Comprehensive Notes in Undergraduate Physics

Castaly Fan

Department of Physics, University of Florida, Gainesville, FL 32608

Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854

July 14, 2023

Abstract

Physics plays a crucial role in most of modern science-related areas. This is a collective note in physics for college students which covers the topics including classical mechanics, thermal physics (thermodynamics and statistical mechanics), electromagnetism (including optics), and modern physics (special relativity and quantum mechanics).

The note is designed for college students in physics major, but it also suitable for ambitious high school students or the graduate students who need to recap basic concepts. The topics particularly focus on the junior and senior years lectures in undergraduate physics, but also covers a few graduate-level concepts (e.g. Hamiltonian mechanics, spherical harmonics in QM, etc.).

Notice the notations are mostly consistent with the convention in most of popular textbooks, while the readers should expect some differences among chapters. The note was started compiling as of the summer in 2021 and completed in July 2023.

Contents

1 Preliminaries **4**
- 1.1 Introduction . . . 4
- 1.2 Coordinate systems . . . 4
- 1.3 Some useful identities . . . 5
- 1.4 Calculus in three dimensions . . . 6
- 1.5 Vector calculus and fundamental theorems . . . 6
- 1.6 The Dirac delta function . . . 8

2 Classical Mechanics **10**
- 2.1 Newton's laws of motion . . . 10
- 2.2 Projectile motion and drag . . . 11
- 2.3 Energy and work . . . 13
- 2.4 Momentum and rotation . . . 15
- 2.5 Conservation laws . . . 20
- 2.6 Gravity and Kepler's laws . . . 23
- 2.7 Fluids . . . 26
- 2.8 Oscillations . . . 28
- 2.9 Waves . . . 33
- 2.10 Lagrangian mechanics . . . 38
- 2.11 Hamiltonian mechanics . . . 41
- 2.12 Two-body problem . . . 44
- 2.13 Non-inertial mechanics . . . 48

3 Thermal Physics **52**
3.1 Ideal gas law . . . 52
3.2 Kinetic theory of gases . . . 52
3.3 Energy, heat, and work . . . 54
3.4 The first law of thermodynamics . . . 55
3.5 Entropy and the second law . . . 57
3.6 Engines and refrigerators . . . 60
3.7 Interactions in thermodynamics . . . 64
3.8 Thermodynamic free energy . . . 65
3.9 Boltzmann statistics . . . 66
3.10 Quantum statistics . . . 69

4 Electromagnetism **73**
4.1 Charges, forces, and fields . . . 73
4.2 E-fields and the Gauss's law . . . 75
4.3 Electric potentials and boundary conditions . . . 77
4.4 Work and electric potential energy . . . 80
4.5 Conductors . . . 81
4.6 Current and resistance . . . 83
4.7 Circuit-1: emf, Kirchhoff's rules, and RC circuit . . . 85
4.8 Magnetic fields . . . 89
4.9 Multipole expansion . . . 92
4.10 Electromagnetic induction . . . 95
4.11 Circuit-2: RL, LC, RLC circuits, and AC . . . 97
4.12 Electromagnetism in matter . . . 103
4.13 Maxwell's equations . . . 105
4.14 Electromagnetic waves . . . 108
4.15 Radiation . . . 110
4.16 Conservation laws . . . 111
4.17 Potential and fields . . . 114

5 Optics **117**
5.1 Reflection and refraction . . . 117
5.2 Images and lens . . . 119
5.3 Interference . . . 123
5.4 Diffraction . . . 125
5.5 Dispersion and scattering . . . 126

6 Modern Physics: Special Relativity **129**
6.1 The two postulates . . . 129
6.2 Time and length in relativity . . . 129
6.3 Lorentz transformation . . . 131
6.4 Minkowski spacetime . . . 132
6.5 Relativistic energy and momentum . . . 134

7 Modern Physics: Quantum Mechanics **136**
7.1 Foundation-1: Blackbody radiation . . . 136
7.2 Foundation-2: Photoelectric effect . . . 137
7.3 Foundation-3: Matter waves . . . 138
7.4 The wave function . . . 138
7.5 Formalism in quantum mechanics . . . 140
7.6 The infinite square well . . . 142
7.7 The delta potential barrier . . . 144
7.7.1 Bound state . . . 144
7.7.2 Scattering state . . . 145

7.8 The finite square well . . . 147
7.8.1 Bound state . . . 147
7.8.2 Scattering state . . . 148
7.9 Harmonic oscillators . . . 149
7.10 3-D Schrödinger equation (spherically symmetric potential) . . . 152
7.10.1 The radial part . . . 153
7.10.2 The angular part . . . 154
7.11 The hydrogen atom . . . 156
7.12 Angular momentum and spin . . . 160

8 Epilogue **164**

1 Preliminaries

1.1 Introduction

Physics is a subject studying how the world works. Humans are always curious about the accurate "picture" of the world. What physicists do is to construct a more accurate picture of our universe. From the essential things, the particles, the atoms, to our daily lives such as the motion of a car or the construction of a telescope, then to the motion of planets, the black holes, the galaxies, even to the origin of our universe – everything is related to physics.

Nowadays, we have four subjects that all physics students must learn in college: classical mechanics, electromagnetism, thermal physics, and modern physics (special relativity and quantum mechanics). We will also reach classical optics, nuclear and particle physics, and astrophysics to some extent. As for the more fundamental things in physics, we will consider the four fundamental interactions: strong interaction, electromagnetic interaction, weak interaction, and gravitational interaction (sorted from stronger to weaker). The subjects to study these forces are very popular in recent decades. For the strong, weak, and electromagnetic interactions, they can be described in the framework of quantum field theory (QFT), whereas for the gravity it can be best described by general relativity (GR) currently. It is a huge goal for physicists to unify these theories. Why is it important? Well, once we know that ultimate theory, it means we shall unravel the origin of the universe because all interactions were unified at the beginning of the Big Bang.

Before moving on the new page, here is the list of some useful physical constants:

Table 1: Fundamental physical constants.

Symbol	Name	Value
G	Newtonian constant	6.674×10^{-11} $\mathrm{m^3 kg^{-1} s^{-2}}$
N_A	Avogadro constant	6.022×10^{23} $\mathrm{mol^{-1}}$
k_B	Boltzmann constant	1.381×10^{-23} J $\mathrm{K^{-1}}$
m_p	proton mass	1.672×10^{-27} kg
m_e	electron mass	9.109×10^{-31} kg
e	elementary charge	1.602×10^{-19} C
k	Coulomb constant	9×10^9 kg $\mathrm{m^3 s^{-2} C^{-2}}$
ε_0	vacuum electric permitivity	8.854×10^{-12} F $\mathrm{m^{-1}}$
μ_0	vacuum magnetic permeability	$4\pi \times 10^{-7}$ H $\mathrm{m^{-1}}$
c	speed of light	3×10^8 m $\mathrm{s^{-1}}$
h	Planck constant	6.626×10^{-34} J $\mathrm{Hz^{-1}}$
$\hbar$	reduced Planck constant	1.054×10^{-34} J s

1.2 Coordinate systems

In physics, the coordinate system is a foundation to describe any kinds of motions. The most ordinary system is the **Cartesian coordinate** (x, y, z). Another one is the **cylindrical coordinate system** (ρ, ϕ, z), which is used to describe the objects with axial symmetry. The other one is the **spherical coordinate system** (r, θ, ϕ), just as its name, it would be useful in describing any objects with spherical symmetry. For the conversion between the Cartesian coordinate with the other two systems, it can be shown as the following:

$$\begin{cases} x &= \rho\sin\phi \\ y &= \rho\cos\phi \\ z &= z \end{cases} \quad \begin{cases} \rho &= \sqrt{x^2+y^2} \\ \phi &= \arctan\frac{y}{x} \\ z &= z \end{cases} \qquad \text{(cylindrical coordinates)} \tag{1.1}$$

$$\begin{cases} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{cases} \quad \begin{cases} r &= \sqrt{x^2+y^2+z^2} \\ \theta &= \arccos\frac{z}{r} \\ \phi &= \arctan\frac{y}{x} \end{cases} \qquad \text{(spherical coordinates)} \tag{1.2}$$

Both of them are extended from the polar coordinate system, (r, θ). In both physics and engineering, it would be useful to express a coordinate system by only the length (r) and the angle (θ). Note for the spherical coordinate system, usually, r is called the radial distance, θ is called the polar angle, and ϕ is called the azimuthal angle. If you take a slice of both of them as a 2D plane, you will notice they are in the polar coordinate system.

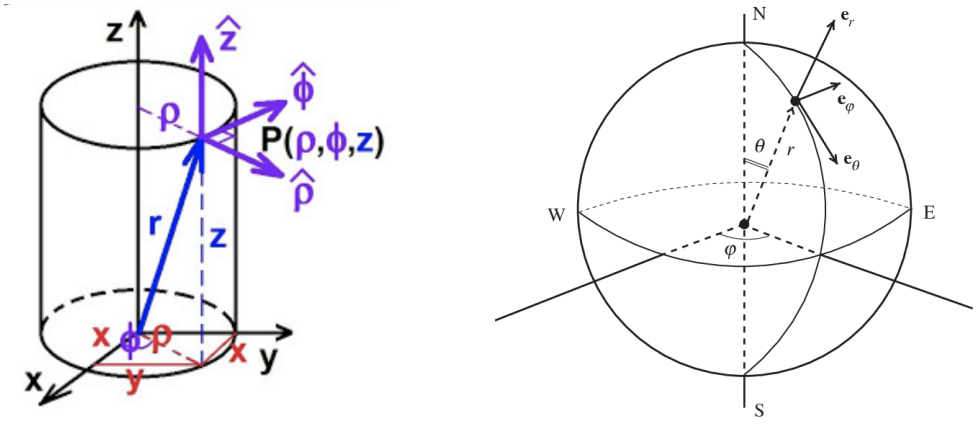

Figure 1: The cylindrical coordinate system (left) and the spherical coordinate system (right). *Source*: http://www.ilectureonline.com/lectures/subject/PHYSICS/35/247. (left) DOI: 10.1098/rspa.2017.0063 (right).

1.3 Some useful identities

Here is a collection of some trigonometric identities:

$$\begin{aligned} &\sin^2 x + \cos^2 x = 1, \quad \sec^2 x - \tan^2 x = 1 \\ &\cos^2 x = \frac{1+\cos 2x}{2}, \quad \sin^2 x = \frac{1-\cos 2x}{2} \\ &\sin 2x = 2\sin x\cos x = \frac{2\tan x}{1-\tan^2 x} \\ &\cos 2x = \cos^2 x - \sin^2 x = 2\cos^2 x - 1 = \frac{1-\tan^2 x}{1+\tan^2 x} \\ &\sin(x \pm y) = \sin x\cos y \pm \cos x\sin y \\ &\cos(x \pm y) = \cos x\cos y \mp \sin x\sin y \end{aligned} \tag{1.3}$$

For the hyperbolic functions,

$$\cosh x = \frac{e^x + e^{-x}}{2}, \quad \sinh x = \frac{e^x - e^{-x}}{2} \tag{1.4}$$

The relation between the trigonometric functions and the exponential functions can be related by the polar coordinate system:

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta} \quad \text{so that} \quad \boxed{e^{i\pi} = -1} \tag{1.5}$$

, which is the famous **Euler's formula**.

In addition to the trigonometry, Taylor expansion is also a crucial tool for approximation.

$$\begin{aligned}
\cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - ... \cong 1 \quad \text{(for small angle)} \\
\sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - ... \cong x \quad \text{(for small angle)} \\
e^x &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + ... \\
(1+x)^n &= 1 + nx + \frac{n(n-1)}{2!}x^2 + ... \quad \text{(binomial series)}
\end{aligned} \tag{1.6}$$

1.4 Calculus in three dimensions

Calculus is a fairly useful mathematical tool in physics and engineering. The reason is, everything in the world is not really in the shape of a straight line, a square, or a triangle. Conversely, we see many curved objects, a river, a cap, and our Earth is also a (approximately) spherical object. Calculus enables us to integrate over a curved object by summing all tiny elements. Depending on what object we would like to take an integral, the corresponding line and volume elements can be denoted by

$$\begin{cases} d\mathbf{r} &= d\rho\hat{\boldsymbol{\rho}} + \rho\, d\phi\hat{\boldsymbol{\phi}} + dz\hat{\mathbf{z}} \\ dV &= \rho\, d\rho\, d\phi\, dz \end{cases} \quad \text{(cylindrical coordinates)}$$

$$\begin{cases} d\mathbf{r} &= dr\,\hat{\mathbf{r}} + r\, d\theta\hat{\boldsymbol{\theta}} + r\sin\theta\, d\phi\hat{\boldsymbol{\phi}} \\ dV &= r^2\sin\theta\, dr\, d\theta\, d\phi \end{cases} \quad \text{(spherical coordinates)}$$

To get a more clearer understanding of all of them, I suggest to take a look at Figure 2.

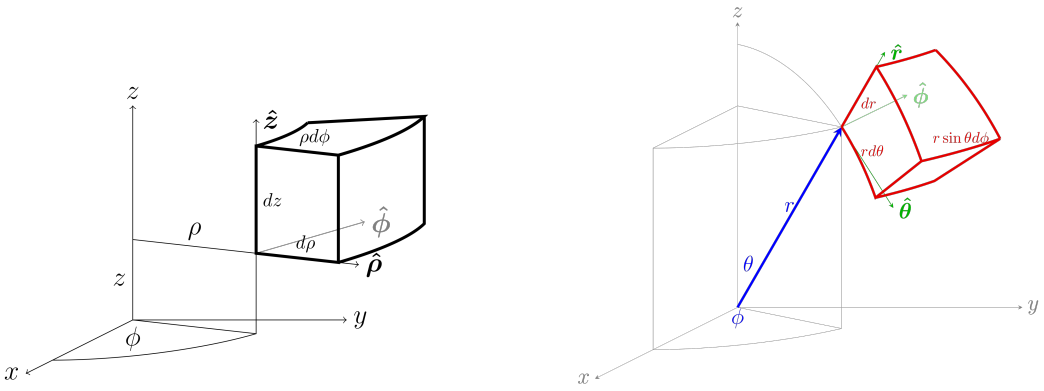

Figure 2: An integral element of the cylindrical coordinate system (left) and the spherical coordinate system (right). *Source*: Wikipedia.

1.5 Vector calculus and fundamental theorems

One of the most appealing topic about mathematical tools in physics might be the various properties stemming from vector calculus. In the discussion of vector fields, these properties would be important. That's why we can see a lot of nabla notation "∇" in electromagnetism.

For a scalar field or a vector field, there are three properties can sufficiently describe them: **gradient** (how and in what directions a scalar field increases), **divergence** (how a vector field expands or converges), and **curl** (how a vector field rotate with respect to a source). For example, gradient is useful in calculating how temperature in a room increases or decreases from a source; divergence is useful in calculating the

surrounding field of an electric charge; curl is useful in calculating how much the magnetic field generated from a straight wire with current (recall the right-hand rule).

Usually, the gradient is denoted by $\text{grad}(f) = \nabla f$, the divergence is denoted by $\text{div}(\mathbf{f}) = \nabla \cdot \mathbf{f}$, and the curl is denoted by $\text{curl}(\mathbf{f}) = \nabla \times \mathbf{f}$. The mathematical expressions of these three quantities are listed as the following (in different coordinate systems):

$$\begin{aligned}
\boldsymbol{\nabla} f &= \frac{\partial f}{\partial x}\hat{\mathbf{x}} + \frac{\partial f}{\partial y}\hat{\mathbf{y}} + \frac{\partial f}{\partial z}\hat{\mathbf{z}} \quad \text{(Cartesian)} \\
&= \frac{\partial f}{\partial \rho}\hat{\boldsymbol{\rho}} + \frac{1}{\rho}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}} + \frac{\partial f}{\partial z}\hat{\mathbf{z}} \quad \text{(cylindrical)} \\
&= \frac{\partial f}{\partial r}\hat{\boldsymbol{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}} \quad \text{(spherical)}
\end{aligned} \tag{1.7}$$

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \mathbf{f} &= \frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} + \frac{\partial f_z}{\partial z} \quad \text{(Cartesian)} \\
&= \frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho f_\rho) + \frac{1}{\rho}\frac{\partial}{\partial \phi}(f_\phi) + \frac{\partial}{\partial z}(f_z) \quad \text{(cylindrical)} \\
&= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 f_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}(\sin\theta f_\theta) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}(f_\phi) \quad \text{(spherical)}
\end{aligned} \tag{1.8}$$

$$\begin{aligned}
\boldsymbol{\nabla} \times \mathbf{f} &= \left(\frac{\partial f_z}{\partial y} - \frac{\partial f_y}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{\partial f_x}{\partial z} - \frac{\partial f_z}{\partial x}\right)\hat{\mathbf{y}} + \left(\frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y}\right)\hat{\mathbf{z}} \quad \text{(Cartesian)} \\
&= \left[\frac{1}{\rho}\frac{\partial}{\partial \phi}(f_z) - \frac{\partial}{\partial z}f_\phi\right]\hat{\boldsymbol{\rho}} + \left[\frac{\partial}{\partial z}f_\rho - \frac{\partial}{\partial \rho}f_z\right]\hat{\boldsymbol{\phi}} + \frac{1}{\rho}\left[\frac{\partial}{\partial \rho}(\rho f_\phi) - \frac{\partial}{\partial \phi}(f_\rho)\right]\hat{\boldsymbol{z}} \quad \text{(cylindrical)} \\
&= \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial \theta}(\sin\theta f_\phi) - \frac{\partial f_\theta}{\partial \phi}\right]\hat{\boldsymbol{r}} + \left[\frac{1}{r\sin\theta}\frac{\partial f_r}{\partial \phi} - \frac{1}{r}\frac{\partial}{\partial r}(r f_\phi)\right]\hat{\boldsymbol{\theta}} + \frac{1}{r}\left[\frac{\partial}{\partial r}(r f_\theta) - \frac{\partial f_r}{\partial \theta}\right]\hat{\boldsymbol{\phi}} \quad \text{(spherical)}
\end{aligned} \tag{1.9}$$

Some identities for the second derivatives:

$$\begin{aligned}
&\boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \times \mathbf{f}) = 0 \\
&\boldsymbol{\nabla} \times (\boldsymbol{\nabla} f) = 0 \\
&\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{f}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{f}) - \nabla^2 \mathbf{f}
\end{aligned} \tag{1.10}$$

From now on I would like to change the bold symbol "$\boldsymbol{\nabla}$" to "∇". But the reader should bear in mind that the nabla notation is always a vector operator (not a "vector").

In vector calculus, there are three fundamental theorems that are helpful to calculate the fields:

1. **Gradient Theorem**:
$$\int_a^b (\nabla f) \cdot d\mathbf{l} = f(\mathbf{b}) - f(\mathbf{a}) \tag{1.11}$$

2. **Stokes Theorem**:
$$\boxed{\iint_S (\nabla \times \mathbf{f}) \cdot d\mathbf{S} = \oint_{\partial S} \mathbf{f} \cdot d\mathbf{l}} \tag{1.12}$$

Incidentally, the **Green's theorem** is a 2-dimensional case of the Stokes theorem,

$$\iint_A \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx\, dy = \oint_{\partial A} (f\, dx + g\, dy) \tag{1.13}$$

3. **(Gauss's) Divergence Theorem**:

$$\boxed{\iiint_V (\nabla \cdot \mathbf{f})\, dV = \oiint_{\partial V} \mathbf{f} \cdot d\mathbf{S}} \tag{1.14}$$

These three theorems, especially the Stokes's and Gauss's, are fairly useful in electromagnetism. Calculating the flux passing through a space or a surface requires these theorems. To get insight of them, taking Stokes theorem (1.12) as an example, the formula tells us that taking an integral of the curl of a vector field through a surface is equivalent to taking an integral of the field through its 1-dimensional boundary (imagine the open surfaces like a circle or a cap, or see Figure 3). Similarly, the divergence theorem (1.14) tells us that taking an integral of the divergence of a vector field on a volume is equivalent to taking an integral of the field through its 2-dimensional boundary, that is, its surrounding surface (imagine the closed surfaces like a cube or a doughnut). If it is still hard to visualize, please think about a faucet: when you turn on the valve, the water will come out. When you turn more and the water come out in a greater amount. That is to say, your 1-dimensional behavior (turn on/off the valve along with a circular path) determines the strength of flux (the amount of water), and this is what the Stokes theorem shows!

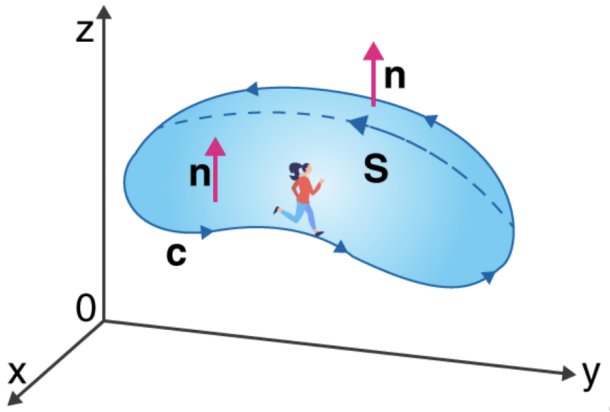

Figure 3: The illustration of the Stokes theorem: the curl (normal vector) of the vector field through a surface depends on its boundary (the circle at the bottom), and the direction is abiding by the right-hand rule.

1.6 The Dirac delta function

A lot of mathematical tools would be used in physics, particularly for those more abstract objects. For example, in electrostatics or mechanics, a point charge and a point particle are often used to describe the fundamental phenomena. In fact, these "point-like" objects are the idealized objects. If we plot out a function of density, say $\delta(x)$, on an x-y plane, then we will find that everywhere is zero except for where the "point charge" (or a point particle) locates – at that point the density (y-value) reaches infinity. Given that the point charge locates at $x = 0$, we have

$$\delta(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ \infty & \text{if } x = 0 \end{cases} \tag{1.15}$$

This is the **Dirac delta function** (Figure 4).

The integral over the whole region would be 1,

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1$$

Given an ordinary function $f(x)$, we will have the following two identities:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\delta(x)\,dx &= f(0)\\ \int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx &= f(a)\end{aligned} \tag{1.16}$$

The second line is nothing special but just consider a displacement a.

In three dimensions, Eq (1.16) turns into

$$\iiint_{\text{all space}} f(\mathbf{r})\delta^3(\mathbf{r}-\mathbf{a})dV = f(\mathbf{a}) \tag{1.17}$$

where $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$. Moreover,

$$\nabla \cdot \left(\frac{\hat{\mathbf{r}}}{r^2}\right) = 4\pi\delta^3(\mathbf{r}) \tag{1.18}$$

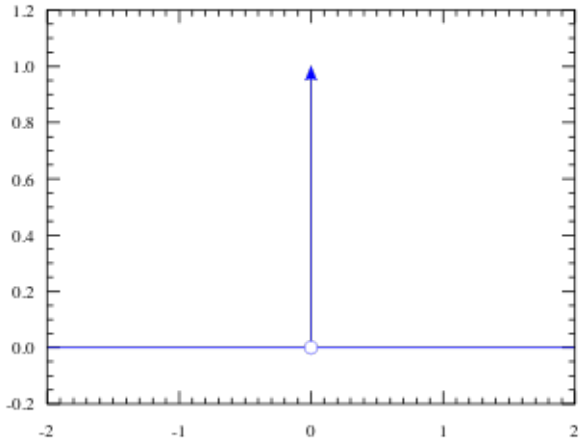

Figure 4: The Dirac δ-function. *Source*: Wikipedia

For sure, there are still many properties of the δ-function. For instance, given a constant k, we will have

$$\delta(kx) = \frac{\delta(x)}{|k|} \tag{1.19}$$

Besides, consider the Laplace transform, we also have

$$\int_0^{\infty} \delta(t-a)e^{-st}\,dt = e^{-sa}$$

Recall that the Laplace transform is a transform from a real variable t to a complex variable s,

$$F(s) = \mathcal{L}[f(t)] = \int_0^{\infty} f(t)e^{-st}\,dt \tag{1.20}$$

Laplace transform is important in engineering, though we will not often use this in the later sections.

2 Classical Mechanics

2.1 Newton's laws of motion

In classical mechanics, Newton's laws of motion are the cores of physics. Consider a motion of an object (or a particle), it is necessary to know its position. The derivative of the position x is the velocity v, and the derivative of the velocity v is the acceleration a . All of these are with respect to the time t. There are four equations describing the motion of these parameters:

$$\begin{aligned} x &= x_0 + \bar{v}t \\ v &= v_0 + at \\ \Delta x &= v_0 t + \frac{1}{2}at^2 \\ \Delta x &= vt - \frac{1}{2}at^2 \\ \Delta x &= \frac{1}{2}(v_0 + v)t \\ v^2 &= v_0^2 + 2a\Delta x \end{aligned} \tag{2.1}$$

These all can be derived straightforwardly by simple calculations.

The three laws are:

1. **Newton's First Law (the principle of inertia)**: An object will move with a constant velocity if there is no external forces.

2. **Newton's Second Law**: The **force** is determined by the mass m and the acceleration $\mathbf{a}$ of an object,

$$\boxed{\mathbf{F} = m\mathbf{a}} \tag{2.2}$$

3. **Newton's Second Law**: Given an action, that is, a force exerts from the object 1 on the object 2, then the force from the object 2 acts on the object 1 must be in the same magnitude but opposite directions, that is, $\mathbf{F}_{12} = -\mathbf{F}_{21}$

According to the second law, the force can be written as

$$\mathbf{F} = m\mathbf{a} = m\frac{d\mathbf{v}}{dt} = \dot{\mathbf{p}} \tag{2.3}$$

, where $\dot{\mathbf{p}} = d\mathbf{p}/dt$ is the **momentum**. That is, a force is the derivative of the momentum.

When solving the force problem, it is more often to employ the second law. Also, **friction** is also a common topic, which is defined by

$$f_{s,\max} = \mu_s N \quad \text{(maximum static friction)} \tag{2.4}$$

and

$$f_k = \mu_k N \quad \text{(kinetic friction)} \tag{2.5}$$

Here N is the normal force that is perpendicular to the plane. As for μ is the coefficient of the friction. As you learned in the high school period, it is important to sketch a free-body diagram when solving any problems about force and motion (see Figure 5). You can see that we also often consider the gravitational force

$$\mathbf{F}_g = m\mathbf{g} \tag{2.6}$$

and g here is the free-fall acceleration. Sometimes when we consider the wright of an object, it can be defined by $W = mg$. Consider a situation such as shown in the top of Figure 5, a block on a plane without any vertical motion implies that $N = F_g = mg$.

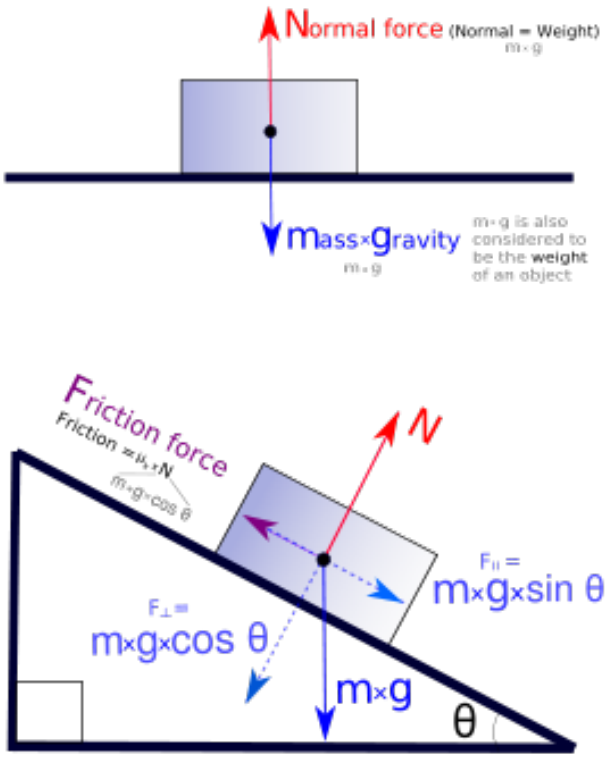

Figure 5: The free-body diagram of a block on a table (top) and on a inclined plane (bottom). *Source*: http://www.physics.unlv.edu/ jeffery/astro/mechanics/

In two-dimensional polar coordinates, the second law would be changed. The velocity has the following form:

$$\mathbf{v} = \dot{\mathbf{r}} = \dot{r}\hat{\mathbf{r}} + r\dot{\phi}\hat{\boldsymbol{\phi}} \tag{2.7}$$

Taking a derivative to get the acceleration, we realize the force will be in the form of

$$\mathbf{F} = F_r\hat{\mathbf{r}} + F_\phi\hat{\boldsymbol{\phi}} \tag{2.8}$$

with the radial and the angular components

$$F_r = m(\ddot{r} - r\dot{\phi}^2) \quad \text{and} \quad F_\phi = m(r\ddot{\phi} + 2\dot{r}\dot{\phi})$$

This would be useful in the cases such as the motion of a pendulum, etc.

2.2 Projectile motion and drag

Now, let us extend our case to 2-dimensional. The most common topic might be the projectile motion. Typically, we can still sketch a free-body diagram for the moving object (a ball, etc). But now we have not only an x direction but also a y direction. Given an initial velocity, it should be in the following form:

$$\mathbf{v}_0 = v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}}$$

, where $v_x = v_0\cos\theta_0$ and $v_y = v_0\sin\theta_0$. There would be an angle θ between the horizontal and the vertical directions. Under an ideal situation, we will consider the horizontal velocity $\mathbf{v_x}$ is constant while the vertical

velocity $\mathbf{v_y}$ is varying. Also, at the vertex of the projectile parabola, $\mathbf{v_y} = 0$. An entire process of an object in flight is shown in Figure 6.

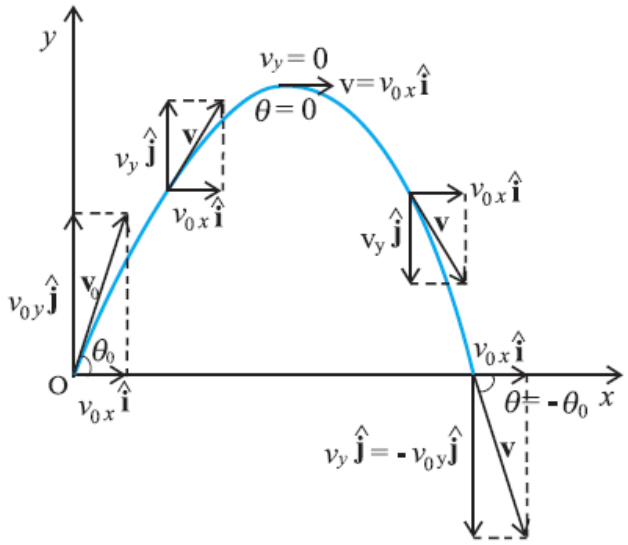

Figure 6: The free-body diagram of a projectile motion *Source*: https://sites.google.com/site/123iitphysics/

Some useful equations:

$$\Delta x = v_0 \cos\theta_0 t, \quad \Delta y = v_0 \sin\theta_0 t - \frac{1}{2}gt^2 \tag{2.9}$$

$$v_y = v_0 \sin\theta_0 - gt, \quad v_y^2 = (v_0 \sin\theta_0)^2 - 2g\Delta y \tag{2.10}$$

As usual, $\Delta x = x - x_0$ and $\Delta y = y - y_0$. These can be derived from Eq (2.1). First, let us think about the maximum height. To derive the expression, we can consider $v_y = 0$ and then plug into the second equation of Eq (2.10), so you will get

$$h = \frac{(v_0 \sin\theta_0)^2}{2g} \tag{2.11}$$

Similarly, consider the two equations in Eq (2.9) with $\Delta y = 0$, we can work out an expression for the horizontal range,

$$R = \frac{v_0^2 \sin 2\theta_0}{g} \tag{2.12}$$

This is the distance for an object that complete the flight until reaching the initial height. If you try to derive this expression by hand, you might also find the flying time

$$t = \frac{2v_0 \sin\theta_0}{g} \tag{2.13}$$

Besides, from Eq (2.9), you can also derive the trajectory of this projectile motion:

$$y = x \tan\theta_0 - \frac{gx^2}{2(v_0 \cos\theta_0)^2} \tag{2.14}$$

Hence, you might notice that Eq (2.9) and Eq (2.10) are conceptually helpful for deriving out all the other related expressions.

Last, let us consider the air resistance. In general, we called air resistance as **drag** in physics. Usually we can neglect these resistive forces when solving essential problems. But once we are going to consider their effects, a drag force can be written as

$$f = bv + cv^2 + ...$$

with the coefficients b and c. Usually, two approximate (linear and quadratic) terms are good enough. Consider the direction of a linear drag, it will be $\mathbf{f} = -b\mathbf{v}$ (air resistance must be in an opposite direction), and the equation of motion would be

$$m\dot{\mathbf{v}} = m\mathbf{g} - b\mathbf{v} \tag{2.15}$$

For the vertical motion, we can figure out the terminal speed

$$v_{\text{ter}} = \frac{mg}{b} \tag{2.16}$$

Similarly, for the quadratic case, it will become

$$v_{\text{ter}} = \sqrt{\frac{mg}{c}}$$

Imagine you throw down a ball vertically, its velocity will gradually become a constant when $F_g = f$ and the initial acceleration will be 0. Such a velocity is the **terminal velocity**.

In the fluid dynamics, the drag force is dependent on the air density ρ and the effective cross-sectional area A:

$$f = \frac{1}{2}C_d A\rho v^2 \tag{2.17}$$

, where C_d is the drag coefficient. In this way, the terminal speed is

$$v_{\text{ter}} = \sqrt{\frac{2mg}{C_d A\rho}} \tag{2.18}$$

2.3 Energy and work

Energy is ubiquitous in the universe. In our daily life, we see that energy can be in different forms. The most basic definition is the **kinetic energy**

$$T = \frac{1}{2}mv^2 \tag{2.19}$$

It is also usual to denote as K, but in the future topics, I think T would be more professional. Now, let us introduce the definition of the **work**:

$$W = \mathbf{F} \cdot \mathbf{r} \tag{2.20}$$

It is simply the inner product of the force and the displacement. Equivalently,

$$W = \int \mathbf{F} \cdot d\mathbf{r} \tag{2.21}$$

This formula implies that, if you have a plot with the displacement Δr and the force $F(r)$, the work done by this force is equal to the area under its function. Now you can try to take a derivative of Eq (2.19) with respect to time, and you will find that $dT = \mathbf{F} \cdot d\mathbf{r}$. This follows the **work-kinetic energy theorem**:

$$\boxed{\Delta T = \int \mathbf{F} \cdot d\mathbf{r} = W} \tag{2.22}$$

When it comes to the work, the power can also be defined:

$$P = \frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v} \tag{2.23}$$

The mechanical energy $E = T + U$ indicates that we still have to define another quantity – **potential energy**. Given an object moving from a point a to b, the potential energy can be defined by its force

$$\boxed{\Delta U = -\int_a^b \mathbf{F}(\mathbf{r})\, d\mathbf{r} = -W} \tag{2.24}$$

The gravitational force at a certain height can be written as $F_g(h) = -mg$, and we can directly derive the potential energy of gravity:

$$U(h) = mgh \tag{2.25}$$

For the block-spring system, the force $F_k(x) = -kx$, and then the elastic potential energy can be defined as

$$U(x) = \frac{1}{2}kx^2 \tag{2.26}$$

Moreover, you might have notice that the sum of kinetic energy (2.22) and potential energy (2.24) would be equal to 0. This gives a conclusion that if a force acting an object is conservative, then the mechanical energy

$$E = T + U \tag{2.27}$$

would never change. We shall discuss the conservation laws later.

Then, the force such as gravity, elastic force, or electric force – they depends on the displacement instead of the path. This leads to a very important concept that the **conservative force** must be *path independent* (Figure 7). From the definition of the potential energy (2.24), you can simply rearrange and find that

$$F(r) = -\frac{dU(r)}{dr}$$

Here I did not use any vector's notations. But if you consider a more general case, particularly in 3-dimensional case, you will need to consider the vector forms and the gradient in the first line of Eq (1.7). In this way, we shall obtain the following expression:

$$\boxed{\mathbf{F} = -\nabla\, U} \tag{2.28}$$

This shows that the force is the gradient of potential energy. By the way, here is also a useful relation between the derivative of a function and its gradient:

$$df = \nabla f \cdot d\mathbf{r} \tag{2.29}$$

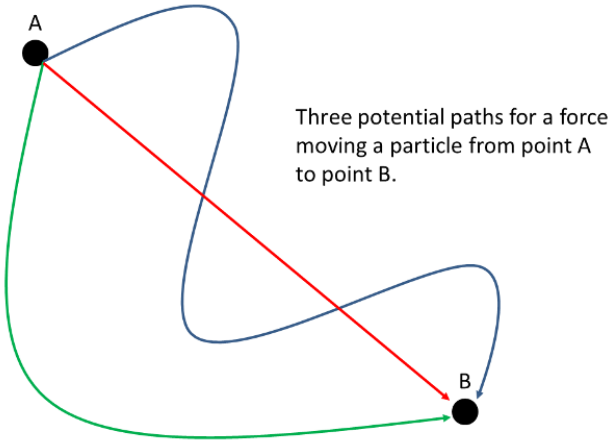

Figure 7: A conservative force must satisfy path independence, that is, depends on the displacement between point A and B (red) rather than the other paths *Source*: https://study.com/academy/lesson/conservative-forces-examples-effects.html

Next, from the second line of Eq (1.10), we can replace the scalar function f by our potential energy U. In this way,

$$\nabla \times (\nabla U) = \boxed{\nabla \times \mathbf{F} = 0} \tag{2.30}$$

Thus, for a conservative force, its curl must be zero. When we employ the Stokes's theorem (1.12), it also gives

$$W = \oint \mathbf{F} \cdot d\mathbf{r} = 0 \tag{2.31}$$

That is, given a particle moving from Point A to B along with a closed path, the work done by the force would be zero. Once any kind of force obeys these conditions, then the force is conservative.

There are also some exceptions, for example, friction is a non-conservative force because it transfers mechanical energy into thermal energy to some extent. Imagine a rock falling into a elastic spring, it should be rebounded and we can calculate the change of the kinetic / potential energy. In any case, the sum should be equal to the initial potential energy at the momentum we release from the air. However, if a rock directly falling into the ground, the friction of the table transfers a few amount of energy into thermal energy, and the mechanical energy also transfers into the sound or the deformation of your table – the scenario is not as simple as what we had for kinetic energy and potential energy in the ideal spring's case.

2.4 Momentum and rotation

The momentum's definition is as the following:

$$\mathbf{p} = m\mathbf{v} \tag{2.32}$$

, which can also be related to force by Eq (2.3),

$$\boxed{\mathbf{F} = \dot{\mathbf{p}}} \tag{2.33}$$

We are first going to discuss the linear momentum. To begin with, let's define the **center of mass** (com or CM) of a system. For the position of a two-particle system, it can be written as

$$x_{\text{CM}} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}$$

Similarly, the velocity's expression can be obtained by replacing x as v. In general, the position's expression for several particles can be defined as

$$\mathbf{R}_{\text{CM}} = \frac{1}{M} \sum_{i=1}^{n} m_i \mathbf{r}_i \tag{2.34}$$

where M is the total mass of the system. For a continuously distributed body, a more generalized expression is given by

$$\mathbf{R}_{\text{CM}} = \frac{1}{M} \int \mathbf{r}\, dm = \frac{1}{M} \int \rho \mathbf{r}\, dV \tag{2.35}$$

Recall $dm = \rho\, dV$. This expression can be used to calculate the CM of a sphere, a cone, etc. Back to the linear momentum, in the CM framework we shall have

$$\mathbf{p} = M\dot{\mathbf{R}}_{\text{CM}} \quad \text{and} \quad \mathbf{F}_{\text{ext}} = M\ddot{\mathbf{R}}_{\text{CM}}$$

Next, a common application is **collision**. The momentum with a small change can be expressed as $d\mathbf{p} = \mathbf{F}(t)\, dt$. Taking the integral and we can immediately define a new quantity as $\mathbf{J}$ as the following:

$$\mathbf{J} = \int_{t_i}^{t_f} \mathbf{F}(t)\, dt = \Delta\mathbf{p} \tag{2.36}$$

In a more straightforward form, the **impulse** $\mathbf{J} = \mathbf{p}_f - \mathbf{p}_i$. Above is also similar to the work-energy theorem. Sometimes we don't know how the force varies but we know the average force during a time interval, $\bar{F}$. In this way, the magnitude of the impulse can be written as

$$J = \bar{F}\Delta t \tag{2.37}$$

Then, if we would like to consider the rotation, we have to define the rotational variables first. The angular position can be defined by the arc-length and the circular radius:

$$\theta = \frac{s}{r}$$

Converting to the unit, we would say one revolution (rev) is equal to 2π radians (rad). Similar to the linear case, once we know the position, we can next define the (instantaneous) angular velocity

$$\omega = \frac{d\theta}{dt}$$

And then we can get the (instantaneous) angular acceleration

$$\alpha = \frac{d\omega}{dt}$$

Consequently, we can rewrite the Eq (2.1) in the angular form:

$$\begin{aligned}
\theta &= \theta_0 + \bar{\omega}t \\
\omega &= \omega_0 + \alpha t \\
\Delta\theta &= \omega_0 t + \frac{1}{2}\alpha t^2 \\
\Delta\theta &= \omega t - \frac{1}{2}\alpha t^2 \\
\Delta\theta &= \frac{1}{2}(\omega_0 + \omega)t \\
\omega^2 &= \omega_0^2 + 2\alpha\Delta\theta
\end{aligned} \tag{2.38}$$

In physics, we often use the linear variables to describe angular variables, which is especially related by the position (radius) R. For example, the position will be

$$s = \theta R \tag{2.39}$$

and the speed will be

$$v = \omega R \tag{2.40}$$

Note that $\mathbf{v} \perp \mathbf{R}$, which means the velocity is always in tangential direction. Before determining the acceleration's case, we can introduce the period. It is straightforward because a time rotating around a circle is just the distance ($2\pi R$) divided by speed (v), and this follows:

$$T = \frac{2\pi R}{v} = \frac{2\pi}{\omega} \tag{2.41}$$

Or equivalently,

$$\omega = \frac{2\pi}{T} = 2\pi f \tag{2.42}$$

where f is the frequency. The acceleration has two components: radial component a_r and tangential component a_t. The tangential component is defined as

$$a_t = \alpha R \tag{2.43}$$

, while the radial component is

$$a_r = \frac{v^2}{R} = \omega^2 R \tag{2.44}$$

We have seen that the angular variables θ, ω, and α are corresponding to the linear variables x, v, and a. For the mass of a rotational rigid body, it has to be generalized, too. Now, let us introduce a new quantity, the momentum of inertia I:

$$I = \sum m_i r_i^2 \tag{2.45}$$

In fact, this is analogous to the "mass" in linear cases. The definition can be derived when we write the kinetic energy in the angular form. The kinetic energy will be

$$K = \frac{1}{2} I\omega^2 \tag{2.46}$$

You might realize that this is fairly similar to what we defined previously, $T = \frac{1}{2}mv^2$. The moment of inertial of a continuous body is

$$I = \int r^2 \, dm \tag{2.47}$$

You can then derive many cases as shown in Figure 8. Besides, if we have already known the momentum of inertia about a parallel axis through the body's CM, I_{CM}, we can simply get the momentum of inertia shifted by a distance h from the CM by the following rule:

$$I = I_{\text{CM}} + Mh^2 \tag{2.48}$$

This is called the **parallel-axis theorem**.

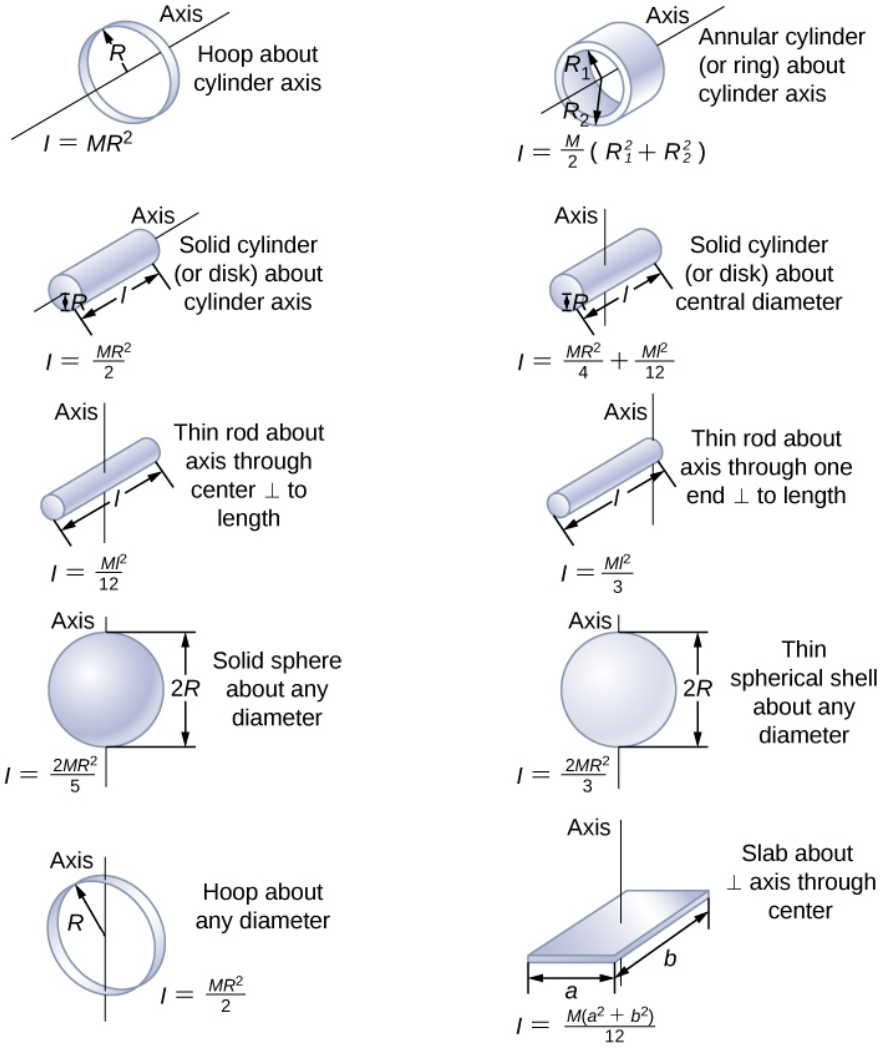

Figure 8: The moment of inertias of some rigid bodies.
Source: https://cnx.org/contents/lQXiC0Me@65.2:PHHFYWlM@3/Moment-of-Inertia

In Newton's laws of motion, we usually describe "force" in physics. As for the rotational case, we shall determine the **torque**, which is a turning / twisting action due to the force $\mathbf{F}$:

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \tag{2.49}$$

If you only consider the magnitude, then

$$\tau = Fr \sin\theta$$

and θ is the angle between the force and the moment arm. As mentioned, the moment of inertia I plays a role in the mass m in Newton's second law. Now, the torque τ is analogous to the force F. Formally, the rotational version of Newton's second law is shown as the following:

$$\boxed{\boldsymbol{\tau} = I\boldsymbol{\alpha}} \tag{2.50}$$

This is analogous to $\mathbf{F} = m\mathbf{a}$. In the same fashion, the work and the power can be rewritten as

$$W = \int_{\theta_1}^{\theta_2} \tau \, d\theta$$

and

$$P = \boldsymbol{\tau} \cdot \boldsymbol{\omega}$$

Incidentally, sometimes we are interested in the motion of "rolling". In effect, this is just the combination of the rotational motion and the translational motion. The kinetic energy of rolling is:

$$K = \frac{1}{2} I_{\mathrm{CM}} \omega^2 + \frac{1}{2} M v_{\mathrm{CM}}^2 \tag{2.51}$$

For any inclined plane, the linear acceleration can be calculated based on

$$a_{\mathrm{CM,x}} = -\frac{g \sin\theta}{1 + I_{\mathrm{CM}}/MR^2} \tag{2.52}$$

where R is the radius of the rolling object.

We can move on to the next topic – **angular momentum**. The definition is simply the cross product of the momentum and the position vector:

$$\boldsymbol{\ell} = \mathbf{r} \times \mathbf{p} = m(\mathbf{r} \times \mathbf{v}) \tag{2.53}$$

For the magnitude, it will be $\ell = mvr \sin\theta$. The reader should bear in mind that the angular momentum is the analog of the linear momentum $\mathbf{p}$. Therefore, $\mathbf{F} = \dot{\mathbf{p}}$ implies that

$$\boxed{\boldsymbol{\tau} = \dot{\boldsymbol{\ell}}} \tag{2.54}$$

Indeed,

$$\dot{\boldsymbol{\ell}} = \dot{\mathbf{r}} \times \mathbf{p} + \mathbf{r} \times \dot{\mathbf{p}} = m(\dot{\mathbf{r}} \times \dot{\mathbf{r}}) + \mathbf{r} \times \dot{\mathbf{p}} = \mathbf{r} \times \mathbf{F} = \boldsymbol{\tau}$$

Here we employed the mathematical fact that $\dot{\mathbf{r}} \times \dot{\mathbf{r}} = 0$.

Now, let us consider a rigid body. What we would like to do is considering many segments (particles) that make up with a body. Therefore, the angular momentum can be written as

$$\mathbf{L} = \sum \boldsymbol{\ell}_i = \sum m\mathbf{r}_i \times \mathbf{v}_i = \sum m\mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i) = \boldsymbol{\omega} \sum m_i \mathbf{r}_i^2 = I\boldsymbol{\omega}$$

If we use the "BAC-CAB" rule in mathematics,

$$A \times (B \times C) = B(A \cdot C) - C(A \cdot B) \tag{2.55}$$

We shall finally derive out the angular momenta for three components:

$$\begin{aligned} L_x &= I_{xx}\omega_x + I_{xy}\omega_y + I_{xz}\omega_z \\ L_y &= I_{yx}\omega_x + I_{yy}\omega_y + I_{yz}\omega_z \\ L_z &= I_{zx}\omega_x + I_{zy}\omega_y + I_{zz}\omega_z \end{aligned} \tag{2.56}$$

They satisfy the following expressions:

$$\begin{aligned} I_{xx} &= \sum m_i(y_i^2 + z_i^2) \\ I_{xy} &= -\sum m_i x_i y_i \end{aligned} \tag{2.57}$$

Some readers might notice that it would be more convenient to write the angular momentum as a more compact form:

$$L_i = \sum_{i=1}^{3} I_{ij}\omega_j \tag{2.58}$$

where the moment of inertia can be in the following tensor's form:

$$\mathbf{I} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix}$$

As for the angular momentum $\mathbf{L}$ and the angular velocity $\boldsymbol{\omega}$, they are also have three components x, y, z. This gives a crucial expression in the rotational motion of rigid bodies:

$$\boxed{\mathbf{L} = \mathbf{I}\boldsymbol{\omega}} \tag{2.59}$$

Given a 3-dimensional example, we are interested in finding the **principal axes** of a rigid body. Then, how to find the principal axes? Actually, finding the principal axes means we need to find a diagonalized form of moment of inertia's tensor like

$$\mathbf{I} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \tag{2.60}$$

First, we notice there are three eigenvalues λ_1, λ_2, and λ_3, corresponding to three **principal moments**. Secondly, what we learned from linear algebra suggests that these three moments must be orthogonal to each other. Thus, three principal axes have to be perpendicular to each other. The way to work out this form is starting from Eq (2.59) and write down the eigenvalue equation as the following:

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega} = \lambda\boldsymbol{\omega} \tag{2.61}$$

, which means

$$(\mathbf{I} - \lambda\mathbf{1})\boldsymbol{\omega} = 0$$

Accordingly, we have to solve the following characteristic equation:

$$\det(\mathbf{I} - \lambda\mathbf{1}) = 0 \tag{2.62}$$

Finally we can work out the matrix form like Eq (2.60).

