6**

E. New an E. Couch K. Chinnapared A. E ton A. Prakash and R. Torrence Metric of a rotating charged ass *M y* **6**

H. B. Nielsen and P. Olesen Vorte line odels for dual strings *N* B **61**

A. J. Nie i Ha iltonian approach to knotted solitons a contri uted chapter in *de no* A. Stasiak V. Katritch and L. H. Kau an eds orld Scienti c Singapore

R. Osser an *A ey o M n ce* Dover New York NY

H. Pagels and E. To oulis Vacuu of the quantu Yang Mills theory and agnetostatics *N* B **143**

T. H. Parker and C. H. Tau es On itten s proof of the positive energy theore *Co n M y* **84** .

S. Paul and A. Khare Charged vortices in an A elian Higgs odel with Chern Si ons ter *y Le* B **17** .

R. Penrose Naked singularities *Ann Ne Yo Ac d c* **224** .

M. Peyrard Nonlinear dyna ics and statistical physics of DNA *Non ne y* **17** R R .

M. Peyrard and A. R. Bishop Statistical echanics of a nonlinear odel for DNA denaturation *y e Le* **62** .

H. Pollard *M e c n od o on o Ce e Mec n c* Prentice Hall Englewood Cli s New Jersey .

A. M. Polyakov Particle spectru in quantu eld theory *Le* **20** .

A. M. Polyakov Quantu geo etry of osonic strings *y Le* B **103** .

A. M. Polyakov Quantu geo etry of fer ionic strings *y Le* B **103** .

R. Portugal and S. L. Sautu Applications of Maple to general relativity *Co p e y c Co n c on* **105** .

M. K. Prasad and C. M. So er eld E act classical solutions for the t Hooft onopole and the Julia Zee dyon *y e Le* **35** **B 108** H .

E. Radua and M. S. Volkov Stationary ring solitons in eld theory knots and vortons *y ep* **468** .

A. I. M. Rae *n Mec n c* Institute of Physics Pu lishing Bristol and Philadelphia .

V. A. Ru akov and A. N. Tavkhelidze Sta le ano alous states of superdense atter

T. H. R. Skyr e Particle states of a quantized eson eld *oc oy oc* A **262**

T. H. R. Skyr e A uni ed eld theory of esons and aryons *N a y* **31**

T. H. R. Skyr e The origins of Skyr ions *n e n Mod y* A **3**

J. C. Slater The theory of co ple spectra *y e* **34**

J. Spruck and Y. Yang Proof of the Julia Zee theore *Co n M y* **291**

G. t Hooft Magnetic onopoles in uni ed gauge theories *N a y* B **79**

G. t Hooft Co putation of the quantu e ects due to a four di ensional pseudoparticle *y e* D **14**

G. t Hooft On the phase transition towards per anent quark con ne ent *N a y* B **138**

G. t Hooft A property of electric and agnetic u in nona elian gauge theories *N a y* **B 153**

G. t Hooft Topology of the gauge condition and new con ne ent phases in non A elian gauge theories *N a y* B **190**

C. H. Tau es Ar itrary *N* vorte solutions to the rst order Ginz urg Landau equations *Co n M y* **72**

C. H. Tau es On the equivalence of the rst and second order equations for gauge theories *Co n M y* **75**

L. H. Tho as The calculation of ato ic elds *M oc C d e o op c oc* **23**

A. F. Vakulenko and L. V. Kapitanski Sta ility of solitons in S^2 nonlinear σ odel *o y Do* **24**

H. J. de Vega and F. Schaposnik Electrically charged vortices in non A elian gauge theories with Chern Si ons ter *y e Le* **56**

H. J. de Vega and F. Schaposnik Vortices and electrically charged vortices in non A elian gauge theories *y e* D **34**

R. M. ard *ene e y* U. Chicago Press Chicago and London

J. H. C. hitehead An e pression of Hopf s invariant as an integral *oc N Ac d c A* **33**

F. ilczek Quantu echanics of fractional spin particles *y e Le* **49**

F. ilczek *c on c nd Anyon pe cond o* orld Scienti c Singapore

E. itten A new proof of the positive energy theore *Co n M y* **80**

T. T. u and C. N. Yang Concept of nonintegra le phase factors and glo al for ulation of gauge elds *y e* D **12**

T. T. u and C. N. Yang Dirac onopole without strings Monopole har onics *N a y* B **107**

Y. Yang On the glo al ehavior of sy etric Skyr ions *Le M y* **19**

Y. Yang *o on n e d eo y nd Non ne An y* Springer Monographs in Mathe atics Springer Verlag New York and Berlin

S. T. Yau Cala i s conjecture and so e new results in alge raic geo etry *oc N Ac d c* **74**

I. Zahed and G. E. Brown The Skyr e odel *y epo* **142**

B. Zwie ach *A Co e n n eo y* Ca ridge U. Press Ca ridge U. K.