In the topic about rigid dynamics, the frame of reference of a rotating body would be different to the ordinary frame of reference. Typically, it would be more convenient to take the three principal axes as the rotating body's frame of reference (Figure 9). The dominant equation is the **Euler's equation**:

$$\dot{\mathbf{L}} + \boldsymbol{\omega} \times \mathbf{L} = \boldsymbol{\tau} \tag{2.63}$$

The form in three components is:

$$\begin{aligned} \lambda_1\dot{\omega}_1 - (\lambda_2 - \lambda_3)\omega_2\omega_3 &= \tau_1 \\ \lambda_2\dot{\omega}_2 - (\lambda_3 - \lambda_1)\omega_3\omega_1 &= \tau_2 \\ \lambda_3\dot{\omega}_3 - (\lambda_1 - \lambda_2)\omega_1\omega_2 &= \tau_3 \end{aligned} \tag{2.64}$$

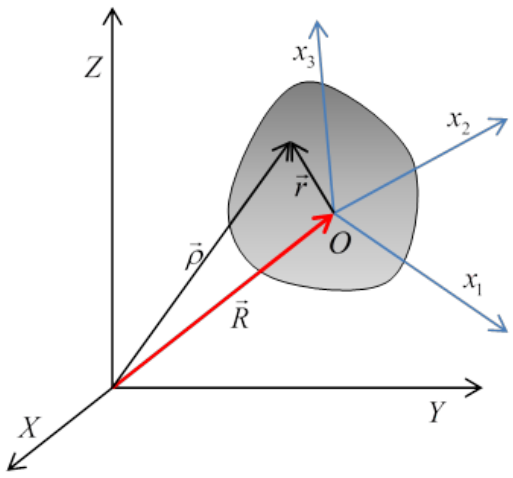

Figure 9: The space frame (XYZ) and the body frame $(x_1x_2x_3)$. *Source*: https://galileoandeinstein.phys.virginia.edu/

In many cases the torque is zero. The differences among the principal moments also decide the stability. This requires us solve the differential equations of the zero-torque Euler's equation. An interesting example is tossing a book: you will find that the smallest and the largest principal moment are stable, whereas the intermediate principal moment is always unstable (Figure 10). This shows the usage of the Euler's equation to solve the object with rotating symmetry.

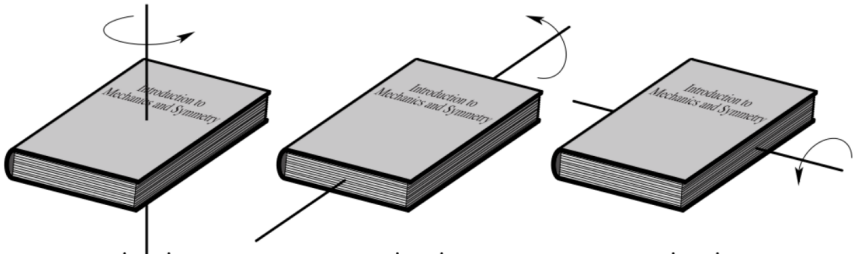

Figure 10: When you toss a book in the air, you can make it easier to rotate about the shortest (left) and the longest (middle) axis. However, it would be harder for the case of the intermediate axis (right) since it is unstable. *Source*: Jerrold E. Marsden (1999) Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems. Publisher: Springer.

2.5 Conservation laws

Conservation laws are the cores in physics. First, let's take a look at Eq (2.33). If we consider several particles, the expression can be generalized:

$$\mathbf{F}_{\text{ext}} = \dot{\mathbf{p}} \tag{2.65}$$

where $\mathbf{F}_{\text{ext}}$ represents the external forces from the other objects (perhaps is gravity, etc). We can conclude that: if there is no external force in a given system, $\mathbf{F}_{\text{ext}} = 0$, then the momentum $\mathbf{p}$ must be constant. This is the **conservation of momentum**.

In the problem about **collision**, we usually consider the conservation of linear momentum by writing down the identity:

$$\mathbf{p}_i = \mathbf{p}_f \tag{2.66}$$

For example, given a two-particle system, the momentum conservation implies that

$$m_1 v_1 + m_2 v_2 = m_1 v_1' + m_2 v_2'$$

, where v_1 and v_2 are initial velocity whereas v_1' and v_2' are final velocity. Similarly, the kinetic energy's conservation gives

$$\frac{1}{2}mv_1^2 + \frac{1}{2}mv_2^2 = \frac{1}{2}mv_1'^2 + \frac{1}{2}mv_2'^2$$

In most of the cases, these two expressions can sufficiently help us solve the one-dimensional collision problem. Depends on the situations, you can set either one object's velocity as zero, etc. But for the completely inelastic collision, imagine two cars on-head collide with each other and combine together, the momentum will be:

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2)v'$$

, which means two initially different masses combine with each other. The final velocity v' of the CM frame will be:

$$\mathbf{v}_{\text{CM}} = \frac{\mathbf{p}_1 + \mathbf{p}_2}{m_1 + m_2} = \frac{m_1\mathbf{v}_1 + m_2\mathbf{v}_2}{m_1 + m_2} \tag{2.67}$$

Here I revive the vector notation. Back to the elastic collision, given the expression about the momentum's conservation and kinetic energy's conservation, we shall figure out the following magnitude:

$$\begin{aligned} v_1' &= \frac{m_1 - m_2}{m_1 + m_2}v_1 + \frac{2m_2}{m_1 + m_2}v_2 \\ v_2' &= \frac{2m_1}{m_1 + m_2}v_1 + \frac{m_2 - m_1}{m_1 + m_2}v_2 \end{aligned} \tag{2.68}$$

For the two-dimensional case, we should consider the angles after collision. For instance, a billiard ball collide with another (stationary) one, as shown in Figure 11. In this case, we also need to consider the conservation of momentum and kinetic energy. By the conservation of linear momentum, we can write down the expressions of two components:

$$\begin{aligned} m_1 v_1 &= m_1 v_1' \cos\theta_1 + m_2 v_2' \cos\theta_2 \quad (x\text{-component}) \\ 0 &= -m_1 v_1' \sin\theta_1 + m_2 v_2' \sin\theta_2 \quad (y\text{-component}) \end{aligned}$$

And also, the conservation kinetic energy yields

$$\frac{1}{2}m_1 v_1^2 = \frac{1}{2}m_1 v_1'^2 + \frac{1}{2}m_2 v_2'^2$$

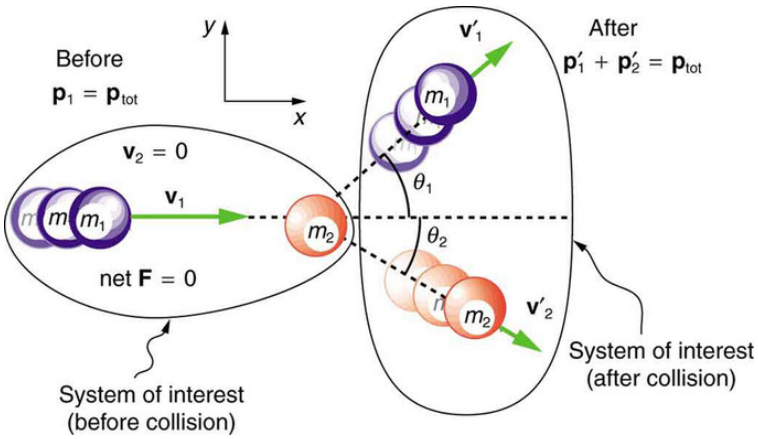

Figure 11: The 2-dimensional collision of two billiard balls. *Source*: https://courses.lumenlearning.com/boundless-physics/chapter/collisions/

Another application is the rocket. As shown in Figure 12, a rocket can eject a mass dm of some combustion product and accelerate a little bit more due to Newton's third law. After some derivations, we will obtain

$$-\frac{dM}{dt}v_{\text{rel}} = Ma \tag{2.69}$$

where the velocity of rocket relative to a mass dm is $v_{\text{rel}} = v + dv - u$, and u here is the velocity of dm relative to frame. Solving the differential eqution of dv, we will also get

$$v_f - v_i = v_{\text{rel}} \ln \frac{M_i}{M_f} \tag{2.70}$$

These two are sometimes called the rocket equations.

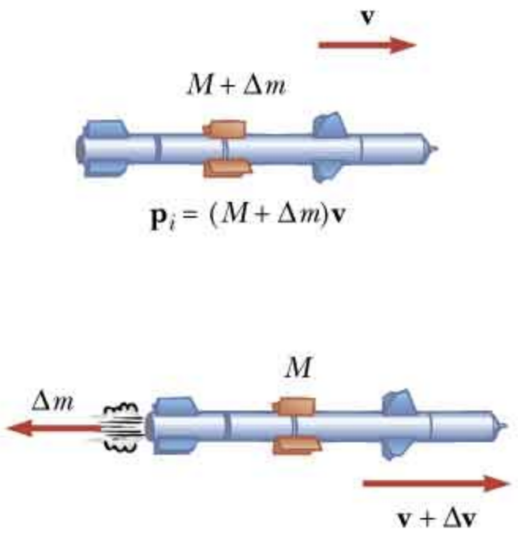

Figure 12: An accelerating rocket ejects a mass can be solved by the conservation of momentum. *Source*: http://www.ux1.eiu.edu/ cfadd/1350/09Mom/Rock.html

The second conservation law is the **conservation of angular momentum**. Similar to the linear momentum's case, the conservation demands $\dot{\mathbf{L}} = 0$, that is to say, $\mathbf{L}$ must be a constant. We can also write as

$$I_i\boldsymbol{\omega}_i = I_f\boldsymbol{\omega}_f \tag{2.71}$$

For the system of multiple particles,

$$\boldsymbol{\tau}_{\text{ext}} = \dot{\mathbf{L}} \tag{2.72}$$

If the net external torque is zero, $\boldsymbol{\tau}_{\text{ext}}$, the system's total angular momentum

$$\mathbf{L} = \sum \mathbf{r}_i \times \mathbf{p}_i$$

must be constant.

Finally, the **conservation of energy** is in the similar fashion. If this is established, it suggests that the mechanical energy we previously defined

$$E = T + U$$

would be constant during a given time interval. In a more general case, we will consider the time-dependent potential energy $U(\mathbf{r}, t)$. Taking the derivative of the kinetic energy,

$$dT = \frac{dT}{dt}dt = m\dot{\mathbf{v}} \cdot \mathbf{v}\, dt = \mathbf{F} \cdot d\mathbf{r}$$

As for the potential energy, the derivative (in 3 dimensions) is

$$dU = \underbrace{\frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy + \frac{\partial U}{\partial z}dz}_{=\nabla U \cdot d\mathbf{r}} + \frac{\partial U}{\partial t}dt = -\mathbf{F} \cdot d\mathbf{r} + \frac{\partial U}{\partial t}dt$$

Notice that we have derived $dT = \mathbf{F} \cdot d\mathbf{r}$. After plugging in and rearrange, we shall obtain the following relation:

$$\frac{dE}{dt} = \left(\frac{\partial U}{\partial t}\right)_{\mathbf{r}} \tag{2.73}$$

where $dE = d(T + U)$. This shows that the mechanical energy E is conserved when U is independent of time, that is, $\partial U/\partial t = 0$. From here, let's collect the conservation laws we had so far and their essences:

- The conservation of **momentum** $\iff$ **Translational** invariance
- The conservation of **angular momentum** $\iff$ **Rotational** invariance
- The conservation of **energy** $\iff$ **Time** invariance

These are the fundamental observations of the **Noether's theorem**, which is the mathematical description of the conservation laws.

2.6 Gravity and Kepler's laws

Gravity is a force that is in proportional to the masses but inversely proportional to the distance. Newton's law of gravitation gives

$$\boxed{F_g = \frac{GMm}{r^2}} \tag{2.74}$$

For the gravitational acceleration, the expression becomes

$$g = \frac{GM}{R^2} \tag{2.75}$$

Given the mass and the radius of Earth, the gravitational acceleration is 9.8 m/s^2.

Given the definition of potential energy (2.24), we can derive that

$$W = \int_R^\infty \mathbf{F} \cdot d\mathbf{r} = -\frac{GMm}{R}$$

Note the lower limit R represents the radial distance from Earth's center. Recall that this is the work required to move the object from a distance R to infinity. Therefore, the gravitational potential energy is

$$U = -\frac{GMm}{r} \tag{2.76}$$

Consider the escape speed, it means that the required speed when the kinetic energy and the potential energy can cancel each other. Taking R as a radius of an object (a planet, etc.), we have

$$K + U = \frac{1}{2}mv^2 - \frac{GMm}{R} = 0$$

Hence, the escape speed is

$$v = \sqrt{\frac{2GM}{R}} \tag{2.77}$$

If we consider the orbit, we should first define the **centripetal force**:

$$\begin{aligned} F_c = ma_c &= m\frac{v^2}{R} \\ &= m\omega^2 R = m\frac{4\pi^2 R}{T^2} \end{aligned} \tag{2.78}$$

Here R is the radius of a circular orbit and m is an object that rotates around the orbit, such as the planet. For the velocity of an orbiting object, we should consider the situation that $F_g = F_c$, that is

$$\frac{GMm}{R^2} = \frac{mv^2}{R}$$

This yields

$$v = \sqrt{\frac{GM}{R}} \tag{2.79}$$

In a more daily example, when you sit on a roller coaster moving around the circular orbit, the centripetal force would be equal to the gravity at the highest point:

$$\frac{mv^2}{R} = mg$$

Thus,

$$v = \sqrt{gR} \tag{2.80}$$

is the velocity when you experience at the top of the circular orbit.

Here I would like to have a brief introduction of the **Kepler's laws**. The motion of planets obeys Kepler's three laws:

1. **The law of orbits**: All planets move around an elliptical orbit with a star at a focus.
2. **The law of areas**: The rate for a planet sweep out an area A with respect to time t is constant,

$$\boxed{\frac{dA}{dt} = \text{const.}} \tag{2.81}$$

3. **The law of periods**: The square of period T is proportional to the cube of (semimajor) axis of the orbit R,

$$\boxed{T^2 \propto R^3} \tag{2.82}$$

The first law simply addresses the shape of a planetary orbit. The Sun locates at a focus. For the nearest point from the Sun, it is called the **perihelion**. For the farthest point, it is called the **aphelion** (Figure 13).

The second law can be derived in different ways, but here is a simple method. Given a planet moves with a time interval dt, it should sweep out a wedge with an area dA. Since the angle $d\theta$ is small, the opposite edge of the wedge is $R \sin d\theta \cong R\, d\theta$. In this way, we can approximate the area of a wedge to a triangle:

$$dA = \frac{1}{2} R(R\, d\theta)$$

Divide by the given time interval,

$$\frac{dA}{dt} = \frac{1}{2} R \left(R \frac{d\theta}{dt} \right) = \frac{1}{2} Rv$$

Recall $R\, d\theta/dt = R\omega = v$. Then, we know that the angular momentum is

$$L = mRv$$

Note that the orbital radius must be perpendicular to the moving (tangential) velocity, so $\sin\theta = 1$. Now, combine with the previous expression, we obtain

$$\frac{dA}{dt} = \frac{L}{2m} \tag{2.83}$$

It is noteworthy that L/m is called the **specific angular momentum**. The rule shows the equal area in the equal time interval (Figure 14), and also, in this process the angular momentum is conserved.

Last, the third law is also easy to be derived (though there are several ways to achieve it, here I shall show the simplest one). The simplest way is to consider our orbit as a circular orbit with a radius R. Then we should make the centripetal force equal to the gravitational force, $F_c = F_g$:

$$m\frac{4\pi^2 R}{T^2} = \frac{GMm}{R^2}$$

Eliminating the mass of planet m, we can easily arrive at the following formula:

$$T^2 = \frac{4\pi^2}{GM}R^2 \tag{2.84}$$

This is exactly the law of periods. From this derivation we also see that the Kepler's third law does not depend on the mass of planet m, but depends on the mass of the star M. We will discuss more detailed derivations in the later sections.

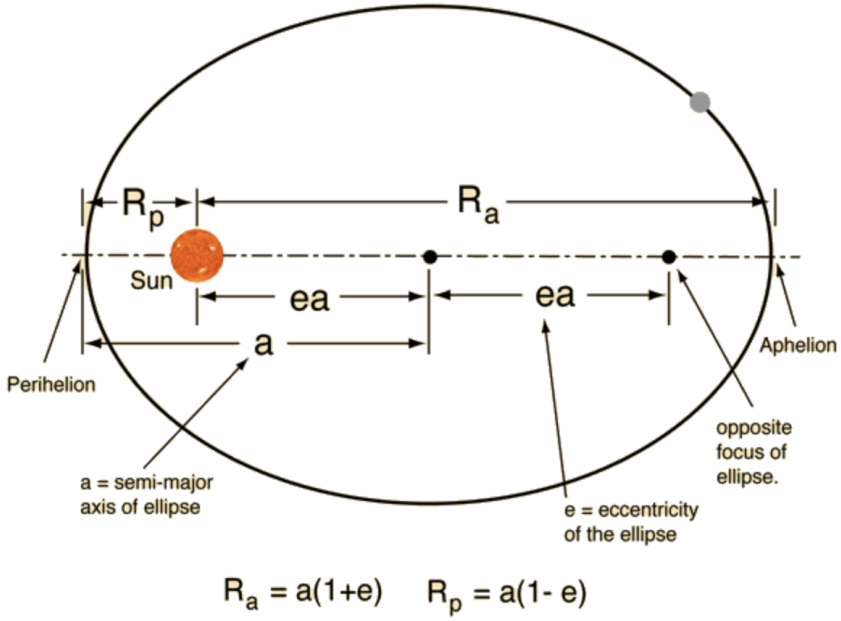

Figure 13: The first Kepler's law: A planet has an elliptical orbit with the Sun locates at the focus. *Source*: http://dmr-astronomersclub.blogspot.com/2012/09/answer-to-question-79-keplers-law-of.html

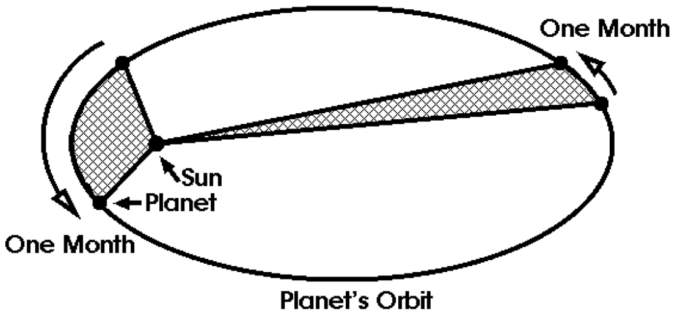

Figure 14: The second Kepler's law: Given an equal time interval, the shaded areas are equal. Therefore, a planet would travel faster at perihelion but slower at aphelion. *Source*: https://yourclasses.in/physics-gravitation-2

2.7 Fluids

In undergraduate physics, we would typically not cover too many topics about fluids. Still, this is an essential topic for many useful research areas nowadays, for example, hydrodynamics or continuum mechanics.

Fluids not only include liquid, but also include air. One of the most fundamental quantities is the pressure,

$$P = \frac{\Delta F}{\Delta A} \tag{2.85}$$

The unit is the pascal (Pa). In mostly ordinary situations,

$$1 \text{ atm} = 1.013{\times}10^5 \text{ Pa} = 1.013 \text{ bar} = 760 \text{ torr}$$

Imagine a tank of water. We are going to consider two different points with different depths (y_1, y_2), with the forces F_2 (upward) and F_1 (downward). In addition to the downward F_1, we must also consider the gravity, mg. We can write down

$$F_2 = F_1 + mg$$

Due to $m = \rho V = \rho A(y_1 - y_2)$,

$$P_2 A = P_1 A + \rho A g(y_1 - y_2)$$

This gives the following equation:

$$P_2 = P_1 + \rho g(y_1 - y_2) \tag{2.86}$$

If we set P_1 to P_0 as an atmospheric pressure on the surface, where$y_1 = 0$, then the equation becomes

$$P = P_0 + \rho g h \tag{2.87}$$

where h is the depth below the surface. But it would be more convenient to write down the differential equation. Therefore, as shown in Figure 15, we can write down the equation like (based on the free-body diagram) and use the similar method to work out:

$$\boxed{\frac{dP}{dh} = -\rho g} \tag{2.88}$$

This is the equation of the **hydrostatic equilibrium**. Such a situation can not only be calculated for our atmosphere, but also be useful to describe the condition that the star won't collapse itself.

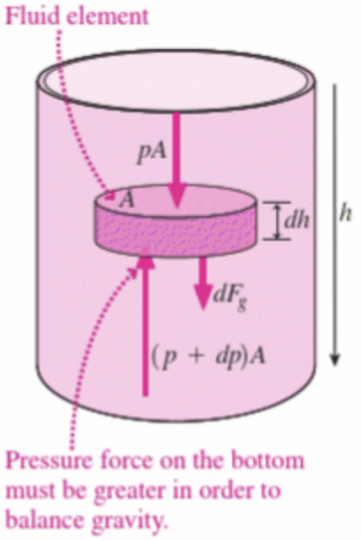

Figure 15: The diagram for hydrostatic equilibrium.
Source: Chegg.com

An application of pressure is the **Pascal's principle**. It can be described by the following equation:

$$\frac{F_1}{A_1} = \frac{F_2}{A_2} \tag{2.89}$$

This implies that, given an enclosed vessel with liquid, we input a tiny force on a piston with a smaller area can generate a stronger output force from another piston with a larger area on the other side.

In addition, for a body fully or partially submerged in a water, there would be a upward force with the following magnitude:

$$F_b = m_f g \tag{2.90}$$

This is called the **buoyant force**. Note m_f represents the mass of the fluid, instead of the body. For a floating body, the force can be written as

$$F_b = F_g = m_f g$$

If $F_b < F_g$, the body will submerge. Furthermore, according to the **Archimedes' principle**, the buoyant force acting on a body is equal to the weight of the liquid displaced by the body.

Consider an ideal fluid (steady, incompressible, frictionless, irrotational flow), the sum of pressure, gravitational term, and kinetic energy density will be a constant:

$$\boxed{P + \rho g y + \frac{1}{2}\rho v^2 = \text{const.}} \tag{2.91}$$

This is the famous **Bernoulli's equation**, which is useful to compare a flow in different locations. If we consider the elevation of a tube (with a fluid flow inside) as 0 (see Figure 16), that is, $y = 0$, then the equation can be rewritten as

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2 \tag{2.92}$$

By rearranging this equation, we have

$$P_1 - P_2 = \frac{\rho}{2}(v_2^2 - v_1^2)$$

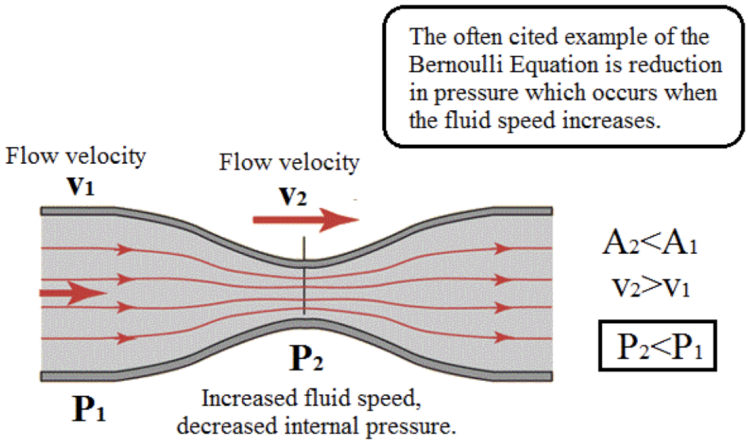

Figure 16: Bernoulli's principle indicates that the velocity increases when a flow passing through a smaller diameter's tube, and the pressure would decrease instead. *Source*: https://socratic.org/questions/a-5-7-diameter-horizontal-pipe-gradually-narrows-to-3-6-cm-the-the-water-flows-t

Meanwhile, the continuity equation demands the mass flow rate to be

$$\rho A_1 v_1 = \rho A_2 v_2 \tag{2.93}$$

Given a situation that $A_2 < A_1$, the velocity

$$v_2 = \frac{A_1}{A_2} v_1$$

implies $v_2 > v_1$. Plugging into the Eq (2.92), we have

$$P_1 - P_2 = \frac{\rho}{2} v_1^2 \left(\frac{A_1^2}{A_2^2} - 1 \right)$$

Therefore, this shows that P_2 decreases as v_2 increases.

2.8 Oscillations

The simplest model is a block-spring system (Figure 17). The force of a spring depends on the displacement x and the spring's constant k:

$$F(x) = -kx \tag{2.94}$$

This is the **Hooke's law**. The spring constant's strength depends on the way of connection,

$$k_{\text{tot}} = k_1 + k_2 \quad \text{for parallel}$$
$$\frac{1}{k_{\text{tot}}} = \frac{1}{k_1} + \frac{1}{k_2} \quad \text{for series}$$

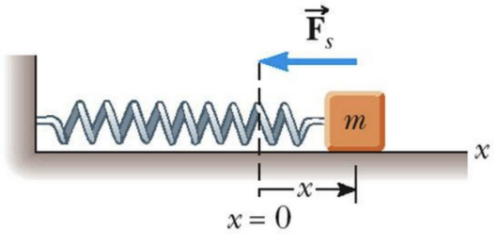

Figure 17: A block-spring system.
Source: https://www.learnpick.in/

Now, consider the Newton's second law, we have:

$$m\ddot{x} + kx = 0$$

This gives the following second-order differential equation for acceleration:

$$\boxed{\ddot{x} + \omega^2 x = 0} \tag{2.95}$$

In fact, we have assumed

$$\ddot{x} = -\frac{k}{m} x = -\omega^2 x \tag{2.96}$$

Here we have introduced

$$\omega = \sqrt{\frac{k}{m}} \tag{2.97}$$

The period is

$$T = 2\pi \sqrt{\frac{m}{k}} \quad \text{(spring)} \tag{2.98}$$

In fact, for a simple pendulum, the corresponding period is

$$T = 2\pi\sqrt{\frac{\ell}{g}} \quad \text{(simple pendulum)} \tag{2.99}$$

As for a physical pendulum, which has a non-uniform distribution of mass, a better expression is

$$T = 2\pi\sqrt{\frac{I}{mgh}} \quad \text{(physical pendulum)} \tag{2.100}$$

where I is the moment of inertia and h is a distance from the pivot point. You might notice that Eq (2.99) is a special situation that $I = mh^2 = ml^2$ under Eq (2.100).

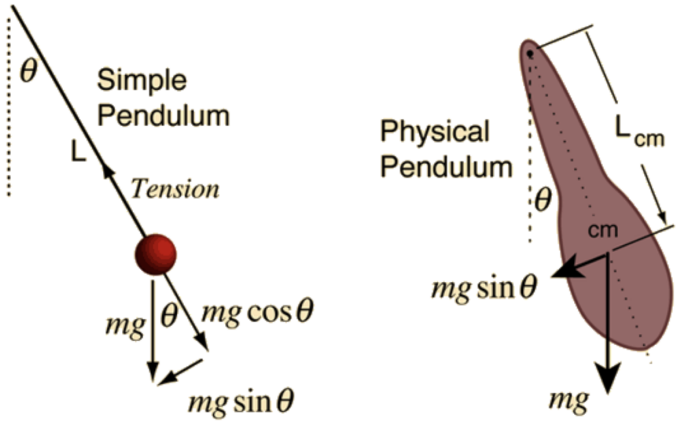

Figure 18: The free-body diagram of a simple pendulum and a physical pendulum. Note that the notation is slightly different from the text. *Source*: hyperphysics.com

Back to the differential equation (2.95), this is an equation for us to study the **simple harmonic motion (SHM)**. The solution to this differential equation is

$$x(t) = A\cos(\omega t - \phi) \tag{2.101}$$

where A is the amplitude, which is actually a constant when we solve the differential equation. The phase shift ϕ sometimes can be written with positive sign, while the negative ϕ shifts the cosine curve rightward. By taking the derivatives, you can find the expression of velocity

$$v(t) = -A\omega\sin(\omega t - \phi)$$

and acceleration

$$a(t) = -A\omega^2\cos(\omega t - \phi)$$

Armed with these, you can simply derive the energy of motion:

$$E = T + U = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kA^2 \tag{2.102}$$

Again, since k and A are constants, our energy E is constant so that this shows that the spring's force is conservative.

Next, we are going to discuss the **damped oscillations**. Due to the existence of resistive forces (air resistance, etc), our spring will slow down and the oscillation will die out eventually. Therefore, we must include the resistive force $-bv$ in our equation,

$$F = m\ddot{x} = -kx - b\dot{x}$$

Then, Eq (2.95) becomes

$$m\ddot{x} + b\dot{x} + kx = 0 \tag{2.103}$$

Now we assume

$$\beta = \frac{b}{2m}$$

is the **damping constant** and

$$\omega_0 = \sqrt{\frac{k}{m}}$$

is the **natural frequency**. In this way, our differential equation becomes

$$\boxed{\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0} \tag{2.104}$$

Let me introduce some steps for solving this second-order homogeneous differential equation. First, as usual, we should guess $x(t) = e^{rt}$ and note the general solution should be in the form of

$$x(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t}$$

Substituting $x(t) = e^{rt}$ and its derivatives into Eq (2.104), you will have the following auxiliary equation:

$$r^2 + 2\beta r + \omega_0^2 = 0$$

Solving this quadratic equation allows us to get

$$r_1 = -\beta + \sqrt{\beta^2 - \omega_0^2} \quad \text{and} \quad r_2 = -\beta - \sqrt{\beta^2 - \omega_0^2}$$

Thus, our general solution would be

$$x(t) = e^{-\beta t}\left(C_1 e^{\sqrt{\beta^2-\omega_0^2}t} + C_2 e^{-\sqrt{\beta^2-\omega_0^2}t}\right) \tag{2.105}$$

Unlike the previous case, now we have to discuss the different situations of our damping constant β:

- **Undamped oscillation** ($\beta = 0$): In this case, Eq (2.105) reduces to the previous (undamped) case,

$$x(t) = C_1 e^{i\omega_0 t} + C_2 e^{-i\omega_0 t} \tag{2.106}$$

- **Underdamped oscillation** ($\beta < \omega_0$): Since $\beta < \omega_0$, the exponential components in Eq (2.105) turns to be imaginary, so that

$$\sqrt{\beta^2 - \omega_0^2} = i\sqrt{\omega_0^2 - \beta^2} = i\omega_1$$

, where

$$\omega_1 = \sqrt{\omega_0^2 - \beta^2}$$

Then, the solution becomes

$$x(t) = e^{-\beta t}\left(C_1 e^{i\omega_0 t} + C_2 e^{-i\omega_0 t}\right)$$

There are two factors here: the one is $e^{-\beta t}$, which indicates the exponentially decaying; the other is $C_1 e^{i\omega_0 t} + C_2 e^{-i\omega_0 t}$, which is, in fact, the simple harmonic motion again and we can replace this with $A\cos(\omega_1 t - \phi)$. The solution can be rewritten as

$$x(t) = Ae^{-\beta t}\cos(\omega_1 t - \phi) \tag{2.107}$$

The decay parameter is β. The only difference here is the natural frequency has been replaced by a slightly lower frequency ω_1 .

- **Critical damping** ($\beta = \omega_0$): When the damping constant equal to the natural frequency, $\beta = \omega_0$, Eq (2.105) reduces to

$$x(t) = e^{-\beta t}$$

However, a good general solution has to include two parts. Mathematically, this is the repeated root, and what we can do is to find its partner. Soon we found that $x(t) = te^{-\beta t}$ is also a solution. In this way, the general solution for critical damping is

$$x(t) = C_1 e^{-\beta t} + C_2 t e^{-\beta t} \tag{2.108}$$

The decay parameter is $\beta = \omega_0$.

- **Overdamped oscillation** ($\beta > \omega_0$): In this case, Eq (2.105) is exactly the general solution:

$$x(t) = e^{-\beta t}\left(C_1 e^{\sqrt{\beta^2-\omega_0^2}t} + C_2 e^{-\sqrt{\beta^2-\omega_0^2}t}\right) \tag{2.109}$$

The decay parameter is $\beta - \sqrt{\beta^2 - \omega_0^2}$. The overdamping approaches to zero slower and has less obvious oscillation.

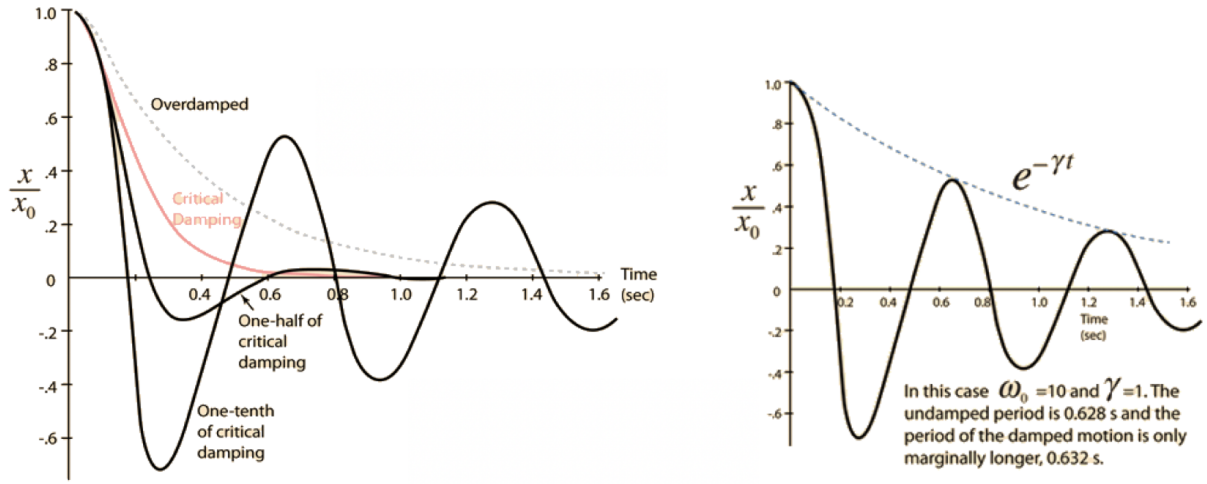

Figure 19: The oscillation curves of overdamping, critical damping, and underdamping (left); for the underdamping, we see the exponentially decay with respect to time (right). *Source*: hyperphysics.com

Finally, let us consider the case such that we impose an extra force on our oscillator. Different from the purely damped case, now we have to include a force in our equation of motion:

$$m\ddot{x} + b\dot{x} + kx = F(t) \tag{2.110}$$

Designating $f(t) = F(t)/m$, our differential equation becomes:

$$\boxed{\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = f(t)} \tag{2.111}$$

This is the second-order non-homogeneous differential equation. Indeed, this sounds more complicated. For the general solution, we should not only consider its own homogeneous solution $x_h(t)$, but also include its particular solution $x_p(t)$:

$$x(t) = x_p(t) + x_h(t) \tag{2.112}$$

For the homogeneous part, as we have shown previously,

$$x_h(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} \tag{2.113}$$

But for the particular solution, we might need to guess a solution and make some calculations. The steps are more complicated so that I shall skip the details. At the end, we will arrive at the following expression:

$$x_p(t) = A\cos(\omega t - \phi) \tag{2.114}$$

Combine Eq (2.113) and (2.114) together, our general solution will be:

$$\begin{aligned} x(t) &= A\cos(\omega t - \phi) + C_1 e^{r_1 t} + C_2 e^{r_2 t} \\ &= \underbrace{A\cos(\omega t - \phi)}_{\text{steady-state}} + \underbrace{A_{\text{tr}} e^{-\beta t}\cos(\omega_1 t - \phi_{\text{tr}})}_{\text{transient}} \end{aligned} \tag{2.115}$$

Here, you should notice that the second line has two terms, where the latter one has a factor $e^{-\beta t}$ that implies the oscillation will die out. Such a solution is called the **transient solution**. Yet, it depends on the initial condition only; for a long-term behavior, it is actually dominated by the cosine term (Figure 20).

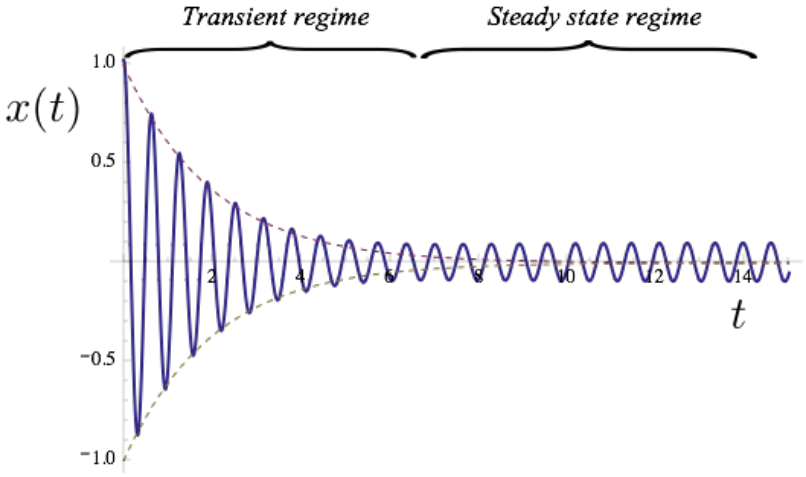

Figure 20: The driven damped oscillation has two parts: the initial condition affects the transient solution, whereas the long-term behavior is governed by the cosine term. *Source*: https://click-to-continue.online/

In the derivation, the squared amplitude is

$$A^2 = \frac{f_0^2}{(\omega_0^2 - \omega^2)^2 + 4\beta^2\omega^2} \tag{2.116}$$

First, you might find that the amplitude dependes on the driving force f_0. Besides, when ω and ω_0 are close and β is tiny, the denominator becomes small and the amplitude becomes dramatically large. This causes a very obvious response, which is called the **resonance** (Figure 21). The maximum of response is occurred when

$$\omega = \sqrt{\omega_0^2 - 2\beta^2}$$

If we put $\omega \approx \omega_0$ into Eq (2.116), we shall find that the maximum amplitude will be

$$A_{\max} \approx \frac{f_0}{2\beta\omega_0} \tag{2.117}$$

Conspicuously, when β is smaller, the maximum amplitude will be larger. Furthermore, sometimes we would like to have a very narrow resonance. A way to see its extent is to introduce the so-called **quality factor**:

$$Q = \frac{\omega_0}{2\beta} \tag{2.118}$$

If Q is larger, then it means the resonance would be sharper. Resonance is a crucial issue in engineering, because it means that despite the driving force is small, once the driving frequency ω is matched with the natural frequency ω_0, it will cause a drastic response and damage of the entire structure.

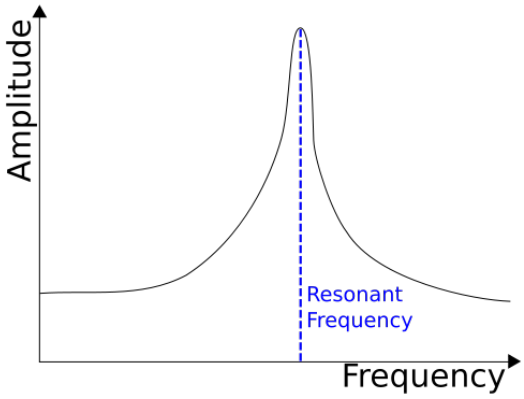

Figure 21: The sharp resonance will happen when the driving frequency ω is very close to the natural frequency ω_0. *Source*: wikibooks.org

2.9 Waves

Some of the fundamental quantities about waves have been defined in the section about momentum and rotation, for instance, the period (2.41) and the angular frequency (2.42). For the frequency, it is simply the reciprocal of the period,

$$f = \frac{1}{T} = \frac{\omega}{2\pi}$$

From the previous section, you might notice that waves are usually described by trigonometric functions,

$$u(x,t) = A\sin\left(kx - \omega t + \phi\right) \tag{2.119}$$

Here, we call k the wave number

$$k = \frac{2\pi}{\lambda} \tag{2.120}$$

and the wave speed can be defined by

$$v = \frac{\lambda}{T} = f\lambda = \frac{\omega}{k} \tag{2.121}$$

In general, we are going to discuss the transverse waves. If the traveling direction points toward $-x$, then the wave becomes $u(x,t) = A\sin\left(kx + \omega t\right)$. In fact, these sinusoidal waves are the solutions to the **wave equation**

$$\boxed{\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}} \tag{2.122}$$

, where the wave speed

$$c = \sqrt{\frac{T}{\mu}} \tag{2.123}$$

is determined by the tension of a string T and its linear mass density μ. Let me go through the derivation of this equation. As shown in Figure 22, we can write down the force expressions with horizontal and vertical components:

$$T_x = T\cos\theta_2 - T\cos\theta_1 \cong 0$$

$$T_y = T\sin\theta_2 - T\sin\theta_1 = dm\, a = \mu\, dx\, \frac{\partial^2 u}{\partial t^2}$$

Dealing with both expressions, we have

$$\frac{T\sin\theta_2}{T\cos\theta_2} - \frac{T\sin\theta_1}{T\cos\theta_1} = \tan\theta_2 - \tan\theta_1 = \frac{\mu\, dx}{T_x}\frac{\partial^2 u}{\partial t^2}$$

Recall

$$\tan\theta_1 = \left.\frac{\partial u}{\partial x}\right|_x \quad \text{and} \quad \tan\theta_2 = \left.\frac{\partial u}{\partial x}\right|_{x+dx}$$

Accordingly, we have

$$\frac{1}{dx}\left[\left.\left(\frac{\partial u}{\partial x}\right)\right|_x - \left.\left(\frac{\partial u}{\partial x}\right)\right|_{x+dx}\right] = \frac{\mu}{T_x}\frac{\partial^2 u}{\partial t^2}$$

Rearrange and then we obtain:

$$\boxed{\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}} \tag{2.124}$$

This is exactly the wave equation we have had in Eq (2.122), and now c is just replaced by v, though the expression remains the same as Eq (2.123).

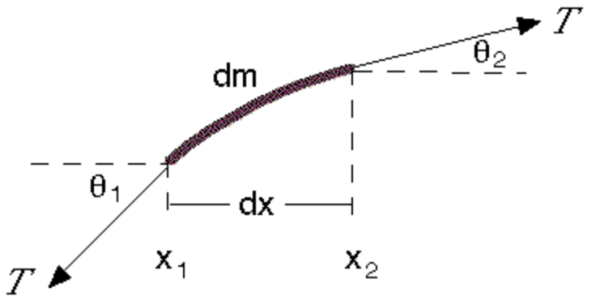

Figure 22: The derivation of the transverse wave from a string.
Source: https://www.animations.physics.unsw.edu.au/

Waves can transmit and reflect, and also, have the superposition. When a wave being transmitted between two fixed ends, the superposition leads to

$$u'(x,t) = u_1(x,t) + u_2(x,t) = A\sin(kx - \omega t) + A\sin(kx + \omega t) = 2A\sin kx\,\cos\omega t$$

Under this circumstance, the waves generate the nodes and antinodes, and they are called the **standing waves**. Consider the position x and the integer n ranging from $0, 1, 2, ...$, we have

$$kx = n\pi \quad (n = 0, 1, 2, ...)$$

Note $k = 2\pi/\lambda$, so this gives an expression of nodes

$$x = n\frac{\lambda}{2} \quad \text{(nodes)} \tag{2.125}$$

Similarly, when we consider

$$kx = \left(n + \frac{1}{2}\right)\pi \quad (n = 0, 1, 2, ...)$$

Then this gives an expression of antinodes

$$x = \left(n + \frac{1}{2}\right)\frac{\lambda}{2} \quad \text{(antinodes)} \tag{2.126}$$

You might also refer to Figure 23. Given a distance L between two ends, you can readily find out the relation between L and λ:

$$L = \frac{n\lambda}{2} \quad (n = 1, 2, 3, ...) \tag{2.127}$$

The n here indicates the **harmonic numbers** of the standing waves, and typically the first harmonic ($n = 1$) is called the fundamental mode (the second harmonic ($n = 2$) is called the first overtone, and so forth). The frequency is also limited by the position of nodes which is determined by the harmonic numbers,

$$f = \frac{v}{\lambda} = \frac{nv}{2L} \quad (n = 1, 2, 3, ...) \tag{2.128}$$

This is called the **resonant frequency**.

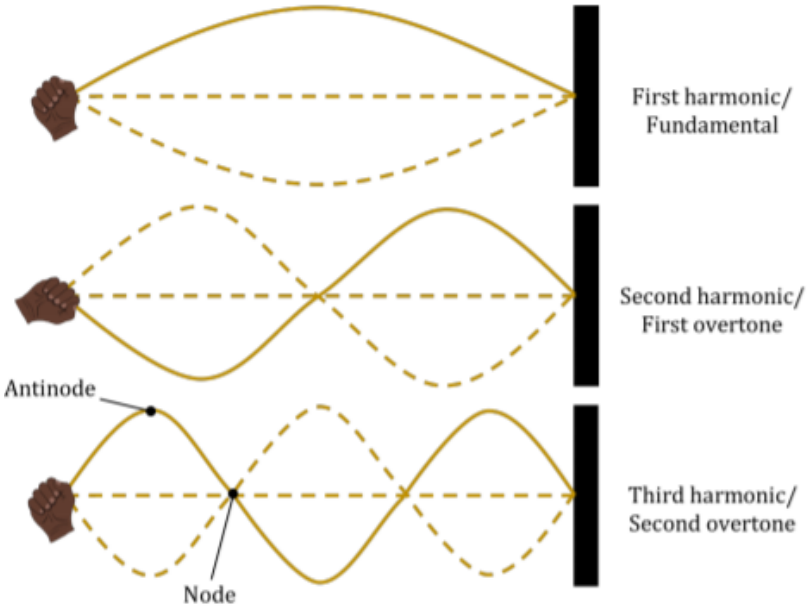

Figure 23: The derivation of the transverse wave from a string. *Source*: https://phys.libretexts.org/

If we now consider a sound wave traveling through a pipe with two open ends, the two open ends are corresponding to the antinodes (instead of nodes), while the expressions are entirely the same:

$$\lambda = \frac{2L}{n} \quad \text{and} \quad f = \frac{nv}{2L} \quad \text{for } n = 1, 2, 3, ... \quad \text{(two open ends)} \tag{2.129}$$

However, what if we consider a sound wave traveling through a closed pipe (that is, with only one open end)? Since there must be an antinode at the open end and a node at the closed end, the expressions have some changes:

$$\lambda = \frac{4L}{n} \quad \text{and} \quad f = \frac{nv}{4L} \quad \text{for } n = 1, 3, 5, ... \quad \text{(one open end)} \tag{2.130}$$

That is to say, the harmonic numbers allow odd numbers only. For the interference of waves, I might discuss in detail in the section of optics. To summarize, you might want to navigate Figure 24.

Sometimes, when two waves with slightly different frequencies, once both are matched with each other, a **beat** would arise. In this way, the beat frequency is

$$f_{\text{beat}} = f_1 - f_2 \tag{2.131}$$

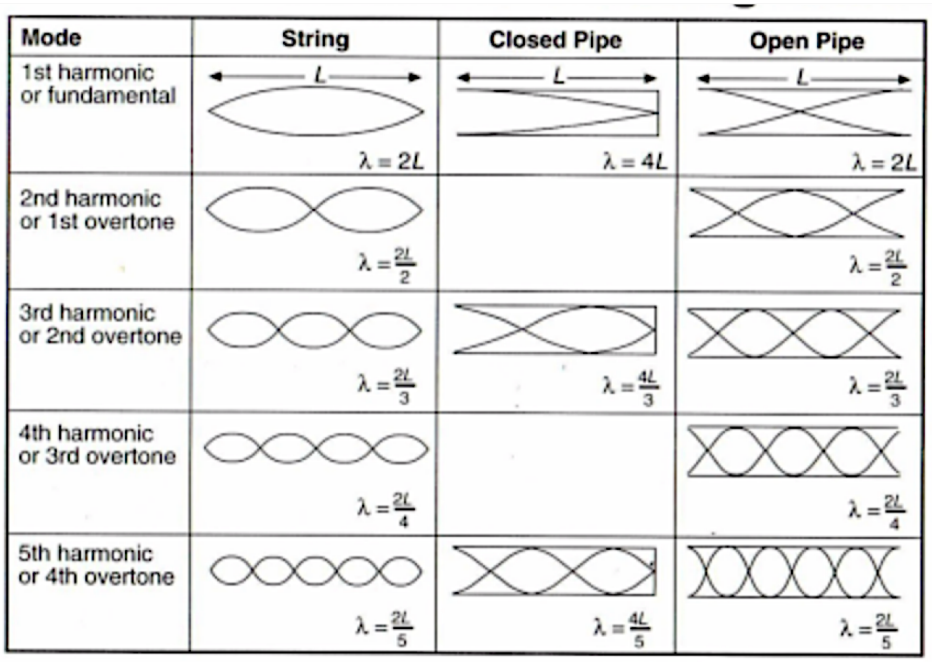

Figure 24: A comparison table for the vibration modes of a string, a sound wave in an open pipe and a closed pipe. *Source*: http://tghsl3physics.weebly.com/standing-waves.html

Now, let us explore more topics about the sounds waves. First, to define the speed of sound, we have to think about what we had for the wave speed in Eq (2.123),

$$v = \sqrt{\frac{T}{\mu}}$$

This is used to describe the wave carried by a string. It means that we have to change the tension T to something else. Indeed, assume a sound wave travels from a point source in 3 dimensions, we should introduce a quantity called the **bulk modulus**:

$$B = -V\frac{dP}{dV} \tag{2.132}$$

, where P is the pressure and V is the initial volume. Then, armed with the volume density of air ρ, we can have the formula for the speed of sound:

$$v = \sqrt{\frac{B}{\rho}} \tag{2.133}$$

This also suggests a property that the sound speed is the fastest in solid but slowest in air. There is also a relation of the sound speed and the temperature T through a dry air:

$$v = 331 + 0.6\,T \quad \text{m/s} \tag{2.134}$$

Given a temperature of 20°C, the sound speed is 343 m/s.

For the **intensity**, given P as the power of the source and A is the surface area that a sound wave transmits, we have

$$I = \frac{P}{A} = \frac{P}{4\pi r^2} \tag{2.135}$$

Incidentally, for a string wave, the average power is given by

$$\bar{P} = \frac{1}{2}A^2\mu v\omega^2$$

This is derived from the sinusoidal expressions of waves. In this fashion, the sound wave's intensity is given by

$$I = \frac{1}{2} A^2 \rho v \omega^2$$

Note A here represents the displacement amplitude. Sometimes we want to know the scale of loudness. A useful way is to define the **sound levels**:

$$\beta = 10 \log \frac{I}{I_0} \quad [\text{dB}] \tag{2.136}$$

Here dB (decibel) is the unit of sound levels. For the sound of winds the value is around 20 dB, whereas for the firework it can reach 140 dB.

Have you ever noticed that a when an ambulance moves toward you, the siren's frequency would increase; when it moves away from you, the siren's frequency would decrease? This is the **Doppler effect**, which points out the relative velocity of source and observer can influence the observed frequency. For the case of sound waves, I would suggest the reader to memorize this useful formula:

$$\boxed{f' = \frac{v \pm v_o}{v \mp v_s} f} \tag{2.137}$$

Here, f' is the observed frequency whereas f is the original frequency; v is the speed of sound (343 m/s through the air), v_o is the moving speed of the observer, and v_s is the moving speed of the source. For the sign convention, I list the rules below:

- $v + v_o$: The observer moves *toward* the source.
- $v - v_o$: The observer moves *away from* the source.
- $v - v_s$: The source moves *toward* the observer.
- $v + v_s$: The source moves *away from* the observer.

In short, toward means f' increases (shift up), whereas away means f' decreases (shift down).

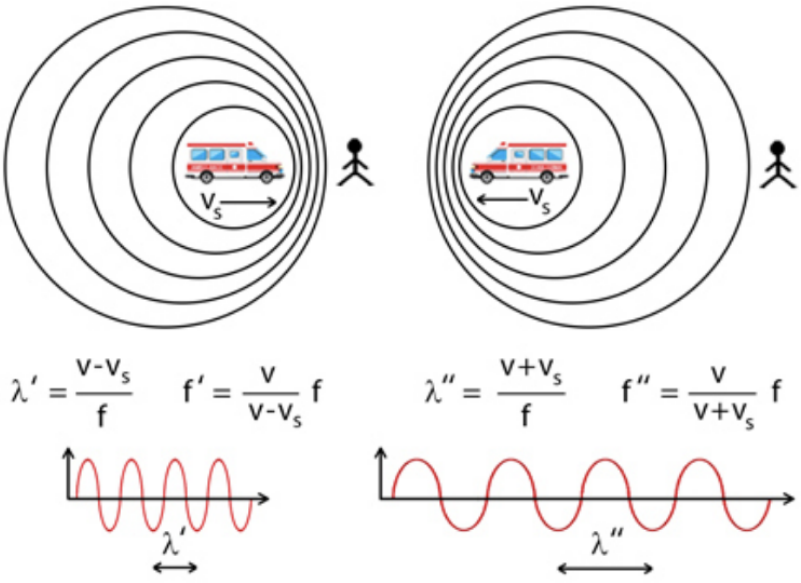

Figure 25: The Doppler effect. In the illustration, it shows the case of the stationary observer. When the source moves toward, the frequency increases, and *vice versa*. *Source*: https://www.sciencefacts.net/doppler-effect.html

Then, some readers might be curious about a scenario that the source's speed v_s is larger than the speed of sound v (that is, $v_s > v$). In this case, the v_s is called the supersonic speed, and our Doppler's formula no longer applies. Instead, it results in the **shock waves**, and you will find an angle along with the V-shape wavefronts:

$$\theta = \sin^{-1}\left(\frac{v}{v_s}\right) \tag{2.138}$$

This is called the Mach cone angle.

2.10 Lagrangian mechanics

Basically, we have covered all of the essential topics about classical mechanics. From now on, I shall introduce some advanced topics, though they are fairly useful before delving into in-depth knowledge in physics.

In physics, we are often interested in finding the "extrema". Hence, the **calculus of variations** is a crucial tool for us. Given a two points on a piece of paper, what is the shortest path between them? This looks like a silly question, because you can simply draw a straight line between them, and this is definitely the shortest path. Yet, how could you prove this is the shortest path mathematically? This becomes our ultimate goal. To begin with, let's consider the same problem: the shortest path between two points. Assume the tiny segment of an arbitrary path is denoted by

$$dS = \sqrt{dx^2 + dy^2} = \sqrt{dx^2\left(1 + \frac{dy^2}{dx^2}\right)} = \sqrt{1 + y'(x)^2}dx$$

Now you might notice that the unknown is no longer a variable, but a function $y = y(x)$. Next, the total length becomes

$$S = \int_1^2 dS = \int_{x_1}^{x_2} \sqrt{1 + y'(x)^2}\, dx \tag{2.139}$$

In the variational problem, we often write the following integral at the beginning (this form is often called a **funtional**):

$$S = \int_{x_1}^{x_2} f[y(x), y'(x), x]dx \tag{2.140}$$

The **Euler-Lagrange equation**

$$\boxed{\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = 0} \tag{2.141}$$

allows us to find the extrema of a funtional. To see how it works, let us consider our example of the shortest path between two points. From Eq (2.139), our funtional is

$$f[y(x), y'(x), x] = (1 + y'^2)^{\frac{1}{2}}$$

According to Eq (2.141),

$$\frac{\partial}{\partial y}(1 + y'^2)^{\frac{1}{2}} = 0$$

and

$$\frac{\partial}{\partial y'}(1 + y'^2)^{\frac{1}{2}} = \frac{y'}{(1 + y'^2)^{\frac{1}{2}}}$$

Then you will notice

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = 0$$

, which means $\partial f/\partial y'$ must be a constant, so we can write it as

$$\frac{\partial f}{\partial y'} = k$$

Now, rearrange this and you should find that y' can also be defined as a constant, say a:

$$y' = \sqrt{\frac{k^2}{1-k}} = a$$

Taking an integral of y' with respect to x, we obtain

$$y = \int y'\,dx = ax + b \tag{2.142}$$

Without a doubt, $y = ax + b$ means a straight line, which is exactly the shortest path between two points! A similar method can be used to calculate the path of light, and eventually you will verify that light will follow the shortest path between two points, which is called the **Fermat's principle**.

To describe a physical system, we should reset our coordinate system to the one that satisfies the calculus of variations. Thus, we should introduce the generalized coordinates:

$$(q_1, \dot{q}_1, ..., q_n, \dot{q}_n, t)$$

Then, we can define the **Lagrangian**, which is simply the kinetic energy minus the potential energy:

$$\mathcal{L} = T - V \tag{2.143}$$

Assume the 1-dimensional case, the Lagrangian in generalized coordinates is $\mathcal{L} = \mathcal{L}(x, \dot{x}, t)$. Recall the kinetic energy depends on the velocity $\dot{x}$ and the potential energy depends on the position x, according to the Euler-Lagrange equation, we have

$$\frac{\partial \mathcal{L}}{\partial x} = -\frac{\partial U}{\partial x} = F_x$$

and

$$\frac{\partial \mathcal{L}}{\partial \dot{x}} = \frac{\partial T}{\partial \dot{x}} = p_x$$

Recall the relation that $F_x = \dot{p}_x$, we can work out

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}}$$

, which is the **Lagrange's equation**:

$$\boxed{\frac{\partial \mathcal{L}}{\partial q_i} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_i}} \tag{2.144}$$

From this point, you can understand that Lagrange equation is an analogy of Newton's second law. According to Eq (2.140), we can define the **action** integral:

$$S = \int_{t_1}^{t_2} \mathcal{L}\,dt \tag{2.145}$$

Note the action must be stationary when following the actual path. In other words, the variation of action must be zero,

$$\boxed{\delta S = 0} \tag{2.146}$$

This is the **Hamilton's principle**, which is the physical version of the **least action principle** in the variational calculus.

The Lagrange's equation is the core of **Lagrangian mechanics**. Some readers might have confused: why is Lagrangian mechanics useful? The first reason is that you can choose any coordinate system as you want, on the basis of the generalized coordinates. Secondly, you can solve the problem directly by energies of a motion. In some more complicated problem, Newtonian mechanics drives us to sketch an intrigue free-body diagram, vectors by vectors, etc. But in Lagrangian mechanics, you only need to focus on the energy of a system, which is a scalar. This brings about another reason – the problem of **constraints**. Given an example of a pendulum, a string is a constraint because it limits the motion within a certain range (Figure 26). In

Newtonian mechanics, it would be an extremely tricky case to deal with the problems with constraint forces, while in Lagrangian mechanics we can just consider the energy changes within a system and readily work out the equations of motion (I shall go over this example soon). All of these, in short, are in the framework of calculus of variations.

The number of independent coordinates to describe a motion in a system gives the definition of **degree of freedom**. For instance, a particle that can move anywhere in three dimensions has 3 degrees of freedom; for a system comprised of N particles, it has $3N$ degrees of freedom. If there are two particles in the space and connected with a string, in this case, the degree of freedom is $3 \times 2 - 1$, where I minus 1 in the end means a constraint (a string) of this system. If a system has n degrees of freedom that can be described by n generalized coordinates $(q_1, ..., q_n)$, then such a system is **holonomic**. Mathematically, a system with holonomic constraint means

$$f(q_1, q_2, ..., q_n, t) = 0$$

, which usually depends on the position and the time. If a bead lies on a sphere, the radius of the sphere r and its distance from the center of a sphere d has to be $r^2 - d^2 = 0$, which is holonomic. However, if this bead will fall off the spherical surface, it means $r^2 - d^2 > 0$, and the system becomes non-holonomic. Nevertheless, in undergraduate level, we often discuss the holonomic systems.

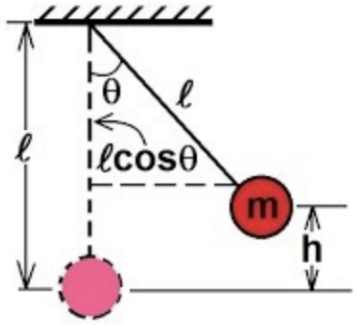

Figure 26: The free-body diagram of a simple pendulum. *Source*: http://www.ilectureonline.com/

Before moving on the next topic, let me illustrate a simple example to see how powerful the Lagrangian mechanics is. Given a simple pendulum in Figure 26 as an example. Here, we shall consider a constraint (a string) in our system, which has a length $l = \sqrt{x^2 + y^2}$. For convenience, it would be better to use the (generalized) polar coordinates, (r, θ) where $r = l$ in this case, so the velocity $v = r\omega = l\dot{\theta}$. The first step is to write down the kinetic energy

$$T = \frac{1}{2}mv^2 = \frac{1}{2}ml^2\dot{\theta}^2$$

as well as the potential energy

$$U = mgh = mgl(1 - \cos\theta)$$

Note that h represents difference that a mass swings in a motion, and you should notice that we tried to consider things based on the two variables: l and θ. In this way, we can straightforwardly write down the Lagrangian of this system:

$$\mathcal{L} = T - U = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos\theta) \tag{2.147}$$

Now, we can employ the Lagrange equation (2.144) to solve the problem:

$$\frac{\partial \mathcal{L}}{\partial \theta} = -mgl\sin\theta$$

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}}\right) = \frac{d}{dt}(ml^2\dot{\theta}) = ml^2\ddot{\theta}$$

Put an equal sign between them, we derive the equation of motion:

$$-mgl\sin\theta = ml^2\ddot{\theta} \tag{2.148}$$

At a first glance you might not have feelings about this equation. Soon, you will find that the left-hand side is the torque due to gravity, ml^2 is the moment of inertia, and $\ddot{\theta}$ is just the angular acceleration α. Well, this is undeniably the same equation as Eq (2.50),

$$\tau = I\alpha$$

From this example, we found the exactly same equation of Newtonian mechanics by the method of solving the Lagrangian of energies in a system.

2.11 Hamiltonian mechanics

In Lagrangian mechanics, the generalized coordinates depend on the position and the velocity,

$$(q_1, ..., q_n, \dot{q}_1, ..., \dot{q}_n)$$

Yet, in physics, the relations between position and momentum are often useful as well. Therefore, in **Hamiltonian mechanics**, the basic variables are

$$(q_1, ..., q_n, p_1, ..., p_n)$$

The motion would then be discussed in the 2-dimensional **phase space** with generalized positions and momenta. Typically, we give different names of spaces with different coordinates:

- **Configuration space**: n coordinates $(q_1, ..., q_n)$
- **State space**: $2n$ coordinates $(q_1, ..., q_n, \dot{q}_1, ..., \dot{q}_n)$
- **Phase space**: $2n$ coordinates $(q_1, ..., q_n, p_1, ..., p_n)$

Note the phase space is a more generalized space with all possible states therein. The Hamiltonian is defined by

$$\mathcal{H} = \sum_{i=1}^{n} p_i\dot{q}_i - \mathcal{L} \tag{2.149}$$

This is, mathematically, called the **Legendre transformation**, which converts a set of variables (like x, v) into another set of conjugate variables (like x, p). In 1-dimensional case, it is

$$\mathcal{H} = p\dot{q} - \mathcal{L} \tag{2.150}$$

, where $\mathcal{H} = \mathcal{H}(q, p)$, $\dot{q} = \dot{q}(q, p)$, so $\mathcal{L} = \mathcal{L}[q, \dot{q}(q, p)]$. Another thing you might have noticed is that

$$p_i = \frac{\partial\mathcal{L}}{\partial\dot{q}_i} \tag{2.151}$$

is the **generalized momentum**, as what we have mentioned when deriving Eq (2.144). Given a Lagrangian

$$\mathcal{L} = \frac{1}{2}M(q)\dot{q}^2 - U(q)$$

Note that I here indicated the kinetic energy depends on a more complicated function $M(q)$. Taking a derivative with respect to $\dot{q}$, we can get the generalized momentum

$$p = \frac{\partial\mathcal{L}}{\partial\dot{q}} = M(q)\dot{q}$$

Substituting p back to Eq (2.150), you will find $p\dot{q} = M(q)\dot{q}^2 = 2T$, whereas $\mathcal{L} = T - U$, and so that

$$\mathcal{H} = T + U \tag{2.152}$$

Thus, in any natural system, the Hamiltonian can be regarded as a total energy of a system.

Let us label out the variables carried by the Hamiltonian:

$$\mathcal{H}(q,p) = p\dot{q}(q,p) - \mathcal{L}[q,\dot{q}(q,p)] \tag{2.153}$$

Next, taking a derivative with respect to q and p. Using some chain rules, we will get:

$$\frac{\partial \mathcal{H}}{\partial q} = p\frac{\partial \dot{q}}{\partial q} - \left(\frac{\partial \mathcal{L}}{\partial q} + \underbrace{\frac{\partial \mathcal{L}}{\partial \dot{q}}}_{=p} \frac{\partial \dot{q}}{\partial q} \right) = \underbrace{-\frac{\partial \mathcal{L}}{\partial q} = -\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}}}_{\text{Lagrange equation}} = -\frac{d}{dt}p = -\dot{p}$$

and

$$\frac{\partial \mathcal{H}}{\partial p} = \left(\dot{q} + p\frac{\partial \dot{q}}{\partial p} \right) - \underbrace{\frac{\partial \mathcal{L}}{\partial \dot{q}}}_{=p} \frac{\partial \dot{q}}{\partial p} = \dot{q}$$

These lead to a set of two equations:

$$\dot{q} = \frac{\partial \mathcal{H}}{\partial p} \quad \text{and} \quad \dot{p} = -\frac{\partial \mathcal{H}}{\partial q} \tag{2.154}$$

In several dimensions, they can be expressed as

$$\boxed{\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}} \quad \text{and} \quad \boxed{\dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i}} \tag{2.155}$$

They are the **Hamilton's equation**.

Let me go through an example. Given an Atwood machine as shown in Figure 27, as before, we can write down the kinetic energy

$$T = \frac{1}{2}mv^2 = \frac{1}{2}(m_1 + m_2)\dot{x}^2$$

and the potential energy

$$U = -mgh = -m_1 gx - m_2 gy = -m_1 gx - (-m_2 gx) = -(m_1 - m_2)gx$$

Then, we have to figure out the generalized momentum

$$p = \frac{\partial \mathcal{L}}{\partial \dot{x}} = \frac{\partial T}{\partial \dot{x}} = (m_1 + m_2)\dot{x}$$

We can rewrite the velocity in terms of the momentum:

$$\dot{x} = \frac{p}{m_1 + m_2}$$

Armed with this, we can substitute $\dot{x}$ to the kinetic energy's term and construct a (one-dimensional) Hamiltonian:

$$\mathcal{H} = T + U = \frac{p^2}{2(m_1 + m_2)} - (m_1 - m_2)gx$$

Taking derivatives, we obtain

$$\dot{x} = \frac{\partial \mathcal{H}}{\partial p} = \frac{p}{m_1 + m_2} \quad \text{and} \quad \dot{p} = -\frac{\partial \mathcal{H}}{\partial x} = (m_1 - m_2)g \tag{2.156}$$

In fact, this gives the famous result of the acceleration of the Atwood machine:

$$\ddot{x} = \frac{\dot{p}}{m_1 + m_2} = \frac{m_1 - m_2}{m_1 + m_2} g \tag{2.157}$$

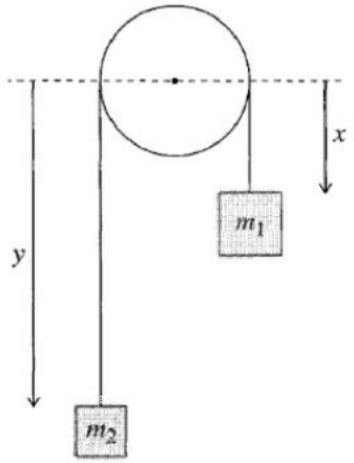

Figure 27: The free-body diagram of an Atwood machine.
Source: https://slideplayer.com/slide/17904696/

You might also notice that the Lagrange's equation is a second-order differential equation with n variables, whereas the Hamilton's equations are the first-order differential equations with $2n$ variables. This implies that Hamilton's equations are easier to be solved, while the steps are more than solving a Lagrange's equation. Nevertheless, Hamiltonian formalism can give us a clearer mathematical picture of classical mechanics. Besides, both equations have the following relations:

$$\frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t} \tag{2.158}$$

Furthermore, when solving the Hamilton's equations, sometimes you will find the situation such that

$$\frac{\partial \mathcal{L}}{\partial q_i} = \dot{p}_i = 0$$

In this case, q_i represents a **ignorable coordinate**, which can greatly simplify the complexity of a system. To find the ignorable coordinates, one can employ the **canonical transformation**

$$(\mathbf{q}, \mathbf{p}) \to (\mathbf{Q}, \mathbf{P}) \tag{2.159}$$

Or, we have a change with

$$\mathbf{Q} = \mathbf{Q}(\mathbf{q}, \mathbf{p}) \quad \text{and} \quad \mathbf{P} = \mathbf{P}(\mathbf{q}, \mathbf{p}) \tag{2.160}$$

, and the Hamilton's equations are invariant under such transformations. In fact, this is the transformation of points in a phase space. You can think of the phase-space description as a core of the Hamiltonian mechanics. In addition, the state-space orbit is also helpful to study the chaotic phenomenon (Figure 28).

Last, the **Liouville's theorem** shows that the volume of a moving closed surface in the phase space must be constant in time. In other words, as time goes by, any point that is initially in such a surface should remain inside. That is to say, the volume of any closed surface has to be a constant when moving around in the phase space (Figure 29). This leads to some subjects such as the symplectic geometry to study the topology about a phase space.

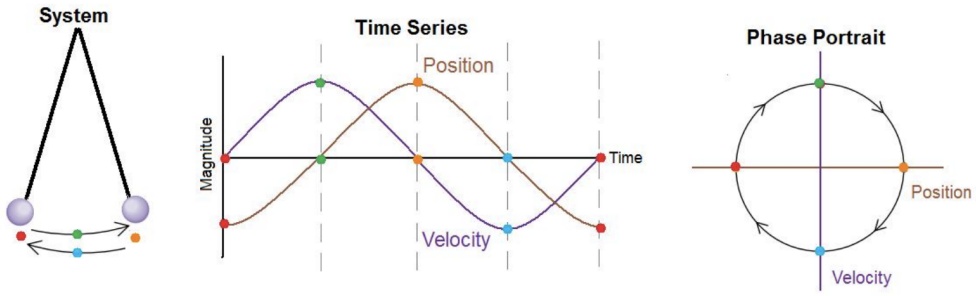

Figure 28: The phase-space orbit of the motion of a simple pendulum. Note in this case we plot the axes with position and velocity. *Source*: Wikipedia

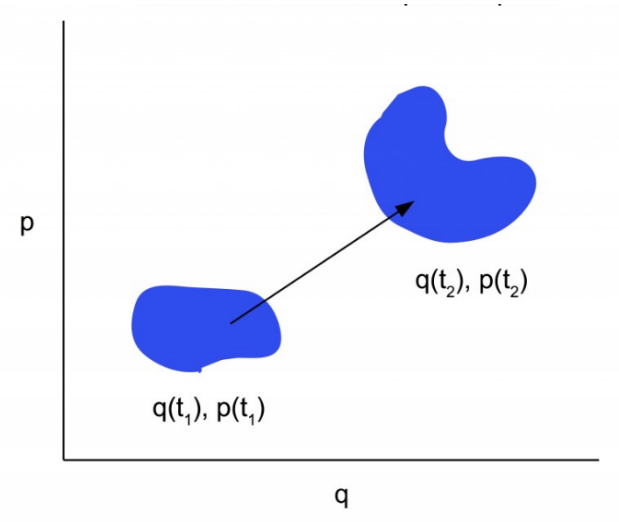

Figure 29: A two-dimensional illustration of the phase space. According to the Liouville's theorem, the volume of a closed surface would not change as it moves around with respect to time. *Source*: http://gershonwolfe.com/

2.12 Two-body problem

Two-body problem is important when we study the motion of planets or the motion of electrons within an atom. In Section 2.6, we have had a first glance of the Kepler's laws. In this section, we are going to discuss the details of the energies and the shapes or orbits.

The first concept I shall introduce is the reduced mass,

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{2.161}$$

The kinetic energy of a two-body system has the following form:

$$T = \frac{1}{2} M \dot{\mathbf{R}}^2 + \frac{1}{2} \mu \dot{\mathbf{r}}^2$$

Then, the Lagrangian will be

$$\mathcal{L} = \frac{1}{2} M \dot{\mathbf{R}}^2 + \left(\frac{1}{2} \mu \dot{\mathbf{r}}^2 - U(\mathbf{r}) \right) = \mathcal{L}_{\text{cm}} + \mathcal{L}_{\text{rel}} \tag{2.162}$$

That is, such a system can be viewed as two parts: one with the CM coordinate $\mathbf{R}$ and the other with the relative coordinate $\mathbf{r}$ (as shown in Figure 30). Since the Lagrangian $\mathcal{L}$ is independent of $\mathbf{R}$, the $\mathbf{R}$ is ignorable

so that $\dot{\mathbf{R}}$ has to be a constant. In this way, it would be more convenient to choose a frame that $\dot{\mathbf{R}} = 0$, and the Lagrangian of the CM frame vanishes, $\mathcal{L}_{\rm cm} = 0$.

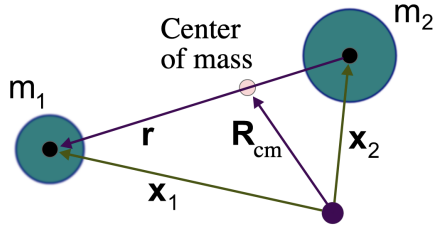

Figure 30: A two-body system. *Source*: Wikipedia

Now, let us solve the equations of motion. First, we can intuitively choose the polar coordinates because of the circular motion we are studying. According to Eq (2.7), we can write down the Lagrangian in the polar language:

$$\mathcal{L} = \mathcal{L}_{\rm rel} = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\phi}^2) - U(r) \tag{2.163}$$

We can immediately solve the angular equation (ϕ-equation)

$$\frac{\partial \mathcal{L}}{\partial \dot{\phi}} = \mu r^2 \dot{\phi} = \ell = \text{const.} \tag{2.164}$$

as well as the radial equation (r-equation)

$$\frac{\partial \mathcal{L}}{\partial r} = \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{r}} \quad \Longrightarrow \quad \mu r\dot{\phi}^2 - \frac{dU}{dr} = \mu\ddot{r} \tag{2.165}$$

Note that in Eq (2.164) we have confirmed that the term $\mu r^2\dot{\phi}$ is exactly the angular momentum ℓ, which should be a constant owing to the conservation law of angular momentum. Next, let us discuss the radial equation (2.165). The reader might have noticed that our two-body problem has been reduced to an one-dimensional problem. We can identify that $-dU/dr$ should be the (actual) **central force**, but for the term $\mu r\dot{\phi}^2$, it is a centrifugal force pointing outward. That is to say, for the radial equation,

$$\mu\ddot{r} = \underbrace{-\frac{dU}{dr}}_{\text{central force}} + \underbrace{\mu r\dot{\phi}^2}_{\text{centrifugal force}}$$

Since

$$\dot{\phi} = \frac{\ell}{\mu r^2}$$

, the centrifugal force can be written as

$$F_{\rm cf} = \mu r\dot{\phi}^2 = \frac{\ell^2}{\mu r^3} \tag{2.166}$$

Recall that in one dimension $F_{\rm cf} = -dU_{\rm cf}/dr$, from this we can infer that the centrifugal potential energy is

$$U_{\rm cf} = \frac{\ell^2}{2\mu r^2} \tag{2.167}$$

Collecting these terms, our radial equation can be rewritten as

$$\mu\ddot{r} = -\frac{dU_{\rm eff}(r)}{dr} \tag{2.168}$$

where

$$U_{\rm eff}(r) = U(r) + U_{\rm cf}(r) = -\frac{Gm_1m_2}{r} + \frac{\ell^2}{2\mu r^2} \tag{2.169}$$

is called the **effective potential energy** (Figure 31).

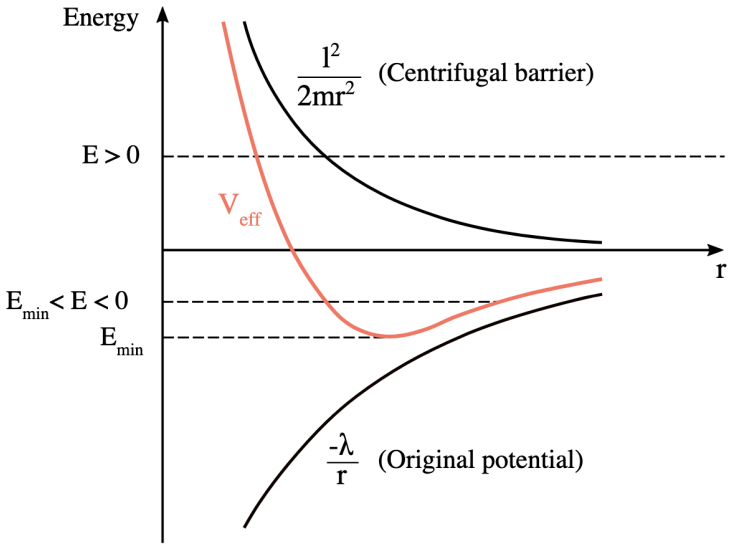

Figure 31: The effective potential compared to the gravitational (original) potential energy and the centrifugal potential energy. *Source*: https://physref.com/Classical-Mechanics/Central-forces/

We have seen that the radial equation is

$$\mu\ddot{r} = F(r) + \frac{\ell^2}{\mu r^3} \tag{2.170}$$

Alternatively, we can define a new variable $u = 1/r$. Doing some calculations, we will get a differential equation of $u(\phi)$:

$$\ddot{u}(\phi) = -u(\phi) - \frac{\mu}{\ell^2 u(\phi)^2} F \tag{2.171}$$

This can be referred to an orbit's equation. Now, let us define $\gamma = Gm_1m_2 = F/u^2$ in gravitational case, we will find that

$$\ddot{u}(\phi) = -u(\phi) - \frac{\mu\gamma}{\ell^2} \tag{2.172}$$

The general solution can be written as

$$u(\phi) = \frac{\mu\gamma}{\ell^2} + A\cos\phi = \frac{\mu\gamma}{\ell^2}(1 + e\cos\phi) \tag{2.173}$$

Here, A is a constant,

$$e = \frac{A\ell^2}{\mu\gamma}$$

and

$$c = \frac{\ell^2}{\mu\gamma}$$

are the parameters we just introduced. Recall the fact that $r = 1/u$, we will get the equation of Kepler's orbits:

$$\boxed{r(\phi) = \frac{c}{1 + e\cos\phi}} \tag{2.174}$$

It is easy to show that at $\phi = 0$

$$r_p = \frac{c}{1 + e} \tag{2.175}$$

is the **perihelion**, while $\phi = \pi$

$$r_a = \frac{c}{1-e} \tag{2.176}$$

is the **aphelion**. If we now assign

$$a = \frac{c}{1-e^2} \quad \text{and} \quad b = \frac{c}{\sqrt{1-e^2}}$$

, and also, $d = ae$, then you can derive that the orbital equation is in the shape of an ellipse:

$$\frac{(x+d)^2}{a} + \frac{y^2}{b} = 1 \tag{2.177}$$

These reiterate the statements of the Kepler's laws, and for sure, you can derive the rest of them by these expressions. For instance, the third law will be

$$\boxed{T^2 = \frac{4\pi^2}{GM}a^3} \tag{2.178}$$

Note the circular radius R in Eq (2.84) has been replaced by the semimajor axis a. Besides, the ratio of the major axis to the minor axis is given:

$$\frac{b}{a} = \sqrt{1-e^2} \tag{2.179}$$

From here, you might be aware that a is the semimajor axis whereas b is the semiminor axis. Moreover, e represents the **eccentricity**.

Ultimately, we can relate the eccentricity to the effective potential energy we had. At the closest distance r_p (perihelion), the orbiting object (a planet or a comet) has the energy that is equal to the effective potential:

$$E = U_{\text{eff}}(r_p) = -\frac{\gamma}{r_p} + \frac{\ell^2}{2\mu r_p^2} \tag{2.180}$$

Again,

$$r_p = \frac{c}{1+e} = \frac{\ell^2}{\mu\gamma(1+e)}$$

and you can remind yourself that

$$c = \frac{\ell^2}{\mu G m_1 m_2}$$

Substitute the r_p's expression into Eq (2.180), we will obtain the energy's equation based on the eccentricity e:

$$\boxed{E = \frac{\mu\gamma^2}{2\ell^2}(e^2-1)} \tag{2.181}$$

Once the eccentricity $e \geq 0$ and $E \geq 0$, the orbit would be unbounded. The possible shapes and the energies of Kepler orbits are shown as the following:

Table 2: The possible results of the Kepler orbits.

Eccentricity	Energy	Shape	Bounds
$e = 0$	$E < 0$	circle	bounded
$0 < e < 1$	$E < 0$	ellipse	bounded
$e = 1$	$E = 0$	parabola	unbounded
$e > 1$	$E > 0$	hyperbola	unbounded

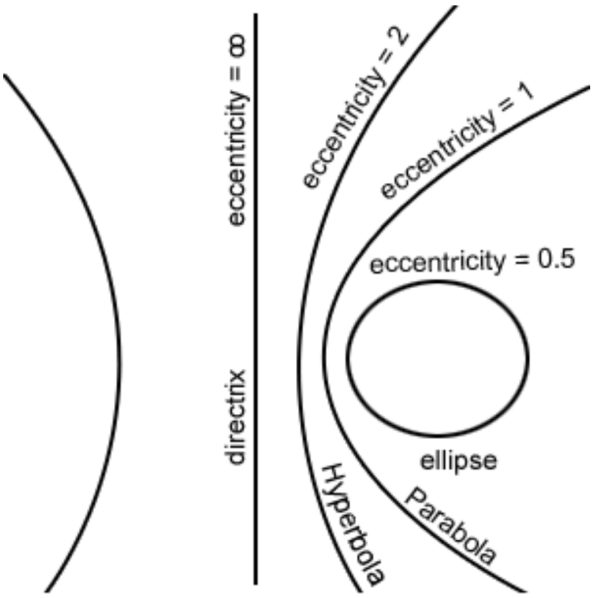

Figure 32: The eccentricities and the shapes of the Kepler orbits. *Source*: https://www.pinterest.com/pin/258182991122441256/

2.13 Non-inertial mechanics

In the inertial frame, the frame of reference does not have acceleration or rotation. In the last topic about classical mechanics, I shall discuss some ideas about the non-inertial frame. First, let us consider a frame that includes the acceleration (but without rotation). One can write down the Newton's second law by

$$m\ddot{\mathbf{r}} = \mathbf{F} + \mathbf{F}' = \mathbf{F} - m\mathbf{A} \tag{2.182}$$

, where $\mathbf{F}'$ represents the **inertial force** and $\mathbf{A}$ is the acceleration vector in the frame of reference. For example, if we consider a simple pendulum in a train, then we might have to consider an acceleration in an opposite direction of where the train moves.

Another interesting example is the tides. Taking a look at Figure 33, we assume the mass of the Moon is M and there is a mass on the Earth's surface is m (you can imagine it as a segment of the ocean on Earth, where it is shown as a square in Figure 33). You can then determine the inertial force

$$\mathbf{F}' = m\mathbf{A} = -m\frac{GM}{r^2}\hat{\mathbf{r}}$$

According to our new second law (2.182), the equation is

$$m\ddot{\mathbf{r}} = \mathbf{F} - m\mathbf{A} = \left(m\mathbf{g} - \frac{GMm}{s^2}\hat{\mathbf{s}} + \mathbf{F}_{\text{other}}\right) + m\frac{GM}{r^2}\hat{\mathbf{r}}$$

, where $\mathbf{F}_{\text{other}}$ is the non-gravitational force such as the buoyant force of the sea water, etc. Now, we can define the **tidal force**:

$$\mathbf{F}_{\text{tide}} = -GMm\left(\frac{\hat{\mathbf{s}}}{s^2} - \frac{\hat{\mathbf{r}}}{r^2}\right) \tag{2.183}$$

That is to say, the non-inertial Newton's second law of the Earth-Moon system can be written as

$$m\ddot{\mathbf{r}} = m\mathbf{g} + \mathbf{F}_{\text{tide}} + \mathbf{F}_{\text{other}} \tag{2.184}$$

Besides, the non-inertial acceleration also reflects the fact that the opposite side of the Earth has the corresponding tidal bulge, as shown in Figure 34.

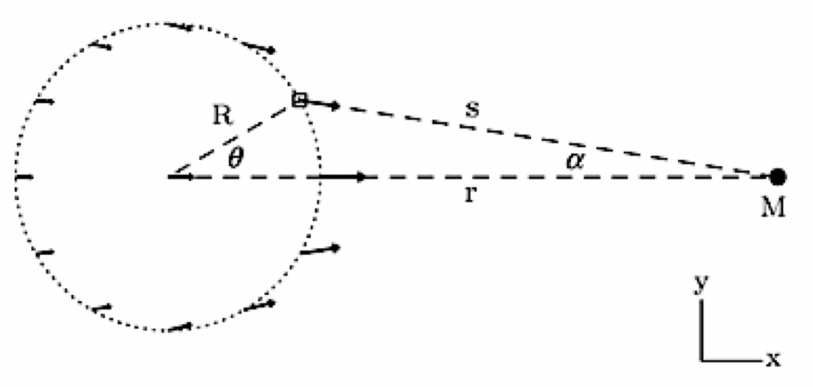

Figure 33: The force diagram of the tides. Consider our Earth-Moon system, the Earth is on the left side and the Moon is on the right side. *Source*: Keeton C. (2014) Tidal Forces. In: Principles of Astrophysics. Undergraduate Lecture Notes in Physics. Springer, New York, NY.

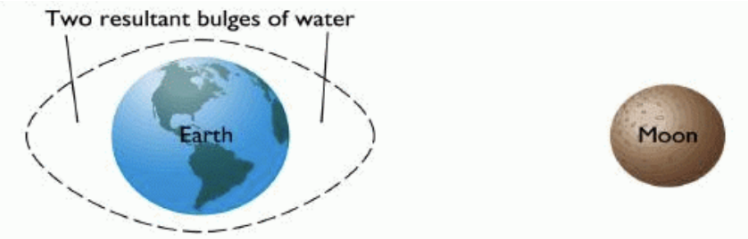

Figure 34: The gravitational attraction arises the tides on the Earth, and the corresponding non-inertial force causes the bulge on the other side of the Earth. *Source*: https://www.lockhaven.edu/ dsimanek/scenario/tides.htm

Now, we start thinking the case of rotation. We define S_0 as the inertial frame and S as the non-inertial frame (as shown in Figure 35). You can think the non-inertial frame is the one who stay on a merry-go-round and the inertial frame is where an observer on the ground. Given an arbitrary vector $\mathbf{v}$, which might be a velocity or a position, one can have a useful relation between S_0 and S:

$$\left(\frac{d\mathbf{v}}{dt}\right)_{S_0} = \left(\frac{d\mathbf{v}}{dt}\right)_{S} + \boldsymbol{\omega} \times \mathbf{v} \tag{2.185}$$

This relation not only consider the time derivative, but also make a relation between the ordinary frame to the non-inertial frame.

I shall go through some derivations about the non-inertial Newton's second law. First, in the ordinary frame,

$$\mathbf{F} = m\left(\frac{d^2\mathbf{r}}{dt^2}\right)_{S_0}$$

According to Eq (2.185), replaced an arbitrary vector $\mathbf{v}$ by a position vector $\mathbf{r}$, the acceleration term becomes

$$\begin{aligned}
\left(\frac{d^2\mathbf{r}}{dt^2}\right)_{S_0} &= \left(\frac{d}{dt}\right)_{S_0}\left(\frac{d\mathbf{r}}{dt}\right)_{S_0} \\
&= \left(\frac{d}{dt}\right)_{S_0}\left[\left(\frac{d\mathbf{r}}{dt}\right)_{S} + \boldsymbol{\omega} \times \mathbf{r}\right] \\
&= \left(\frac{d}{dt}\right)_{S}\left[\left(\frac{d\mathbf{r}}{dt}\right)_{S} + \boldsymbol{\omega} \times \mathbf{r}\right] + \boldsymbol{\omega} \times \left[\left(\frac{d\mathbf{r}}{dt}\right)_{S} + \boldsymbol{\omega} \times \mathbf{r}\right]
\end{aligned}$$

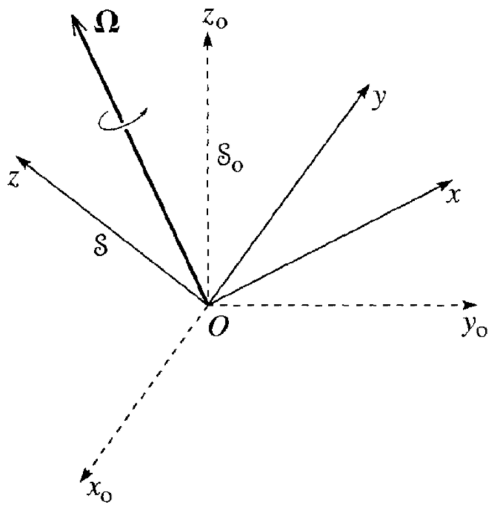

Figure 35: The original frame S_0 and the non-inertial frame S with angular velocity. *Source*: https://mhjensen.github.io/Physics321/doc/pub/frames/html/frames.html

Then, we can take a notation in the rotating frame like

$$\dot{\mathbf{r}} = \left(\frac{d\mathbf{r}}{dt}\right)_S$$

In this way, you can make your expression more concise:

$$\left(\frac{d^2\mathbf{r}}{dt^2}\right)_{S_0} = \ddot{\mathbf{r}} + 2\boldsymbol{\omega} \times \dot{\mathbf{r}} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{2.186}$$

This is the acceleration term of the rotating frame from our point of view (so the subscript is S_0). Thus, the Newton's second law in the non-inertial frame will be:

$$m\ddot{\mathbf{r}} = \mathbf{F} + 2m\boldsymbol{\omega} \times \dot{\mathbf{r}} + m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{2.187}$$

Sometimes, in the case of non-uniformly rotating frame, we have an additional term of the tangential acceleration:

$$\mathbf{a} = -\frac{d\boldsymbol{\omega}}{dt} \times \mathbf{r}$$

In addition, we can insert the inertial force from Eq (2.182). With adding two more terms, Eq (2.187) will be:

$$\boxed{m\ddot{\mathbf{r}} = \mathbf{F} + m\mathbf{A} + m\dot{\boldsymbol{\omega}} \times \mathbf{r} + 2m\boldsymbol{\omega} \times \dot{\mathbf{r}} + m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})} \tag{2.188}$$

, and we can then identify different terms therein:

$$m\ddot{\mathbf{r}} = \mathbf{F} + \underbrace{m\mathbf{A}}_{\text{inertial force}} + \underbrace{m\dot{\boldsymbol{\omega}} \times \mathbf{r}}_{\text{Euler force}} + \underbrace{2m\boldsymbol{\omega} \times \dot{\mathbf{r}}}_{\text{Coriolis force}} + \underbrace{m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})}_{\text{centrifugal force}}$$

Here, let me address a brief introduction for each of them:

- **Inertial force**: The force appear in the acceleration frame,

$$\mathbf{F}' = -m\mathbf{A} \tag{2.189}$$

- **Euler force**: The force appear in the non-uniformly rotating frame,

$$\mathbf{F}_{\text{Euler}} = -m\dot{\boldsymbol{\omega}} \times \mathbf{r} \tag{2.190}$$

- **Coriolis force**: A force that always perpendicular to the velocity of a moving object. This is also a force that affects the direction of a straight path in a rotating frame (e.g. the winds on the Earth).

$$\mathbf{F}_{\text{Coriolis}} = -m\dot{\boldsymbol{\omega}} \times \mathbf{r} \tag{2.191}$$

- **centrifugal force**: A force in the opposite direction to the centripetal force.

$$\mathbf{F}_{\text{cf}} = -m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \tag{2.192}$$

Note that all of these are the **fictitious force**. Only one exception, that is, the $\mathbf{F}$ term in Eq (2.188) is an actual force. For the list above, the first one I called inertial force is used to describe the accelerating frame. But sometimes you might also see the authors use "inertial force" to refer to all kinds of fictitious forces. As for the Euler force, the Coriolis force, and the centrifugal force are basically the effects in the rotating frame. Again, considering the accelerating frame and the rotating frame together, the non-inertial force has the following expression:

$$\boxed{m\ddot{\mathbf{r}} = \mathbf{F} + \mathbf{F}' + \mathbf{F}_{\text{Euler}} + \mathbf{F}_{\text{Coriolis}} + \mathbf{F}_{\text{cf}}} \tag{2.193}$$

3 Thermal Physics

3.1 Ideal gas law

Thermodynamics is a subject to study the macroscopic phenomena of heat and energy. If the case is in microscopic, then it's about the branch of **statistical mechanics**. In the first topic of thermodynamics, let us first discuss the gases.

An ideal gas should obey the following identity:

$$\boxed{PV = nRT} \tag{3.1}$$

This is the well-known **ideal gas law**. Here, P is the pressure, V is the volume of gas, T is the temperature. In addition to these three quantities, given N as the number of molecules,

$$n = \frac{N}{N_A} \tag{3.2}$$

, where $N_A = 6.02 \times 10^{23}$ mol^{-1} is the Avogadro's number. As for R is the gas constant,

$$R = 8.31 \text{ J/mol·K}$$

Armed with this, we can derive a more fundamental quantity, the **Boltzmann constant**,

$$k = k_B = \frac{R}{N_A} = 1.38 \times 10^{-23} \text{ J/K}$$

Using the relation (3.2), we can find another expression of the ideal gas law:

$$\boxed{PV = NkT} \tag{3.3}$$

Practically, this equation is more useful especially when discussing the microscopic effects.

3.2 Kinetic theory of gases

From the microscopic view, the energy of a gas is carried by the motion of a large number of particles. It seems difficult for us to deal with a many-particle system. Nevertheless, we can derive the macroscopic quantities by some easy methods. This is about the **kinetic theory of gases**, and here we assume the gases are ideal, which has to obey the ideal gas law. In the simplest case, we assume a 1-mole gas in a box with an edge of L, and the motion is dominated by the monatomic particles (such as He, Ne) with the speed v, as shown in Figure 36.

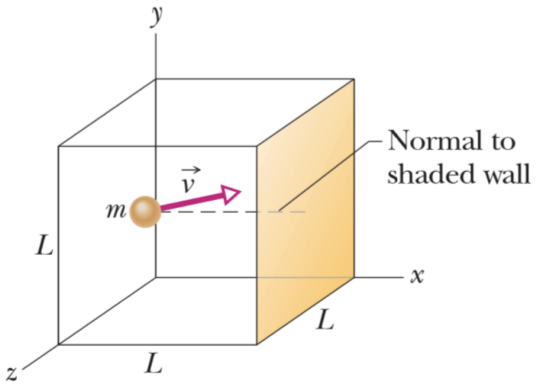

Figure 36: An illustration of a particle of an ideal gas in a box. We assume the gas particle moves and collides the wall elastically. *Source*: David Halliday (2015), Principles of Physics, 10th Ed, Wiley published.

Let us just consider the case of x-direction. Given a particle with a mass of m moving toward the wall with a distance L, it will be collided back with the same speed but in an opposite direction (remember this is an ideal gas). Thus, the change of momentum can be written as

$$\Delta p_x = -mv_x - mv_x = -2mv_x$$

As for the force, which is the variation of the momentum with respect to a given time $\Delta t = 2L/v_x$:

$$F_x = \frac{\Delta p_x}{\Delta t} = \frac{-2mv_x}{\frac{2L}{v_x}} = \frac{mv_x^2}{L}$$

Soon we can define the pressure of such a gas. Note the pressure is a macroscopic quantity, which means we have to consider the summation of all particles' effects:

$$P = \frac{F}{A} = \frac{F_x}{L^2} = \frac{mv_{x1}^2 + mv_{x2}^2 + ... + mv_{xN}^2}{L^3} = \sum_{i=1}^{N} \frac{mv_{xi}^2}{L^3} \tag{3.4}$$

Now, since $v^2 = v_x^2 + v_y^2 + v_z^2$,

$$v_x^2 = \frac{v^2}{3}$$

Also, due to $N = nN_A$, and meanwhile we define $L^3 = V$ and $N_A m = M$ (that is, the total mass of a gas), our expression of pressure becomes

$$P = \frac{nN_A m\bar{v_x^2}}{L^3} = \frac{nMv_{\text{rms}^2}}{3V} \tag{3.5}$$

Here, as you might notice, we have introduced a new quantity called the **root-mean-square (rms) speed**:

$$v_{\text{rms}} = \sqrt{\frac{3PV}{nM}} = \sqrt{\frac{3RT}{M}} \tag{3.6}$$

Recall that $k = R/N_A$, thus we have

$$\boxed{v_{\text{rms}} = \sqrt{\frac{3kT}{m}}} \tag{3.7}$$

Then, we can determine the average kinetic energy:

$$\boxed{\bar{K} = \frac{1}{2}mv_{\text{rms}}^2 = \frac{3}{2}kT} \tag{3.8}$$

This tells us that the kinetic energy of an ideal gas is dependent on the temperature only. From the previous chapter, you might have seen a lot of types of energy are in the forms of

$$\frac{1}{2}mv^2, \quad \frac{1}{2}kx^2, \quad \frac{1}{2}I\omega^2$$

According to thermal physics, the **equipartition theorem** implies that the average energy at temperature T is $kT/2$. To generalize, if a system has N molecules with f degrees of freedom, then the total thermal energy will be

$$\boxed{U = f \cdot N \cdot \frac{1}{2}kT} \tag{3.9}$$

For example, in our simplest case the monatomic molecules have translational motion of three directions only, so the degrees of freedom is $f = 3$. This is corresponding to Eq (3.8).

Sometimes it would be helpful to calculate the average distance that a moving particle collides with another particle. This is the **mean free path**. More accurate, it indicates a distance that a moving particle changes its energy or direction. In general, the mean free path can be expressed as

$$l = \frac{1}{n\sigma} \tag{3.10}$$

, where n is the number of particles in a given volume and σ is the effective cross-sectional area for collision. Alternatively, the mean free path can be written as

$$l = \frac{V}{\sqrt{2}\pi d^2 N} = \frac{kT}{\sqrt{2}\pi d^2 P} \tag{3.11}$$

This is based on the ideal gas law:

$$n = \frac{N}{V} = \frac{P}{kT} \tag{3.12}$$

3.3 Energy, heat, and work

First, **temperature** is the most essential quantity when you start studying thermal physics. Simply speaking, it is what you measure from a thermometer. Technically, it is a measure when an object exchange energy spontaneously. If two objects contacted each other for a while, both will come to the same temperature, and such a situation is called **thermal equilibrium**. In thermal physics and several areas of physics, we usually discuss temperature T in the Kelvin scale (K). But in daily life, you may use Celsius scale

$$T_C = T - 273.15^\circ$$

and Fahrenheit scale

$$T_F = \frac{9}{5}T_C + 32^\circ$$

Note at -40 degrees the Celsius and Fahrenheit scales are the same, $-40°\text{C} = -40°\text{F}$. In addition, the thermal expansion shows that the size of an object is related to its temperature. We typically have the linear expansion

$$\Delta L = \alpha L_0 \Delta T$$

and the volume expansion

$$\Delta V = \gamma V_0 \Delta T$$

The coefficients of expansion for both are α and γ, respectively, and $\gamma = 3\alpha$.

As for **energy**, this is the most fundamental quantity in all kinds of dynamics. What you can do is to have in mind that the total amount of energy in the universe never changes. When two objects have different temperature, there will be spontaneous flows of energy between them. Such a flow is called the **heat**, denoted by Q. The amount of heat to raise the temperature is the **heat capacity**, which is

$$C = \frac{Q}{\Delta T} \tag{3.13}$$

Given a unit mass of m, the **specific heat capacity** is

$$c = \frac{C}{\Delta T} \tag{3.14}$$

Sometimes you will need to calculate the absorbed heat, then it can be expresseed as

$$Q = mc\Delta T \tag{3.15}$$

Also, you might have to solve the final temperature after two objects reach the thermal equilibrium, and it can be expressed as:

$$T_f = \frac{m_1 c_1 T_1 + m_2 c_2 T_2}{m_1 c_1 + m_2 c_2}$$

During the phase of transformation, such as melting or vaporizing, the formula becomes

$$Q = mL \tag{3.16}$$

, where L is the **latent heat**, and for water there are two different values – the heat of vaporization

$$L_v = 539\,\text{cal/g} = 2266\,\text{kJ/kg}$$

and the heat of fusion

$$L_f = 79.5\,\text{cal/g} = 333\,\text{kJ/kg}$$

Remember that 1 cal $\approx$ 4.18 Joules. Let me quickly introduce three processes of heat transfer:

- **Conduction**: Transfer caused by direct contact.
- **Convection**: Transfer caused by the bulk motion of gas or liquid.
- **Radiation**: Transfer caused by the emission of electromagnetic waves.

Another similar but different concept is **work**, which is the energy transfer within a system. Unlike the heat, which is a spontaneous flow, work typically requires something drive the transfer of energy. Consider a thermal system of a bucket of gas with a piston, we can set $dW = \mathbf{F} \cdot d\mathbf{x} = PA\ dx = P\ dV$, and then the work can be defined as

$$\boxed{W = -\int_{V_i}^{V_f} P(V)\ dV} \tag{3.17}$$

In this case, since the volume of gas will decrease when we push the piston, the work done by gas should be negative. For constant volume

$$W = 0 \tag{3.18}$$

, whereas for constant pressure

$$W = P\Delta V \tag{3.19}$$

3.4 The first law of thermodynamics

In the previous section, we see that both work and heat are about the energy in transit. To calculate the total energy in transit (e.g. how much energy enters a system), we can determine the following identity:

$$\boxed{\Delta U = Q + W} \tag{3.20}$$

Here U represents the internal energy of a system. This is the **first law of thermodynamics**. Here W is the work done *on* a system, so if we assign W as the work done *by* a system, the expression becomes $\Delta U = Q - W$.

Now, combined the first law with the heat capacities, we also have the following two types: first, assume the volume doesn't change so that $W = 0$, we have

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V \tag{3.21}$$

Then, if the pressure is fixed, the first law demands $\Delta Q = \Delta U - (-P\Delta V)$, so

$$C_P = \left(\frac{\partial U}{\partial T}\right)_P + P\left(\frac{\partial V}{\partial T}\right)_P \tag{3.22}$$

Recall the energy $U = NfkT/2$, the heat capacity at constant volume can be written as

$$C_V = \frac{Nfk}{2} \tag{3.23}$$

For instance, in the case of a monatomic gas, $f = 3$ so that

$$C_V = \frac{3}{2}Nk = \frac{3}{2}nR$$

For a 1-mole gas $C_V = 12.5$ J/K. As for the capacity at constant pressure, you can work out the relationship between both ones:

$$C_P = C_V + Nk = C_V + nR \tag{3.24}$$

In this case, a monatomic gas has

$$C_P = \frac{5}{2}Nk = \frac{5}{2}nR$$

Let us discuss the processes of compression work. If a system's temperature doesn't change during a process, that is, T is constant, then according to $P = NkT/V$, the work done can be written as

$$W = -\int_{V_i}^{V_f} P(V)\, dV = -NkT \int_{V_i}^{V_f} \frac{1}{V}\, dV = NkT \ln \frac{V_i}{V_f} \quad \text{(isothermal)} \tag{3.25}$$

This is an **isothermal process**. Recall that in Eq (3.9) it shows that U is proportional to T, so the internal energy ΔU turns out to be zero, $\Delta U = 0$. Then, the heat of a system given by the first law (3.20) will be

$$Q = \Delta U - W = 0 - NkT \ln \frac{V_i}{V_f} = NkT \ln \frac{V_f}{V_i} \quad \text{(isothermal)} \tag{3.26}$$

Yet, if there is no heat flows through the system, $Q = 0$, then $\Delta U = W$. This implies

$$\frac{f}{2} Nk\, dT = -P\, dV$$

Rearrange this can lead to a differential equation:

$$\frac{f}{2} \frac{dT}{T} = -\frac{dV}{V}$$

Taking integrals on both sides, we have

$$\frac{f}{2} \ln \frac{T_f}{T_i} = -\ln \frac{V_f}{V_i}$$

Consequently,

$$V_i T_i^{\frac{f}{2}} = V_f T_f^{\frac{f}{2}}$$

To express in a more compact way, we let $\gamma = \frac{f+2}{f}$, and we will get

$$PV^\gamma = \text{const.} \tag{3.27}$$

Such a process without heat flows is called the **adiabatic process**, and the adiabatic exponent can also be written as

$$\gamma = \frac{C_P}{C_V}$$

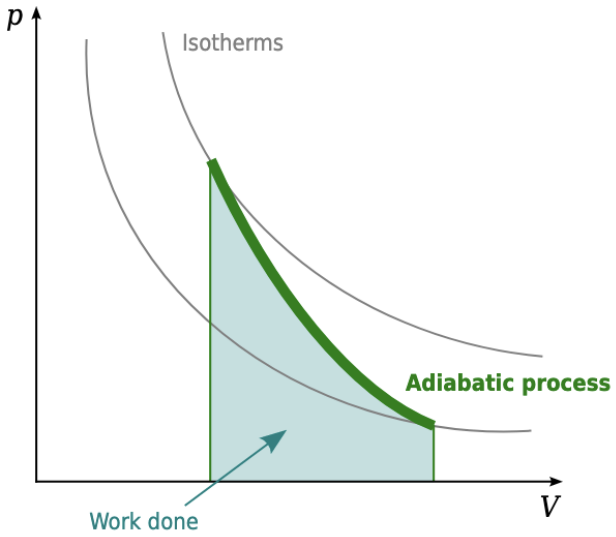

Figure 37: The PV diagram of adiabatic process, where you can see the adiabat starts from a lower temperature and ends on a higher temperature during the adiabatic compression process. *Source*: Wikipedia.

In addition to the isothermal and the adiabatic process, there are still some special cases of the first law. For instance, for an **isochoric process**, that is, the volume V is constant, $W = 0$. Therefore,

$$\Delta U = Q \quad \text{(isochoric)} \tag{3.28}$$

If a system is in a closed cycle, the total amount of exchanged heat and work will be zero. Thus, because of $\Delta U = 0$, we have

$$Q = -W \quad \text{(cyclical)} \tag{3.29}$$

This is named the **cyclical process**.

Last, if a process is not only adiabatic, but also has no work done within a system, the process has $Q = W = 0$ so that

$$\Delta U = 0 \quad \text{(free expansion)} \tag{3.30}$$

This process is called the **free expansion**. In term of what we got for the adiabatic expansion, we have

$$P_i V_i = P_f V_f \tag{3.31}$$

All of these processes can be collected as the following:

Table 3: The special cases for the first law of thermodynamics, $\Delta U = Q + W$.

Process	Variable	Condition	Consequence
Isothermal	$T = \text{const.}$	$\Delta U = 0$	$Q = NkT \ln V_f/V_i$
Isobaric	$P = \text{const.}$	$W = -P\Delta V$	$\Delta U = Q - P\Delta V$
Isochoric	$V = \text{const.}$	$W = 0$	$\Delta U = Q$
Adiabatic	$Q = 0$	$Q = 0$	$\Delta U = W$
Cyclical	$\Delta U = 0$	$\Delta U = 0$	$Q = -W$
Free expansion	$Q = W = 0$	$Q = W = 0$	$\Delta U = 0$

3.5 Entropy and the second law

In this section we are going to discuss the most interesting topic in thermodynamics. You can also view this section as an introduction about statistical mechanics. As a start, let us take a first glance at the combinatorics. Let's toss a coin. Doing several times, you will totally get these results (H means head and T means tail):

HH
HT
TH
TT

Simply speaking, you have 1 state for both heads (HH), 1 state for both tails (TT), and 2 states for one head with one tail (HT and TH). In this case, we have totally 4 states, which are called microstates. But when you are told that there are one state of HH, two states of either H or T, and one state of TT – then these are called macrostates. The number of microstates based on the given macrostates is called the **multiplicity**. In this case, we have the following combinations:

1H 1H, $\Omega = 1$
1H 1T, $\Omega = 2$
1T 1T, $\Omega = 1$

, where the total multiplicity is $1 + 2 + 1 = 4$. If we have N coins, the multiplicity of n heads' macrostates is defined as

$$\Omega(N, n) = \binom{N}{n} = \frac{N!}{n!(N-n)!} \tag{3.32}$$

In our example of coins, if we want to find the multiplicity of the macrostates with 1 head, then you will find that the value $\Omega(N, n) = 2$, which is corresponding to 1H 1T and 1T 1H.

Now, we consider a the oscillation of atoms in a 3-dimensional solid. All atoms have the same frequency and cannot interact with each other. Such a model is the **Einstein solid**. According to quantum physics, each atom might have certain energy units. We can consider the simplest case: a solid made up of 2 oscillators with the maximum energy unit of 2. This means that you can have either 0, 1, or 2 for the number of energy units. You can list the results as the following:

$$\begin{array}{cc} 0 & 0 \end{array} \quad \longrightarrow \quad \Omega(0) = 1$$

$$\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \quad \longrightarrow \quad \Omega(1) = 2$$

$$\begin{array}{cc} 0 & 2 \\ 1 & 1 \\ 2 & 0 \end{array} \quad \longrightarrow \quad \Omega(2) = 3$$

In general, given N oscillators and q energy units, we have the following expression:

$$\Omega(N, q) = \binom{N+q-1}{q} = \frac{(N+q-1)!}{q!(N-1)!} \tag{3.33}$$

Therefore, in our simplest case, the multiplicity of $N = 2$ oscillators with $q = 2$ energy units has $\Omega(2, 2) = 3$, as we listed above.

Figure 38: The Einstein model of a solid. In this diagram we have 10 oscillators. Given q dots and $N - 1$ separation lines, we are interested in how many choices, that is, $\Omega(N, q)$. *Source*: Wikipedia.

You might feel a little bit boring so far. What if we consider a system of a large number of particles, like 10^{23}? Without doing this tricky case, let's just assume a system of two Einstein solids A and B, and both can interact with each other weakly (i.e. flows from B to A), assuming $N_A = 200$, $N_B = 100$, and $q = 100$. Then you find that the values are already much larger than your expectations because of many ways of combinations, also, the multiplicity runs drastically to the maximum in the middle, as shown in Table 4 and Figure 39. Basically, when the two systems interact up to the maximum of total multiplicity Ω_{total} (in this case it happens when $q_A = 67$ and $q_B = 33$), the system is in thermal equilibrium. The energy flows spontaneously from B to A. When it reaches the maximum, it means that you will observe that both A and B have almost the same amount of states. In other words, the flow stops when our system is near its most likely macrostate. Looking at the rightest column, you should see that the probability to find the most likely states ($\sim 10^{95}$) is much larger than those other states (10^{58}, 10^{60}, or even 10^{81}, etc.) For sure, when the system comes to stop, there are some small fluctuations from A to B around $q_A = 67$, while all of these are probabilistic. Overall, the system is *irreversible*, which is an important implication of what we are going to discuss.

Table 4: The results of macrostates and multiplicities of a two-Einstein solid's system, with $N_A = 200$, $N_B = 100$, and $q = 100$.

q_A	Ω_A	q_B	Ω_B	Ω_{total}
0	1	100	4.53×10^{58}	4.53×10^{58}
1	200	99	2.28×10^{58}	4.55×10^{60}
2	20100	98	1.14×10^{58}	2.29×10^{62}
⋮	⋮	⋮	⋮	⋮
66	2.25×10^{63}	34	5.40×10^{31}	1.21×10^{95}
67	8.91×10^{63}	33	1.28×10^{31}	1.23×10^{95}
68	3.50×10^{64}	32	3.45×10^{30}	1.21×10^{95}
⋮	⋮	⋮	⋮	⋮
99	9.27×10^{80}	1	100	9.27×10^{82}
100	2.77×10^{81}	0	1	2.77×10^{81}

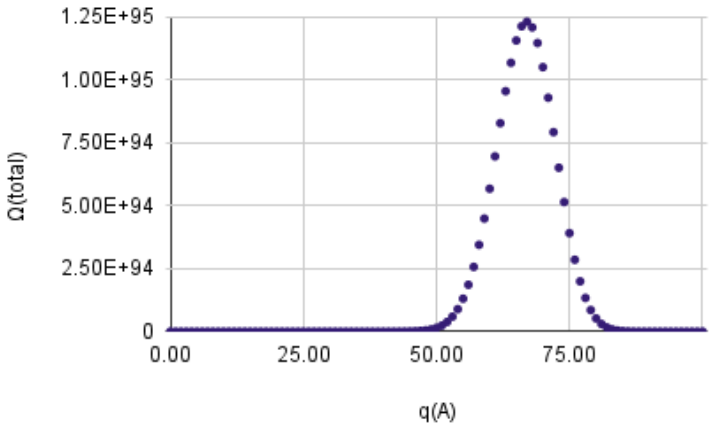

Figure 39: The plot of the multiplicities when the energy flows spontaneously from B to A. Note that the system comes to equilibrium and stops at the maximum, that is, the most likely macrostates.

Nevertheless, you should be in mind that this example only considers 101 macrostates. In the case of real world, more often than not, there are at least 10^{20} oscillators within a solid. It would be a draconian task to calculate so many possible states. Thus, we should use the **Stirling's formula**,

$$N! \approx N^N e^{-N} \sqrt{2\pi N}$$

, to deal with the case when $N \gg 1$. It can be written as a more useful way:

$$\ln N! \approx N \ln N - N \tag{3.34}$$

Thanks to this formula, we can eventually find an approximation when $q \gg N$:

$$\ln \Omega \approx N \ln \frac{q}{N} + N + \frac{N^2}{q}$$

, where the last term can be omitted because $q \gg N$. Then exponentiate both sides we will have

$$\Omega(N, q) \approx \left(\frac{eq}{N}\right)^N \quad (q \gg N) \tag{3.35}$$

If $q \ll N$, it becomes

$$\Omega(N, q) \approx \left(\frac{eN}{q}\right)^q \quad (q \ll N)$$

In this way, we can define a quantity that can be related to thermodynamics:

$$\boxed{S = k \ln \Omega} \tag{3.36}$$

This quantity is called the **entropy**. In fact, it represents how many ways to arrange states in a system, with the unit of J/K. Since k is a constant, we can simply associate the entropy with the multiplicity. Recall the example we have discussed just now: the particles in a given system will spontaneously rearrange until it reaches the maximum multiplicity, and the process must be irreversible (except for the tiny fluctuations). Note the system we are talking about is *isolated*, that is, we don't expect it will exchange energies with outside environment. To sum up, the implication can be simply stated as: for an isolated system, *the entropy must tend to increase*,

$$\Delta S \geq 0 \tag{3.37}$$

– this is the **second law of thermodynamics**.

A simple way is to think of the entropy as a kind of *disorder*. For example, you might notice that everything – including a flower, a building, you, your friends, the Earth, and even the universe – tends to perish. You can say that humans might be able to intervene these, just like you will not keep your bedroom messy but clean up the stuffs. However, when you try to do this, the entropy has been transferred to the energies generated in your bodies. Thus, it would be impossible to decrease the entropy. Some people have assumed that the universe is an isolated system, and it follows that the universe will die when it reaches the maximum entropy, perhaps takes at least 10^{1000} years. This theory is called the "heat death" of the universe. The second law of thermodynamics also tells us the direction of time, which is irreversible.

In statistical mechanics, the entropy of a monatomic ideal gas can be expressed as

$$S = kN \left[\ln\left(\frac{V}{N}\left(\frac{4\pi mU}{3Nh^2}\right)^{\frac{3}{2}}\right) + \frac{5}{2}\right] \tag{3.38}$$

This is the Sackur-Tetrode equation, where h is the Planck constant – we will discuss in the later chapters. For instance, in the case of isothermal expansion, it becomes

$$\Delta S = kN \ln \frac{V_f}{V_i}$$

In thermodynamics, another useful formula of the change of the entropy is

$$\boxed{\Delta S = \int \frac{dQ}{T}} \tag{3.39}$$

For the isothermal case,

$$\Delta S = \frac{Q}{T} \tag{3.40}$$

3.6 Engines and refrigerators

The **heat engine** is an example of the entropy in reality. A heat engine can generate energy by absorbing heat from the environment and doing work. We can define the efficiency:

$$e \equiv \frac{W}{Q_h} \quad \text{(engine)} \tag{3.41}$$

Here, W is corresponding to the benefit and Q_h is corresponding to the cost, basically. As shown in Figure 40 (left), the difference of temperatures corresponding to different amount of heats. When the heat flows

from higher temperature T_h to lower temperature T_c, the absorbed heat Q_h and the lost heat Q_c would be generated. Then there would be work W done by the engine. Mathematically, it must follow the identity

$$W = Q_h - Q_c$$

Therefore,

$$e = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \quad \text{(engine)} \tag{3.42}$$

Meanwhile, the second law implies that the entropy of a system will increase,

$$S_c \geq S_h$$

Alternatively,

$$\frac{Q_c}{Q_h} \geq \frac{T_c}{T_h}$$

, which means

$$e \leq 1 - \frac{T_c}{T_h} \quad \text{(engine)} \tag{3.43}$$

This suggests that the efficiency of an engine will be higher when the temperature difference is more notable.

If we reverse the direction: let the energy flows from the cold reservoir to the hot reservoir and absorb work, then we should define a similar quantity but with a different name – the coefficient of performance:

$$\text{COP} \equiv \frac{Q_c}{W} \quad \text{(refrigerator)} \tag{3.44}$$

Note the cost now is in the work W whereas the benefit now is the Q_c. As before, we know that $W = Q_h - Q_c$, which gives

$$\text{COP} = \frac{Q_c}{Q_h - Q_c} \quad \text{(refrigerator)} \tag{3.45}$$

The second law suggests that

$$S_h \geq S_c$$

This is equivalent to

$$\frac{Q_h}{Q_c} \geq \frac{T_h}{T_c}$$

In turn, we can conclude that

$$\text{COP} \leq \frac{T_c}{T_h - T_c} \quad \text{(refrigerator)} \tag{3.46}$$

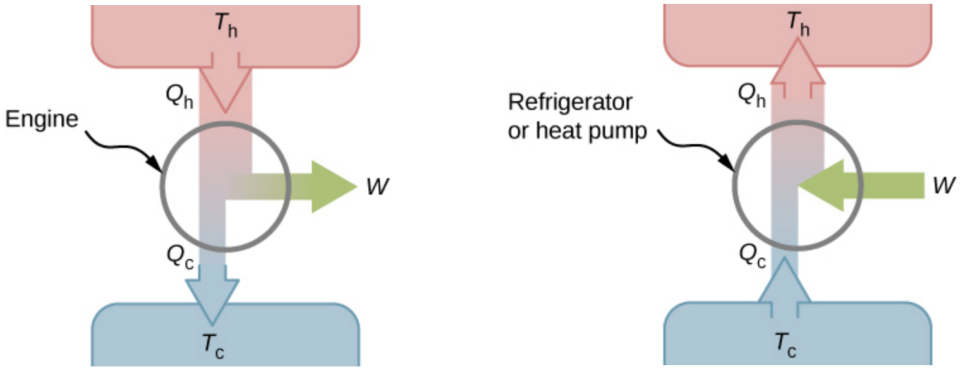

Figure 40: The heat engine (left) and the refrigerator (right). *Source*: https://phys.libretexts.org/

Let me briefly introduce some ideal and real-world examples. In ideal case, the most notable example for a heat engine is the **Carnot cycle**. Such an engine's all processes are reversible, and there is not waste

during energy transfers. As shown in Figure 41 (a), the cycle is consist of two isothermal and two adiabatic processes. The entropy changes has to be $\Delta S = 0$ because of the complete cycle, it implies that

$$\Delta S = \frac{Q_h}{T_h} - \frac{Q_c}{T_c} = 0$$

, where Q is supposed to be a positive value. This gives

$$\frac{Q_h}{T_h} = \frac{Q_c}{T_c} \tag{3.47}$$

Since $W = Q_h - Q_c$, by using the formula of Eq (3.42) and Eq (3.43), the efficiency of a Carnot engine turns out that

$$e = 1 - \frac{Q_c}{Q_h} = 1 - \frac{T_c}{T_h} \quad \text{(Carnot engine)} \tag{3.48}$$

In order to achieve an 100% efficiency, we must require $T_c \to 0$ and $T_h \to \infty$, which is impossible. Thus, the second law also shows that the perfect engine cannot exist.

If we now replace the two adiabats with the constant-volume curves and plot out the P-V diagram, then this becomes a **Stirling engine** (Figure 42). Unlike the Carnot engine, the Stirling engine doesn't have a perfect efficiency because the entropy changes occur in all processes when the reversible heat transfers.

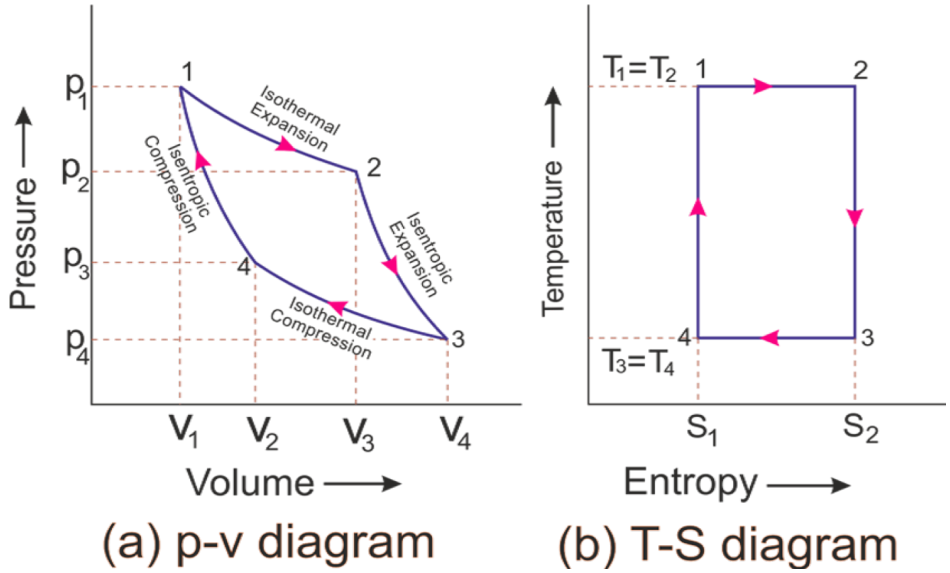

Figure 41: The P-V diagram and the T-S diagram of a Carnot engine. *Source*: https://www.mechanicaltutorial.com/4-stages-of-carnot-cycle-improving-thermal-efficiency.

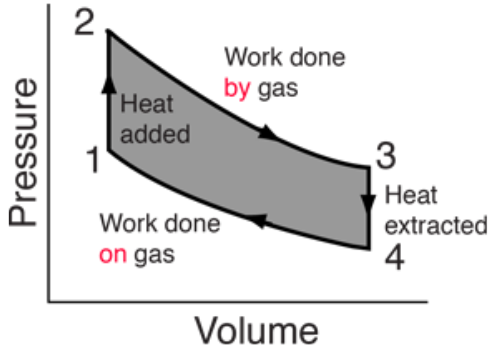

Figure 42: The P-V diagram of a Stirling engine. *Source*: Hyperphysics.com.

There are also some real-world example. For heat engines, one of the most useful engine is the steam engine, which was built based on the **Rankine cycle**. As shown in Figure 43, let me briefly illustrate how it works:

1. From 1 to 2: The liquid water is compressed via a pump, and tends to be in higher pressure.
2. From 2 to 3: Under a constant pressure, the liquid water is heated in the boiler and turns out to be steam.
3. From 3 to 4: The steam activates the turbine, generates work, and enters the adiabatic process, which makes it expands and cools down.
4. From 4 to 1: The mixed water and steam enters the condenser and becomes back to the liquid water.

To calculate the efficiency of a steam engine, we should introduce another important quantity in thermodynamics, the **enthalpy**:

$$\boxed{H = U + PV} \tag{3.49}$$

To understand this quantity, you can imagine that a magician try to create an item in the air. To achieve, the magician must have an energy of the item itself, but also have an additional energy PV to push the atmosphere away. Back to our main topic, the Rankine cycle has the following efficiency:

$$e = 1 - \frac{Q_c}{Q_h} = 1 - \frac{H_4 - H_1}{H_3 - H_2} \tag{3.50}$$

It is noteworthy that the Rankine cycle supports a wide amount of energy in today's world. More examples are shown in Table 5.

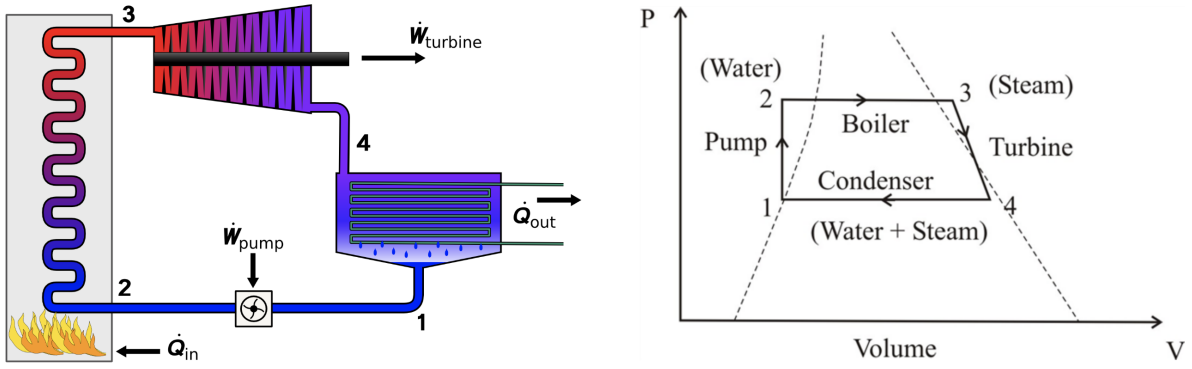

Figure 43: The schematic diagram of a steam engine (left) and the P-V diagram of the corresponding Rankine cycle (right). *Source*: https://physics.stackexchange.com/questions/591279/work-in-rankine-cycle

Table 5: Some examples about thermodynamic cycles.

Cycle	Compression	Heat absorbed	Expansion	Heat expelled	Applications
Carnot cycle	isentropic	isothermal	isentropic	isothermal	Carnot engine
Stirling cycle	isothermal	isochoric	isothermal	isochoric	Stirling engine
Rankine cycle	adiabatic	isobaric	adiabatic	isobaric	steam engines
Otto cycle	isentropic	isochoric	isentropic	isochoric	petrol engines
Diesel cycle	adiabatic	isobaric	adiabatic	isochoric	Diesel engine

As for the refrigerator in real world, as shown in Figure 44, we have the four steps:

1. From 1 to 2: The fluid or a gas is compressed adiabatically, so the temperature and pressure has been raised.

2. From 2 to 3: In the condenser, the heat is released toward the hot reservoir, and the gas becomes liquefied.

3. From 3 to 4: After passing through the throttle, its pressure and temperature becomes lower.

4. From 4 to 1: It absorbs heat from the cold reservoir and turns to be a gas in the evaporator.

Because the absorbed heat is $Q_c = H_1 - H_4$ and the expelled heat is $Q_h = H_2 - H_3$, the coefficient of performance will be

$$\text{COP} = \frac{Q_c}{Q_h - Q_c} = \frac{H_1 - H_4}{H_2 - H_3 - H_1 + H_4}. \tag{3.51}$$

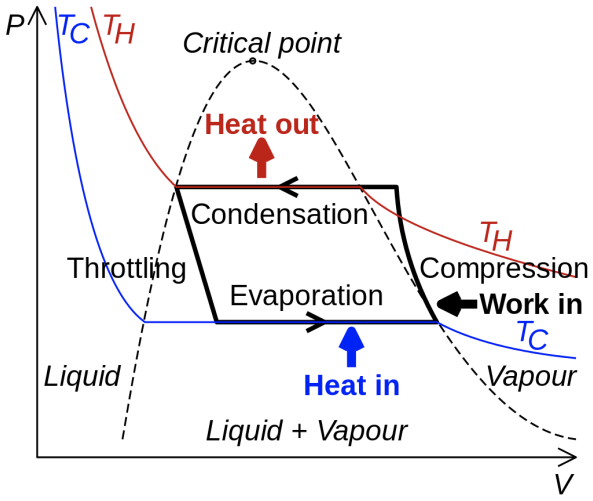

Figure 44: The P-V diagram of a refrigerator. *Source*: Wikipedia.

3.7 Interactions in thermodynamics

As an intermission section, I would like to introduce a simple but very useful identity:

$$\boxed{dU = T\,dS - P\,dV + \mu\,dN} \tag{3.52}$$

This identity tell us some useful things about the interactions in thermodynamics.

In general, we have three different cases. The first type of interaction is related to the thermal equilibrium of two objects. At thermal equilibrium, the governing variable is the energy U, and we have

$$\frac{\partial S_1}{\partial U_1} = \frac{\partial S_2}{\partial U_2}$$

Your knowledge has told you that both objects should have the same temperature. This permits us to define a quantity from the above's equation:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{V,N} \tag{3.53}$$

Note the volume and the number of particles are fixed. In other words, the formal definition of temperature in thermodynamics is:

$$T = \left(\frac{\partial U}{\partial S}\right)_{V,N} \tag{3.54}$$

The second case is the mechanical equilibrium. In this scenario, the energy and volume can exchange, instead. It means that the governing variable is the pressure. Therefore, at mechanical equilibrium,

$$\frac{\partial S_1}{\partial V_1} = \frac{\partial S_2}{\partial V_2}$$

Similar to the thermal equilibrium, we can now define the relation of pressure:

$$P = T\left(\frac{\partial S}{\partial V}\right)_{U,N} \tag{3.55}$$

In mechanical equilibrium, the pressures will be the same.

The other case is called the diffusive equilibrium. Previously we do not allow the exchange of particles. However, in the diffusive case, the exchanged quantity is the number of particles, which implies

$$\frac{\partial S_1}{\partial N_1} = \frac{\partial S_2}{\partial N_2}$$

at equilibrium. We can now determine a new quantity called the **chemical potential**:

$$\mu = -T\left(\frac{\partial S}{\partial N}\right)_{U,V} \tag{3.56}$$

Combined Eqs (3.53), (3.55), and 3.56, we found that

$$dS = \left(\frac{\partial S}{\partial U}\right)_{V,N} dU + \left(\frac{\partial S}{\partial V}\right)_{U,N} dV + \left(\frac{\partial S}{\partial N}\right)_{U,V} dN = \frac{1}{T}dU + \frac{P}{T}dV - \frac{\mu}{T}dN$$

This gives us the thermal identity (3.52):

$$dU = T\,dS - P\,dV + \mu\,dN$$

Hence, this relation contains all information about the interactions in thermodynamics.

3.8 Thermodynamic free energy

Here we comes into the advanced topics that the general physics course will not cover typically. The first topic I would like to elaborate is about the free energy in thermodynamics. The first law of thermodynamics tells us that

$$W = U - Q$$

If we are going to create an item out of nothing, we must provide work, and there is also the heat from environment to the system. Given the environment has a temperature of T, we can then define the **(Helmholtz) free energy**:

$$\boxed{F \equiv U - TS} \tag{3.57}$$

If we also consider an environment with constant P, then we can define the work to create a system out of nothing, which is also the work you should have for recovering when you annihilate it. This is the **Gibbs free energy**:

$$\boxed{G \equiv U - TS + PV = H - TS} \tag{3.58}$$

Recall the enthalpy is $H = U + PV$. The relationships of U, H, F, and G are shown in Figure 45.

The reader should bear in mind of the thermodynamic identity (3.52). Armed with this, we can derive the similar cases for enthalpy, free energy, and Gibbs free energy. For the derivative of enthalpy, we have $dH = dU + P\,dV + V\,dP$ (here we used the chain rule). Therefore, combined with (3.52), we give

$$dH = T\,dS + V\,dP + \mu\,dN \tag{3.59}$$

The Helmoholtz free energy's derivative is $dF = dU - T\,dS - S\,dT$, combined with dU's expression, we also get

$$dF = -S\,dT - P\,dV + \mu\,dN \tag{3.60}$$

Similarly, the Gibbs free energy's derivative is given by $dG = dH - P\,dS - S\,dT$, then thermodynamic identity is

$$dG = -S\,dT + V\,dP + \mu\,dN \tag{3.61}$$

From Eqs (3.60) and (3.61), we can recognize some useful identities as the following:

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,N} \quad P = -\left(\frac{\partial F}{\partial V}\right)_{T,N} \quad \mu = \left(\frac{\partial F}{\partial N}\right)_{T,V}, \tag{3.62}$$

and,

$$S = -\left(\frac{\partial G}{\partial T}\right)_{P,N} \quad V = -\left(\frac{\partial G}{\partial P}\right)_{T,N} \quad \mu = \left(\frac{\partial G}{\partial N}\right)_{T,P} \tag{3.63}$$

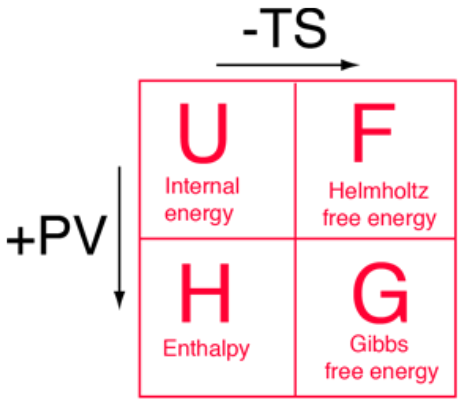

Figure 45: The relationships of four similar quantities. *Source*: Hyperphysics.com.

3.9 Boltzmann statistics

Now, we are going to discuss the so-called statistical mechanics. Here I will launch a short introduction and not go over too many details.

Boltzmann statistics is a classical analysis of the thermodynamic systems that involving in the exchanges of energy only (without the considerations of exchanging particles). As a start, we knew that the Ω is the multiplicity, which corresponds to the microstates. We can say that the probability of finding state 1 and state 2 has the following relations:

$$\frac{P_2}{P_1} = \frac{\Omega_2}{\Omega_1} = \frac{e^{S_2/k}}{e^{S_1/k}} = e^{\frac{S_2 - S_1}{k}}$$

Meanwhile, thermodynamics tells us that

$$dS = \frac{1}{T}(dU + P\,dV - \mu\,dN)$$

Since we only consider the exchange of energy, we can throw the last term $\mu\,dN$ away; as for $P\,dV$, it is also negligible compared to dU. Hence, the identity becomes

$$dS = \frac{1}{T}dU$$

The change of entropy is then given by

$$\Delta S = S_2 - S_1 = \frac{1}{T}(U_2 - U_1) = -\frac{1}{T}(E_2 - E_1)$$

, where E is the atomic energy. Therefore, the probability relation can be written as

$$\frac{P_2}{P_1} = e^{(E_2 - E_1)/kT} = \frac{e^{-E_2/kT}}{e^{-E_1/kT}} \tag{3.64}$$

Here, $e^{-E_i/kT}$ is called the **Boltzmann factor**. Now, it would be convincing to introduce a constant

$$\frac{1}{Z} = \frac{P_i}{e^{-E_i/kT}}$$

In this way, the probability distribution of a given state i has the following form:

$$\boxed{P_i = \frac{1}{Z} e^{-E_i/kT}} \tag{3.65}$$

This is the famous **Boltzmann distribution**. Then, you might be slightly confused about the meaning of Z. The total probability of finding the atom in a certain state must be 1, which means

$$\sum_i P_i = \frac{1}{Z} \sum_i e^{-E_i/kT} = 1$$

From here we can give a definition of Z:

$$Z = \sum_i e^{-E_i/kT}, \tag{3.66}$$

that is, the summation of all Boltzmann factors. Sometimes we will use $\beta \equiv 1/kT$ for notation. In fact, Z is called the **partition function**. It only depends on the temperature which also relates the microscopic states to the macroscopic phenomenon.

We denote $E = n\epsilon$ as the energy of the state occupied by n particles. Since Boltzmann's analysis does not consider multiple particles, we can simply set $n = 1$ as a single particle, and the probability of finding the particle in a certain energy state will be

$$P_i = \frac{1}{Z_1} e^{-\epsilon/kT} \tag{3.67}$$

The occupancy follows the definition

$$\bar{n} = \sum_n n P_n \tag{3.68}$$

It means that Boltzmann's distribution is

$$\bar{n} = N P_n = \frac{N}{Z_1} e^{-\epsilon/kT}$$

Here we have to use a formula about the chemical potential:

$$\mu = -kT \ln \frac{Z_1}{N} \tag{3.69}$$

Then, we will soon find that

$$\bar{n} = e^{\mu/kT} e^{-\epsilon/kT} = e^{(\mu - \epsilon)/kT} = e^{-(\epsilon - \mu)/kT}$$

Or equivalently,

$$\boxed{\bar{n}_{\mathrm{B}} = \frac{1}{e^{(\epsilon - \mu)/kT}}} \tag{3.70}$$

This is also a common form of the Boltzmann distribution.

When we consider the probability of the moving molecules, what we can do so far might be calculating the root-mean-square speed. Now, I would like to introduce the so-called distribution function $D(v)$ of speed v. The distribution depends on the probability of a molecule/particle with speed v as well as the number of velocity vectors. The former is proportional to the Boltzmann factor, say $e^{-mv^2/2kT}$ (where we substituted $E_i = mv^2/2$); the latter is proportional to $4\pi v^2$, where we constructed a 3-dimensional velocity space (imagine a sphere with a radius v) so that the number of velocity vectors is related to its surface area $4\pi v^2$. We can simply write down what we had so far:

$$D(v) = C \cdot 4\pi v^2 \cdot e^{-mv^2/2kT}$$

Taking the integral over the distribution, the probability of finding the molecule/particle at some speed must be 1:

$$\int_0^\infty D(v)\, dv = 4\pi C \int_0^\infty v^2 e^{-mv^2/2kT}\, dv = 1$$

Evaluating the integral with the constant C, we obtain the distribution function as the following:

$$\boxed{D(v) = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} 4\pi v^2 e^{-mv^2/2kT}} \tag{3.71}$$

This is the **Maxwell-Boltzmann distribution**.

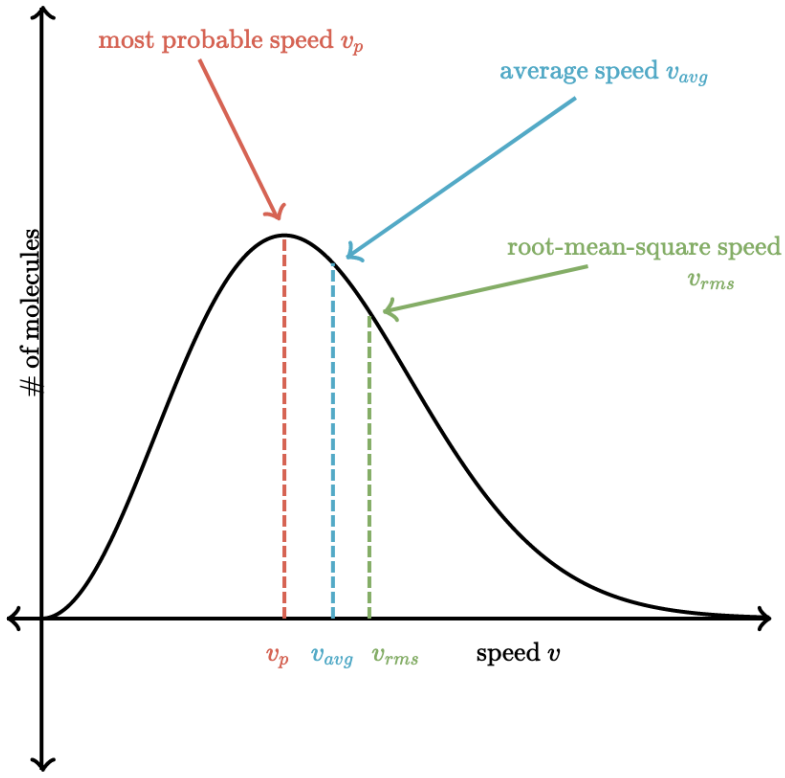

Figure 46: The Maxwell-Boltzmann distribution. Note that the left side of the curve is parabolic, whereas the right part of the curve dies exponentially. *Source*: Khan Academy.

Taking the derivative $dD(v)/dv = 0$, we shall find that the maximum of the distribution occurs at

$$v_{\max} = \sqrt{\frac{2kT}{m}} \tag{3.72}$$

, which is the most probable speed. As for the average speed, which is different, what we need to do is to sum over all the possible speeds, $\bar{v} = \sum v\, D(v)\, dv$, and then we get

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}} \tag{3.73}$$

Note that both of them are different from the root-mean-square speed

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}}$$

In comparison, $v_{\max} < \bar{v} < v_{\text{rms}}$. The distribution can be referred to Figure 46.

3.10 Quantum statistics

In the previous section, we only consider the exchange of energy. What if we not only consider the exchange of energy, but also the exchange of particles? In this way, the thermodynamic identity gives

$$dS = -\frac{1}{T}(dU - \mu\, dN)$$

The term $P\, dV$ is so small that we can throw away. But unlike the Boltzmann statistics, here we keep the term $\mu\, dN$. This means the change in entropy becomes

$$S_2 - S_1 = -\frac{1}{T}(E_2 - E_1 - \mu N_2 + \mu N_1)$$

Note the RHS has a minus sign due to the small system. As what we did in the previous section, the probability's ratio turns out to be

$$\frac{P_2}{P_1} = \frac{e^{-(E_2 - \mu N_2)/kT}}{e^{-(E_1 - \mu N_1)/kT}}$$

Here, $e^{-(E-\mu N)/kT}$ is called the **Gibbs factor**. We are convinced to write down the probability's expression in a similar way:

$$P_i = \frac{1}{\mathcal{Z}} e^{-(E_i - \mu N_i)/kT} \tag{3.74}$$

, where the **grand partition function**

$$\mathcal{Z} = \sum_i e^{-(E_i - \mu N_i)/kT} \tag{3.75}$$

Now, let me make a short introduction about what we are going to discuss. In quantum mechanics (which will be discussed afterwards), some particles can share a state with another same particle, but some cannot. For those that can share with the same particle, they are called **bosons**; for those that cannot share with the identical particle in a given state, they are called **fermions**. The well-known **Pauli exclusion principle** indicates that two identical fermions cannot occupy the same state.

Then, we can derive some useful distributions. As a starting point, let us consider the Boltzmann's case (3.70). In quantum statistics, things become more interesting. First, the probability's expression (3.67) turns into the following form:

$$P_n = \frac{1}{\mathcal{Z}} e^{-n(\epsilon - \mu)/kT} \tag{3.76}$$

Next, we shall discuss the two different cases. The first case is for fermions, which can only have either 0 or 1 particle occupied a certain state. Thus, let $n = 0, 1$

$$\mathcal{Z} = 1 + e^{-(\epsilon - \mu)/kT} \tag{3.77}$$

The occupancy followed by Eq (3.68) can be written as

$$\bar{n} = 0 \cdot P_0 + 1 \cdot P_1 = \frac{1}{\mathcal{Z}} e^{-(\epsilon - \mu)/kT} = \frac{e^{-(\epsilon - \mu)/kT}}{1 + e^{-(\epsilon - \mu)/kT}}$$

This can be equivalently expressed as

$$\boxed{\bar{n}_{\mathrm{FD}} = \frac{1}{e^{(\epsilon-\mu)/kT}+1}} \tag{3.78}$$

, which is called the **Fermi-Dirac distribution**. Note the results are between 0 and 1: when $\epsilon \gg \mu$, it goes to 0; when $\epsilon \ll \mu$, it goes to 1.

As for the case of bosons, things become different. Without the restriction of Pauli exclusion principle, the n can be any integer for bosons, and the partition function can be written as

$$\begin{aligned}\mathcal{Z} &= 1 + e^{-(\epsilon-\mu)/kT} + e^{-2(\epsilon-\mu)/kT} + ... \\ &= 1 + e^{-(\epsilon-\mu)/kT} + [e^{-(\epsilon-\mu)/kT}]^2 + ... \\ &= \frac{1}{1-e^{-(\epsilon-\mu)/kT}}\end{aligned} \tag{3.79}$$

Note that in the last line we used the following Taylor approximation:

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + ... = \frac{1}{1-x} \tag{3.80}$$

Having the grand partition function $\mathcal{Z}$, we can determine the occupancy according to Eq (3.68):

$$\bar{n} = 0 \cdot P_0 + 1 \cdot P_1 + 2 \cdot P_2 + ...$$

Well, you might get stuck with this. To evaluate the sum above, we can define $u \equiv (\epsilon - \mu)/kT$. In this way, substituting Eq (3.76), we will find that

$$\begin{aligned}\bar{n} &= \sum_n n P_n = \sum_n n \frac{1}{\mathcal{Z}} e^{-nu} \\ &= -\frac{1}{\mathcal{Z}} \sum_n n e^{-nu} = -\frac{1}{\mathcal{Z}} \sum_n \frac{\partial}{\partial u} e^{-nu} \\ &= -\frac{1}{\mathcal{Z}} \frac{\partial \mathcal{Z}}{\partial u}\end{aligned}$$

Continue the calculations,

$$\begin{aligned}\bar{n} &= -\frac{1}{\mathcal{Z}} \frac{\partial \mathcal{Z}}{\partial u} \\ &= -(1-e^{-u}) \frac{\partial}{\partial u} \left(\frac{1}{1-e^{-u}} \right) \\ &= (1-e^{-u})(1-e^{-u})^{-2}(e^{-u}) \\ &= \frac{e^{-u}}{1-e^{-u}} \\ &= \frac{e^{-(\epsilon-\mu)/kT}}{1-e^{-(\epsilon-\mu)/kT}}\end{aligned}$$

Hence, we obtain the occupancy for bosons:

$$\boxed{\bar{n}_{\mathrm{BE}} = \frac{1}{e^{(\epsilon-\mu)/kT}-1}} \tag{3.81}$$

This is called the **Bose-Einstein distribution**. Note when $\epsilon \gg \mu$ it goes to 0, but when $\epsilon \approx \mu$ it goes to infinity. It is noteworthy that when $(\epsilon - \mu)/kT \gg 1$, three distributions merge into the same trend, as shown in Figure 47.

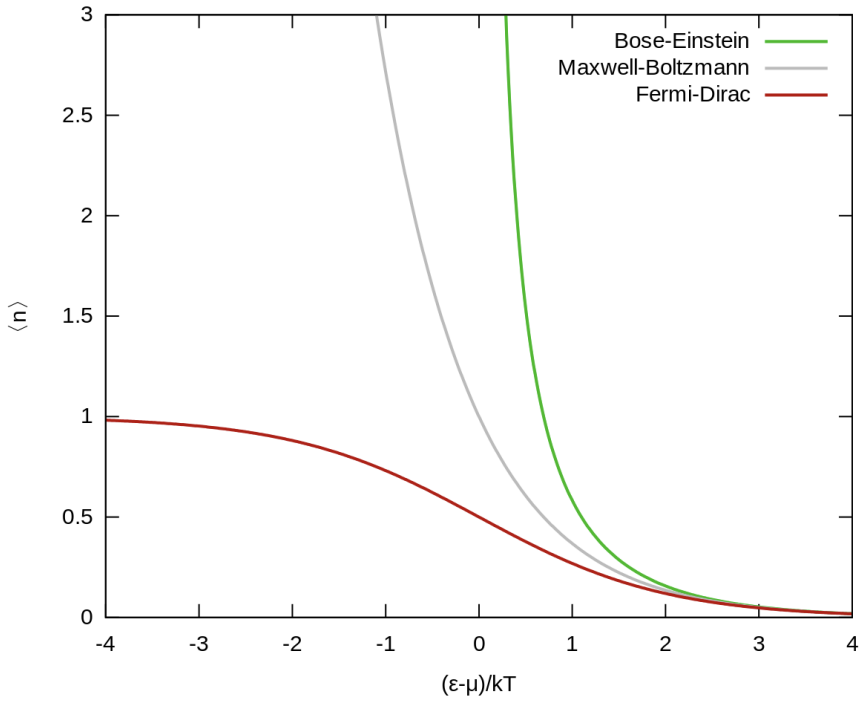

Figure 47: The Botzmann, Fermi-Dirac, and Bose-Einstein distributions. Three curves converge when $(\epsilon - \mu)/kT \gg 1$. *Source*: Wikipedia.

Before moving toward the end of our chapter, let me discuss some properties about the Fermi-Dirac distribution at a very low temperature. When $T \to 0$, the distribution will become a step function (as shown in Figure 48), and the states of a single particle were either occupied or unoccupied. In such a scenario, the threshold is

$$\epsilon_{\mathrm{F}} = \mu \tag{3.82}$$

, which is called the **Fermi energy**. The corresponding temperature is called the **Fermi temperature**:

$$T_{\mathrm{F}} = \frac{\epsilon_{\mathrm{F}}}{k} \tag{3.83}$$

Sometimes the situation of a cold gas that all states above ϵ_{F} are unoccupied whereas all states below ϵ_{F} are occupied is called degenerate gas. Note the term "**degenerate**" in quantum mechanics means a set of states having the same energy. Nevertheless, the meaning of the degenerate gas is not really related to the term as it is described in quantum mechanics. Still, we should introduce some techniques to illustrate a set of allowed states. In quantum mechanics, given a box of a free electron's wavefunctions (the reasons would be discussed in the later chapter), the energies are given by

$$\epsilon = \frac{p^2}{2m} = \frac{h^2}{8mL}(n_x^2 + n_y^2 + n_z^2) \tag{3.84}$$

where L is the length of the box (don't worry about the derivations – we will discuss in quantum's chapter). For the allowed states, they can be illustrated as a 3-dimensional n-space, where the states can be filled the region we considered and the axes are n_x, n_y, and n_z. It would be more convenient to consider the region of a eighth of a sphere, that is, the first octant of a sphere. The Fermi energy is then

$$\epsilon_{\mathrm{F}} = \frac{h^2 n_{\mathrm{max}^2}}{8mL^2} = \frac{h^2}{8m}\left(\frac{3N}{\pi V}\right)^{2/3} \tag{3.85}$$

Here, $V = L^3$ is the volume of the box. As for the N it represents the total number of occupied states, which is derived from the volume of the eighth-sphere. Then, the average energy of the electrons will be derived:

$$U = \frac{3}{5}N\epsilon_{\mathrm{F}} \tag{3.86}$$

This expression is derived by taking the integral over the spherical elements of the n-space. Last, I would like to introduce another quantity (but without derivations), the **degeneracy pressure**:

$$P = -\frac{\partial U}{\partial V} = \frac{2N\epsilon_{\mathrm{F}}}{5V} = \frac{2U}{3V} \tag{3.87}$$

This pressure is important in nuclear physics or astrophysics. As we know, the electrostatic force pulls protons and electrons together in a atom. The degenerate pressure can then prevent matter from collapsing under such a force. Note this is unrelated to the electromagnetic forces, but purely related to the Pauli exclusion principle.

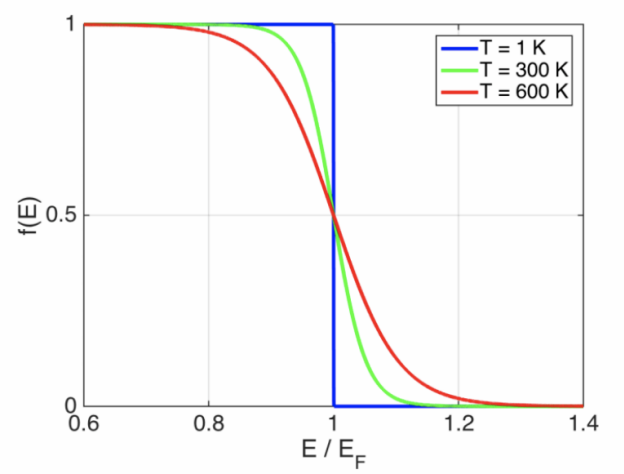

Figure 48: The Fermi-Dirac distribution behaves as a step function when $T \to 0$.
Source: https://bklein.ece.gatech.edu/laser-photonics/filling-the-states/

Many of these concepts are significant in condensed matter physics. In Planck's blackbody radiation (which will be discussed later), the electromagnetic waves can be viewed as the photon gas in a box; similarly, in the so-called **Debye model**, the oscillations of atoms can be viwed as the **phonon** gas in a box. They both obey the Bose-Einstein statistics (that is, they are bosons). You might have an impression that we mentioned the Einstein's solid model consider the solid as a set of non-interacting, independent quantum oscillators. For the Debye model of solids, it is a revision of the Einstein's model because it gives a more accurate prediction at a lower temperature, where the Einstein's model did not.

4 Electromagnetism

4.1 Charges, forces, and fields

Electromagnetism is an important branch of physics. As a start, we are going to discuss **electrostatics** in the first several sections, and we will mention some topics about the electrodynamics afterwards.

As you might have learned during middle school, many objects carry the **electric charges**. When you rub a plastic rod with fur and then put it close to a glass rod, you will find both rods attract each other. On the other hand, if the two cases are made glass rods and both are rubbed with a silk cloth, then both rods would repel each other. Why do they happen? In order to analyze such a phenomenon, physicists (such as Benjamin Franklin) define the positive and negative charges. We say that after rubbing with fur / silk cloth, the plastic rod would carry the negative charges whereas the glass rod carries the positive charges. The charges with opposite signs make them attract together. As for the other case, both glass rods carry the positive charges so that they repel each other.

Today, we realized that the charge is a fundamental, intrinsic property in physics. In other words, charge is quantized, and the quantized object is called an **electron**. The charge that an electron carries is called the elementary charge,

$$e \cong 1.602 \times 10^{-19}\ \mathrm{C}$$

where C is the unit Coulomb. The conservation of charge implies that the charge of an electron is -1 so that a proton is $+1$ because the total charge of a *neutral* atom should be 0. But one should be curious about how the "rubbing" helps some neutral object such as a plastic rod generate charges. Therefore, we should be in mind that the process of "rubbing" move the electrons (negative charges) of atoms to another atoms. For example, when a plastic rod was rubbed with fur, the electrons of fur have been moved to the plastic rod's atoms, so the charge of fur becomes "relatively" positive and the charge of the plastic rod becomes "relatively" negative. In macroscopic view, these phenomena can be expressed as the moving of positive and negative charges. But in microscopic view they are due to the transfer of electrons between objects.

The attraction or repulsion of charges can be described by the **Coulomb's law**. The force between charges is

$$\boxed{\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{Qq}{r^2}\hat{\mathbf{r}}} \tag{4.1}$$

The constants have been mentioned in Table 1. The vector $\mathbf{r}$ describes the force between two charges. Sometimes (especially in introductory textbooks), we use k to describe the constant:

$$k = \frac{1}{4\pi\epsilon_0}$$

Note Eq (4.1) is fairly similar to the gravitational force (2.74) (and now we use its vector form), since both forces are inversely proportional to r^2.

The electric force is a non-contact force, and the interaction from any charge can be propagated via the **electric field**. Given a small test charge q_0, it would experience the field from a source with a magnitude of

$$\boxed{\mathbf{E} = \frac{\mathbf{F}}{q_0}} \tag{4.2}$$

This is the definition of an E-field. Alternatively, for a charged particle q placed in an electric field, there would be a force

$$\mathbf{F} = q\mathbf{E} \tag{4.3}$$

acts on the particle. Besides, if we consider the charge q as a source, the electric field can be written as

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}} \tag{4.4}$$

Assume a positive charge is the source, the field's direction starts from itself which can be illustrated as the **electric lines**. But for the negative charge, the direction of E-field would point into the source, as shown in Figure 49.

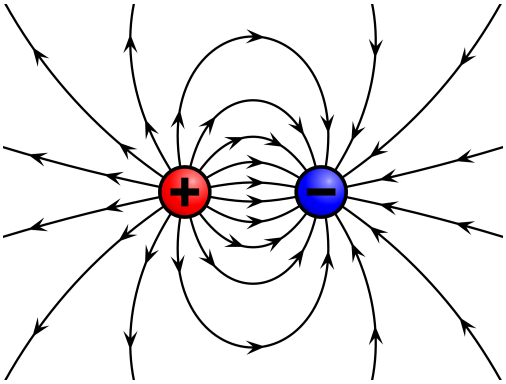

Figure 49: The electric field lines of two source charges (namely, a **dipole**). *Source*: Wikipedia

For the continuous charge distribution, we need to consider the small segment of charge, denoted by dq. The field's expression will be:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{1}{r^2}\hat{\mathbf{r}}\,dq \tag{4.5}$$

Then, we can discuss the three cases of such a distribution: a line charge, a surface charge, or a volume charge:

- **Line charge**: The charge element is $dq = \lambda\,dl$, where λ is the charge per unit length. The corresponding electric field is
$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int \frac{\lambda}{r^2}\hat{\mathbf{r}}\,dl \tag{4.6}$$
- **Surface charge**: The charge element is $dq = \sigma\,dA$, where σ is the charge per unit area. The corresponding electric field is
$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\iint \frac{\sigma}{r^2}\hat{\mathbf{r}}\,dA \tag{4.7}$$
- **Volume charge**: The charge element is $dq = \rho\,dV$, where ρ is the charge per unit volume. The corresponding electric field is
$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\iiint \frac{\rho}{r^2}\hat{\mathbf{r}}\,dV \tag{4.8}$$

Let me demonstrate some useful cases, but without derivations (I would suggest the students derive by hand, using vector analysis, trigonometry, and knowledge of basic calculus):

- For a charged ring:
$$E = \frac{1}{4\pi\epsilon_0}\frac{qz}{(z^2+R^2)^{\frac{3}{2}}} \quad \text{(charged ring)}$$
At a large distance, $z \gg R$ so that
$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{z^2} \quad \text{(far from a charged ring)} \tag{4.9}$$
Note the expression is reduced to Eq (4.4).
- For a charged disk:
$$E = \frac{1}{2\epsilon_0}\left(1 - \frac{z}{\sqrt{z^2+R^2}}\right) \quad \text{(charged disk)}$$
For an infinite sheet (with uniform charge), $R \to \infty$ so that
$$E = \frac{\sigma}{2\epsilon_0} \quad \text{(infinte sheet)} \tag{4.10}$$
Note the expression does not depend on any parameter about the distance.

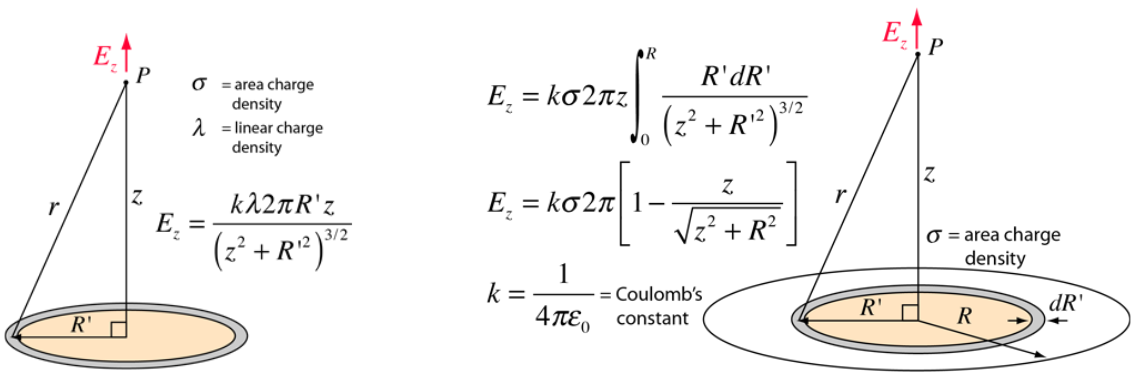

Figure 50: The vector diagrams of a charged ring (left) and a charged disk (right), with some calculation parameters. *Source*: Hyperphysics

4.2 E-fields and the Gauss's law

How much amount of electric field passing a sheet or an object? This drives me introduce another definition – the (electric) **flux**:

$$\Phi_E = \iint \mathbf{E} \cdot d\mathbf{A} \tag{4.11}$$

Notice the inner produce implies that we can calculate the flux by projecting it onto a plane. For a flat surface and a uniform field, the flux would be

$$\Phi_E = EA\cos\theta$$

Now, please imagine a sphere encloses a point charge q. Then, what would be the flux? Simply speaking, we have

$$\Phi_E = \oiint \mathbf{E} \cdot d\mathbf{A} = E \cdot A = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \cdot (4\pi r^2) = \frac{q}{\epsilon_0}$$

This can be generalized to any closed surface. Formally, the expression is

$$\boxed{\oiint \mathbf{E} \cdot d\mathbf{A} = \frac{Q_{\text{enc}}}{\epsilon_0}} \tag{4.12}$$

, where Q_{enc} is the total enclosed charge inside a surface. This is the famous **Gauss's law**, which is very helpful when we cope with some geometrically complicated cases (but with symmetry) of the charged surface. The enclosed charge can be written in the form of the charge density ρ:

$$Q_{\text{enc}} = \iiint_V \rho \, dV$$

Next, using the divergence theorem (1.14) and above's charge expression, we have

$$\iiint_V (\nabla \cdot \mathbf{E})\, dV = \oiint_S \mathbf{E} \cdot d\mathbf{A} = \frac{Q_{\text{enc}}}{\epsilon_0} = \iiint_V \frac{\rho}{\epsilon_0}\, dV$$

Hence, from the first and the last terms, we obtain the differential form of the Gauss's law:

$$\boxed{\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}} \tag{4.13}$$

This is the first equation of the Maxwell's equations, which will be discussed later. Let us consider a charge q enclosed by a spherical Gaussian surface with a radius r, the area is $4\pi r^2$, so the Gauss's law gives

$$E(4\pi r^2) = \frac{q}{\epsilon_0}$$

Therefore,

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}$$

Again, this brings us back to the Coulomb's law.

Now, let us see some examples of the Gauss's law:

- For a **conducting surface**, the enclosed area is simply A,

$$\oiint \mathbf{E} \cdot d\mathbf{A} = E \cdot A = \frac{Q_{\text{enc}}}{\epsilon_0} = \frac{\sigma A}{\epsilon_0}$$

  Therefore,

$$E = \frac{\sigma}{\epsilon_0} \quad \text{(conducting surface)} \tag{4.14}$$

- For the case with **planar symmetry**, we should consider two sides of a sheet,

$$\oiint \mathbf{E} \cdot d\mathbf{A} = 2(E \cdot A) = \frac{Q_{\text{enc}}}{\epsilon_0} = \frac{\sigma A}{\epsilon_0}$$

  This gives

$$E = \frac{\sigma}{2\epsilon_0} \quad \text{(charged sheet)} \tag{4.15}$$

  Note this is agree with our previous result in Eq (4.10).

- For the case with **spherical symmetry**, we typically consider two cases.
  - **Conducting sphere**: First, when $r \geq R$, as we derived shortly before,

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \quad \text{(spherical shell at } r \geq R\text{)} \tag{4.16}$$

    However, when $r < R$, the Gaussian surface encloses nothing so that

$$E = 0 \quad \text{(spherical shell at } r < R\text{)} \tag{4.17}$$

  - **Uniformly charged sphere**: The case would be the same for $r \geq R$:

$$E = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2} \quad \text{(sphere with uniform charge at } r \geq R\text{)} \tag{4.18}$$

    Yet, when $r < R$, we should take the ratio such that

$$\frac{q'}{q} = \frac{V'}{V} = \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi R^3} \quad \text{so that} \quad q' = \frac{r^3}{R^3}q$$

    Then the Gauss's law gives

$$\oiint \mathbf{E} \cdot d\mathbf{A} = E \cdot A = E(4\pi r^2) = \frac{1}{\epsilon_0}\frac{r^3}{R^3}q$$

    Hence, this yields

$$E = \frac{1}{4\pi\epsilon_0}\frac{qr}{R^3} \quad \text{(sphere with uniform charge at } r < R\text{)} \tag{4.19}$$

- For the case with **cylindrical symmetry**, it can be used to derive the E-field of an infinitely long, straight line. Thus, we consider the cylinder as a Gaussian surface this time (r is the radius and h is the height),

$$\oiint \mathbf{E} \cdot d\mathbf{A} = E \cdot A = E(2\pi r \cdot h) = \frac{Q_{\text{enc}}}{\epsilon_0} = \frac{\lambda h}{\epsilon_0}$$

  In this way, we can work out that

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \quad \text{(infinitely long, straight wire)} \tag{4.20}$$

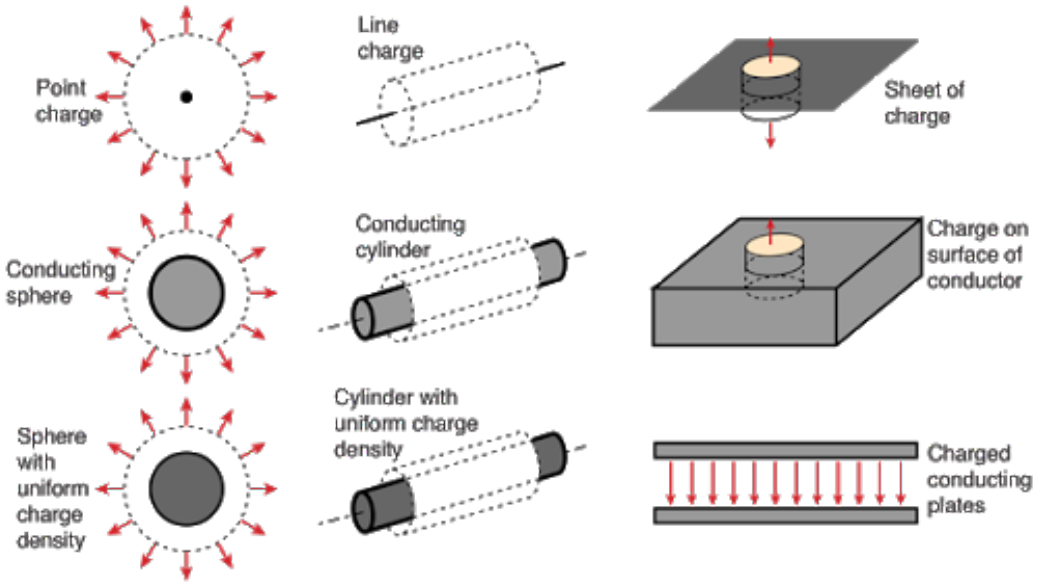

Figure 51: Different cases of the Gaussian surfaces (the dash lines). *Source*: Hyperphysics

The electric force generated from an electric field is also a conservative force. According to Eq (2.30), this implies the electric field should be curl-free. Indeed, the integral we take around a closed path has to be zero:

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \tag{4.21}$$

The Stokes' theorem gives

$$\nabla \times \mathbf{E} = 0 \tag{4.22}$$

Note we currently only consider the case in electrostatics, which means this rule would be modified if we consider the existence of the magnetic fields.

4.3 Electric potentials and boundary conditions

Let me first introduce the formal definition of **electric potential**:

$$\boxed{V(\mathbf{r}) = -\int \mathbf{E} \cdot d\mathbf{l}} \tag{4.23}$$

The potential difference between two points and be defined as

$$\Delta V = V_b - V_a = -\int_a^b \mathbf{E} \cdot d\mathbf{l} \tag{4.24}$$

Yes, as you expected, this is just the **voltage** between two points, and the unit of them is commonly V (volt). Consider a positive charge, to find its potential, we can move a test charge to infinity. The potential difference is given by

$$\Delta V = 0 - V = -\int E\, dr = -\int_r^\infty \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\, dr = -\frac{1}{4\pi\epsilon_0}\frac{q}{r}$$

Then, we can derive out the potential due to a charged particle:

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{r} \tag{4.25}$$

Another point I should emphasize is that the electric force, field, and potential all obey the superposition principle:

- $\mathbf{F} = \sum_{i=1}^{n} \mathbf{F}_i = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + ...$
- $\mathbf{E} = \sum_{i=1}^{n} \mathbf{E}_i = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + ...$
- $V = \sum_{i=1}^{n} V_i = V_1 + V_2 + V_3 + ...$

If we are going to discuss the potential of continuous charge distribution, as what we dealt with for the electric field's case in (4.5), we have

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{1}{r}\, dq$$

Specifically, in the case of a volume charge, Eq (4.8) becomes

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \iiint \frac{\rho}{r}\, dV \tag{4.26}$$

The gradient theorem (1.11) suggests that

$$V(b) - V(a) = \int_a^b (\nabla V) \cdot d\mathbf{l} = -\int_a^b \mathbf{E} \cdot d\mathbf{l}$$

The last two terms explicitly gives

$$\boxed{\mathbf{E} = -\nabla V} \tag{4.27}$$

That is to say, the electric field is the gradient of the scalar potential. Another property, which is important as well. From Eq (4.13), we can rewrite the term $\mathbf{E}$ by what we derived just now (4.27):

$$\nabla \cdot \mathbf{E} = \nabla \cdot (-\nabla V) = -\nabla^2 V = \frac{\rho}{\epsilon_0}$$

This yields the **Poisson's equation**:

$$\boxed{\nabla^2 V = -\frac{\rho}{\epsilon_0}} \tag{4.28}$$

If we consider a region that $\rho = 0$ (no charge), the Poisson's equation reduces to the famous **Laplace's equation**:

$$\boxed{\nabla^2 V = 0} \tag{4.29}$$

Note if there is no charge everywhere, the potential would be reasonably 0. However, in some scenarios, there would be no charge in a region but with charge in other places, then it would be helpful for us to apply the Laplace's equation. This equation can help us solve the potential function of a given system. We derived the Poisson's and Laplace's equation from the divergence of E-field, $\nabla \cdot \mathbf{E}$, but what if the curl of E-field? Well, you find that

$$\nabla \times \mathbf{E} = \nabla \times (-\nabla V) = 0$$

This comes from the second condition we defined in (1.10). As we expected, the curl of the E-field is supposed to be zero – which is verified by the potential again.

I will not demonstrate how to derive the potential from the Laplace's equation, which requires a bunch of mathematical techniques and you will encounter again when we discuss the solutions in quantum mechanics. But I should give you a sense that when we solve the case involved 3-dimensional spherical coordinates, more often than not, our target function (the potential) will be decomposed into two parts:

$$V(r, \theta) = R(r)\Theta(\theta)$$

One is the radial part and the other is the angular part. Without too many derivations, from the definition of the spherical coordinates, we will obtain two equations:

- The radial equation:

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = l(l+1)R$$

The general solution will be

$$R(r) = Ar^l + \frac{B}{r^{l+1}}$$

, where A and B are the arbitrary constants.

- The angular equation:

$$\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) = -l(l+1)\sin\theta\Theta$$

The general solution will be

$$\Theta(\theta) = P_l(\cos\theta)$$

This form is the **Legendre polynomial**, where the $P_l(x)$ can be expressed by the **Rodrigues formula**:

$$P_l(x) = \frac{1}{2^l l!}\left(\frac{d}{dx}\right)^l (x^2-1)^l \tag{4.30}$$

For instance, $P_0(x) = 1$, $P_1(x) = x$, and so forth.

Thus, the general form of the potential will be

$$V(r,\theta) = \sum_{l=0}^{\infty}\left(A_l r^l + \frac{B_l}{r^{l+1}}\right)P_l(\cos\theta) \tag{4.31}$$

When solving the differential equations, it is very often to consider the **boundary conditions**. I should introduce the boundary conditions in the electrostatics right now. Let us split into the two different cases:

- The electric field: The normal component is discontinuously and obeys the following condition:

$$E^{\perp}_{\text{above}} - E^{\perp}_{\text{below}} = \frac{\sigma}{\epsilon_0}$$

This is derived from the Gauss's law with $Q_{\text{enc}} = \sigma A$ of a given surface region. On other hand, the tangential component has to be continuous:

$$E^{\parallel}_{\text{above}} - E^{\parallel}_{\text{below}} = 0$$

We can conclude in a more compact form:

$$\mathbf{E}_{\text{above}} - \mathbf{E}_{\text{below}} = \frac{\sigma}{\epsilon_0}\hat{\mathbf{n}} \tag{4.32}$$

- The electric potential: Unlike the electric field, the potential is continuous across any boundary:

$$V_{\text{above}} - V_{\text{below}} = 0$$

Nevertheless, due to the discontinuity of $\mathbf{E}$, the gradient of V is not really continuous (because $\mathbf{E} = -\nabla V$), so that

$$\nabla V_{\text{above}} - \nabla V_{\text{below}} = -\frac{\sigma}{\epsilon_0}\hat{\mathbf{n}} \tag{4.33}$$

Last but not least, the boundary conditions in differential equations have the three major types:

1. **Dirichlet boundary condition**: The physical constraint of an unknown function y we put directly on the boundaries (f is a specified scalar function).

$$y = f$$

2. **Neumann boundary condition**: The derivative of such unknown function.

$$\frac{\partial y}{\partial n} = f$$

3. **Robin boundary condition**: The mixed condition of the first and the second boundary conditions.

$$ay + b\frac{\partial y}{\partial n} = f$$

These conditions are not only useful in the electromagnetism, but also powerful in many areas of mathematical physics.

4.4 Work and electric potential energy

As every cases of work we have seen, like Eq (2.24), the work done by an electric force is

$$W = \int \mathbf{F} \cdot d\mathbf{l} = -q \int \mathbf{E} \cdot d\mathbf{l} = q\Delta V = -\Delta U$$

Looking at the last two terms, if we set the reference point at infinity (or say, we consider the work when moving a test charge from infinity to the source), we find the expression for **electric potential energy**:

$$\boxed{U = qV} \tag{4.34}$$

This is similar to $\mathbf{F} = q\mathbf{E}$ in Eq (4.3). In fact, in a similar fashion, this indicates that the potential follows the definition:

$$V = \frac{U}{q_0} \tag{4.35}$$

This is also similar to $\mathbf{E} = \mathbf{F}/q_0$ in Eq (4.2), where q_0 is the test charge. That is to say, when we bring a positive test charge from infinity to the source, the potential can be considered as the electric potential energy per unit charge. Note the work $W = \mathbf{F} \cdot \mathbf{r}$ and $F = qE$, so that the electric potential energy between two points with a distance d can be expressed as

$$U = qEd$$

This can be considered as the work we have done when bringing a positive charge from a negative plate (lower potential) to a positive plate (higher potential) – which implies you are required to spend more force to achieve that, so the work would be positive. As for the electric potential it can be directly written as

$$V = Ed$$

Going back to Eq (4.25), the electric potential energy within a two-particle system can be written as

$$U = \frac{1}{4\pi\epsilon_0}\frac{Qq}{r} \tag{4.36}$$

Without a doubt, the work-energy theorem and the conservation law demand

$$\Delta U = q\Delta V \quad \text{and} \quad \Delta K = -q\Delta V \tag{4.37}$$

The energy of an electron with an elementary charge $e = 1.602 \times 10^{-19}$ C passing 1 V potential difference is

$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$$

This unit is called the **electrovolt** (eV), which is fairly common in particle physics. For example, you might have heard the Higgs boson has the energy of $\sim$ 125 GeV; you also heard the current energy limit of our LHC (Large Hadron Collider) is around 13 TeV. Well, 13 TeV $\sim 2.08 \times 10^{-6}$ J seems very small – indeed, that's what you would call into question. However, we should bear in mind that in the LHC or any particle experiments, there would be trillions of particle beams. In this way, the total energy we require will be very huge.

Then, let us consider the energy of a configuration of point charges. The total work to assemble n charges is

$$W = \frac{1}{2}\sum_{i=1}^{n} q_i V_i$$

Armed with this expression, the work of the continuous charge distribution will be found. For example, assume a volume charge density ρ, the work becomes

$$W = \frac{1}{2}\iiint \rho V \, d\tau. \tag{4.38}$$

Since I have introduced the potential as V, to avoid any confusion, I am going to use $d\tau$ to replace the original notation dV. Doing the operation to switch into $\rho = \epsilon_0(\nabla \cdot \mathbf{E})$ and some integrating methods, we will obtain

$$\boxed{W = \frac{\epsilon_0}{2}\iiint E^2 \, d\tau} \tag{4.39}$$

This is the energy of a continuous charge distribution.

4.5 Conductors

In this section, we would take a glance at the conductors. The first concept is the induced charges, which means when you put a positive charge close to a conductor (without contact), there would be the negative charges gathered on the side of the conductor where you move a charge closer, whereas the positive charges would be pushed to the other side of a conductor. A demonstration is shown as Figure 52.

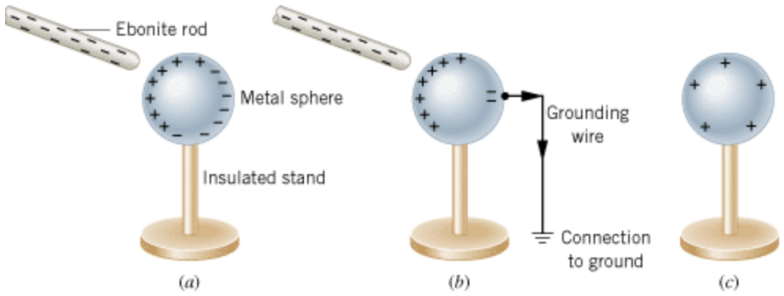

Figure 52: (a) When we move a rod with negative charges closer to the conducting sphere, the positive charges would be attracted and the negative charges would be separated away due to the induced charges; (b) next, the negative charges leave the sphere after grounding; (c) finally, the sphere carry a positive net charge only. *Source*: http://demo.webassign.net/

Another concept is the main topic in this section. Up to now, some readers have possibly noticed the proportionality among $\mathbf{E}$, V, and q: A charge Q is proportional to $\mathbf{E}$, and so it would be proportional to V. We can define the proportional constant as the **capacitance**:

$$\boxed{C \equiv \frac{Q}{V}} \tag{4.40}$$

, and the unit is 1 frad (1 F), which is Coulomb per volt. Thus we can also have

$$Q = CV \tag{4.41}$$

Next, out first application is the case of a parallel-plate capacitor. Note in our denotation, we assumed the potential (4.23) is taken in direction of moving a charge from the charged plate with negative charge to positive charge. However, to consider how the potential done by an electric field, we should consider the path in an opposite direction, since the E-field must point from positive to negative charges (as shown in Figure 53). Therefore, the potential difference turns into

$$V = \int_{-}^{+} \mathbf{E} \cdot d\mathbf{l} = E \int_{0}^{d} dl$$

In the meantime, Gauss's law suggests that

$$\oiint \mathbf{E} \cdot d\mathbf{A} = E \cdot A = \frac{Q}{\epsilon_0}$$

Therefore, we shall find that the capacitance of a parallel-plate conductor is

$$C = \frac{\epsilon_0 A}{d} \tag{4.42}$$

Note the capacitance in this case does not depend on the charge but only depend on its geometry.

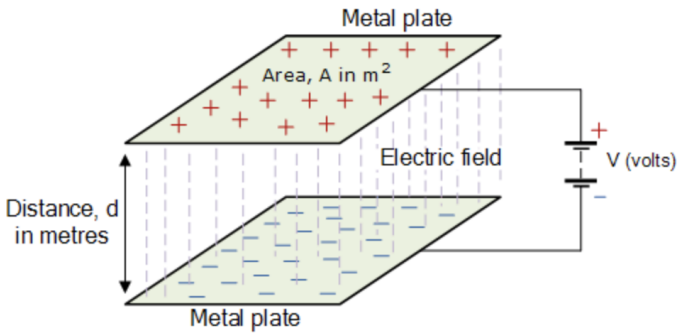

Figure 53: A parallel-plate capacitor. *Source*: https://www.electronics-tutorials.ws/

The capacitors assembled *in parallel* or *in series* have differences. When the capacitors are in parallel, the potential difference V would be the same across the circuit:

$$Q = Q_1 + Q_2 + ... = C_1 V + C_2 V + ... = (C_1 + C_2 + ...)V$$

Therefore, the equivalent capacitance of a parallel combination should be

$$C_{\text{eq}} = \sum_{i=1}^{n} C_i \quad \text{(in parallel)} \tag{4.43}$$

On the other hand, for the capacitors in series, it has an identical charge Q, and then

$$V = V_1 + V_2 + ... = \frac{Q}{C_1} + \frac{Q}{C_2} + ... = \left(\frac{1}{C_1} + \frac{1}{C_2} + ...\right) Q$$

This implies that the equivalent capacitance of a series combination would be

$$\frac{1}{C_{\text{eq}}} = \sum_{i=1}^{n} \frac{1}{C_i} \quad \text{(in series)} \tag{4.44}$$

Don't worry too much if you feel confused now. We will go through more details in the later sections.

Finally, let us see what would be changed when we calculate the energy of a capacitor. First, we set $dW = q\,dV$ and $dV = q/C$. This allows me to write down

$$dW = q\,dV = \frac{q}{C}\,dq$$

Taking an integral, we find

$$W = \int dW = \int_0^Q \frac{q}{C}\,dq = \frac{Q^2}{2C}$$

Plugging $Q = CV$, we conclude that

$$\boxed{W = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}CV^2} \tag{4.45}$$

Moreover, we determine the **energy density** is supposed to be

$$u = \frac{U}{\tau}$$

, where τ is the volume. Then, our knowledge so far tells us

$$u = \frac{U}{Ad} = \frac{CV^2}{2Ad} = \frac{1}{2}\epsilon_0\left(\frac{V}{d}\right)^2$$

Note we have applied Eq (4.42) above. Recall $\mathbf{E} = -\nabla V$, in 1-dimensional case, this is equivalent to

$$\mathbf{E} = -\frac{\partial V}{\partial r} = -\frac{\Delta V}{\Delta r}$$

Thus, this yields

$$u = \frac{1}{2}\epsilon_0 E^2 \tag{4.46}$$

, which is surprisingly agree with our result in Eq (4.39).

4.6 Current and resistance

The **electric current** is defined by

$$I \equiv \frac{Q}{t} \tag{4.47}$$

The unit of a current is ampere (A), which is equivalent to 1 A = 1 C/s. In other words,

$$Q = \int dq = \int I\,dt \tag{4.48}$$

Meanwhile, we should introduce another fundamental quantity called the **current density**. For example, the surface current density means the current (denoted by a vector $\mathbf{I}$) per unit width:

$$\mathbf{K} \equiv \frac{d\mathbf{I}}{dl} = \sigma\mathbf{v}$$

But another most common quantity is the volume current density:

$$\mathbf{J} \equiv \frac{d\mathbf{I}}{dA} = \rho\mathbf{v} \tag{4.49}$$

Incidentally, from the relation $Q = ne(Al)$, we can derive the drift speed

$$v_d = \frac{I}{neA} = \frac{J}{ne}$$

From the definition (4.49) we can rewrite our expression of the current

$$\boxed{I = \iint \mathbf{J} \cdot d\mathbf{A}} \tag{4.50}$$

Some readers might have noticed that we can employ the divergence theorem. Let's do it:

$$\oiint \mathbf{J} \cdot d\mathbf{A} = \iint (\nabla \cdot \mathbf{J})\, d\tau = -\frac{d}{dt} \iiint \rho\, d\tau = -\iiint \frac{\partial \rho}{\partial t} d\tau$$

Note the minus sign appeared because the charge decreases when the current flows outside from the volume enclosed by a surface. Then, we obtain a very crucial equation:

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0} \tag{4.51}$$

This is the **continuity equation**, which maintains the conservation of charge.

Now, let us consider the current density in a substance. The current's motion should be determined by how much force we impose on, and how fast it moves once we push it depends on the material. Therefore, we can have the following relation:

$$\mathbf{J} = \sigma \mathbf{F}$$

, where $\mathbf{F}$ is the force and σ here is not the surface density – it is called the **conductivity** instead. Another proportionality factor is also often to see, which is called the **resistivity**:

$$\rho = \frac{1}{\sigma} \tag{4.52}$$

Notice both conductivity and the resistivity only depend on the materials. Without considering the magnetic field, the current density can be related to the corresponding E-field:

$$\boxed{\mathbf{J} = \sigma \mathbf{E}} \quad \text{or} \quad \boxed{\mathbf{E} = \rho \mathbf{J}} \tag{4.53}$$

In fact, this is a general form of the well-known **Ohm's law**. However, you should be more familiar with its another form:

$$\boxed{V = IR} \tag{4.54}$$

They, in fact, describe the same phenomenon. But we here use R to represent the **resistance**, which is simply the slope of an I-V plot:

$$R \equiv \frac{V}{I} \tag{4.55}$$

Note we already had the identity $V = Ed$. Considering the potential's expression we derived between two points, $V = Ed$, now we replace d with the length of a wire L. This means

$$E = \frac{V}{L} = \frac{IR}{L}$$

At the same time we also know the current density $J = I/A$. Using $E = \rho J$ in Eq (4.53), we can find out that

$$R = \rho \frac{L}{A} \tag{4.56}$$

This identity suggests that the resistance is in proportional to the length of a wire but inversely proportional to the cross-sectional area. Straightforwardly, when the area is larger, it means more current can easily pass through so that the resistance becomes smaller.

Then we can write down the power within a circuit. Since $dq = I\, dt$,

$$dU = V\, dq = V I\, dt$$

In addition, we have the Ohm's law $V = IR$. These allow to write down the complete relation of power ($P = dW/dt = dU/dt$):

$$\boxed{P = IV = I^2 R = \frac{V^2}{R}} \tag{4.57}$$

The first term can be used to calculate the electrical energy transfer; the last two terms can be used to calculate the dissipation caused by the resistance.

4.7 Circuit-1: emf, Kirchhoff's rules, and RC circuit

An important application of electromagnetism is the electronics. One thing to characterize the electric energy across a circuit is the **electromotive force (emf)**, defined by

$$\mathcal{E} = \frac{dW}{dQ} \tag{4.58}$$

That is, the work done per unit charge. Note this is not really a force, though the name is misleading. Strictly speaking, it is equivalent to the integral of a force per unit charge:

$$\boxed{\mathcal{E} \equiv \oint \mathbf{f} \cdot d\mathbf{l} = -\int \mathbf{E} \cdot d\mathbf{l} = V} \tag{4.59}$$

Consider a single-loop circuit with a current,

$$dW = \mathcal{E}\, dQ = \mathcal{E} I\, dt = I^2 R\, dt$$

Therefore, for a single-loop circuit, the emf is equal to the product of the current and the resistance:

$$\mathcal{E} = IR \tag{4.60}$$

This means $\mathcal{E} - IR = 0$. If there is a real battery connected with a circuit, the potential difference across such a battery would be

$$V = \mathcal{E} - IR \tag{4.61}$$

Then, as what we did before, we can also determine the rate of energy transfer:

$$P = IV = I(\mathcal{E} - IR) = I\mathcal{E} - I^2 R = P_{\text{emf}} - P_{\text{r}}$$

, where the two terms are the power of emf device

$$P_{\text{emf}} = I\mathcal{E} \tag{4.62}$$

and the internal dissipation rate

$$P_{\text{r}} = I^2 R \tag{4.63}$$

Now, some readers might question: How could we define a resistance if the circuit has more than one resistors? This is involved in the situations of combined in series or in parallel, as we did for capacitors. But before doing this, let me introduce two useful rules for calculating the current and the voltage within a circuit. Because the connections in parallel are usually involved in multiloop circuits, we need to develop a handy law for calculations. The rules are called the **Kirchhoff's circuit laws**:

1. **Kirchhoff's junction rule**: Based on the conservation of charge, the currents entering a given junction has be equal to the currents leaving the junction, or equivalently,

$$\sum_{i=1}^{n} I_i = 0 \tag{4.64}$$

2. **Kirchhoff's loop rule**: Based on the fact that electric potential is conservative, the sum of the voltages around a loop must be equal to 0,

$$\sum_{i=1}^{n} V_i = 0 \tag{4.65}$$

These can help us find the voltages, currents, or resistances of a circuit. Note the direction does matter as well, as shown in Figure 54. For the right diagram, you can also imagine the voltage starting from V_s and

passing across $V_1, V_2, ..., V_7$, so that $V_s - V_1 - V_2 - V_3 - V_4 - V_5 - V_6 - V_7 = 0$, which is the same as what it has shown.

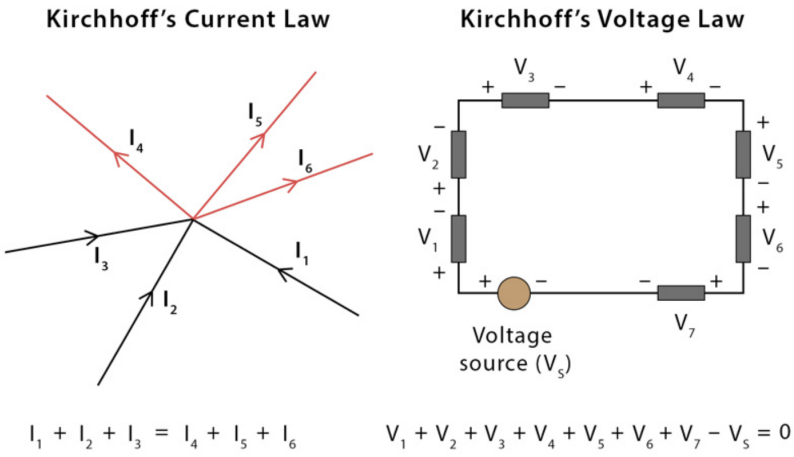

Figure 54: The Kirchhoff's junction rule of currents (left) and the loop rule of voltages (right). *Source*: https://www.sciencefacts.net/kirchhoffs-law.html

An example is shown as Figure 55. According to the Kirchhoff's rules, you should find that

$$I_1 = I_2 + I_3 \quad \text{(at Junction b)}$$
$$V_1 - I_1 R_1 - I_2 R_2 = 0 \quad \text{(across Loop abefa)}$$
$$V_2 - (-I_3 R_4) - (-I_3 R_3) - I_2 R_2 = 0 \quad \text{(across Loop ebcde)}$$

It is noteworthy that in this example, the current I_3 through R_3, R_4 is in an opposite direction to the Loop ebcde's current caused by the battery. This demands us to take minus signs ahead I_3. By solving these equations, you should find that $I_1 = 3$ A, $I_2 = 5$ A, and $I_3 = -2$ A.

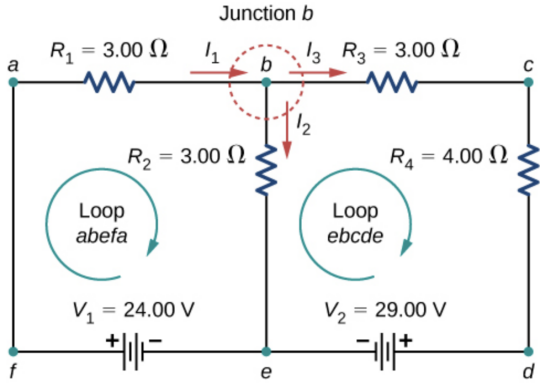

Figure 55: The resistances connected in series (left) and in parallel (right). The former is according to the Kirchhoff's loop rule and the latter is the junction rule. *Source*: https://openpress.usask.ca/physics155/chapter/6-3-kirchhoffs-rules/

For a circuit combined in series, in the simplest case, a circuit should be a single loop. We have to demand the sum of all potential differences V across the resistors are equal to V itself. According to the Kirchhoff's loop rule,

$$\mathcal{E} - V_1 - V_2 - ... = \mathcal{E} - IR_1 - IR_2 - ... = \mathcal{E} - IR_{\text{eq}} = 0$$

Hence,

$$R_{\text{eq}} = \sum_{i=1}^{n} R_i \quad \text{(in series)} \tag{4.66}$$

However, for the parallel combination, we should consider the junctions in a circuit. In this way, the Kirchhoff's junction rule suggests that

$$I = I_1 + I_2 + ... = \frac{V}{R_1} + \frac{V}{R_2} + ... = \left(\frac{1}{R_1} + \frac{1}{R_2} + ...\right) V = \frac{1}{R_{\text{eq}}} V$$

Accordingly,

$$\frac{1}{R_{\text{eq}}} = \sum_{i=1}^{n} \frac{1}{R_i} \quad \text{(in parallel)} \tag{4.67}$$

Note these are different from what we derived for capacitor's cases. For resistors, what I would suggest to bear in mind is that for series there would be the same current through all resistors, while for parallel there would be the same potential difference across all resistors.

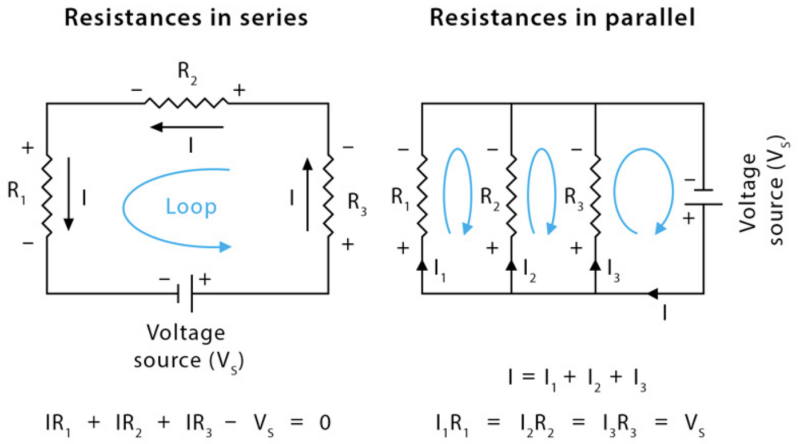

Figure 56: The resistances connected in series (left) and in parallel (right). The former is according to the Kirchhoff's loop rule and the latter is the junction rule. *Source*: https://www.sciencefacts.net/kirchhoffs-law.html

Armed with the basic concepts regarding the resistance and the capacitance, we can discuss an example of a time-varying circuit – the **RC circuit**. This is a simple circuit composed of a electric source (battery), a resistor, and a capacitor (Figure 57). According to what we learned from Kirchhoff's loop rule,

$$\mathcal{E} - V_r - V_c = \mathcal{E} - IR - \frac{Q}{C} = 0$$

From this we can write down a differential equation:

$$R\frac{dQ}{dt} + \frac{Q}{C} = \mathcal{E} \quad \text{(charging)} \tag{4.68}$$

Here, $I = dQ/dt$. As I have denoted, this equation describes a process that a series of cycles come to a fixed value of $\mathcal{E}$, so it is called the charging equation. In addition, we also notice that this would tell us how the current and the charge behaves as time goes by. Solving the differential equation gives two solutions (where I employed the differentiation $I = dQ/dt$ for the second one):

$$Q(t) = C\mathcal{E}\left(1 - e^{-t/RC}\right) \quad \text{and} \quad I(t) = \frac{\mathcal{E}}{R} e^{-t/RC} \quad \text{(charging)} \tag{4.69}$$

Since $V = Q/C$, we can identify that the potential different across the capacitor when charging would be

$$V(t) = \mathcal{E}(1 - e^{-t/RC}) \quad \text{(charging)} \tag{4.70}$$

As a convention, we often introduced the time constant $\tau = RC$ in the equations.

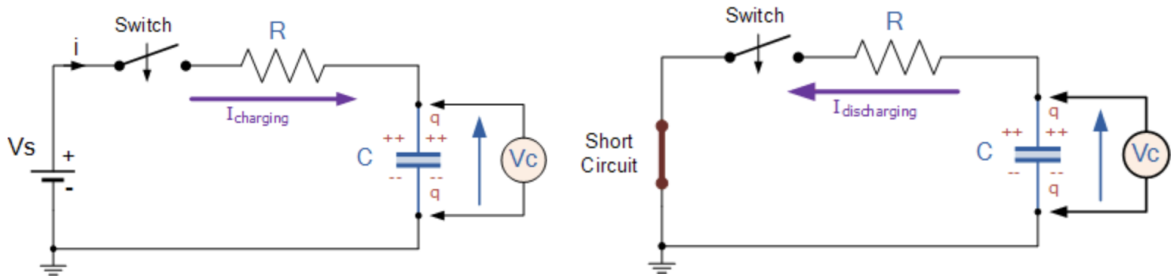

Figure 57: A simple charging (left) and discharging (right) RC circuit. *Source*: https://www.electronics-tutorials.ws/rc/

But if we take the battery away, the case would turn into a discharging circuit. Without the battery, Eq (4.68) becomes

$$R\frac{dQ}{dt} + \frac{Q}{C} = 0 \quad \text{(discharging)} \tag{4.71}$$

Solving the equation gives:

$$Q(t) = Q_0 e^{-t/RC} \quad \text{and} \quad I(t) = -\left(\frac{Q_0}{RC}\right) e^{-t/RC} \quad \text{(discharging)} \tag{4.72}$$

As previously, we often used $\tau = RC$ as a convention. The curves are shown in Figure 58.

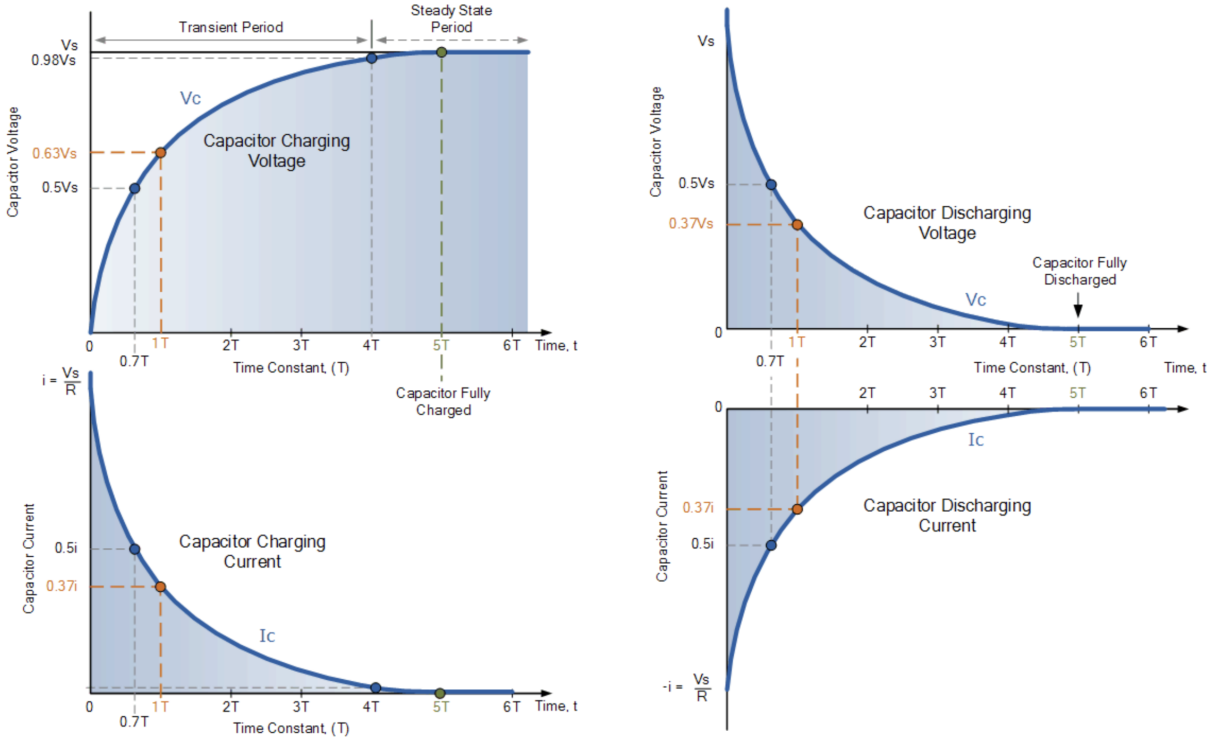

Figure 58: The V-t plot and the q-t plot of a charging circuit (left) and a discharging circuit (right). *Source*: https://www.electronics-tutorials.ws/rc/

4.8 Magnetic fields

Given a charged particle with velocity $\mathbf{v}$, there would be a magnetic force generated by the **magnetic field** $\mathbf{B}$:

$$\mathbf{F}_B = q(\mathbf{v} \times \mathbf{B}) \tag{4.73}$$

For the magnitude, it would be $F_B = qvB\sin\theta$. This is followed by the so-called **Lorentz force law**:

$$\boxed{\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})} \tag{4.74}$$

The magnetic field is in the unit of tesla (T), which is equivalent to 10^4 gauss. It is noteworthy that the work of a magnetic field

$$dW = \mathbf{F}_B \cdot d\mathbf{l} = q(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}\, dt = 0$$

since $(\mathbf{v} \times \mathbf{B})$ is perpendicular to $\mathbf{v}$. That is to say, the magnetic forces do no work! In addition, plugging $q = It$ back to Eq (4.73), we can figure out the magnetic force acting on a current:

$$\mathbf{F}_B = I\mathbf{L} \times \mathbf{B} \tag{4.75}$$

Therefore, for the magnitude it becomes $F_B = ILB\sin\theta$. Furthermore, when a charged particle with a mass m and a charge q is put in an electromagnetic field (where E and B should be perpendicular), it will form a cycloid motion. The corresponding cyclotron frequency is

$$\omega = \frac{qB}{m}$$

and the corresponding radius would be

$$R = \frac{E}{\omega B}$$

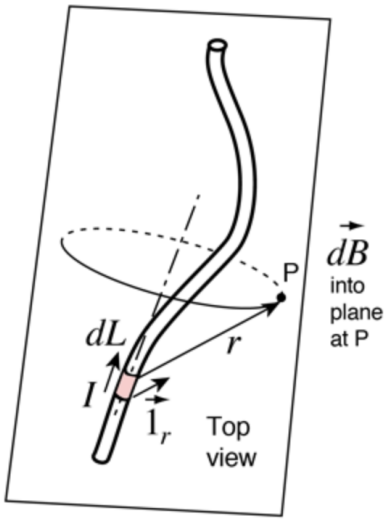

Figure 59: The magnetic field of the wire with a steady current I. *Source*: Hyperphysics.

Back to the topic of magnetic field. A steady current will generate a magnetic field. Then, how much would the magnetic field be generated? The **Biot-Savart law** suggests that the magnetic field of a segment of a wire with a steady current (as shown in Figure 59) is

$$d\mathbf{B} = \frac{\mu_0 I\, d\mathbf{l} \times \mathbf{r}}{4\pi r^2}$$

By integrating this, we can get the expression of the magnetic field:

$$\boxed{\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{I\, d\mathbf{l}\times\hat{\mathbf{r}}}{r^2}} \tag{4.76}$$

For a infinitely long straight wire, the magnitude of a magnetic field reduces to

$$B = \frac{\mu_0 I}{2\pi r} \quad \text{(long straight wire)} \tag{4.77}$$

Note that the directions always obey the right-hand rule.

As what we did for electrostatics, we can try to calculate the curl and the divergence of $\mathbf{B}$. Taking the integral of the magnetic field (4.77) around a closed path,

$$\oint \mathbf{B}\cdot d\mathbf{l} = \frac{\mu_0 I}{2\pi r}\oint d\mathbf{l} = \frac{\mu_0 I}{2\pi r}(2\pi r) = \mu_0 I_{\text{enc}}$$

In fact, this is the famous **Ampère's law**,

$$\boxed{\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 I_{\text{enc}}} \tag{4.78}$$

, which is the magnetostatic analog of the Gauss's law. And, the closed loop we take is also called the Amperian loop. Recall the current can be written by the current density's surface integral,

$$I_{\text{enc}} = \iint \mathbf{J}\cdot d\mathbf{A}$$

Applying the Stokes' theorem, we have

$$\oint \mathbf{B}\cdot d\mathbf{l} = \iint (\nabla\times\mathbf{B})\, d\mathbf{A} = \mu_0 \iint \mathbf{J}\cdot d\mathbf{A}$$

Hence, we find that

$$\boxed{\nabla\times\mathbf{B} = \mu_0\mathbf{J}} \tag{4.79}$$

This is the differential form of the Ampère's law. So far we have seen that the Biot-Savart law is analogous to the Coulomb's law, whereas the Ampère's law is analogous to the Gauss's law in electrostatics. For instance, given a solenoid with a segment of length l, and n is defined by the number of turns per unit length. This means

$$I_{\text{enc}} = I(nl)$$

The Eq (4.78) implies that

$$B(l) = \mu_0 I(nl)$$

Thus, the magnetic field of a solenoid will be

$$B = \mu_0 nI \quad \text{(solenoid)} \tag{4.80}$$

Another example is the magnetic field of a toroid, similarly,

$$B(2\pi r) = \mu_0 IN$$

, where r is the radius of a toroid and N is the total number of turns. Then, we work out the magnetic field of a toriod:

$$B = \frac{\mu_0 NI}{2\pi r} \quad \text{(toroid)} \tag{4.81}$$

Now, how about the divergence of the magnetic field? In a general case, one can consider the volume current, so the Biot-Savart law of a magnetic field can be rewritten as

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\iiint \frac{\mathbf{J}(\mathbf{r}')\times\hat{\mathbf{r}}}{r^2} dV' \tag{4.82}$$

Note we include more details here: the frame of reference for a current distribution is different from the frame of its magnetic field (denoted by primed and unprimed variables). Before taking the divergence, I have to introduce the two mathematical rules:

$$\nabla \cdot (\mathbf{f} \times \mathbf{g}) = \mathbf{g} \cdot (\nabla \times \mathbf{f}) = \mathbf{f} \cdot (\nabla \times \mathbf{g}) \tag{4.83}$$

and

$$\nabla \times (r^n \hat{\mathbf{r}}) = 0 \tag{4.84}$$

Armed with these techniques, we can calculate the divergence:

$$\nabla \cdot \mathbf{B} = \frac{\mu_0}{4\pi} \iiint \nabla \cdot \left(\mathbf{J} \times \frac{\hat{\mathbf{r}}}{r^2} \right) dV'$$

Using the product rule (4.83), the divergence of a term inside the integral will be

$$\nabla \cdot \left(\mathbf{J} \times \frac{\hat{\mathbf{r}}}{r^2} \right) = \frac{\hat{\mathbf{r}}}{r^2} \cdot (\nabla \times \mathbf{J}) - \mathbf{J} \cdot \left(\nabla \times \frac{\hat{\mathbf{r}}}{r^2} \right)$$

This looks a little bit tricky, but don't worry at all! For the first term on the RHS, $(\nabla \times \mathbf{J}) = 0$ because $\mathbf{J} = \mathbf{J}(\mathbf{r}')$ does not depends on the unprimed variables. For the second term, using Eq (4.84), we can assert that this term will vanish. Thus,

$$\boxed{\nabla \cdot \mathbf{B} = 0} \tag{4.85}$$

In other words, the divergence of the magnetic field must be zero. More accurate, $\nabla \cdot \mathbf{B} = 0$ is a mathematical statement suggests that *the magnetic monopole does not exist* (Figure 60).

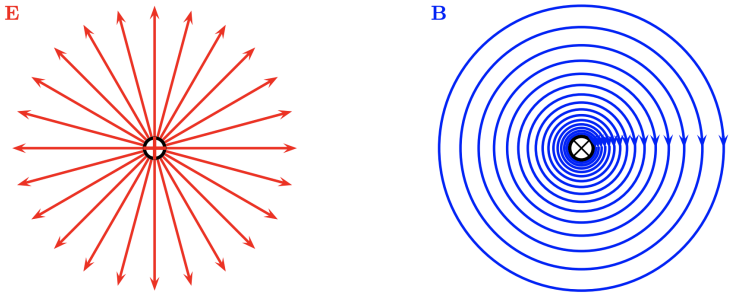

Figure 60: The curl of a pure E-field is zero, but the divergence is not (left). The divergence of a B-field is zero, whereas the curl is not (right). *Source*: https://web2.ph.utexas.edu/.

Some people might be curious that, does a magnetic field have the potential such as the one we have in electrostatics? The answer is, yes. The condition $\nabla \times \mathbf{E} = 0$ enables us to introduce the scalar potential V followed by

$$\mathbf{E} = -\nabla V$$

In a similar fashion, the condition $\nabla \cdot \mathbf{B} = 0$ permits us to have a vector potential, $\mathbf{A}$, followed by

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}} \tag{4.86}$$

If we take the curl of the magnetic field, the identity (1.10) and the Ampère's law demands that

$$\nabla \times \mathbf{B} = \nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J}$$

Now, we introduce the **Coulomb gauge** (which will be discussed in the later sections),

$$\nabla \cdot \mathbf{A} = 0 \tag{4.87}$$

Then we immediately work out the expression as the following:

$$\boxed{\nabla^2 \mathbf{A} = -\mu_0 \mathbf{J}} \tag{4.88}$$

This can be regarded as the magnetic version of the Poisson's equation. The magnetic vector potential can then be expressed explicitly as the following:

$$\boxed{\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iiint \frac{\mathbf{J}(\mathbf{r}')}{r} dV'} \tag{4.89}$$

Note that compared to the field, the vector potential looks not really natural at the beginning. However, in modern field theories, physicists used "potentials" much more often than using the electric or magnetic fields. In some sense, the "potential" might be a more fundamental quantity compared to the fields.

4.9 Multipole expansion

Recall we have mentioned a form of the potential in spherical coordinates as shown in Eq (4.31). Let us define some symbols as shown in Figure 61. As we showed before, the electric potential is defined by

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \iiint \frac{\rho(\mathbf{r}')}{\imath} \, d\tau'$$

, where $\imath = |\mathbf{r} - \mathbf{r}'|$. I shall use this notation several times in the later topics.

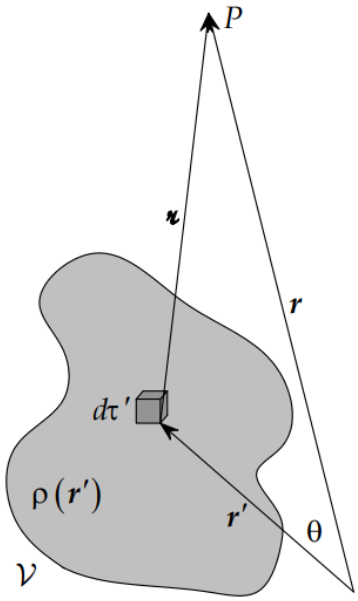

Figure 61: The systematic set-up of the multipole expansion. *Source*: https://phys.libretexts.org/

Doing some expansion techniques in terms of $\imath$, we will find the Legendre polynomials in the end:

$$\frac{1}{\imath} = \frac{1}{r} \sum_{n=0}^{\infty} \left(\frac{r'}{r}\right)^n P_n(\cos\theta)$$

This is the generating function for Legendre polynomials. In this way, the multipole expansion of the electric potential is given by

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \iiint (r')^n P_n(\cos\theta) \rho(\mathbf{r}') \, d\tau' \tag{4.90}$$

Based on the Legendre polynomials (4.30), we can write down in a more explicit way:

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\left[\frac{1}{r}\iiint \rho(\mathbf{r}')d\tau' \right. \\ + \frac{1}{r^2}\iiint r'\cos\theta\,\rho(\mathbf{r}')\,d\tau' \\ \left. + \frac{1}{r^3}\iiint r'^2\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)\rho(\mathbf{r}')d\tau' + ...\right] \tag{4.91}$$

The first term in the bracket is called the electric **monopole**, the second term is called the **dipole**, the third term is called the **quadrupole**, and so forth. Such a method is called the **multipole expansion**.

Without a doubt, the first term of Eq (4.91) is purely the one we get for the potential of a point charge, that is, Eq (4.25),

$$V_{\text{monopole}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} \tag{4.92}$$

For the second term, which is the potential of a dipole,

$$\begin{aligned} V_{\text{dipole}}(\mathbf{r}) &= \frac{1}{4\pi\epsilon_0}\frac{1}{r^2}\iiint r'\cos\theta\,\rho(\mathbf{r}')\,d\tau' \\ &= \frac{1}{4\pi\epsilon_0}\frac{\hat{\mathbf{r}}}{r^2}\cdot\iiint \mathbf{r}'\rho(\mathbf{r}')\,d\tau' \end{aligned} \tag{4.93}$$

The second line is based on the fact that $r'\cos\theta = \hat{\mathbf{r}}\cdot\mathbf{r}'$. One can introduce the concept of the **dipole moment**,

$$\mathbf{p} = \iiint \mathbf{r}'\rho(\mathbf{r}')\,d\tau' \tag{4.94}$$

, so the dipole's potential can be written in a more succinct form:

$$V_{\text{dipole}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2} \tag{4.95}$$

Because this expression can be written in term of (r,θ),

$$V_{\text{dipole}}(r,\theta) = \frac{1}{4\pi\epsilon_0}\frac{p\cos\theta}{r^2}$$

for the electric field of a dipole, we can then write it through the similar variables:

$$\mathbf{E}_{\text{dipole}}(r,\theta) = \frac{1}{4\pi\epsilon_0}\frac{p}{r^3}(2\cos\theta\hat{\mathbf{r}} + \sin\theta\hat{\boldsymbol{\theta}}) \tag{4.96}$$

It can also be proved that the field can be written in a coordinate-free form:

$$\mathbf{E}_{\text{dipole}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{p}{r^3}[3(\mathbf{p}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}] \tag{4.97}$$

Back to Eq (4.94), here what we considered is a more general object with a continuous charge distribution. Still, for the collection of point charges, the dipole moment becomes

$$\mathbf{p} = \sum_{i=1}^{n} q_i\mathbf{r}_i' \tag{4.98}$$

Imagine there is a dipole composed of a positive and a opposite charge, the distance between them can be denoted by d. Hence, the physical dipole moment turns into

$$\mathbf{p} = q\mathbf{d}$$

Since $q = F/E$ and $d = p/q$, we can say the torque acting on a dipole $\tau = Fd\sin\theta$ becomes

$$\tau = pE\sin\theta$$

In the vector form, the net torque becomes

$$\tau = \mathbf{p} \times \mathbf{E} \quad \text{(electric)} \tag{4.99}$$

Using no matter what the right-hand rule or the cross product's rule, the clockwise rotation will lead to the torque with a minus sign, $\tau = -pE\sin\theta$. Therefore, the potential energy on a dipole can be derived:

$$U = -W = -\int \tau\, d\theta = \int pE\sin\theta\, d\theta = -pE\cos\theta$$

Or,

$$U = -\mathbf{p} \cdot \mathbf{E} \quad \text{(electric)} \tag{4.100}$$

This is the potential energy of a dipole.

So far we explored the multipole expansion for electric cases. The same expansion for $1/\mathscr{r}$ and the angle θ represents the angle between $\mathbf{r}$ and $\mathbf{r}'$. But instead of the scalar potential, this time we are going to focus on the vector potential

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi}\oint \frac{1}{\mathscr{r}} d\mathbf{l}'$$

From the Legendre polynomials we used in the previous case, the vector potential can be written in the same way:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi}\sum_{n=0}^{\infty}\frac{1}{r^{n+1}}\oint (r')^n P_n(\cos\theta)\, d\mathbf{l}' \tag{4.101}$$

Accordingly, it can be written in a more explicit form:

$$\begin{aligned}\mathbf{A}(\mathbf{r}) = \frac{\mu_0 I}{4\pi}\Bigg[&\frac{1}{r}\oint d\mathbf{l}' \\ &+ \frac{1}{r^2}\oint r'\cos\theta\, d\mathbf{l}' \\ &+ \frac{1}{r^3}\oint r'^2\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) d\mathbf{l}' + ...\Bigg]\end{aligned} \tag{4.102}$$

This is the multipole expansion for a magnetic potential. A crucial point here is that the first term

$$\mathbf{A}_{\text{monopole}}(\mathbf{r}) = \frac{1}{r}\oint d\mathbf{l}' = 0$$

, which states, again, there is no magnetic monopole in nature. But the magnetic dipole, quadruple, octopole, and so on do exist. For the magnetic dipole, it has the form

$$\begin{aligned}\mathbf{A}_{\text{dipole}}(\mathbf{r}) &= \frac{\mu_0 I}{4\pi r^2}\oint r'\cos\theta\, d\mathbf{l}' \\ &= \frac{\mu_0 I}{4\pi r^2}\oint (\hat{\mathbf{r}}\cdot\mathbf{r}')d\mathbf{l}'\end{aligned} \tag{4.103}$$

A mathematical technique

$$\oint (\hat{\mathbf{r}}\cdot\mathbf{r}')d\mathbf{l}' = -\hat{\mathbf{r}}\times\int d\mathbf{A}'$$

enables us to introduce the magnetic dipole moment

$$\mathbf{m} = I\mathbf{A} \tag{4.104}$$

Hence, the vector potential can be expressed in a more compact form:

$$\mathbf{A}_{\text{dipole}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\hat{\mathbf{r}}}{r^2} \tag{4.105}$$

For sure, we can express the dipole's vector potential in terms of spherical coordinates. Setting $\mathbf{m}$ points in the z-direction, the potential will be

$$\mathbf{A}_{\text{dipole}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{m\sin\theta}{r^2}\hat{\boldsymbol{\phi}}$$

Thus, the magnetic field $\mathbf{B} = \nabla\times\mathbf{A}$ suggests

$$\mathbf{B}_{\text{dipole}}(r,\theta) = \frac{\mu_0}{4\pi}\frac{m}{r^3}(2\cos\theta\hat{\mathbf{r}} + \sin\theta\hat{\boldsymbol{\theta}}) \tag{4.106}$$

For sure, one can prove that the potential can also be written in the coordinate-free form

$$\mathbf{B}_{\text{dipole}}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{1}{r^3}[3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}] \tag{4.107}$$

Let me shed light on some essential concepts. As what we did previously, the torque generated by a magnetic field will be

$$\tau = mB\sin\theta$$

In the vector form, it is

$$\boldsymbol{\tau} = \mathbf{m}\times\mathbf{B} \quad \text{(magnetic)} \tag{4.108}$$

As for the potential energy, similar to the electric case but just replace $\mathbf{p}$ by $\mathbf{m}$:

$$U = -\mathbf{m}\cdot\mathbf{B} \quad \text{(magnetic)} \tag{4.109}$$

4.10 Electromagnetic induction

Consider a rectangular loop with a width of l moved toward or away a magnetic field B with a speed v, the motion emf is

$$\mathcal{E} = vBl \tag{4.110}$$

And, there would be an **induced current** generated through the loop. The direction also obey the right-hand rule, as shown in Figure 62, and you might feel a little weird. Indeed, this sounds incredible, though, we will introduce immediately.

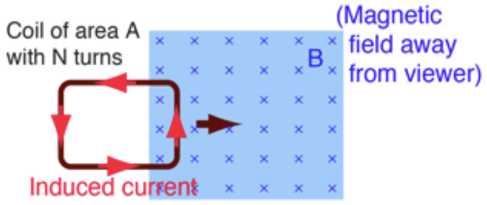

Figure 62: When a loop moving toward a magnetic field, the motion causes the induced current in an opposed direction against the field. *Source*: Hyperphysics.

As what we have in electrostatics, one can have the magnetic flux based on the following definition:

$$\Phi_B = \iint \mathbf{B}\cdot d\mathbf{A} \tag{4.111}$$

The corresponding unit is called weber (Wb), where 1 Wb = 1 T·m^2. The change of magnetic flux (now denoted by Φ) will lead to the induced emf, and so does the induced current. Such a statement can be described by

$$\boxed{\mathcal{E} = -\frac{d\Phi}{dt}} \tag{4.112}$$

This is known as the **Faraday's law**. If we consider a coil of N turns, the law would be modified as

$$\mathcal{E} = -N\frac{d\Phi}{dt}$$

The Faraday's law tells us that a changing magnetic field (with respect to time) induces an electric field. Because

$$\boxed{\oint \mathbf{E}\cdot d\mathbf{l} = -\frac{d\Phi}{dt}} \tag{4.113}$$

, then we also have

$$\mathcal{E} = \oint \mathbf{E}\cdot d\mathbf{l} = -\iint \frac{\partial \mathbf{B}}{\partial t}\cdot d\mathbf{A}$$

Thus, Eq (4.113) is the integral form of the Faraday's law, whereas the differential form of the Faraday's law can be obtained via the Stokes' theorem:

$$\boxed{\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}} \tag{4.114}$$

But a noteworthy question entails: Why do we have a minus sign ahead the equation? This is what the **Lenz's law** indicates: the direction of an induced current must be opposite to the direction that the magnetic flux changes.

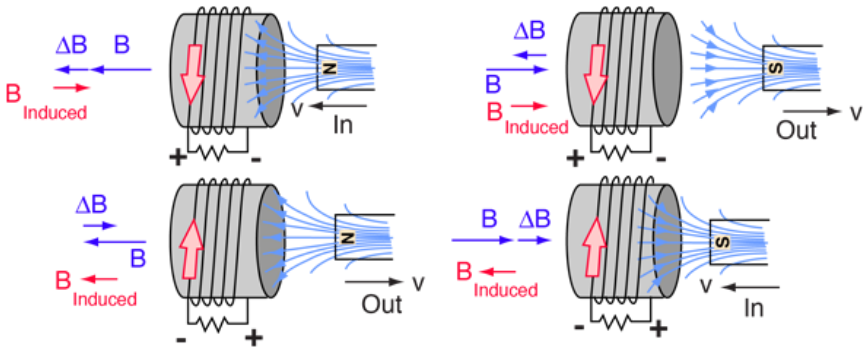

Figure 63: The Lenz's law implies that an induced current's direction opposes to the change of the magnetic flux. *Source*: Hyperphysics.

Before moving on the next topic, I shall introduce a few electronic concepts. So far we have encountered the resistors and the capacitors. Yet, there is also a component called the **inductor**, which can be used to produce the magnetic field. The definition of the **inductance** is related to the flux:

$$L = N\frac{d\Phi}{dI} \tag{4.115}$$

The unit for an inductance is the henry (H), where 1 H = 1 T·m^2/A. Given a solenoid as an example: The magnetic field of a solenoid is $B = \mu_0 nI$, and the total number of turns is defined by $N = nl$ where l is the unit length. Accordingly, the inductance can be determined by

$$L = N\frac{\Phi}{I} = \frac{(nl)(\mu_0 nI\cdot A)}{I} = \mu_0 n^2 Al$$

The fact that

$$L = \mu_0 n^2 Al \quad \text{(solenoid)} \tag{4.116}$$

shows that the inductance only depends on the geometry, just like what we know for the capacitance.

There are two types of the inductance. First, we consider an induced current appeared through a coil itself, such a situation is called the self-induction. To calculate the self-induced emf, we have known that $LI = N\Phi$. According to the Faraday's law,

$$\mathcal{E} = -\frac{d(N\Phi)}{dt} = -L\frac{dI}{dt}$$

In a more general case,

$$\Phi = LI$$

is still held, and we will get the same result for the **self-inductance**:

$$\mathcal{E} = -L\frac{dI}{dt} \tag{4.117}$$

But once we consider the two loops, say loop 1 and loop 2, then the scenarios change. We assign a symbol M_{12} to represent the **mutual inductance**, and the magnetic flux can be describe by

$$\Phi_1 = M_{12}I_2 \quad \text{or} \quad \Phi_2 = M_{21}I_1$$

The mutual inductance is given by the **Neumann formula**:

$$M_{21} = \frac{\mu_0}{4\pi}\oint\oint \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r} = M_{12} \tag{4.118}$$

, where r is the distance between two loops.

Armed with the inductance, in effect, we can figure out the energy in magnetic fields. The work per unit time can be related to the emf, and so be related to the inductance:

$$\frac{dW}{dt} = -\mathcal{E}\frac{dQ}{dt} = -\mathcal{E}I = L\frac{dI}{dt}I$$

Therefore, we will find that

$$\boxed{W = \frac{1}{2}LI^2} \tag{4.119}$$

Doing the further derivations, one can find that the energy will be

$$\boxed{W = \frac{1}{2\mu_0}\iiint B^2\, d\tau} \tag{4.120}$$

Please be aware that this form is fairly similar to the Eq (4.39), which is the energy in electric fields.

4.11 Circuit-2: RL, LC, RLC circuits, and AC

In this section, we are going to consider the circuits with inductors. Previously we have not considered the circuits including magnetic fields, but this time we will. First, the inductors have the similar properties to resistors. When we assemble the inductors in series,

$$L_{\text{eq}} = \sum_{i=1}^{n} L_i \quad \text{(in series)} \tag{4.121}$$

, while in parallel it becomes

$$\frac{1}{L_{\text{eq}}} = \sum_{i=1}^{n} \frac{1}{L_i} \quad \text{(in parallel)} \tag{4.122}$$

In case of any confusion, the collection of the results are shown as Figure 64.

	Capacitor	Resistor	Inductor
Series	$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$	$R = R_1 + R_2$	$L = L_1 + L_2$
Parallel	$C = C_1 + C_2$	$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$	$\frac{1}{L} = \frac{1}{L_1} + \frac{1}{L_2}$
Fundamental Formula	$\Delta V = \frac{Q}{C}$	$\Delta V = IR$	$\mathrm{E}_L = -L\frac{dI}{dt}$

Figure 64: The comparison among the capacitors, resistors, and inductors. *Source*: https://slideplayer.com/slide/5257211/.

Recall for a RC circuit, we first wrote down the equation like Eq (4.68) and then solved it to get Eq (4.69). In the similar way, for the **RL circuit**, we have

$$\mathcal{E} - IR - L\frac{dI}{dt} = 0$$

, which can be rearrange as:

$$L\frac{dI}{dt} + RI = \mathcal{E} \quad \text{(rising RL current)} \tag{4.123}$$

Then our goal to solve the equation:

$$I = \frac{\mathcal{E}}{R}(1 - e^{-Rt/L}) \quad \text{(rising RL current)} \tag{4.124}$$

, where we usually introduce the time constant $\tau = L/R$ so that the exponent can be expressed as $-t/\tau$. On the other hand, if the battery is gone, the current will be decreasing. The differential equation turns out to be

$$L\frac{dI}{dt} + RI = 0 \quad \text{(decaying RL current)} \tag{4.125}$$

Solving the equation, we get

$$I = \frac{\mathcal{E}}{R}e^{-Rt/L} = I_0 e^{-Rt/L} \quad \text{(decaying RL current)} \tag{4.126}$$

Note the initial condition gives $I_0 = I(0) = \mathcal{E}/R$. Also, one can use $\tau = L/R$ as a symbol.

Let us starting considering a more interesting case: the **LC circuit**. In this case, we should write down the differential equation by the conservation of energy. The total energy of an LC circuit is

$$W_{\text{total}} = W_L + W_C = \frac{1}{2}LI^2 + \frac{q^2}{2C}$$

The conservation of energy suggests that $dW/dt = 0$:

$$\frac{dW}{dt} = \frac{d}{dt}\left(\frac{1}{2}LI^2 + \frac{q^2}{2C}\right) = LI\frac{dI}{dt} + \frac{q}{C}\frac{dq}{dt} = 0$$

Rearranging the equation, we have

$$L\frac{d^2q}{dt^2} + \frac{1}{C}q = 0 \quad \text{(LC circuit)} \tag{4.127}$$

Solving the differential equation, we found the solution is no longer in an exponential form, but in a sinusoidal form. This means that the charge or the current of an LC circuit is oscillating:

$$q(t) = Q\cos(\omega t + \phi) \quad \text{and} \quad I(t) = -\omega Q \sin(\omega t + \phi) \quad \text{(LC circuit)} \tag{4.128}$$

Here we just used $I = dq/dt$. Taking the second derivation of q, we have

$$\frac{d^2q}{dt^2} = -\omega^2 Q\cos(\omega t + \phi)$$

Then, substituting this into Eq (4.127), we will get the angular frequency

$$\omega = \frac{1}{\sqrt{LC}} \tag{4.129}$$

In the LC circuit, the charge, current, and potential difference will vary sinusoidally with respect to time. Such a phenomenon is called the **electromagnetic oscillations**, as shown in Figure 65. Specifically, the capacitor dominates the electric motion, while the inductor dominates the magnetic motion. We can also derive the electric energy and the magnetic energy within the oscillations:

$$\begin{aligned} W_E &= \frac{Q^2}{2C}\cos^2(\omega t + \phi) \\ W_B &= \frac{Q^2}{2C}\sin^2(\omega t + \phi) \end{aligned} \tag{4.130}$$

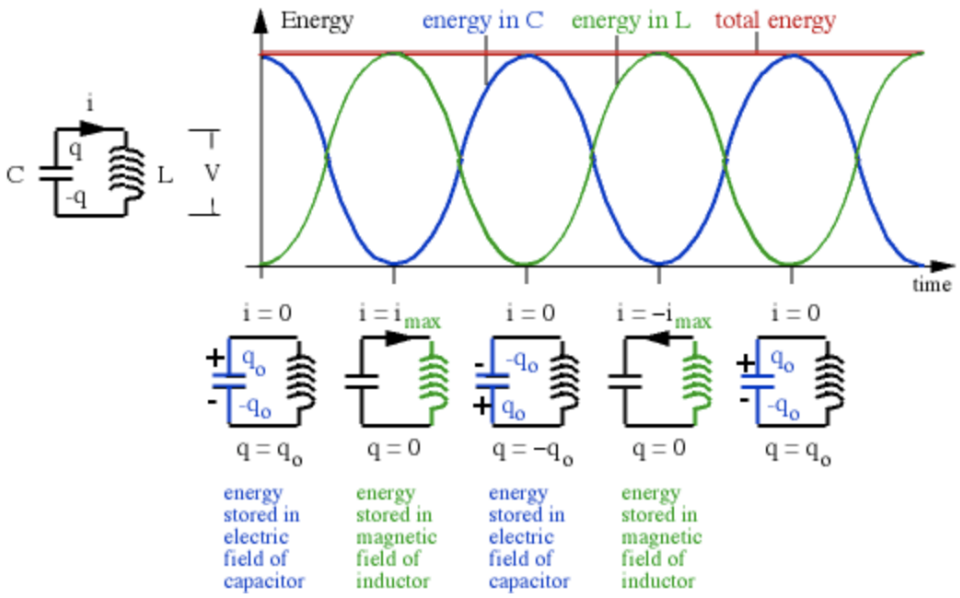

Figure 65: The details of the LC oscillations; different phases are marked in different colors. *Source*: https://www.animations.physics.unsw.edu.au/.

Now, let us consider a more complicated case: the **RLC circuit**. Since we added a resistor, the total energy can be transferred in the form of the thermal energy, which means $dW/dt = -I^2R$:

$$\frac{dW}{dt} = LI\frac{dI}{dt} + \frac{q}{C}\frac{dq}{dt} = -I^2R$$

As a result, the differential equation would be

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = 0 \quad \text{(RLC circuit)} \tag{4.131}$$

In fact, this is analogous to the *damped oscillation* as what we have seen in Section 2.8. Solving the equation gives

$$q(t) = Qe^{-Rt/2L}\cos(\omega' t + \phi) \quad \text{(RLC circuit)} \tag{4.132}$$

, where the damped angular frequency is

$$\omega' = \sqrt{\omega^2 - \left(\frac{R}{2L}\right)^2} = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2} \tag{4.133}$$

Note that ω' is always less than the undamped frequency ω. The Eq (4.132) implies that the charge is not only varying sinusoidally with respect to time, but also decaying exponentially. Indeed, this is the same as what we have seen in a block-spring system.

In effect, the oscillating current is known as the **alternating current (AC)**, which is different from what we introduced in the previous sections, that is, the **direct current (DC)**. Nowadays, most of the electric appliances are supported by the alternating current. One of the best advantages of the AC is that it can be readily transformed to different voltage levels. In a more straightforward view, DC is involved in the electric effect only, while AC is involved in the interaction between both electric and magnetic reactions.

To see the forced (driven) oscillation of the alternating current, one can add an alternating emf

$$\mathcal{E} = \mathcal{E}_{\max}\sin\omega_d t \tag{4.134}$$

, where ω_d is the driven frequency. And, the driven current is

$$I = I_{\max}\sin(\omega_d t - \phi) \tag{4.135}$$

The amplitude $I_{\max}$ will reach the real maximum when $\omega_d \cong \omega$, and such a situation is called the resonance, as what we knew before. To analyze how the forced oscillation happened in a RLC circuit, a convenient way is to decompose it into three parts, and each component is added by an alternating emf $\mathcal{E}$. Here I shall skip the derivations but only show the results.

- **A resistive load**: The voltage and the current amplitude are related by

$$V_R = I_R R \tag{4.136}$$

- **A capacitive load**: The voltage and the current amplitude are related by

$$V_C = I_C X_C \tag{4.137}$$

  , where the capacitive reactance is

$$X_C = \frac{1}{\omega_d C} \tag{4.138}$$

- **An inductive load**: The voltage and the current amplitude are related by

$$V_L = I_L X_L \tag{4.139}$$

  , where the inductive reactance is

$$X_L = \omega_d L \tag{4.140}$$

Let us assemble all of them in series together, as shown in Figure 66. The emf can be rewritten as

$$\mathcal{E} = \mathcal{E}_R + \mathcal{E}_C + \mathcal{E}_L$$

Next, Figure 67 suggests that

$$\mathcal{E}_{\max} = V_R^2 + (V_L - V_C)^2 = (I_{\max}R)^2 + [I_{\max}(X_L - X_C)]^2$$

In this way, the current amplitude $I_{\max}$ can be written as

$$I_{\max} = \frac{\mathcal{E}_{\max}}{\sqrt{R^2 + (X_L - X_C)^2}} = \frac{\mathcal{E}_{\max}}{Z} \tag{4.141}$$

, where the denominator can be defined as the **impedance**:

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{4.142}$$

Just as its name, this is a quantity contains the resistance, capacitive reactance, and inductive reactance.

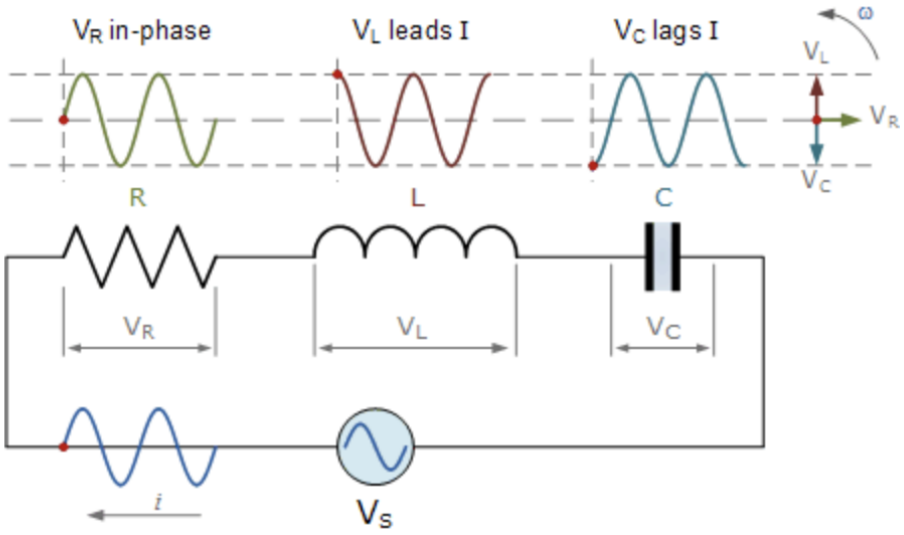

Figure 66: A setup of the series RLC circuit. The situations that the instantaneous voltage across V_R, V_L, and V_C are also shown. *Source*: https://www.electronics-tutorials.ws/

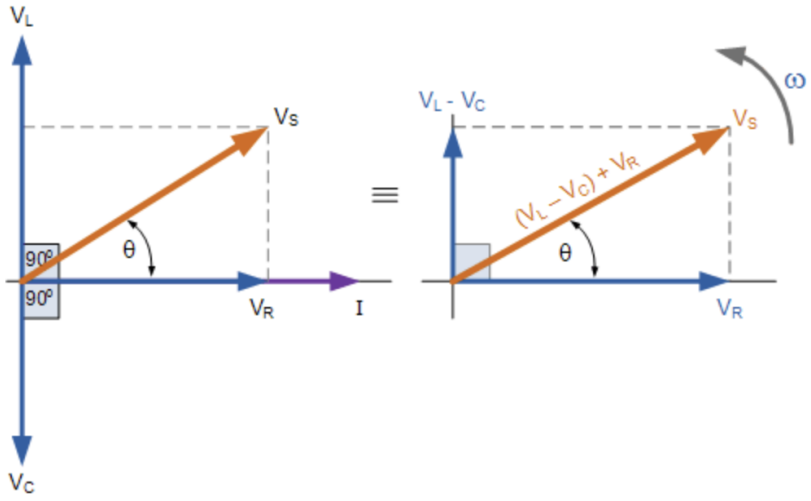

Figure 67: The phasor diagram for the RLC circuit. *Source*: https://www.electronics-tutorials.ws/

Meanwhile, we can define the phase constant (though it is shown as θ in the figure)

$$\phi = \tan^{-1}\left(\frac{X_L - X_C}{R}\right) \tag{4.143}$$

Then, three situations can be discussed:

- When $\phi > 0$, $X_L > X_C$, the circuit is more *inductive* than capacitive.
- When $\phi < 0$, $X_L < X_C$, the circuit is more *capacitive* than inductive.
- When $\phi = 0$, $X_L = X_C$, the circuit is in *resonance*.

In the third scenario, as we mentioned, the resonance appears when

$$\omega_d = \omega = \frac{1}{\sqrt{LC}}$$

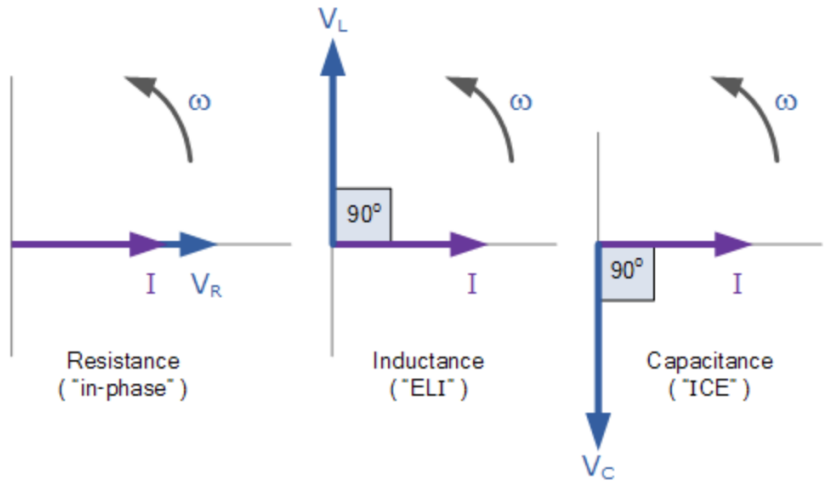

Figure 68: The voltage vectors in three different cases. *Source*: https://www.electronics-tutorials.ws/

The last topic in this section is the **transformer**, which is a significant application of the alternating currents. A simple transformer has the primary coil (p) and the secondary (s) coil. There is a handy relationship among each other:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s} \tag{4.144}$$

As for the equivalent resistance can be described by

$$R_{\text{eq}} = \left(\frac{N_p}{N_s}\right)^2 R \tag{4.145}$$

The principle of a transformer is to transfer the electric power by varying the magnetic field.

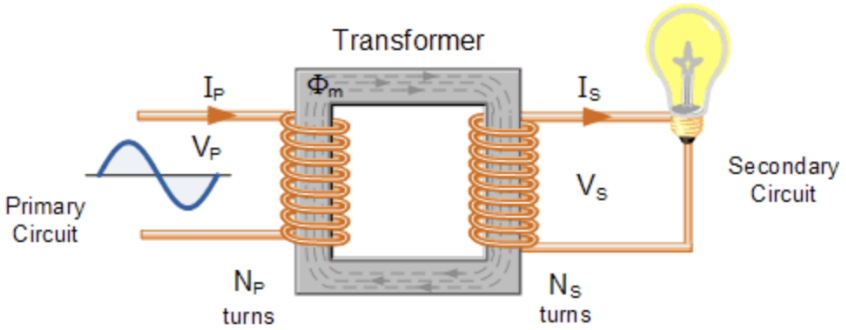

Figure 69: A simple transformer. *Source*: https://www.electronics-tutorials.ws/

4.12 Electromagnetism in matter

So far, we delved into electromagnetism without considering the situations in matter. A first glace might be the topics that we had a short discussion about the electric / magnetic moment. In fact, it is necessary to study the electromagnetic phenomena in matter. Let us start from the molecular scale, when many dipoles has been put in a field, they will point in the same direction as the field's. Such a phenomenon is called the **polarization**, which can be denoted by a capital symbol $\mathbf{P}$. More accurate, $\mathbf{P}$ represents the dipole moment per unit volume. As a reminder, in the section before we mentioned that the torque is

$$\tau = \mathbf{p} \times \mathbf{E} \tag{4.146}$$

As for the force, we can express in a general form:

$$\mathbf{F} = (\mathbf{p} \cdot \nabla)\mathbf{E} \tag{4.147}$$

For a single dipole, the potential can be written as

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p} \cdot \hat{\mathbf{r}}}{r^2}$$

Since $\mathbf{p} = \mathbf{P}d\tau'$, we have

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \iiint \frac{\mathbf{P}(\mathbf{r}') \cdot \hat{\mathbf{r}}}{r^2} \, d\tau'$$

After doing some mathematical tricks, one can verify that

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \oiint \frac{\sigma_b}{r} dA' + \frac{1}{4\pi\epsilon_0} \iiint \frac{\rho_b}{r} d\tau' \tag{4.148}$$

, where the surface charge

$$\sigma_b = \mathbf{P} \cdot \hat{\mathbf{n}} \tag{4.149}$$

and the volume charge

$$\rho_b = -\nabla \cdot \mathbf{P} \tag{4.150}$$

are the quantities of the **bound charges**. They are produced due to the polarization effect. For others they are called the **free charges**, denoted by ρ_f. For a **dielectric**, it is not a conductor but has the effect of the polarization. That is to say, when we put it in a field, the molecules and atoms will automatically align. These can be thought of as the bound charges ρ_b within the dielectric and σ_b on its surface. For the other charges that are not related to the dielectric itself (sometimes come from outside), they are the free charges. Specifically,

$$\rho = \rho_b + \rho_f$$

Now, the Gauss's law tells us that

$$\epsilon_0(\nabla \cdot \mathbf{E}) = \rho = \rho_b + \rho_f = -\nabla \cdot \mathbf{P} + \rho_f$$

Rearranging the equation gives

$$\nabla \cdot (\epsilon_0 \mathbf{E} + \mathbf{P}) = \rho_f$$

In a more concise form, it becomes

$$\boxed{\nabla \cdot \mathbf{D} = \rho_f} \tag{4.151}$$

, where the **electric displacement** is defined as

$$\boxed{\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}} \tag{4.152}$$

This can be viewed as the Gauss's law in matter. As usual, we have a corresponding integral form

$$\oiint \mathbf{D} \cdot d\mathbf{A} = Q_{f,\text{enc}} \tag{4.153}$$

The polarization is proportional to **E**,

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \tag{4.154}$$

, where χ_e is the **electric susceptibility**. Similarly,

$$\mathbf{D} = \epsilon \mathbf{E} \tag{4.155}$$

, where $\epsilon = \epsilon_0(1 + \chi_e)$ is the **permittivity** in matter. Sometimes we define $\epsilon_r = 1 + \chi_e$ as the **dielectric constant** (or **relatively permittivity**). Another property that can be found in a linear dielectric is that

$$\rho_b = -\left(\frac{\chi_e}{1 + \chi_e}\right) \rho_f \tag{4.156}$$

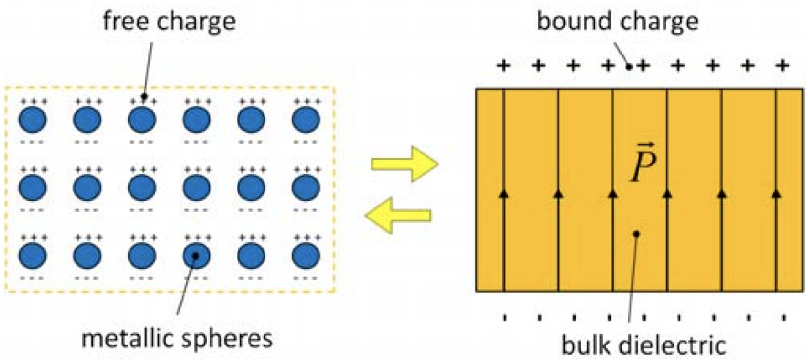

Figure 70: The microscopic (left) and the macroscopic (right) view of a dielectric. *Source*: Progress In Electromagnetics Research B, Vol. 64, 83-101, 2015

A totally similar fashion appear in magnetism in matter. For example, the torque

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B} \tag{4.157}$$

and the force acting on a dipole

$$\mathbf{F} = \nabla(\mathbf{m} \cdot \mathbf{B}) \tag{4.158}$$

The magnetic vector potential can be written as

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \hat{\mathbf{r}}}{r^2}$$

Since $\mathbf{m} = \mathbf{M} d\tau'$, where **M** is the **magnetization** (namely, the magnetic dipole moment per unit volume), we have

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iiint \frac{\mathbf{M}(\mathbf{r}') \times \hat{\mathbf{r}}}{r^2} \, d\tau'$$

After doing some mathematical tricks, one can verify that

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \oiint \frac{\mathbf{K}_b(\mathbf{r}')}{r} dA' + \frac{\mu_0}{4\pi} \iiint \frac{\mathbf{J}_b(\mathbf{r}')}{r} d\tau' \tag{4.159}$$

, where the surface charge

$$\mathbf{K}_b = \mathbf{M} \times \hat{\mathbf{n}} \tag{4.160}$$

and the volume charge

$$\mathbf{J}_b = \nabla \times \mathbf{M} \tag{4.161}$$

are the quantities of the **bound currents**. They are, indeed, analogous to the bound charges σ_b and ρ_b. Therefore, the current in matter can be described as

$$\mathbf{J} = \mathbf{J}_b + \mathbf{J}_f$$

The Ampère's law permits

$$\mathbf{J} = \frac{1}{\mu_0}(\nabla \times \mathbf{B}) = \mathbf{J}_b + \mathbf{J}_f = \nabla \times \mathbf{M} + \mathbf{J}_f$$

Rearrange the quantities, we obtain

$$\nabla \times \left(\frac{1}{\mu_0}\mathbf{B} - \mathbf{M}\right) = \mathbf{J}_f$$

Now, let us write the expression in a more compact way:

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J}_f} \tag{4.162}$$

, where we introduced the so-called **auxiliary field** (H-field):

$$\boxed{\mathbf{H} = \frac{1}{\mu_0}\mathbf{B} - \mathbf{M}} \tag{4.163}$$

This is analogous to the electric displacement $\mathbf{D}$ in electrostatics in matter. For sure, we have the corresponding integral form:

$$\oint \mathbf{H} \cdot d\mathbf{l} = I_{f,\text{enc}} \tag{4.164}$$

The magnetic field can generate the effect of the magnetization within paramagnetic and diamagnetic materials. We should note that the proportional relation

$$\mathbf{M} = \chi_m \mathbf{H} \tag{4.165}$$

, where χ_m is called the **magnetic susceptibility**. Please notice this is proportional to $\mathbf{H}$ instead of $\mathbf{B}$ (which is different from the previous case). For linear media, we have

$$\mathbf{B} = \mu \mathbf{H} \tag{4.166}$$

The **permeability** is $\mu = \mu_0(1 + \chi_m)$, and we thus have

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu_0(1 + \chi_m)\mathbf{H} \tag{4.167}$$

For the bound current it can be written as

$$\mathbf{J}_b = \chi_m \mathbf{J}_f \tag{4.168}$$

4.13 Maxwell's equations

From the beginning to now, we have already four laws that can almost dominate stories in the electromagnetism:

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \quad \text{(Gauss's law)} \\
\nabla \cdot \mathbf{B} &= 0 \quad \text{(Gauss's law for magnetism)} \\
\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \quad \text{(Faraday's law)} \\
\nabla \times \mathbf{B} &= \mu_0 \mathbf{J} \quad \text{(Ampère's law)}
\end{aligned}$$

Everything looks good, yet, one thing brings us attention. First, the divergence of a curl has to zero. Indeed,

$$\nabla \cdot (\nabla \times \mathbf{E}) = \nabla \cdot \left(-\frac{\partial \mathbf{B}}{\partial t}\right) = -\frac{\partial}{\partial t}(\underbrace{\nabla \cdot \mathbf{B}}_{=0}) = 0$$

This is surely correct since the divergence of a magnetic field has to be zero. However, when we do the same trick:

$$\nabla \cdot (\nabla \times \mathbf{B}) = \mu_0(\underbrace{\nabla \cdot \mathbf{J}}_{\neq 0})$$

The LHS is zero, but it has not reason that the RHS has to be zero, except for a steady current. In fact, one can employ the continuity equation (4.51):

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial t}\epsilon_0(\nabla \cdot \mathbf{E}) = -\nabla \cdot \left(\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right)$$

This gives

$$\nabla \cdot \mathbf{J} + \nabla \cdot \left(\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right) = 0$$

In other words, if we replace $\nabla \cdot \mathbf{J}$ with $\nabla \cdot (\mathbf{J} + \epsilon_0 \, \partial \mathbf{E}/\partial t)$, the answer that $\nabla \cdot (\nabla \times \mathbf{B}) = 0$ does make sense. Hence, the Ampère's law has to be modified as

$$\boxed{\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}} \tag{4.169}$$

This is also called the **Ampère-Maxwell law**. The law implies that the changing magnetic field would also induce an electric field. The additional term

$$\mathbf{J}_d = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{4.170}$$

, which can also be written as

$$I_d = \epsilon_0 \frac{d\Phi_E}{dt} \tag{4.171}$$

, is called the **displacement current**. The integral form will be

$$\boxed{\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 I_{\text{enc}} + \iint \left(\mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \cdot\right) dA} \tag{4.172}$$

Hence, we are already able to write down a set of equations that can sufficiently describe all phenomena within electromagnetism:

$$\begin{cases} \nabla \cdot \mathbf{E} = \dfrac{\rho}{\epsilon_0} \quad \text{(Gauss's law)} \\ \nabla \cdot \mathbf{B} = 0 \quad \text{(Gauss's law for magnetism)} \\ \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} \quad \text{(Faraday's law)} \\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} \quad \text{(Ampère-Maxwell law)} \end{cases} \tag{4.173}$$

These are the well-known **Maxwell's equations**. The integral form of the Maxwell's equations is;

$$\begin{cases} \oiint \mathbf{E} \cdot d\mathbf{A} = \dfrac{Q_{\text{enc}}}{\epsilon_0} \quad \text{(Gauss's law)} \\ \oiint \mathbf{B} \cdot d\mathbf{A} = 0 \quad \text{(Gauss's law for magnetism)} \\ \oint \mathbf{E} \cdot d\mathbf{l} = -\dfrac{\partial \Phi_B}{\partial t} \quad \text{(Faraday's law)} \\ \oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 I_{\text{enc}} + \mu_0 \epsilon_0 \dfrac{\partial \Phi_E}{\partial t} \quad \text{(Ampère-Maxwell law)} \end{cases} \tag{4.174}$$

Indeed, Eq (4.173) is a beautiful set of the formulas describing the whole stories about electromagnetism. But, this can be applied to the ideal situations such as vacuum. For matter, we have to consider the bound charge density $\rho_b = -\nabla \cdot \mathbf{P}$ for electric polarization as well as the bound current $\mathbf{J}_b = \nabla \times \mathbf{M}$ for magnetization. It is worth mentioning that the flow of bound charge (say $\mathbf{J}_p$) can be induced when the

electric polarization changes. This convinces us to have a charge density $\sigma_b = P$, and the change in current implies that

$$dI = \frac{\partial \sigma_b}{\partial t} dA = \frac{\partial P}{\partial t} dA$$

, where A is the cross sectional area. Accordingly, the current density is

$$\mathbf{J}_p = \frac{\partial \mathbf{P}}{\partial t} \tag{4.175}$$

This is sometimes called the **polarization current**, which is a motion of charge generated by the change in electric polarization. In turn, the total charge density in matter

$$\rho = \rho_b + \rho_f = -\nabla \cdot \mathbf{P} + \rho_f$$

does not change, whereas the current density should be written in the three parts:

$$\mathbf{J} = \mathbf{J}_b + \mathbf{J}_f + \mathbf{J}_p = \nabla \times \mathbf{M} + \mathbf{J}_f + \frac{\partial \mathbf{P}}{\partial t}$$

This means that the Ampère-Maxwell law has to be changed:

$$\begin{aligned} \nabla \times \mathbf{B} &= \mu_{\mathbf{J}} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \\ &= \mu_0 \left(\nabla \times \mathbf{M} + \mathbf{J}_f + \frac{\partial \mathbf{P}}{\partial t} \right) + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \end{aligned}$$

Recall the definitions of the electric displacement

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$$

and the H-field

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M}$$

, we can then express the Ampère-Maxwell law as the following:

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t}} \tag{4.176}$$

In fact, the additional term here is called the **displacement current**:

$$\mathbf{J}_d = \frac{\partial \mathbf{D}}{\partial t} \tag{4.177}$$

Note the displacement current is not a "current" generated by electric charges. Instead, it is the change in electric displacement $\mathbf{D}$ with respect to time. As a result, the Maxwell's equations in matter are

$$\begin{cases} \nabla \cdot \mathbf{D} = \rho_f & \text{(Gauss's law)} \\ \nabla \cdot \mathbf{B} = 0 & \text{(Gauss's law for magnetism)} \\ \nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t} & \text{(Faraday's law)} \\ \nabla \times \mathbf{H} = \mathbf{J}_f + \dfrac{\partial \mathbf{D}}{\partial t} & \text{(Ampère-Maxwell law)} \end{cases} \tag{4.178}$$

Their integral form is shown as the following:

$$\begin{cases} \oiint \mathbf{D} \cdot d\mathbf{A} = \dfrac{Q_{\text{enc}}}{\epsilon_0} & \text{(Gauss's law)} \\ \oiint \mathbf{B} \cdot d\mathbf{A} = 0 & \text{(Gauss's law for magnetism)} \\ \oint \mathbf{E} \cdot d\mathbf{l} = -\dfrac{d}{dt} \iint \mathbf{B} \cdot d\mathbf{A} & \text{(Faraday's law)} \\ \oint \mathbf{B} \cdot d\mathbf{l} = I_{f,\text{enc}} + \dfrac{d}{dt} \iint \mathbf{D} \cdot d\mathbf{A} & \text{(Ampère-Maxwell law)} \end{cases} \tag{4.179}$$

4.14 Electromagnetic waves

In Eqs (2.124), we showed the general wave equation where the wave speed $v = \sqrt{T/\mu}$ for a string. Also, we showed the possible solution contains

$$u(z,t) = A\sin(kz - \omega t + \delta)$$

with the angular frequency $\omega = 2\pi v = kv$ and the phase constant δ. The Euler's formula (1.5) allows us to have the complex wave function:

$$\tilde{u}(z,t) = \tilde{A}e^{i(kz-\omega t)}$$

where $\tilde{A} = Ae^{i\delta}$ is the complex amplitude. If we take the real part then it is an actual wave function. Given an incident wave

$$\tilde{u}_0(z,t) = \tilde{A}_0 e^{i(kz-\omega t)}$$

, the reflected wave should be in an opposite direction,

$$\tilde{u}_r(z,t) = \tilde{A}_r e^{i(-kz-\omega t)}$$

, while the transmitted wave would have a different property from the incident k (which is now denoted by k'):

$$\tilde{u}_t(z,t) = \tilde{A}_t e^{i(k'z-\omega t)}$$

In general, the electromagnetic wave is a *transverse* wave. The direction can be expressed in vector form:

$$\tilde{\mathbf{u}}(z,t) = \tilde{A}e^{i(kz-\omega t)}\hat{\mathbf{n}} \tag{4.180}$$

Since such a wave propagates in the three dimensions, we have to notice the polarization vector $\hat{\mathbf{n}}$ determines the direction of vibration, while $\hat{\mathbf{z}}$ is the direction of propagation. In other words,

$$\hat{\mathbf{n}} \cdot \hat{\mathbf{z}} = 0$$

is satisfied. The vertical polarization can be denoted by $\hat{\mathbf{x}}$ whereas the horizontal polarization is $\hat{\mathbf{y}}$.

Now, let us start considering the electromagnetic waves. By taking the curl such that $\nabla \times (\nabla \times \mathbf{E})$ and $\nabla \times (\nabla \times \mathbf{B})$, one can find that

$$\nabla^2 \mathbf{E} = \mu_0\epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad \text{and} \quad \nabla^2 \mathbf{B} = \mu_0\epsilon_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} \tag{4.181}$$

In effect, the three dimensional wave equation is

$$\nabla^2 u = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2}$$

Thus, we see that

$$v = c = \frac{1}{\sqrt{\mu_0\epsilon_0}} \tag{4.182}$$

If you evaluate this, surprisingly, you will find that it is equal to 3×10^8 m/s, which is the familiar value of the speed of light c. The result implies that light might be an electromagnetic wave. This is a significant result and further support the explanation that light is a wave. The fields therefore have the following form:

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 e^{i(kz-\omega t)} \quad \text{and} \quad \tilde{\mathbf{B}}(z,t) = \tilde{\mathbf{B}}_0 e^{i(kz-\omega t)} \tag{4.183}$$

where $\omega = ck$. Furthermore, the Faraday's law

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

suggests that

$$-k(\tilde{E}_0)_y = \omega(\tilde{B}_0)_x \quad \text{and} \quad -k(\tilde{E}_0)_x = \omega(\tilde{B}_0)_y$$

In a more compact form, it becomes

$$\tilde{\mathbf{B}}_0 = \frac{k}{\omega}(\hat{\mathbf{z}} \times \tilde{\mathbf{E}}_0) = \frac{1}{c}(\hat{\mathbf{z}} \times \tilde{\mathbf{E}}_0) \tag{4.184}$$

We see that the E and B are perpendicular to each other. Writing in magnitude,

$$B_0 = \frac{1}{c}E_0 \tag{4.185}$$

In other words, the speed of light c connects the E-field and the B-field. Combined with Eq (4.183), we now use $\mathbf{r}$ to represent the positions and $\hat{\mathbf{k}}$ is the propagation vector, the general form of the electromagnetic waves will be

$$\begin{aligned} \tilde{\mathbf{E}}(\mathbf{r}, t) &= \tilde{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\hat{\mathbf{n}} \\ \tilde{\mathbf{B}}(\mathbf{r}, t) &= \frac{1}{c}\tilde{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}(\hat{\mathbf{k}} \times \hat{\mathbf{n}}) = \frac{1}{c}\hat{\mathbf{k}} \times \tilde{\mathbf{E}} \end{aligned} \tag{4.186}$$

where

$$\hat{\mathbf{n}} \cdot \hat{\mathbf{k}} = 0$$

The EM wave's directions are labelled in Figure 71.

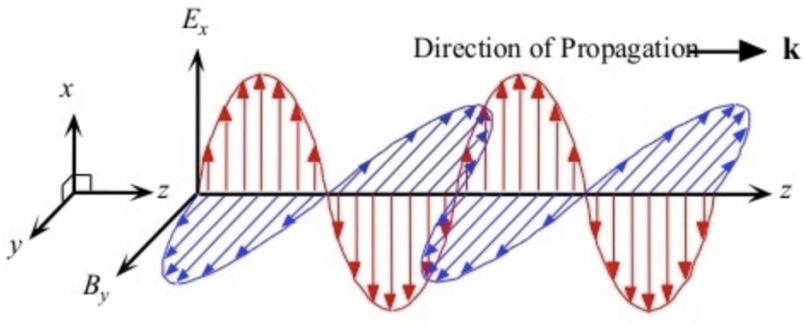

Figure 71: The 3D diagram of the electromagnetic waves. *Source*: S. O. Kasap, 1999, optoelectronics.

Meanwhile, we can calculate the electromagnetic waves in matter. First, the EM wave passing through a plane is shown as Figure 72. We see the incident, reflected, and transmitted electromagnetic waves. The boundary conditions suggest that

$$\tilde{E}_{0,i} + \tilde{E}_{0,r} = \alpha \tilde{E}_{0,t}$$

where

$$\alpha = \frac{\cos\theta_t}{\cos\theta_i}$$

Also,

$$\tilde{E}_{0,i} - \tilde{E}_{0,r} = \beta \tilde{E}_{0,t}$$

where

$$\beta = \frac{\mu_1 v_1}{\mu_2 v_2}$$

These give us the **Fresnel's equation**, which describes the electromagnetic in different matters:

$$\tilde{E}_{0,r} = \left(\frac{\alpha - \beta}{\alpha + \beta}\right)\tilde{E}_{0,i} \quad \text{and} \quad \tilde{E}_{0,t} = \left(\frac{2}{\alpha + \beta}\right)\tilde{E}_{0,i} \tag{4.187}$$

We can then determine the reflection coefficient

$$R = \frac{I_r}{I_i} = \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 \tag{4.188}$$

and the transmission coefficient

$$T = \frac{I_t}{I_i} = \alpha\beta\left(\frac{2}{\alpha+\beta}\right)^2 \tag{4.189}$$

For more details about the refraction and the reflection, we will delve into deeper in the next chapter.

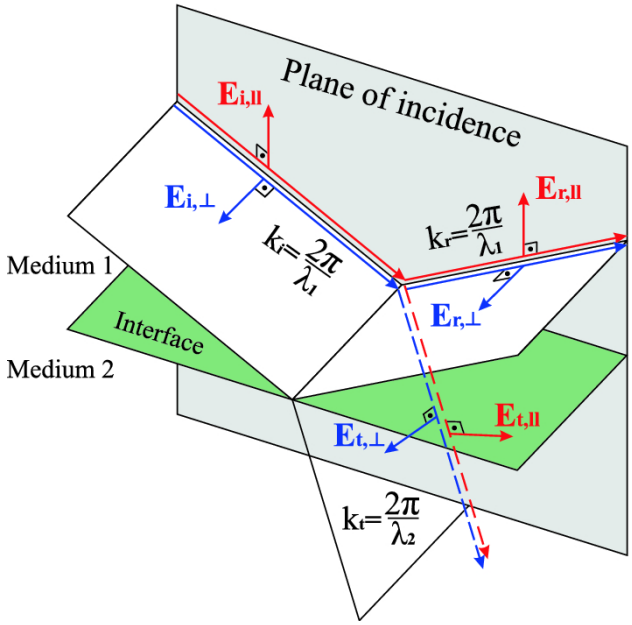

Figure 72: The reflected and the transmitted electromagnetic waves. *Source*: Nickelson L. (2019) Reflection and Transmission of Plane Electromagnetic Waves. Springer, Singapore.

4.15 Radiation

To explore the energy of electromagnetic waves, it is worthwhile to mention a quantity called the **Poynting vector**:

$$\boxed{\mathbf{S} = \frac{1}{\mu_0}(\mathbf{E}\times\mathbf{B})} \tag{4.190}$$

You might be curious about its physical meaning, but I shall explain in the later sections. Armed with our expression of the E-field and the B-field, we can write the Poynting vector as

$$\mathbf{S} = c\epsilon_0 E_0^2 \cos^2(kz - \omega z + \delta)\hat{\mathbf{z}}$$

The **intensity** is equivalent to the $\langle S\rangle$ where we neglect the sinusoidal fluctuations:

$$\boxed{I = \frac{1}{2\mu_0 c}E_0^2 = \frac{\epsilon_0 c}{2}E_0^2} \tag{4.191}$$

This leads to the definition of the **radiation pressure** P, which is $(\Delta P/\Delta t)(1/A)$. The P can be written as

$$P = \frac{1}{2}\epsilon_0 E_0^2 = \frac{I}{c} \tag{4.192}$$

In fact, any point charge q can generate the electromagnetic radiation when it accelerates with a. The effect can be determined by the **Larmor formula**:

$$\boxed{P = \mu_0\frac{q^2a^2}{6\pi c} = \frac{1}{\epsilon_0}\frac{q^2a^2}{6\pi c^3}} \tag{4.193}$$

Assume you are stand on the ground and launch a laser beam (surely, the speed of each photon is c), then you can calculate the radiation power. However, if you move with a velocity $\mathbf{v}$ and under the relativistic situation, the Larmor formula has to be modified. This is determined by Liénard's generalization:

$$P = \frac{\mu_0 \gamma^6 q^2}{6\pi c}\left(a^2 - \left|\frac{\mathbf{v}\times\mathbf{a}}{c}\right|^2\right) \tag{4.194}$$

, where $\gamma = 1/\sqrt{1 - v^2/c^2}$ is a factor in relativity and we will discuss in the later chapter.

At the same time, a charged particle not only radiates but also exerts a force. The reason is that the radiation carries away the energy, which eventually converts into the kinetic energy acting on the charge itself. Therefore, this is sometimes called the radiation reaction. The force is determined by the following equation:

$$\boxed{\mathbf{F} = \frac{\mu q^2}{6\pi c}\dot{\mathbf{a}} = \frac{q^2}{6\pi\epsilon_0 c^3}\dot{\mathbf{a}}} \tag{4.195}$$

This is, formally, called the **Abraham-Lorentz force**. The equation is based on the derivation from the Larmor formula, but we skipped the details here. In case that you feel bored at this point, I would like to mention some strange phenomena based on the Abraham-Lorentz force. The Newton's second law suggets that

$$ma = m\tau\dot{a} = \frac{\mu_0 q^2}{6\pi c}\dot{a}$$

, where the time is

$$\tau = \frac{\mu_0 q^2}{6\pi m c} \tag{4.196}$$

Solving the force's equation by separating variables, we find that

$$a(t) = a(0)e^{t/\tau}$$

This looks nothing special. Yet, for an electron, the time τ is only within 6×10^{-24} s, which is approximately the time that light passing across the size of an electron. The derivation above does not consider any external force, but the acceleration is nonzero. Once we push very slightly on an electron, it will accelerate exponentially and drastically. There is no reason where the drastically increasing speed comes from, so the result is obviously contradictory against the conservation law of energy. To deal with the issue, why don't we modify the time interval from $[0, t]$ to $[t, t']$? Let's try it. We obtain:

$$a(t) = a(t')e^{(t-t')/\tau}$$

Indeed, when $t' \to \infty$, $a(t) = 0$. This means that without external forces, the electron will decelerate until zero. Although this does not violate the conservation law, a more contradictory implication appears: It implies that the acceleration in the future $a(t')$ directly affects the acceleration at a time beforehand, that is, $a(t)$! This challenges the causality. Such an "acausal preacceleration" leaves a problem in our classical electrodynamics (though, it is a "classical" problem), which has not yet been effectively solved until the present day.

4.16 Conservation laws

We have already seen how electromagnetism is constructed. In this section, we are going to analyze the conservation laws entail. First, which is the simplest one – the conservation of charge. The charge as a function of time is defined by

$$Q(t) = \iiint \rho(\mathbf{r}, t)d\tau \tag{4.197}$$

As for the current is defined by Eq (4.50):

$$I(t) = \frac{dQ}{dt} = -\oiint \mathbf{J}\cdot d\mathbf{A}$$

Note the current flows outward so we here put a minus sign ahead. Recall the divergence theorem (1.14) tells us that

$$-\oiint \mathbf{J} \cdot d\mathbf{A} = -\iiint (\nabla \cdot \mathbf{J}) d\tau$$

To write down the equation, we just need to take the derivative with respect to time for Eq (4.197). Relate two equations together, we simply get:

$$\iiint \left(\frac{\partial \rho}{\partial t}\right) d\tau = -\iiint (\nabla \cdot \mathbf{J}) d\tau$$

This gets our familiar result, the continuity equation:

$$\boxed{\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}} \tag{4.198}$$

Indeed, there is nothing special from our derivation of Eq (4.51). What does this equation tells? Imagine a closed surface with some charge density, the rate of the charge density decreases would be equal to the corresponding current density flows outward the surface. Therefore, this equation can be viewed as the conservation of charge.

For the conservation law of electromagnetic energy, it has to be involved in deeper derivations. We showed the electric and the magnetic energy in Eqs (4.39) and (4.120). The total energy per unit volume stored in an electromagnetic field will be

$$u = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2\right) \tag{4.199}$$

The work done on a charge can be mathematically expressed as

$$\mathbf{F} \cdot d\mathbf{l} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}\, dt$$

Recall that $\mathbf{v} \times \mathbf{v} = 0$, so we can throw away the $\mathbf{v} \times \mathbf{B}$ term. Besides, we should apply $q = \rho\, d\tau$ and then $\rho \mathbf{v} = \mathbf{J}$. That is to say,

$$q\mathbf{E} \cdot \mathbf{v}\, dt = \mathbf{E} \cdot \mathbf{J}\, d\tau\, dt$$

Rearranging the expression gives the work done within a volume:

$$\frac{dW}{dt} = \iiint (\mathbf{E} \cdot \mathbf{J}) d\tau$$

Using a series of mathematical tricks, we will ultimately obtain the following equation:

$$\frac{dW}{dt} = -\frac{d}{dt} \iiint \frac{1}{2}\left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2\right) d\tau - \frac{1}{\mu_0} \oiint (\mathbf{E} \times \mathbf{B}) \cdot d\mathbf{A} \tag{4.200}$$

We notice there are two integrals on the RHS. The first integral term is simply the total energy in the electromagnetic fields; the second term represents the rate of the energy transported outward a closed space across its boundary. Using the Poynting vector, the equation will be in a more compact form:

$$\boxed{\frac{dW}{dt} = -\frac{d}{dt} \iiint u\, d\tau - \oiint \mathbf{S} \cdot d\mathbf{A}} \tag{4.201}$$

This is the integral form of the **Poynting theorem**, which states that the work done on a charge is equal to the decrease in energy remaining within the fields. If there's no charge, the work done would be zero, so that

$$\iiint \left(\frac{\partial u}{\partial t}\right) d\tau = -\oiint \mathbf{S} \cdot d\mathbf{A} = -\iiint (\nabla \cdot \mathbf{S}) d\tau$$

Hence, it gives

$$\boxed{\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{S}} \tag{4.202}$$

, which is the continuity equation for electromagnetic energy.

Now, let us take a look at the conservation of momentum. This case is slightly more complicated than the previous one. Nevertheless, I won't go through too many derivations. Starting from the electromagnetic force, we have

$$\mathbf{F} = \iiint (\mathbf{E} + \mathbf{v} \times \mathbf{B})\rho \, d\tau = \iiint (\underbrace{\rho\mathbf{E} + \mathbf{J} \times \mathbf{B}}_{=\mathbf{f}}) d\tau$$

So the force per unit volume is

$$\mathbf{f} = \rho\mathbf{E} + \mathbf{J} \times \mathbf{B}$$

Again, doing many mathematical tricks lets us find that

$$\begin{aligned}\mathbf{f} =& \epsilon_0[(\nabla \cdot \mathbf{E})\mathbf{E} + (\mathbf{E} \cdot \nabla)\mathbf{E}] + \frac{1}{\mu_0}[(\nabla \cdot \mathbf{B})\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{B}] \\ & - \frac{1}{2}\nabla\left(\epsilon_0 E^2 + \frac{1}{\mu_0}B^2\right) - \epsilon_0\frac{\partial}{\partial t}(\mathbf{E} \times \mathbf{B})\end{aligned}$$

This looks fairly messy. Fortunately, for the last term we have the Poynting vector so that $\mathbf{E} \times \mathbf{B} = \mu_0\mathbf{S}$. As for the first three terms, one can introduce the **Maxwell stress tensor** to simply them:

$$T_{ij} \equiv \epsilon_0\left(E_iE_j - \frac{1}{2}\delta_{ij}E^2\right) + \frac{1}{\mu_0}\left(B_iB_j - \frac{1}{2}\delta_{ij}B^2\right) \tag{4.203}$$

, where the **Kronecker delta** notation is

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{4.204}$$

By using the tensor symbol, the Maxwell stress tensor can be written as $\overleftrightarrow{\mathbf{T}}$. Taking the divergence for its jth component, We will find that

$$(\nabla \cdot \overleftrightarrow{\mathbf{T}})_j = \epsilon_0\left[(\nabla \cdot \mathbf{E})E_j + (\mathbf{E} \cdot \nabla)E_j - \frac{1}{2}\nabla_j E^2\right] + \frac{1}{\mu_0}\left[(\nabla \cdot \mathbf{B})B_j + (\mathbf{B} \cdot \nabla)B_j - \frac{1}{2}\nabla_j B^2\right]$$

Accordingly, the force per unit volume will be written as

$$\mathbf{f} = \nabla \cdot \overleftrightarrow{\mathbf{T}} - \mu_0\epsilon_0\frac{\partial \mathbf{S}}{\partial t} \tag{4.205}$$

Taking the integral, the electromagnetic force can be determined formally:

$$\mathbf{F} = \oiint \overleftrightarrow{\mathbf{T}} \cdot d\mathbf{A} - \mu_0\epsilon_0\frac{d}{dt}\iiint \mathbf{S} \, d\tau \tag{4.206}$$

In the static case it reduces to

$$\mathbf{F} = \oiint \overleftrightarrow{\mathbf{T}} \cdot d\mathbf{A} \tag{4.207}$$

Let me simply explain the physical meaning of the Maxwell stress tensor. As its name shows, $\overleftrightarrow{\mathbf{T}}$ represents the force per unit area which is actually the stress. From its components, T_{ij}, for $i = j$ it represents the pressure, whereas for $i \neq j$ it represents the shear. Now we can try to verify the conservation law of momentum. Newton's law tells us that $\mathbf{F} = d\mathbf{p}/dt$, so that

$$\boxed{\frac{d\mathbf{p}}{dt} = \oiint \overleftrightarrow{\mathbf{T}} \cdot d\mathbf{A} - \mu_0\epsilon_0\frac{d}{dt}\iiint \mathbf{S} \, d\tau}$$

This is almost the statement of the conservation of momentum, that is, if the momentum increases, the field momentum (the first integral term) would decrease, otherwise the fields would carry momentum into the volume enclosed by a surface. We can further introduce the momentum density

$$\mathbf{g} = \mu_0\epsilon_0\mathbf{S} = \epsilon_0(\mathbf{E} \times \mathbf{B}) \tag{4.208}$$

As what we did for energy's case, if the momentum does not change, $d\mathbf{p}/dt = 0$, and then

$$\iiint \left(\frac{\partial \mathbf{g}}{\partial t}\right) d\tau = \oiint \overleftrightarrow{\mathbf{T}} \cdot d\mathbf{A} = \iiint (\nabla \cdot \overleftrightarrow{\mathbf{T}}) d\tau$$

Thus, we obtain

$$\boxed{\frac{\partial \mathbf{g}}{\partial t} = \nabla \cdot \overleftrightarrow{\mathbf{T}}} \tag{4.209}$$

, which can be thought of as the continuity equation for electromagnetic momentum.

4.17 Potential and fields

As the last section about electrodynamics, I shall introduce some advanced concepts regarding the potentials and fields. For those who are aspiring to study QFT or theoretical physics in the future, I believe you might want to navigate more details in this section.

We have shown that the electric field is the gradient of a scalar potential:

$$\mathbf{E} = -\nabla V$$

, and the magnetic field is defined by the curl of a vector potential:

$$\mathbf{B} = \nabla \times \mathbf{A}$$

Undoubtedly, these are held under mathematical rules, such as Eq (1.10):

$$\nabla \cdot (-\nabla V) = 0 \quad \text{and} \quad \nabla \cdot (\nabla \times \mathbf{A}) = 0$$

Now, we would like to focus on V and $\mathbf{A}$, and consider the generalization of time-dependent case. Looking at the Maxwell's equations (4.173), the Faraday's law seemingly contains the time's factor. We can simply replace $\mathbf{B}$ by $\nabla \times \mathbf{A}$:

$$\nabla \times \mathbf{E} = -\nabla \times \frac{\partial \mathbf{A}}{\partial t}$$

This gives

$$\nabla \times \underbrace{\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right)}_{=-\nabla V} = 0$$

, where we knew the mathematical rule allows us to replace the term in parenthesis with $-\nabla V$. Therefore, the electric field becomes

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \tag{4.210}$$

This is sufficient to describe the second and the third lines of our Maxwell's equations. But this is not perfect enough! To include the Gauss's law, we can take the divergence $\nabla \cdot \mathbf{E}$ and get:

$$\nabla^2 V + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -\frac{\rho}{\epsilon_0} \tag{4.211}$$

To include the fourth line (the Ampère-Maxwell law), we can take the curl $\nabla \times (\nabla \times \mathbf{A})$ on the LHS and substitute Eq (4.210) for $\mathbf{E}$. Thus, we obtain

$$\nabla \times (\nabla \times \mathbf{A}) = \mu_0 \mathbf{J} - \mu_0 \epsilon_0 \left[\nabla \left(\frac{\partial V}{\partial t}\right) + \frac{\partial^2 \mathbf{A}}{\partial t^2}\right]$$

For the left-hand side, one can employ the third line of Eq (1.10). Doing some rearrangements, we obtain

$$\left(\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2}\right) - \nabla \left(\nabla \cdot \mathbf{A} + \mu_0 \epsilon_0 \frac{\partial V}{\partial t}\right) = -\mu_0 \mathbf{J} \tag{4.212}$$

Indeed, we can conclude that Eqs (4.211) and (4.212) contain all the information of our Maxwell's equations. Yet, we would like to write them in a tidier form.

To achieve this, we should introduce a new concept called the **gauge transformation**. Note this is fairly common and important in the further theories such as QED, string theory, etc. Two classic examples are

$$\begin{aligned} V &\to V' = V - \frac{\partial \varphi}{\partial t} \\ \mathbf{A} &\to \mathbf{A}' = \mathbf{A} + \nabla \varphi \end{aligned} \tag{4.213}$$

, where φ is some scalar function. Theoretically, once we do such a transformation in calculations, the results remain the same. Let's try it. For the magnetic field,

$$\begin{aligned} \mathbf{B}' &= \nabla \times \mathbf{A}' \\ &= \nabla \times (\mathbf{A} + \nabla \varphi) \\ &= \nabla \times \mathbf{A} + \nabla \times (\nabla \varphi) \\ &= \nabla \times \mathbf{A} \\ &= \mathbf{B} \end{aligned}$$

As for the electric field, we have

$$\begin{aligned} \mathbf{E}' &= -\nabla V' - \frac{\partial \mathbf{A}'}{\partial t} \\ &= -\nabla \left(V - \frac{\partial \varphi}{\partial t} \right) - \frac{\partial}{\partial t}(\mathbf{A} + \nabla \varphi) \\ &= -\nabla V + \frac{\partial}{\partial t}\nabla \varphi - \frac{\partial}{\partial t}\mathbf{A} - \frac{\partial}{\partial t}\nabla \varphi \\ &= -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \\ &= \mathbf{E} \end{aligned}$$

We see that both fields are preserved under the gauge transformation. Therefore, we conclude that the magnetic field and the electric field have the **gauge invariance**.

There are two famous examples in electrodynamics for the **gauge fixing**. The first one is the **Coulomb gauge**, which requires

$$\boxed{\nabla \cdot \mathbf{A} = 0} \tag{4.214}$$

Using this condition into Eq (4.211) gives

$$\nabla^2 V = -\frac{\rho}{\epsilon_0} \tag{4.215}$$

, which is exactly our familiar Poisson's equation. Taking an integral by setting $V = 0$ at infinity, we obtain

$$V(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0} \iiint \frac{\rho(\mathbf{r}', t)}{\imath} d\tau' \tag{4.216}$$

, where $\imath = |\mathbf{r} - \mathbf{r}'|$ as before. This looks almost like the same as Eq (4.26). However, we are now considering the time-dependent situation. The expression above indicates that the change of our charge at $\mathbf{r}'$ will immediately affect the potential at $\mathbf{r}$. This does really matter if the distance is considered as large enough: when you change a certain charge density on Earth, its potential will respond immediately at the other side of our galaxy, whatever how the change small is. This implies that the faster-than-light response might violate the causality! Hence, physics demands that the scalar potential is not physically measurable, and the information is carried by the E-field. We see that Eq (4.215) is a good expression for us, and it is an advantage. But for the vector potential, we have

$$\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} + \mu_0 \epsilon_0 \nabla \left(\frac{\partial V}{\partial t} \right)$$

This looks difficult to calculate.

Therefore, we have the second choice, that is, the **Lorenz gauge**:

$$\boxed{\nabla \cdot \mathbf{A} = -\mu_0 \epsilon_0 \frac{\partial V}{\partial t}} \tag{4.217}$$

Substitute into Eqs (4.211) and (4.212), we get

$$\nabla^2 V - \mu_0 \epsilon_0 \frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_0} \tag{4.218}$$

and

$$\nabla^2 \mathbf{A} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \tag{4.219}$$

, respectively. These both expressions look better for us to calculate. One can introduce a new differential operator,

$$\Box^2 \equiv \nabla^2 - \mu_0 \epsilon_0 \frac{\partial^2}{\partial t^2} \tag{4.220}$$

, which is called the **d'Alembertian**. Armed with this, we have a set of concise equations as the following:

$$\boxed{\Box^2 V = -\frac{\rho}{\epsilon_0}} \tag{4.221}$$

and

$$\boxed{\Box^2 \mathbf{A} = -\mu_0 \mathbf{J}} \tag{4.222}$$

In fact, the d'Alembertian has included the three-dimensional coordinates (based on the Laplacian ∇) as well as the time component. This implies that the expressions above can be viewed as the four-dimensional version of the Maxwell's equations (or formally, under the Minkowski spacetime). So far we have talked about almost all the topics that should be covered in undergraduate lectures. After learning the special relativity, we will be able to develop a more concise way to include the Maxwell's equations into a beautiful equation.

5 Optics

We have seen the physical properties of light, which is an electromagnetic wave. In this chapter, we would like to explore more about light. Several physical phenomena are involved in optics. Formally, the first two sections are the topics about **geometrical optics**, as for the last three sections are **physical optics**.

5.1 Reflection and refraction

The **law of reflection** describes the angle of incidence θ_1 has to be equal to the angle of reflection θ_1',

$$\theta_1 = \theta_1' \tag{5.1}$$

In addition, the ray of incidence, the ray of reflection, and the normal have to share the same plane. As for the **law of refraction**, it is involved in different medium. Light in vacuum has a constant speed $v = c$, while it would be slower in glass, water, or some other materials. Therefore, this leads to the different index of refraction, which is noted by n, defined by

$$n = \frac{c}{v}$$

, where v is the speed of light in a medium. In vacuum, $n = 1$, and typically $n_{\text{vacuum}} < n_{\text{air}} < n_{\text{water}} < n_{\text{glass}}$. The **Snell's law** suggests that

$$\boxed{n_1 \sin\theta_1 = n_2 \sin\theta_2} \tag{5.2}$$

where n_1 can be thought of the index of refraction in air, whereas n_2 represents the index of refraction in water. All of these are shown in Figure 73.

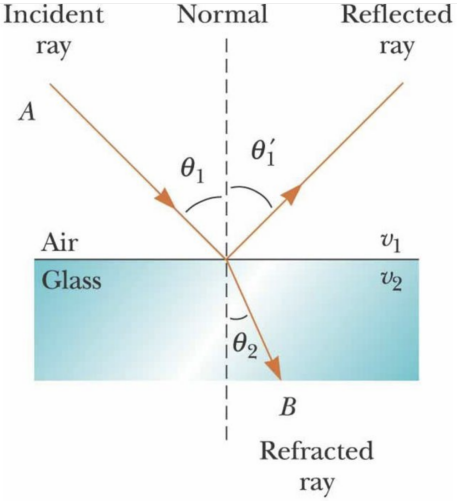

Figure 73: The law of reflection and refraction. *Source*: 2004, Thomson - Brooks/Cole.

Now, if we increase the angle of incidence, the Snell's law follows that the angle of refraction will be close to the maximum where $\theta_c = 90°$. It means that the critical angle θ_c happened when

$$n_1 \sin\theta_c = n_2 \sin 90°$$

Once the incident angle is larger than the critical angle,

$$\theta > \theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \tag{5.3}$$

, then the **total internal reflection** appears. In this situation, the refractive light will tend to reflective so that light will be trapped inside, as shown in Figure 74.

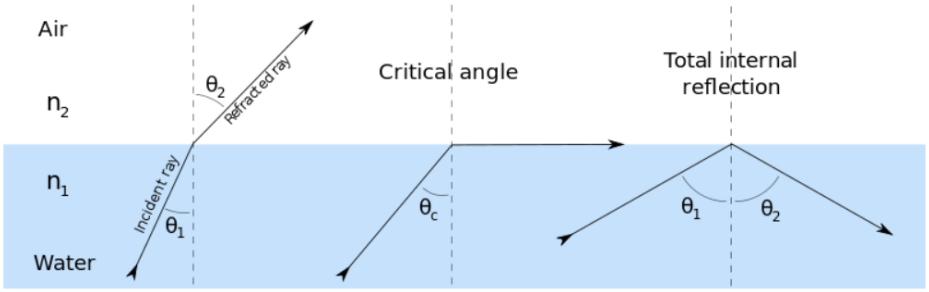

Figure 74: The total internal reflection. *Source*: https://van.physics.illinois.edu/qa/.

One important example is the **optical fiber**, which takes advantage of the total internal reflection to ensure the light's propagation (Figure 75 top). This application greatly influence the construction of Internet, which enables our information transported from one place to another place. Another example is the **mirage**. The cool air has the larger density than the hot air, so that it has a larger refractive index. For example, when you drive a car on the highway in summer, the road has a higher temperature so that the density is smaller. The sunlight from the sky to the ground is then analogous to the light from water to the air. When light reaches the hot air near the ground, the refractive index becomes smaller so that the refractive angle would be larger. Once the refractive angle is large enough, the total internal reflection happens, and you will see the image of the sky is "on the ground" (Figure 75 bottom). Your brain realizes that the sky would not possible appears on the ground, therefore your understanding will be: there must be a puddle of water over there. But when you drive toward you will find nothing there! This is the mirage, which often creates an illusion such as the water in a desert.

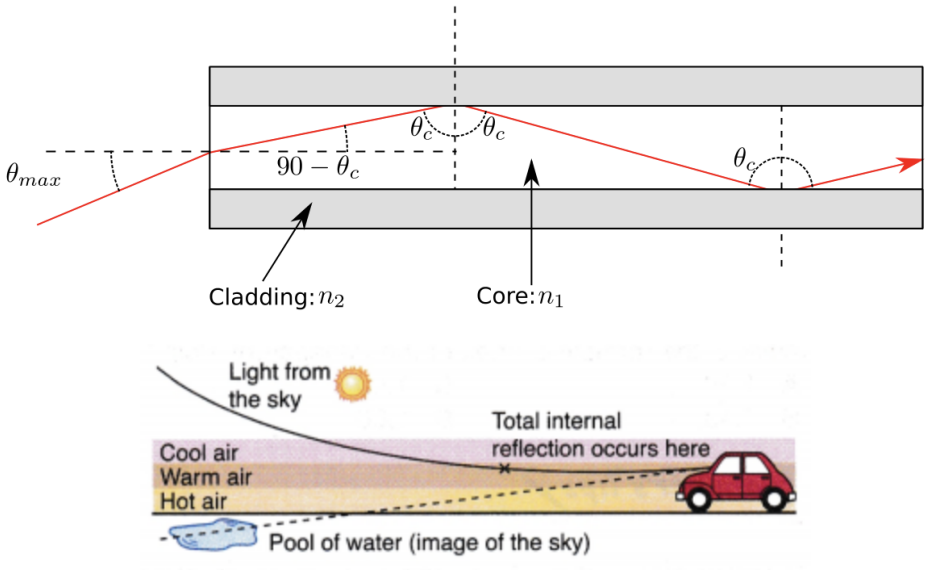

Figure 75: The internal structure of an optical fiber (top) and the illustration of a mirage appears on the highway (bottom). *Source*: https://www.aplustopper.com/total-internal-reflection/.

The other application about the reflection is a sunglasses. This is involved in the polarization of light. Our knowledge about electromagnetism indicates that light will polarize. Typically, an incident light is unpolarized, but when it has been reflected, it would be a polarized ray; as for the refractive ray, it would also be slightly polarized. Note the reflective and the refractive would be perpendicular to each other. The

corresponding incident angle is called the **Brewster angle** (θ_B), which satisfies

$$n_1 \sin\theta_B = n_2 \sin\theta_r = n_2 \sin(90^\circ - \theta_B) = n_2 \cos\theta_B$$

Thus, the Brewster's angle is

$$\theta_B = \tan^{-1}\frac{n_2}{n_1} \tag{5.4}$$

These can be used to manufacture the sunglasses or other improvements in photography.

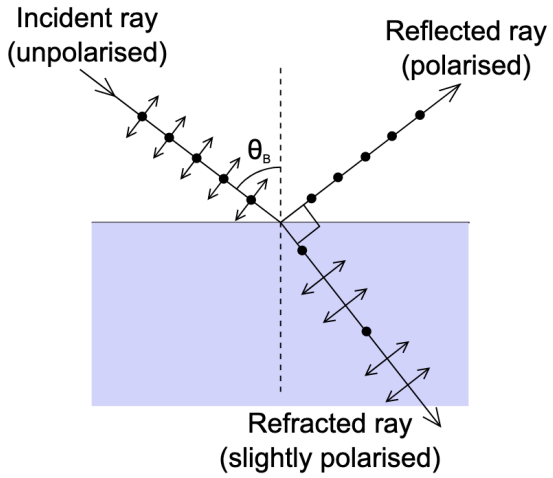

Figure 76: An unpolarized ray incident with the Brewster's angle results in a polarized reflective ray, which is perpendicular to its refractive ray with a slight polarization. *Source*: Wikipedia.

5.2 Images and lens

The simplest case is a plane mirror. All of the incident rays obey the law of reflection. There are two types of images. The first one is just as the movie projected on a screen. You can specifically "see" the image located at a place it should be there. This is called the **real image**. But in some cases, for example, when you are looking at a plane mirror and your partner stands behind you, you will see your partner stands "behind" the mirror in front of you. This is called the **virtual image** because the image does not appear at a right place. A daily example is the tall of a full-length mirror. As shown in Figure 77, all the rays obey the law of reflection. To see the total length of yourself, a mirror must be able to show an image include from your head to your toes. And for sure, the reflective rays must go into your eyes. Using the law of reflection (where I suggest you draw a normal line between the rays), we can calculate the size of the mirror in this case at least has to be

$$\frac{1}{2}(0.1) + \frac{1}{2}(1.5) = 0.8\text{m}$$

But for the **spherical mirrors**, things become more complicated. First, to define the spherical mirror is caved in our flexed out, we should recognize that a spherical mirror have the center of curvature C. A spherical mirror can be viewed as a segment of a sphere. For sure, it has the corresponding radius, called the radius of curvature R. Besides, when the parallel incident rays (which are parallel to the principal axis) come to a concave mirror, for example, it will converge at the **focus** F. The distance from the center of mirror to the focus is called the **focal length** f, which is a half of the radius of curvature:

$$f = \frac{R}{2}$$

That is, $R = 2f$. This is shown in Figure 78. Furthermore, a useful way to judge if an image is real or virtual is to see the image's formed direction. For the real image it will be inverted, while for the virtual image it will be upright. These rules are almost held in any cases.

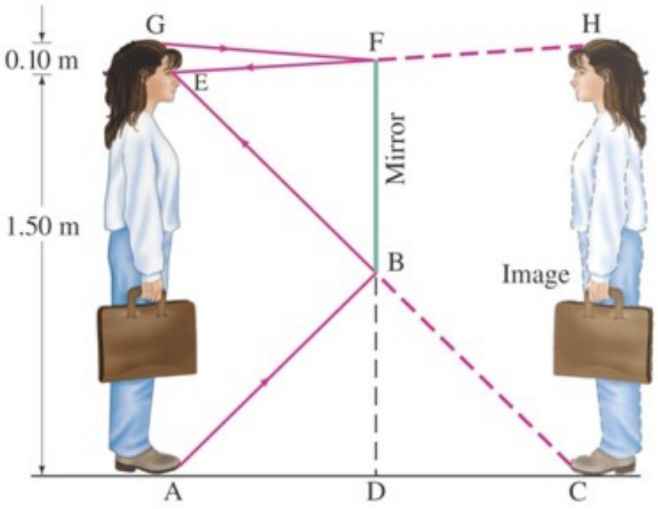

Figure 77: The ray diagram of a full-length mirror. *Source*: https://slideplayer.com/slide/4531026/.

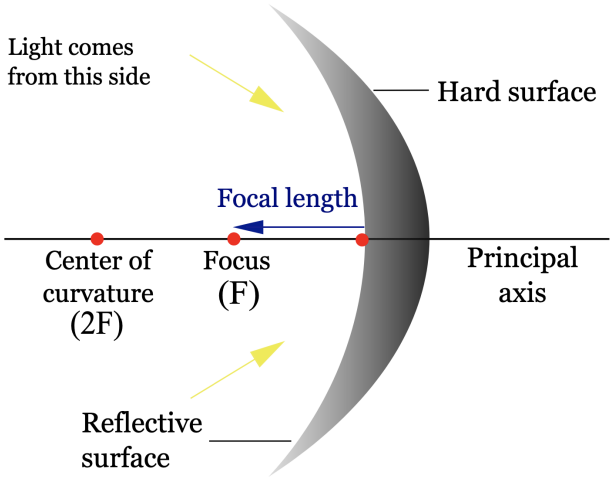

Figure 78: The optical structure of a concave mirror. *Source*: https://slideplayer.com/slide/4531026/.

Too see what location the image will be formed, we should draw the ray diagrams. Step by step, the first step is to sketch a mirror segment, a central axis, and the position of an object (typically, to see how the orientation changes, we like to draw an arrow). Secondly, we can draw a ray which is parallel to the central axis. Then, the ray has to be reflected, here we have the two cases: for a **concave mirror**, it must directly pass through the focus; for a **convex mirror**, it will reflect outward but the ray can be extended to the focus on the other side. Next step, we should draw another ray which directly pass through the focus and reflect in a parallel direction to the central axis. Finally, we can extend the rays we drew, and the image would be located at the intersection between the rays (note the bottom of the object always adhere to the central axis). Although these sound complicated, I have attached Figure 79 for you to refer to. You will find that for a concave mirror the image tends to be larger when we move an object closer to the mirror. As for the convex mirror, the result is always an upright, smaller virtual image. Therefore, the locations of the image depend on the positions of the object (d_o represents the distance between the object and the mirror; d_i represents the distance between the image and the mirror):

Table 6: The image types and positions generated by the spherical mirrors.

Mirror	d_o	d_i	Orientation	Size	Type
Concave	$d_o = \infty$	$d_i = f$	Inverted	Smaller	Real
	$\infty > d_o > 2f$	$f < d_i < 2f$			
	$d_o = 2f$	$d_i = 2f$		Same	
	$2f > d_o > f$	$d_i > 2f$		Larger	
	$d_o = f$	$d_i = \infty$	No image	No image	No image
	$d_o < f$	Behind mirror	Upright	Larger	Virtual
Convex	Anywhere			Smaller	

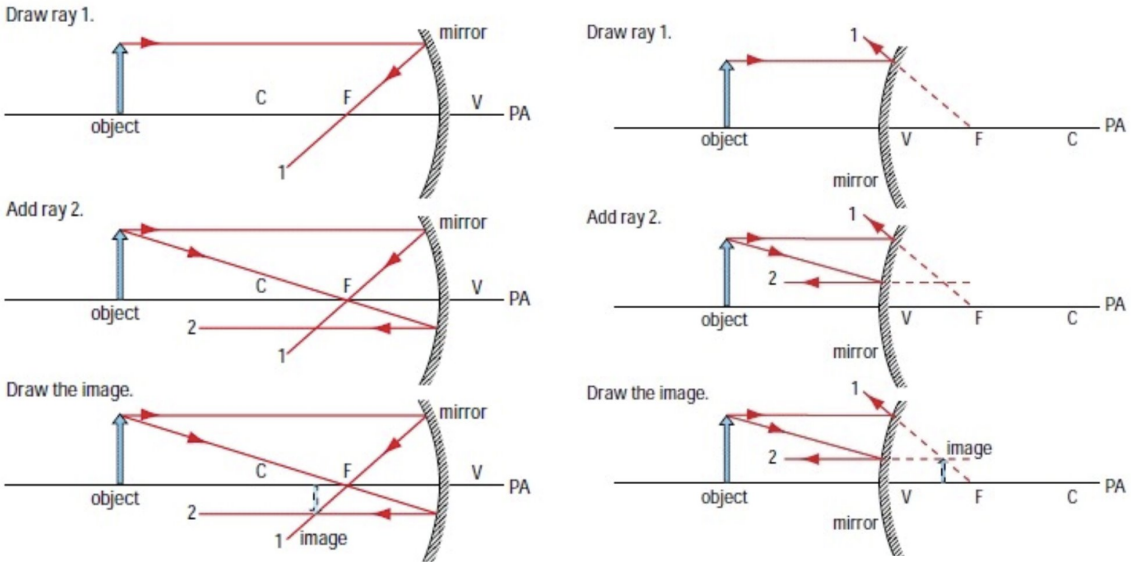

Figure 79: The steps of drawing a ray diagram for the concave mirror (left) and the convex mirror (right). *Source*: https://venngage.net/p/113967/curved-mirrors.

We can determine the magnification by the d_o and d_i:

$$M = \frac{h_i}{h_o} = -\frac{d_i}{d_o} \tag{5.5}$$

Note the if $m > 0$ the image will be in the same orientation, while if $m < 0$ it will be in the opposite orientation.

Now, we have already seen the mirrors. But in our daily applications, the thin lenses might be more common. The convex lens is also called the **converging lens**, whereas the concave lens is also called the **diverging lens**. From their names we can recognize that the former can converge the light ray while the latter one con diverge the light ray. For a converging lens, we have to sketch two rays: one has to be parallel to the central axis and then pass through the focus behind the lens, and the other has to directly pass through the center of lens. In some cases we will need to extend the rays, and such cases the image will be formed on the same side as the object. As for the diverging lens, there will be the one ray passes through the lens with a diverging direction, and the one directly passes through the center of lens. In any case of the diverging lens, the results must be the upright, smaller virtual image, which locates at the same side of lens.

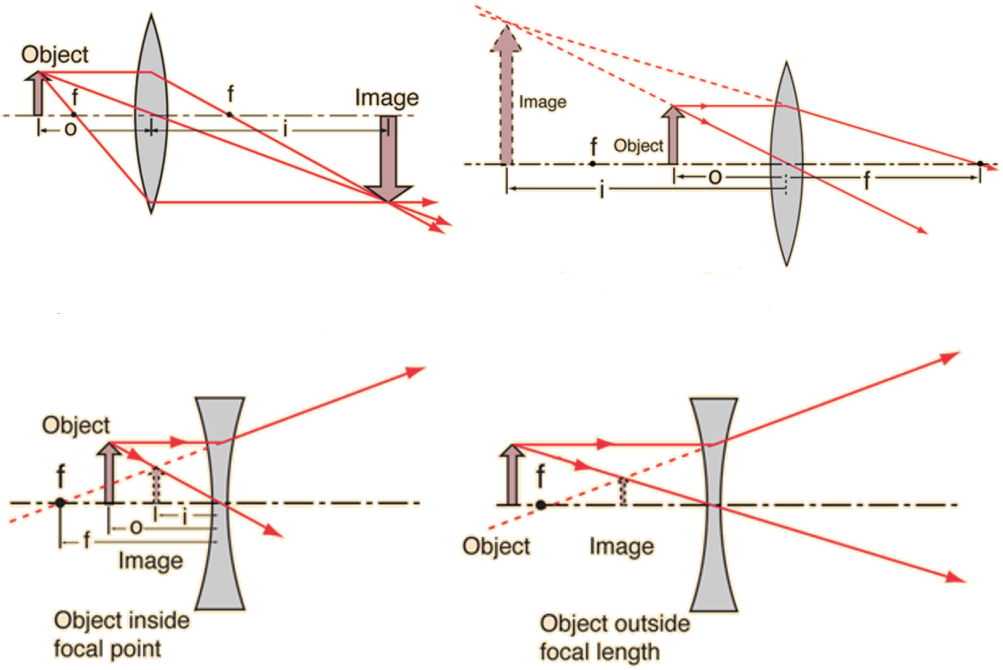

Figure 80: The ray diagrams of the converging lens (top) and the diverging lens (bottom). *Source*: Hyperphysics.

The results are shown below, where we can find there are several similar parts to Table 6:

Table 7: The image types and positions generated by the thin lens.

Mirror	d_o	d_i	Orientation	Size	Type
Converging	$d_o = \infty$	$d_i = f$	Inverted	Smaller	Real
	$\infty > d_o > 2f$	$f < d_i < 2f$			
	$d_o = 2f$	$d_i = 2f$		Same	
	$2f > d_o > f$	$d_i > 2f$		Larger	
	$d_o = f$	$d_i = \infty$	No image	No image	No image
	$d_o < f$	Same side of lens	Upright	Larger	Virtual
Diverging	Anywhere			Smaller	

Despite sketching the ray diagrams is an essence of geometrical optics, it costs a lot of time to draw the rays and images, it is also tough to memorize all the information in Table 6 and Table 7. Thus, we can write down a useful formula, the **lens equation**:

$$\boxed{\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}} \tag{5.6}$$

For the lens in air, we should include the index of refraction n and the radii of curvature R_1, R_2 at the incoming and outgoing sides of lens:

$$(n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \frac{1}{f} \tag{5.7}$$

This is called the **lens-maker's equation**. I suggest the readers remember the lens equation (5.6), which is suitable for both mirror and lens cases. We can define the sign conventions and simply read out the results according to Table 8.

Table 8: The sign convention of the spherical mirrors and thin lenses.

	Spherical mirrors	Thin lenses
$f > 0$	Concave	Convex (converging)
$f < 0$	Convex	Concave (diverging)
$d_o > 0$	In front of the mirror (real image)	Incoming side (real image)
$d_o < 0$	Behind the mirror (virtual image)	Outgoing side (virtual image)
$d_i > 0$	In front of the mirror (real image)	Outgoing side (real image)
$d_i < 0$	Behind the mirror (virtual image)	Incoming side (virtual image)
$M > 0$	Upright image	Upright image
$M < 0$	Inverted image	Inverted image

5.3 Interference

In 17th century, many physicists proposed a model indicating that light is a kind of wave. One representative theory is the **Huygens-Fresnel principle**, which suggests that each point on a wavefront can be regarded as a point source of the spherical secondary wave.

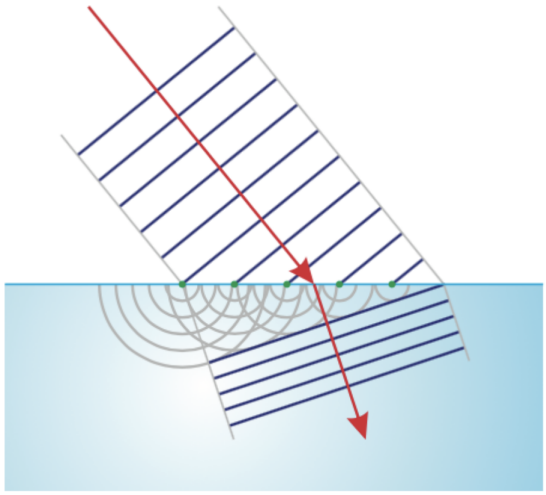

Figure 81: The Huygens-Fresnel principle can predict the wavefronts in the next moment. *Source*: Wikipedia.

When two waves superpose each other, **interference** would appear. The path length difference is denoted by ΔL, and we find that the fully constructive interference is happened when

$$\Delta L = n\lambda, \quad n = 0, 1, 2, ... \quad \text{(constructive)}$$

, whereas the fully destructive interference is happened when

$$\Delta L = n\lambda, \quad n = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, ... \quad \text{(destructive)}$$

Now, we are going to discuss one of the most peculiar experiments in physics. This is the **double-slit experiment**, which has done by Thomas Young in 1801. Simply speaking, this is pretty heuristic when developing quantum mechanics, but at this point I would like to explore its optical meaning. In this experiment, we let light passing through a double-slit, and then we can find a series of bright / dark fringes shown on the screen. The experimental set-up is shown as Figure 82. Note the interference pattern is caused by the path length difference when the light passing through the different slits. We assume the distance between two slits is d, the distance between the slits and the screen is D, and the angle between two light paths is θ. First, the path length difference is determined by

$$\boxed{\Delta L = d\sin\theta = n\lambda} \quad \text{where} \quad n = 0, 1, 2, ... \quad \text{(maxima / bright fringes)} \tag{5.8}$$

$$\boxed{\Delta L = d\sin\theta = \left(n + \frac{1}{2}\right)\lambda} \quad \text{where} \quad n = 0, 1, 2, ... \quad \text{(minima / dark fringes)} \tag{5.9}$$

This also shows the distribution of bright / dark fringes on the screen. Next, we can determine the width among the fringes, which denoted by y. In fact, the trigonometry implies

$$\tan\theta = \frac{y}{D}$$

Since the distance D is much larger, this gives

$$\tan\theta \cong \sin\theta \cong \theta \cong \frac{y}{D}$$

From Eq (5.8), the maximum happened when $d\sin\theta = n\lambda$. Thus, we can derive the distance between the fringes on the screen:

$$\boxed{y = n\frac{\lambda D}{d}} \tag{5.10}$$

The intensity about the interference is given by

$$I = 4I_0\cos^2\frac{\phi}{2} \tag{5.11}$$

, where

$$\phi = \frac{2\pi d\sin\theta}{\lambda}$$

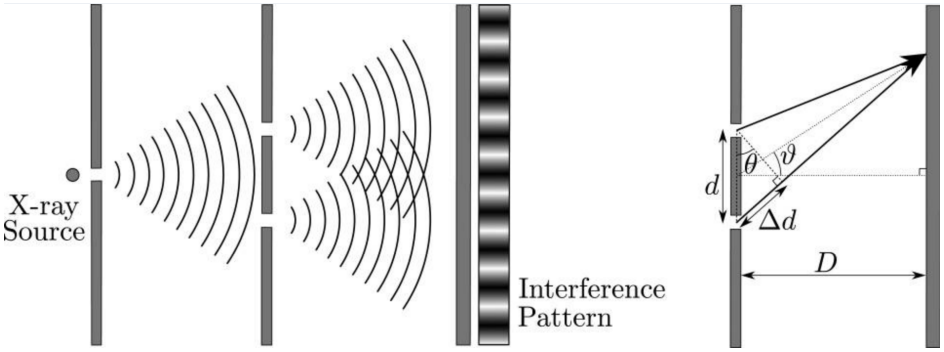

Figure 82: The set-up and the illustration of the double-slit experiment. *Source*: https://www.ncbi.nlm.nih.gov/books/NBK546148/figure/ch9.fig1/.

In fact, there are several types of interference. One of the most common types might be the thin-film interference. A thin film has different layers with different index of refraction. This causes the interference such as the colorful pattern of a bubble soap.

5.4 Diffraction

When light wave passing through two slits, an interference pattern appears. Then, what if we replace a double slit by a single slit? Interestingly, we find the **diffraction** in such an experiment. The central fringe is the brightest and the widest.

The derivation is similar. But here we denote the width of the slit by a. Moreover, the condition of minima is determined by

$$\boxed{a\sin\theta = m\lambda} \quad \text{where} \quad m = 1, 2, 3, ... \quad \text{(minima / dark fringes)} \tag{5.12}$$

The rest of information, including the approximation of the angle, does not change. In this way, we can figure out the distance between the fringes on the screen:

$$\boxed{y = m\frac{\lambda D}{a}} \tag{5.13}$$

The intensity is given by

$$I = I_m\left(\frac{\sin\alpha}{\alpha}\right)^2 \tag{5.14}$$

where

$$\alpha = \frac{\phi}{2} = \frac{\pi a\sin\theta}{\lambda}$$

This also implies that the central maximum would be narrower if the slit is wider. The diffraction pattern is also associated with the ratio between the wavelength and the width. If a/λ is smaller, the diffraction pattern becomes more apparent.

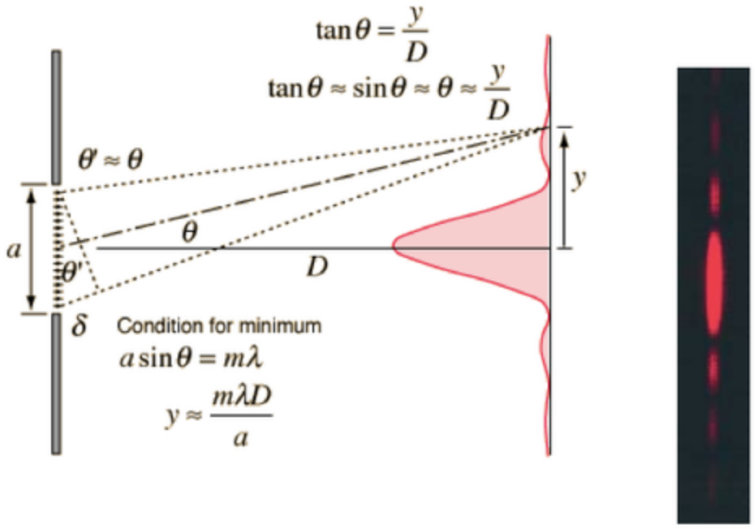

Figure 83: The set-up and the illustration of the single-slit experiment. *Source*: Hyperphysics

An application of diffraction is to calculate the angular resolution, which is determined by the minimum angle of two small objects under the identification of any optical devices. Assume a circular aperture with diameter d, the first minimum suggests that

$$\sin\theta = 1.22\frac{\lambda}{d}$$

But since the angle is small enough, we may have the following approximation:

$$\theta = 1.22\frac{\lambda}{d} \tag{5.15}$$

This is the **Rayleigh's criterion** for resolsvability. As for the single-slit diffraction's case, it becomes

$$\theta = \frac{\lambda}{d}$$

Another important application is the diffraction of X-ray, which is useful for exploring the crystal's physics properties. Given the spacing between atoms is d and the incident wave has the wavelength λ, one can determine the scattering angle by the following formula:

$$2d \sin\theta = m\lambda \quad \text{where} \quad m = 1, 2, 3, ... \tag{5.16}$$

This is known as the **Bragg's law**, as shown in Figure 84. Although this equation looks pretty simple, it gives us a picture that the atom's size, and confirms X-ray diffraction as a new method to explore the crystal properties. That's why Lawrence Bragg and his father William Henry Bragg won the Nobel prize in 1915.

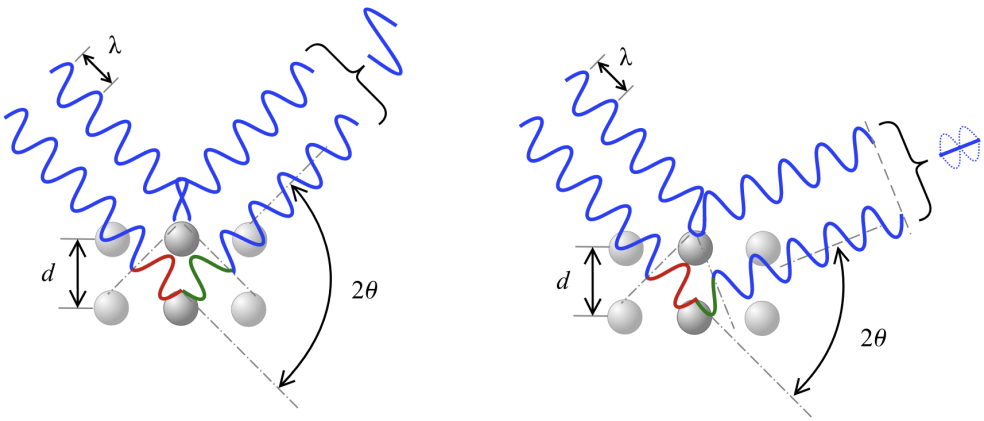

Figure 84: The Bragg's law can help us find the angle of the constructive (left) and the destructive (right) interference within a crystal. *Source*: Wikipedia

5.5 Dispersion and scattering

The simplest example of **dispersion** is the colors we see. A beam of light in fact is a mixture of the electromagnetic waves with different wavelengths. When a light passes through the dispersive prism, the different indices of refraction enables us to identify the lights with different wavelengths, which are corresponding to the different colors we see.

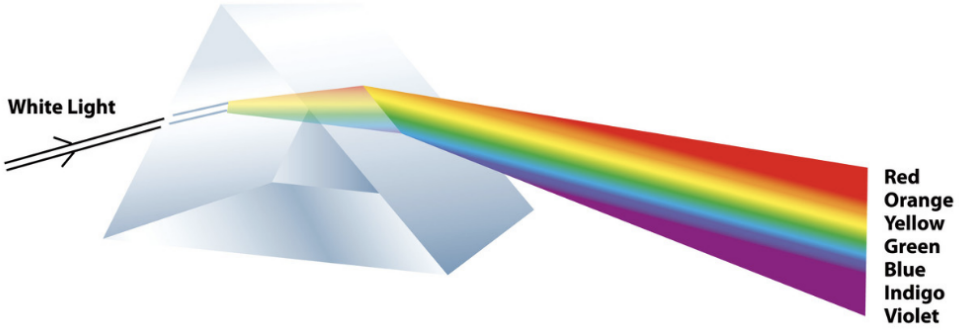

Figure 85: The dispersion of a prism. *Source*: https://illuco.co.kr

The most common phenomenon about dispersion in nature is the **rainbow**. When there are many raindrops in the sky, light will be able to pass through them. Now, the raindrop plays a role in a prism,

which leads to the dispersion so that seven different colors are appeared. Given the refractive index of water, we can find that the rainbow obviously appears when the angle between the incident light and our sight is around 42.5°. Since some of the light has been refracted into our eyes, the rest of light in the raindrop might have the second reflection. This generates the "secondary rainbow", which can be observed occasionally. Note its order of colors would be reversed compared to the primary rainbow, as shown in Figure 86.

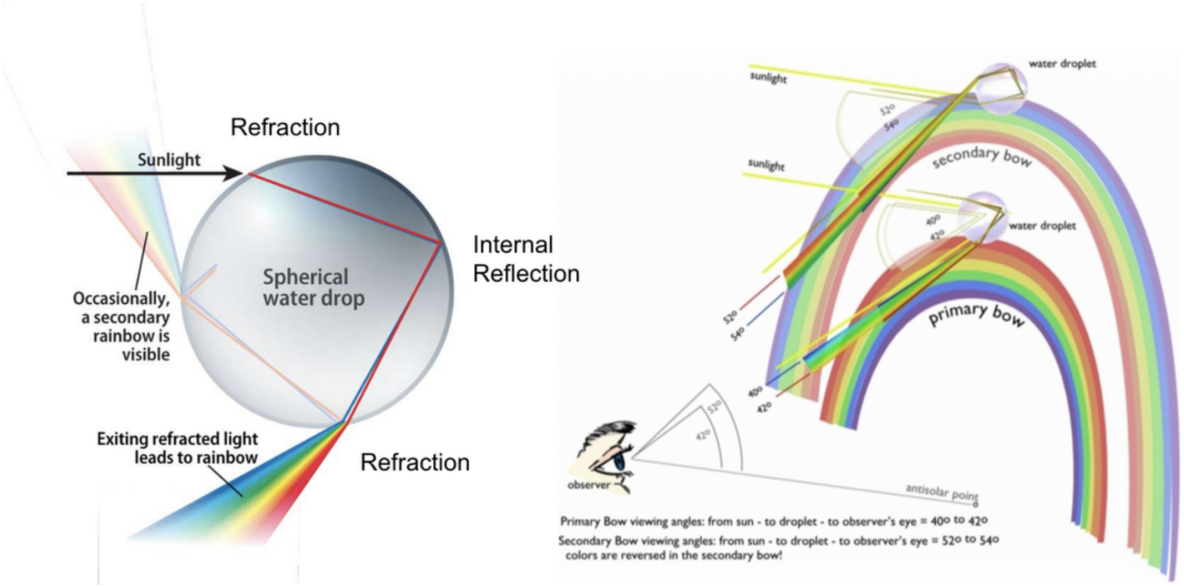

Figure 86: The refraction and the dispersion of light within the raindrops (left) let us see different colors from our eyes, namely, the rainbow. Sometimes we can see the secondary rainbow, which comes from the second reflection within the raindrops (right). Note the order of colors in the primary rainbow and the secondary rainbow are reversed. *Source*: http://roseltac.weebly.com/f2f-artifacts.html

Another optical phenomenon is called the **scattering**. When light pass by small particles, for example, its trajectory would be changed. Have you ever thought why the sky is blue? Similarly, why is the sunset red? Why is the cloud white? In fact, this can be explained by the process called the **Rayleigh scattering**. Physics tells us that the wavelength of blue light is shorter than the red light. According to the Rayleigh scattering, the blue light is easier to be scattered and then come into our eyes. The sunlight is the strongest and directly enters our atmosphere during the daytime, the molecules and atoms of our atmosphere scatter the blue light out and then come into our eyes, that's why we see the sky is blue. Such a procedure is also called the **diffuse sky radiation**. One might wonder: the violet light has the shortest wavelength so it should be easier to be scattered, compared to blue light. Yes, you are right. However, the process also depends on our eyes, instead of the scattering process only. Humans' eyes are more sensitive to identify the blue light compared to the violet one, therefore, even though the violet light has a higher energy, we can still see the sky is in blue color.

As for why the sunset (or sunrise) is red, this is an extended version of the story that the sky is blue. During sunset or sunrise, the sun is near the horizon. Compared to the sunlight during the noon, it has to enter the atmosphere and come into our eyes through a much longer path. The blue light has mostly been scattered out before coming to our eyes, and the remaining light that is able to come into our eyes is either orange or red. That's why the sunrise or the sunset is red. As for the clouds, they are consist of the water droplets, which are typically much larger than the atmospheric particles. When the white sunlight enter the clouds, it would be equally scattered out (instead of scatter certain color out), so that it comes into our eyes in the same color before it has been scattered, which is the white. Hence, we see the clouds are white. But for the moon, it does not have atmosphere. Thus, even if in the daytime, the Rayleigh scattering process and the diffuse sky radiation cannot happen, so the sky on the moon is always dark.

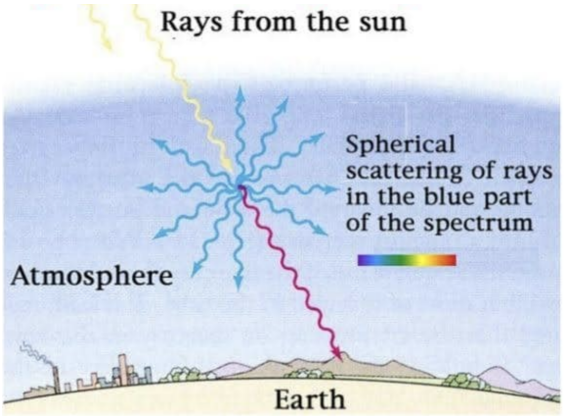

Figure 87: The Rayleigh scattering explains why the sky is blue. *Source*: Mysore Astronomy Community

6 Modern Physics: Special Relativity

From this chapter, we start exploring the two essential cornerstones of modern physics, special relativity and quantum mechanics. We are going to discuss special relativity in this chapter.

6.1 The two postulates

In classical physics, we usually consider the speed that is much slower than the speed of light. More accurate, people have not expected how physics would change as the speed close to light, which is called the relativistic limit. Albert Einstein was the first person addressed this problem. As a start, let me introduce the two postulates in special relativity:

1. The laws of physics are *all look the same* in the inertial reference frames.
2. The speed of light in vacuum is *a constant*, $c \cong 3 \times 10^8$ m/s.

What is the *inertial reference frame*? Simply speaking, it is the frame without considering any inertial forces as we discussed in Section 2.13.

One of the most meaning of (special) relativity is, it indicates space and time are not "absolute" but "relative', depends on the frame of references. In Galilean mechanics, physics, space and time should not change at all. But in Einstein's relativity, it does not. The **simultaneity** would depend on the observer's motion. In other words, if one observes two events happened simultaneously, the other observer might not agree with that. We will see more interesting effect, for example, when you move close to the speed of light, you will feel the time is slower. This sounds actually surprising, while we will prove this immediately.

6.2 Time and length in relativity

In relativity, we have to revise our concept for spacetime. Some fairly interesting phenomena can be obtained from the simple derivations. Assume the path of light in a moving spaceship. As shown in Figure 88 (a), in a moving spaceship, from the astronaut's point of view, nothing is special because the light goes as the path with $2D$. The time interval from releasing to receiving is

$$\Delta\tau = \frac{2D}{c}$$

But in the case of people who stays on the ground, the observation would be slightly different. Since the spaceship is moving, the path for people on the ground to see would be like a triangle, as shown in Figure 88 (b). For the light, the time interval becomes

$$\Delta t = \frac{2S}{c}$$

, and the moving distance would be $2L = v\Delta t$, where v is its moving speed. The simple trigonometry implies that

$$S = \sqrt{D^2 + L^2} = \sqrt{\frac{c^2\Delta\tau^2}{4} + \frac{v^2\Delta t^2}{4}}$$

Plugging $S = c\Delta t/2$ in, we will find that

$$c^2\Delta t^2 = c^2\Delta\tau^2 + v^2\Delta t^2$$

We can get the expression as the following:

$$\Delta t = \frac{\Delta\tau}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{6.1}$$

In other words, for the astronaut in the moving spaceship, the time he experiences is

$$\boxed{\Delta\tau = \frac{\Delta t}{\gamma} = \Delta t\sqrt{1 - \frac{v^2}{c^2}}} \tag{6.2}$$

, where we have defined the Lorentz factor:

$$\gamma \equiv \sqrt{\frac{1}{1-\beta^2}} = \sqrt{\frac{1}{1-(v/c)^2}} \tag{6.3}$$

with $\beta \equiv v/c$. Obviously, the time of the moving clock ($\Delta\tau$) is measured as running slower. This is called the **time dilation**, and the effect would be more apparent under the relativistic limit. For example, if the spaceship is moving in 99.9% of c, that is, $v = 0.999c$, then we can get $\Delta\tau \approx 0.0447\Delta t$. It means that when the spaceship travels 1 year, the time that people experiences on Earth would be over 22 years!

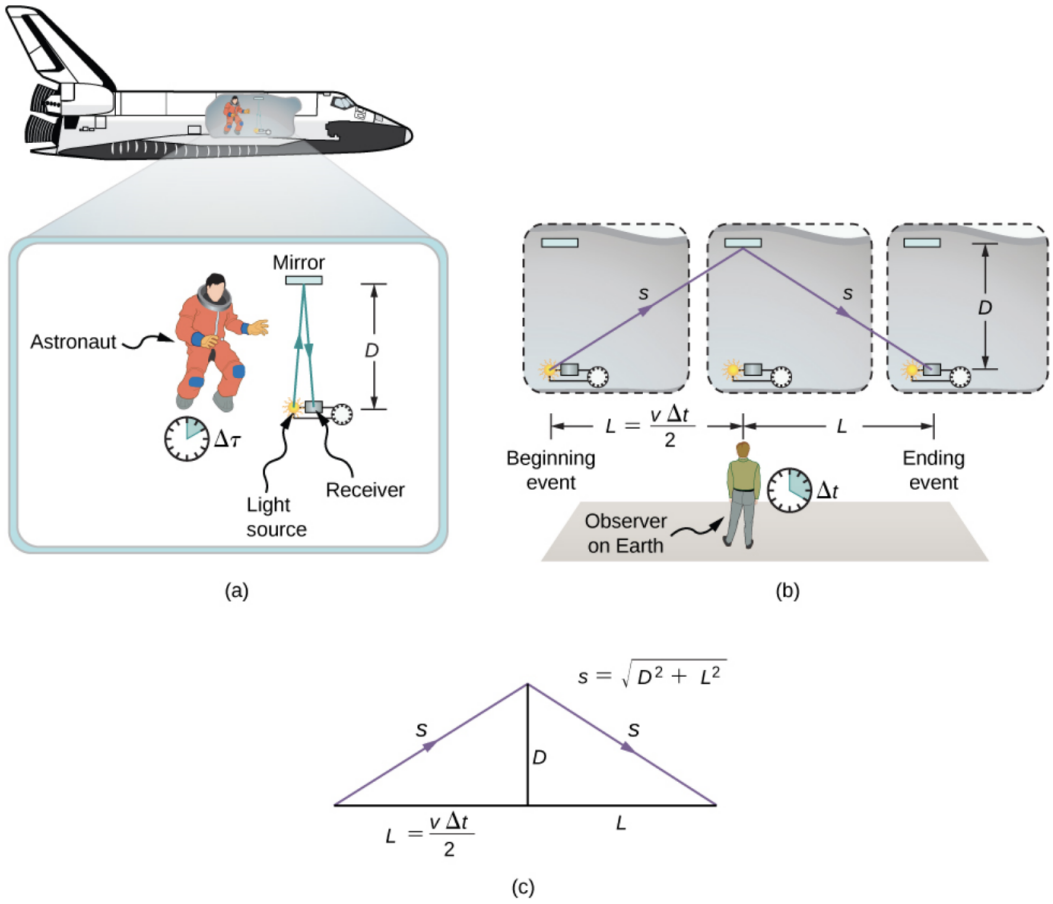

Figure 88: The illustration of the path of light in a moving train from different frames of reference. Here we show the reference frame from the moving train inside (a) and the stationary frame (b). *Source*: https://opentextbc.ca/universityphysicsv3openstax/chapter/time-dilation/

Time dilation implies an incredible scenario as the following: If your older brother goes travel in the spaceship with the speed under relativistic limit, when he finishes the long space trip and comes back, he would find you are much "older" than himself. Then, who is the "older brother"? This is the famous **twin paradox**.

This is not the story yet. The incredible effect can be seen in length. Here we neglect the derivation, as you will see, the result is simple and similar. For the length of a moving object from the point of view of the people on the ground, it will be:

$$\boxed{\Delta L = \frac{\Delta L_0}{\gamma} = \Delta L_0 \sqrt{1 - \frac{v^2}{c^2}}} \tag{6.4}$$

This expression is good enough. Again, we can try our previous example: As the spaceship (the length is ΔL_0) moving with the speed of $v = 0.999c$, the length we observe would be $\Delta L = 0.0447\Delta L_0$. Apparently, the length contracts! If fact, when an object moves with the speed of light, the length would be zero so that it looks disappearing from our sight. This is similar to what you see in some sci-fi plots, and physically it is called the **length contraction** effect.

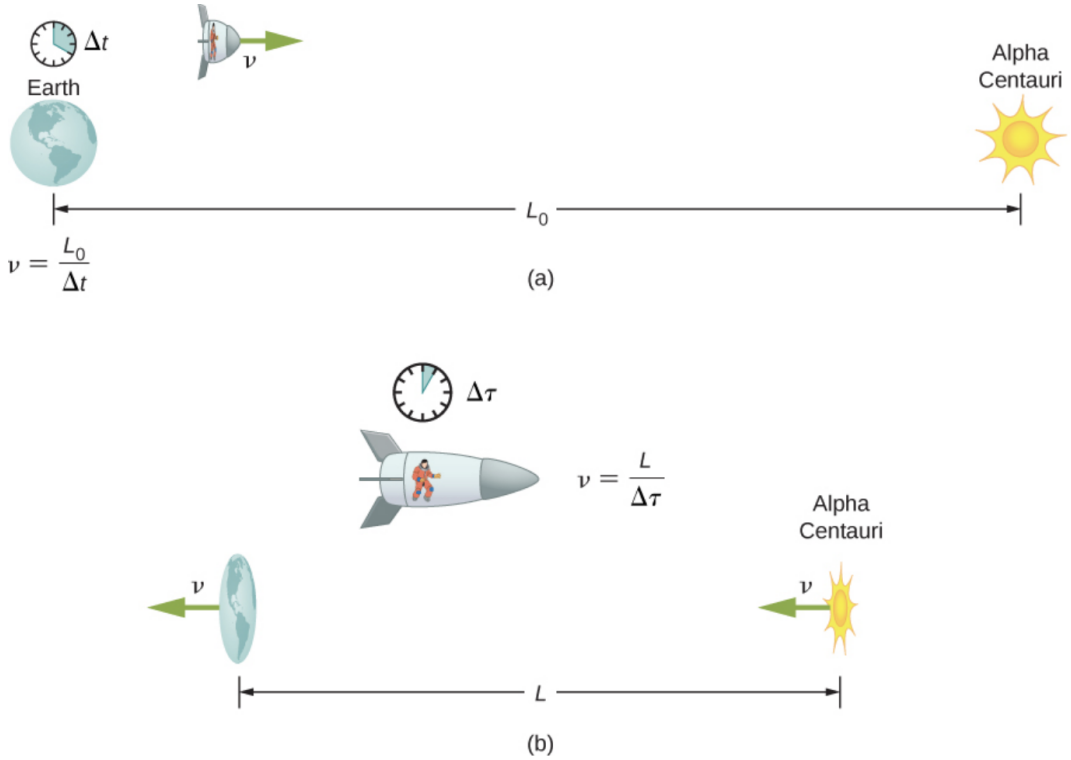

Figure 89: The illustration of the relativistic length contraction, from different points of view. *Source*: https://opentextbc.ca/universityphysicsv3openstax/chapter/length-contraction/

6.3 Lorentz transformation

So far, we have shown a series of revolutions about "space-time" concept. In classical physics, the two arbitrary reference frames can be connected by the **Galilean transformation**:

$$\begin{cases} x' = x - vt \\ y' = y \\ z' = z \\ t' = t \end{cases} \iff \begin{cases} x = x' + vt \\ y = y' \\ z = z' \\ t = t' \end{cases} \tag{6.5}$$

Yet, we saw that physics would have to be significantly changed when $v \to c$. In relativity, consider the moving frame as shown in Figure 90, the reference frames are connected with the **Lorentz transformation**:

$$\begin{cases} x' = \gamma(x - vt) \\ y' = y \\ z' = z \\ t' = \gamma\left(t - \frac{v}{c^2}x\right) \end{cases} \iff \begin{cases} x = \gamma(x' + vt) \\ y = y' \\ z = z' \\ t = \gamma\left(t' + \frac{v}{c^2}x\right) \end{cases} \tag{6.6}$$

The most apparent change here is that we have introduced the Lorentz factor. The first line is corresponding to the length contraction and the last line is corresponding to the time dilation. When $v \ll c$, the Lorentz transformation (6.6) automatically reduces to the Galilean transformation (6.5).

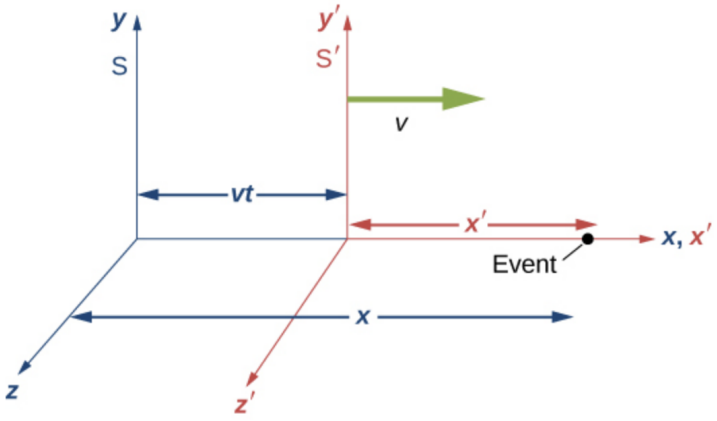

Figure 90: The two frame of references, which can be associated with the Lorentz transformation. *Source*: https://phys.libretexts.org/

To derive out the relativistic velocity, we just need to write down the first and the last lines in Eq (6.6):

$$\Delta x' = \gamma(\Delta x - v\Delta t)$$
$$\Delta t' = \gamma(\Delta t - \frac{v}{c^2}\Delta x)$$

The new velocity is simply the fraction of $\Delta x'$ and $\Delta t'$:

$$u' = \frac{\Delta x'}{\Delta t'} = \frac{\Delta x - v\Delta t}{\Delta t - v\Delta x/c^2}$$

Doing a few arrangement, we arrive at

$$u' = \frac{u - v}{1 - \frac{v}{c^2}u} \tag{6.7}$$

where $u = \Delta x/\Delta t$. In fact, when $v \ll c$ the relative velocity reduces to

$$u' = u - v$$

For the velocity of y and z they can also be obtained in such derivations. To sum up, we have:

$$\begin{cases} u'_x = \dfrac{u_x - v}{1 - u_x v/c^2} \\ u'_y = \dfrac{u_y}{\gamma(1 - u_x v/c^2)} \\ u'_z = \dfrac{u_z}{\gamma(1 - u_x v/c^2)} \end{cases} \tag{6.8}$$

6.4 Minkowski spacetime

Before entering the topic about relativistic mechanics, I shall introduce a few concepts about spacetime. Physically, we are living in a $(3+1)$-dimensional spacetime, with three spatial dimensions and one temporal dimension. Without considering gravitation, this is a flat spacetime, which is formally called the **Minkowski spacetime**. The Lorentz transformation looks good enough, but one should find out a more concise way to include all these information. First, we define the four components:

$$(x^0, x^1, x^2, x^3) = (ct, x, y, z)$$

This set of components is called a **4-vector**. You see that the zeroth component represents the time, whereas the other three components represent the spatial coordinates. Recall the Lorentz factor γ and $\beta = v/c$, using

our new notations, the Lorentz transformation (6.6) can be rewritten as the following:

$$\begin{cases} x'^0 = \gamma(x^0 - \beta x^1) \\ x'^1 = \gamma(x^1 - \beta x^0) \\ x'^2 = x^2 \\ x'^3 = x^3 \end{cases} \tag{6.9}$$

Alternatively, we can write these in matrix form:

$$\begin{pmatrix} x'^0 \\ x'^1 \\ x'^2 \\ x'^3 \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}$$

The matrix in the middle term can be denoted by Λ^μ_ν. Hence, we have the following expression:

$$\boxed{x'^\mu = \sum_{\nu=0}^{3} \Lambda^\mu_\nu x^\mu} \tag{6.10}$$

This is the compact form of the Lorentz transformation. The four-dimensional scalar product usually requests the zeroth component have a minus sign. Therefore, we determine the **covariant** vector:

$$x_\mu = (x_0, x_1, x_2, x_3) = (-x^0, x^1, x^2, x^3) \tag{6.11}$$

The vector with upper indices is called the **contravariant** vector, denoted by x^μ. The minus sign is necessary for a 4-vector's invariance. If one remains the same under Lorentz transformation, then it is **Lorentz invariant**. Note sometimes you will see the convention like $(+,-,-,-)$, and sometimes you will see our version such as $(-,+,+,+)$. Whatever the convention is, the scalar product has to be the same if one is Lorentz invariant. From our current convention, we can formally write down the following expression:

$$\boxed{x_\mu = \sum_{\nu=0}^{3} g_{\mu\nu}\, x^\nu} \tag{6.12}$$

, where

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.13}$$

is the (Minkowski) **metric tensor**. Incidentally, the scalar product in relativity is usually written with **Einstein summation convention**:

$$x_\mu y^\mu = \sum_{\mu=0}^{3} x_\mu y^\mu = x_0 y^0 + x_1 y^1 + x_2 y^2 + x_3 y^3 = -x^0 y^0 + x^1 y^1 + x^2 y^2 + x^3 y^3 \tag{6.14}$$

It is easy to show that

$$x_\mu y^\mu = x^\mu y_\mu \tag{6.15}$$

due to the invariance.

We can define the spacetime interval by doing the scalar product for the 4-vector itself:

$$x^\mu x_\mu = -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 = -c^2t^2 + d^2 \tag{6.16}$$

For x^μ we can determine the following types:

- **Timelike**: $x^\mu x_\mu < 0$. In this case, the reference frames may disagree whether the events happened at the same place, but agree with the time and the order they occurred.
- **Spacelike**: $x^\mu x_\mu > 0$. In this case, the reference frames may disagree with the order that the events occured, but agree that the events are occurred at different locations.
- **Lightlike**: $x^\mu x_\mu = 0$. This is the situation of light. The reference frames will agree with the order that the events occurred.

We can sketch a Minkowski diagram, where the vertical axis represents time and the horizontal axis represents position. Such a diagram can indicate the trajectory of a particle (or an object) in spacetime. Such a trajectory is called the **world line**. The diagram is also called the **light cone**. The boundary is a 45° determined by the photon traveling with the speed of light, where $x^0 = ct$. Basically, particles cannot travel across the spacelike region because things cannot travel faster than light.

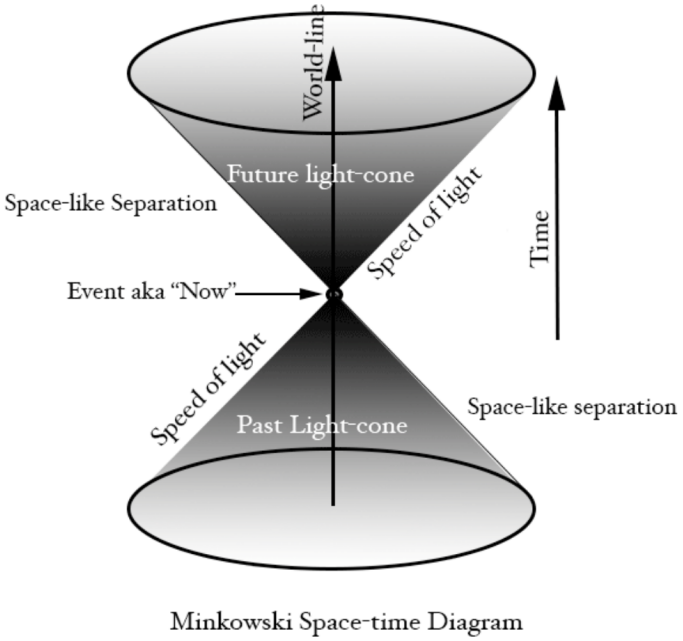

Figure 91: The light cone a diagram for Minkowski spacetime.

6.5 Relativistic energy and momentum

Classically, the momentum $p = mv$ is well defined. But in special relativity, we have to include a Lorentz factor. One can work out an expression by the chain rule:

$$p = m\frac{\Delta x}{\Delta t_0} = m\frac{\Delta x}{\Delta t}\frac{\Delta t}{\Delta t_0} = mv\gamma$$

Therefore, we have the expression for the relativistic momentum:

$$\boxed{\mathbf{p} = \gamma m \mathbf{v}} \tag{6.17}$$

In relativity, things are described by a 4-vector. For the velocity, we can do the same thing. First, according to time dilation's formula Eq (6.2),

$$d\tau = \sqrt{1 - \frac{v^2}{c^2}}dt \tag{6.18}$$

This is, in effect, called the **proper time**. The **4-velocity** is defined as

$$v^\mu = \frac{dx^\mu}{d\tau} \tag{6.19}$$

For the zeroth component, since $x^0 = ct$, we have

$$v^0 = \frac{dx^0}{d\tau} = \frac{c\,dt}{\sqrt{1 - v^2/c^2}dt} = \frac{c}{\sqrt{1 - v^2/c^2}}$$

Therefore, the **4-momentum** can be written as

$$p^\mu = mv^\mu \tag{6.20}$$

For the zeroth component, we have

$$p^0 = mv^0 = \frac{mc}{\sqrt{1 - v^2/c^2}}$$

According to Einstein's convention, energy is associated with the 4-momentum's temporal component. That is to say, $E = p^0 c$ is known as the relativistic energy, which is equivalent to:

$$\boxed{E = \gamma mc^2} \tag{6.21}$$

When $v \to 0$, the Lorentz factor vanishes. In this situation it gives the rest energy:

$$\boxed{E_0 = mc^2} \tag{6.22}$$

This is the well-known equation for the mass-energy equivalence. Since c represents a constant, it means the mass is a form of energy in some sense. Let me know how surprising the result it: given a US dollar bill (approximately 1 gram), it is able to generate the energy that is stronger than the Little Boy atomic bomb (6×10^{13} J), which almost destroyed Hiroshima at the end of World War II. But don't be afraid that your dollar bill will burn and lead to such a disaster, because it is nearly impossible to perfectly convert into whole the energy! Now let us come back to our topic. The total energy is the sum of the rest energy and the kinetic energy:

$$E = E_0 + E_k \tag{6.23}$$

As a result, the kinetic energy can be obtained by

$$E_k = E - mc^2 = mc^2(\gamma - 1)$$

Applying the Taylor expansion for $\sqrt{1 + x}$, we will have

$$E_k = mc^2 \left(\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \right) \cong \frac{1}{2}mv^2 + \frac{3}{8}\frac{mv^4}{c^2} + ...$$

Apparently, we obtain the classical expression of the kinetic energy

$$E_k = \frac{1}{2}mv^2 \tag{6.24}$$

Doing some algebraic procedures, we will ultimately obtain the following relation:

$$\boxed{E^2 = p^2c^2 + m^2c^4} \tag{6.25}$$

This is called the **energy-momentum relation**, which is particularly useful.

7 Modern Physics: Quantum Mechanics

Modern physics is consist of two primary subjects: Relativity and quantum mechanics. Particularly, quantum physics plays a very crucial role in our daily life. From LED, your camera, your television at home, computer, to chemistry, superconductors, even the origin of our universe – all of these have quantum physics involved.

7.1 Foundation-1: Blackbody radiation

An ideal blackbody can absorb all electromagnetic radiations and transfer them to thermal radiations. **Wien's displacement law** tells the relationship between the wavelength and the temperature:

$$\lambda_{max} = \frac{2.898 \times 10^{-3} \mathrm{m} \cdot \mathrm{K}}{T} \tag{7.1}$$

Besides, we know that colors are corresponding to the wavelengths on the spectrum. For instance, blue giant stars usually have higher temperature than red giant stars. Wien also proposed another law to describe blackbody radiation. The law tells us the intensity can be expressed as

$$I(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}}} \tag{7.2}$$

But for the real-world experiment, **Wien's law** performs well high frequency region but does not work under lower frequency region. On the other hand, **Rayleigh–Jeans law** gives the following expression:

$$I(\nu, T) = \frac{2\nu^2 kT}{c^2} \tag{7.3}$$

This is precise at lower energy region. However, when frequency rises, the energy intensity will keep increasing and finally diverges. Such a situation is called the **ultraviolet catastrophe**. According to what we know from experimental data, the intensity curve will not diverge. Therefore, physicists need to find another version to describe the blackbody radiation.

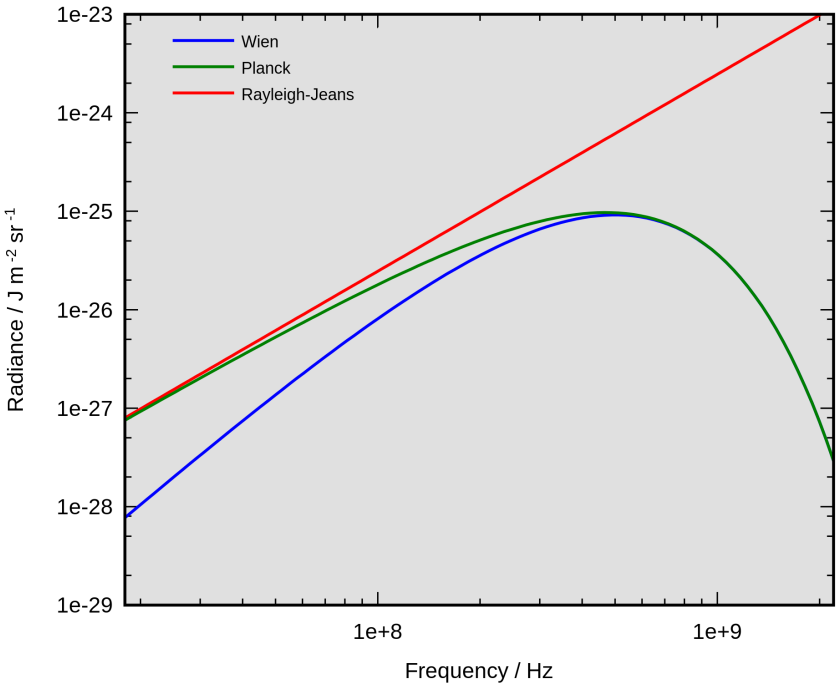

Figure 92: Three predictions of blackbody radiation. Planck's law is the correct one in both low-energy and high-energy limits. *Source*: Wikipedia.

Hence, Max Planck proposed a new picture to describe the blackbody radiation:

$$I(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1} \tag{7.4}$$

This is the **Planck's law**. When $h\nu \gg kT$, it will reduce to the Wien's law Eq (7.2), whereas when $h\nu \ll kT$, it will reduce to the Rayleigh–Jeans law Eq (7.3). This shows that Planck law gives a correct interpretation for blackbody radiation. It is noteworthy that Planck's perspective stems from a new assumption that the energy can be viewed as discrete lumps, and each of them carry a certain amount of energy. The energy is given by

$$\boxed{E = h\nu} \tag{7.5}$$

, where $h \cong 6.626 \times 10^{-34}$ $\mathrm{m^2 kg/s}$ is the Planck constant. This is the first viewpoint of energy's quantization.

7.2 Foundation-2: Photoelectric effect

Many people have known Einstein is a great physicist who proposed relativity. Yet, the success in special or general relativity did not directly let him win the Nobel Prize in Physics. Instead, in 1921, Einstein won the Nobel Prize owing to the discovery of the **photoelectric effect**. Historically, blackbody radiation and photoelectric effect opened a new window of quantum physics.

When electromagnetic wave such as ultraviolet light hits on a metal surface, people observed electrons emitted. Whether the metal surface emits electrons or not depends on the incident light's frequency. When the light's frequency is larger than the threshold frequency (ν_0), the photoelectric effect appears. We define the lowest energy to allow an electron emitted from the surface as the "work function" $W = h\nu_0$. Note the work function depends on the metal surface's material. Then, the energy of incident light is

$$h\nu = W + K \tag{7.6}$$

, where $K = mv^2/2$ is the kinetic energy of emitted electrons. If the incident light's frequency is lower than the threshold frequency, the metal surface will not emit electrons, so the photoelectric effect will not happen.

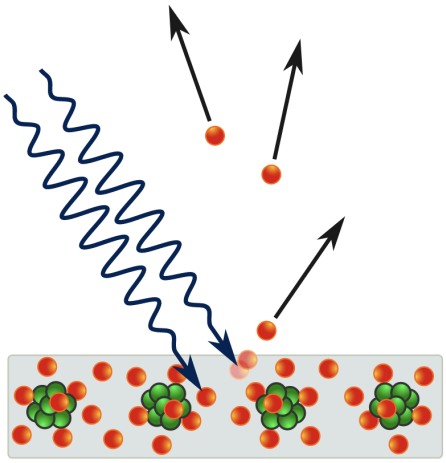

Figure 93: The photoelectric effect: The incident light with a sufficient frequency enable the metal emits electrons. *Source*: Wikipedia.

Einstein asserted that light can be viewed as a bunch of energy lumps, or quanta, called **photons**. They carry the energy

$$E = h\nu$$

A photon carry a certain amount of energy to enable the surface emits an electron. Such a statement revitalizes the perspective that "light is a particle". After the long-term development of optics and Maxwell's beautiful theory of EM, people believed light is a kind of "wave". Now, Einstein claimed light can be viewed as particles, so this has brought people to think about the essence of light again.

7.3 Foundation-3: Matter waves

In 1924, Louis de Broglie proposed that not only light has wave's property, but matters also have such property. Particles have the following wavelength:

$$\boxed{\lambda = \frac{h}{p}} \tag{7.7}$$

This is the matter wave's expression. For photons, the momentum is given by

$$p = \frac{E}{c} = \frac{h}{\lambda}$$

Indeed, although the double-slit experiment of light has been conducted many years ago, people also noticed the same experiment did work for electrons. That is to say, electrons in the double-slit experiment can also produce the interference patterns. De Broglie's description is generally suitable for particles. But for the larger scale as what we see in our daily lives, the wavelength is usually too small to be observed. One crucial point is that a particle have the wave-like and the particle-like behaviors at the same time. This is called the **wave–particle duality**.

De Broglie also hypothesized that the electron's orbit circumference in an atom can be related to the matter wave's wavelength:

$$2\pi R = n\lambda$$

, where R is the orbital radius. From Eq (7.7), $\lambda = h/p = h/mv$, we found that

$$2\pi mvR = nh$$

Also, $mvR = \ell$ is simply the angular momentum. Therefore, we conclude that

$$\ell = n\frac{h}{2\pi} = n\hbar \tag{7.8}$$

Here, $\hbar \cong 1.055 \times 10^{-34}$ $\mathrm{m^2kg/s}$ is called the reduced Planck constant. This constant is quite useful in quantum mechanics. The photon energy can be represented by

$$E = nh\nu = n\hbar\omega \tag{7.9}$$

because $\omega = 2\pi\nu$. Note $n = 1, 2, 3, ...$ for all these cases. In other words, energy can be "quantized" as "oscillators" with an integer amount of small energy lumps. Armed with these fundamental knowledge, we can start learning quantum mechanics.

7.4 The wave function

Now, we can enter the topic about the real "quantum mechanics" (for the previous sections we usually called them "quantum physics"). In classical physics, waves can be described by the wave's equation. Erwin Schrödinger started thinking about a question: Is there any possible equation that can be used to describe the wave-like behavior of particles? Three heuristic techniques motivated him to do that:

1. Energy conservation law in classical mechanics:

$$E = T + V = \frac{p^2}{2m} + V$$

2. Einstein's formula for photon energy:

$$E = h\nu = \hbar\omega$$

3. De Broglie's hypothesis for matter waves:

$$p = \frac{h}{\lambda} = \hbar k$$

, where $k = 2\pi/\lambda$ represents the wave number.

Based on the motivation of classical wave's equation, he introduced a complex plane wave called **wave function** Ψ:

$$\Psi(x,t) = Ae^{i(kx-\omega t)} \tag{7.10}$$

When we take the first derivative with respect to time, we get

$$\frac{\partial \Psi}{\partial t} = -i\omega\Psi$$

Multiply the wave function by energy, we have

$$E\Psi = \hbar\omega\Psi = i\hbar\frac{\partial \Psi}{\partial t}$$

Next, we take a second derivative with respect to position, we get

$$\frac{\partial^2 \Psi}{\partial x^2} = -k^2\Psi$$

Do the same multiplication, this time we used the classical energy conservation law:

$$E\Psi = \frac{p^2}{2m}\Psi + V\Psi = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi$$

Ultimately, we obtain the following equation:

$$\boxed{i\hbar\frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi} \tag{7.11}$$

This is the well-known **Schrödinger equation** – an essential equation in quantum mechanics. It demonstrates how a wave function evolves as time goes by. Its importance is analogous to the "$F = ma$" in classical physics. Sometimes we defined the Hamiltonian $\hat{H}$ to have a more compact form:

$$\boxed{\hat{H}\Psi = E\Psi} \tag{7.12}$$

In fact, we can treat $\hat{H}$ as an operator whereas E as an eigenvalue. Note the **operator** is a specially useful role in quantum mechanics. It is acted as an "operation" for us to calculate certain variables. In the later sections, we shall see how it acts in quantum mechanics.

One of the most important properties for the wave function is the **normalization**. This is even one of the vital properties in QM. The probability (or the probability density) of finding a particle at a certain (x,t) is

$$P = |\Psi(x,t)|^2 = \Psi^*(x,t)\Psi(x,t) \tag{7.13}$$

Normalization means the probability for finding the particle "somewhere" must be 1:

$$\boxed{\int_{-\infty}^{+\infty} |\Psi(x,t)|^2 dx = 1} \tag{7.14}$$

Note this is very straightforward, and we often employ this property to find out the constant ahead a wave function, that is, the normalization constant. For a particle in a quantum state Ψ, the position's expectation value is given by

$$\langle x\rangle = \int_{-\infty}^{+\infty} \Psi^*(x,t)x\Psi(x,t)dx \tag{7.15}$$

When we take the derivative for $\langle x\rangle$, we will find that

$$\langle p\rangle = m\frac{d\langle x\rangle}{dt} = \int_{-\infty}^{+\infty} \Psi^*(x,t)\left(-i\hbar\frac{\partial}{\partial x}\right)\Psi(x,t)dx \tag{7.16}$$

We conclude that the position operator and the momentum operator are defined by

$$\begin{aligned} \hat{x} &= x \\ \hat{p} &= -i\hbar\frac{\partial}{\partial x} \end{aligned} \tag{7.17}$$

That is to say, for a given operator $\hat{O}$, the expectation is

$$\langle\hat{O}\rangle = \int \Psi^* \hat{O} \Psi \, dx \tag{7.18}$$

Such O are **observables**, which represents the variables for people to observe and measure.

Observation is a key concept in QM. Theoretically, when one "observes" a quantum state (e.g. make a measurement), it leads to wave function collapse, specifically, waves collapse into particles to enable us to determine the certain state. This might be slightly metaphysical for someone. Indeed, we have seen that wave functions are probabilistic, which implies quantum world is generally random. When we make an observation, then the states will be determined. In physics, Heisenberg proposed the "uncertainty principle" to describe the uncertainty between the position and the momentum:

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}} \tag{7.19}$$

, where $\Delta x = \sigma_x = \sqrt{\langle x^2\rangle - \langle x\rangle^2}$ and the rule is the same for Δp. It means that when we measure the momentum correctly, then we cannot determine the position correctly. Another version of the uncertainty principle is as the following:

$$\Delta E \Delta t \geq \frac{\hbar}{2} \tag{7.20}$$

Uncertainty principle implies that all physical quantities would be meaningful if and only if one can make measurements for them.

The last thing you should know is an expression relating quantum physics back to classical mechanics:

$$\frac{d\langle p\rangle}{dt} = \left\langle -\frac{\partial V}{\partial x} \right\rangle = \langle F\rangle \tag{7.21}$$

This is almost the same version as what you saw in classical laws of motion. When you calculate the derivative of $\langle p\rangle$ in Eq (7.16), you will obtain the same result. Such a relationship is the so-called **Ehrenfest's theorem**.

Before talking about the applications of the Schrödinger equation, I would like to recap some mathematical formulation about quantum mechanics.

7.5 Formalism in quantum mechanics

Your professors have definitely advised you learn linear algebra well before being engaged in quantum mechanics. The reason is, linear algebra is the natural language of quantum mechanics.

The first tool you have to know is the usage of Dirac notations. Paul Dirac invented two symbols to do mathematical operations in quantum mechanics, called "bra" and "ket" (stemmed from "bracket"). "Bra" can be represented as

$$\langle\alpha| = \begin{pmatrix} a_1^* & a_2^* & \dots & a_n^* \end{pmatrix}$$

which acts as a linear function. On the other hand, "ket" can be represented as

$$|\beta\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

which acts as a vector. In this example, the ket vector constitute a vector space, while the bra vector is consist of the complex conjugate wave functions, where we say they are in a dual space. The inner product can be written as

$$\langle\alpha|\beta\rangle = a_1^* b_1 + a_2^* b_2 + ... + a_n^* b_n \tag{7.22}$$

A ket and a bra can be connected with conjugate (Hermitian) transpose,

$$\langle\psi|^\dagger = |\psi\rangle \quad \text{and} \quad |\psi\rangle^\dagger = \langle\psi|$$

Therefore,

$$\langle\beta|\alpha\rangle = \langle\alpha|\beta\rangle^*$$

These notations describe **quantum states**, so they are sometimes called state vectors.

Mathematically, quantum states can be described on the basis of **Hilbert space**. Or, as what David Griffiths said in his textbook, "wave functions live in Hilbert space". Given two functions $f(x)$ and $g(x)$, the inner produce is defined by

$$\langle f|g\rangle = \int f(x)^* g(x)\, dx \tag{7.23}$$

The inner product of $f(x)$ with itself is

$$\langle f|f\rangle = \int |f(x)|^2\, dx \tag{7.24}$$

Here are some important properties about wave functions:

1. A function is **normalized** if
$$\langle f|f\rangle = 1 \tag{7.25}$$
2. Two functions are **orthogonal** if
$$\langle f|g\rangle = 0 \tag{7.26}$$
3. A set of functions is **orthonormal** if
$$\langle f_m|f_n\rangle = \delta_{mn} = \begin{cases} 1, & \text{if } m = n, \\ 0, & \text{if } m \neq n. \end{cases} \tag{7.27}$$
4. A set of functions is **complete** if
$$f(x) = \sum_{n=1}^{\infty} c_n f_n(x) \tag{7.28}$$
, that is, any other function can be written as a linear combination of them.

In quantum mechanics, observables are described as hermitian operators. How can we say the operators are **hermitian**? Well, this is given by the following condition:

$$\langle f|\hat{O}g\rangle = \langle\hat{O}f|g\rangle \tag{7.29}$$

Also, the hermitian (adjoint) conjugate of an operator is

$$\langle f|\hat{O}g\rangle = \langle\hat{O}^\dagger f|g\rangle \tag{7.30}$$

In this way, $\hat{O} = \hat{O}^\dagger$.

The eigenfunctions of an observable operator are complete. This means the wave function can be expressed as

$$\Psi(x,t) = \sum_n c_n(t)\psi_n(x) \tag{7.31}$$

The orthonormality of eigenfunctions implies the coefficients can be obtained by Fourier's trick:

$$c_n(t) = \langle \psi_n | \Psi \rangle$$

And the total probability must be

$$\sum_n |c_n|^2 = 1 \tag{7.32}$$

Note the wave functions we mentioned so far are typically called the "position-space wave functions" (denoted by $\Psi(x,t)$). They can be transformed into the "momentum-space wave functions" (denoted by $\Phi(p,t)$) by Plancherel's theorem:

$$\begin{aligned} \Psi(x,t) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{\frac{ipx}{\hbar}} \Phi(p,t)\, dp \\ \Phi(p,t) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{-\frac{ipx}{\hbar}} \Psi(x,t)\, dx \end{aligned} \tag{7.33}$$

Such a transform is the well-known Fourier transform. Technically, it is very useful in engineering fields, and it is also a fundamental skills in field theory.

7.6 The infinite square well

We are going to introduce the applications of the Schrödinger equation. Back to Eq (7.11), we see that a general wave function has two parts,

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

Doing separation of variables and plugging into the original equation, you will obtain two equations:

$$i\hbar \frac{d\phi}{dt} = E\phi$$

and

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V\psi = E\psi$$

Solving the first equation gives

$$\phi(t) = e^{-\frac{iEt}{\hbar}} \tag{7.34}$$

Thus, a general form of the wave function can be expressed as

$$\Psi(x,t) = \sum_{n=1}^{\infty} c_n \psi_n(x) e^{-\frac{iE_n t}{\hbar}} = \sum_{n=1}^{\infty} c_n \Psi_n(x,t) \tag{7.35}$$

Note we here simply consider the potential $V(x)$ is time-independent. The applications of quantum mechanics is to explore the solutions of time-independent Schrödinger equation,

$$\boxed{-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V\psi = E\psi} \tag{7.36}$$

We will discuss many cases of solutions based on this second-order differential equation. The first simplest case is the infinite square well.

The infinite square well restricts the potential's boundary conditions in this way:

$$V(x) = \begin{cases} 0, & 0 \le x \le L \\ \infty, & \text{otherwise} \end{cases}$$

This case is often interpreted as a particle in a box (Figure 94). Consider the region inside the well, $0 \le x \le L$, the time-independent Schrödinger equation becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi \tag{7.37}$$

This can be written in a clearer way:

$$\frac{d^2\psi}{dx^2} = -k^2\psi \tag{7.38}$$

, where we have introduced a constant

$$k \equiv \frac{\sqrt{2mE}}{\hbar} \tag{7.39}$$

to enable us to solve this differential equation more straightforward.

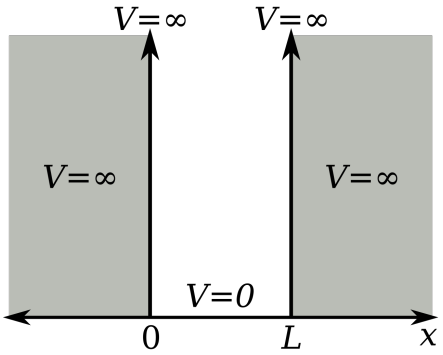

Figure 94: An infinite square well. *Source*: Wikipedia.

The general solution can be written as

$$\psi(x) = A\sin kx + B\cos kx \tag{7.40}$$

Note as what you derived for damped oscillation, you can also write this in exponential form. But we use trigonometric form here for simpler calculations. When solving differential equations, you should bear in mind that boundary conditions are the key roles for us to justify the continuity of a function. In this case, we have to consider two boundary conditions:

- At $x = 0$: $\psi(0) = B$, and then we have
$$\psi(x) = A\sin kx$$
- At $x = L$: $\psi(L) = A\sin kL = 0$, so we have $kL = n\pi$, and then
$$k = n\pi/L$$

Hence, the wave function is taken as the following form:

$$\psi(x) = A\sin\frac{n\pi}{L}x \tag{7.41}$$

Doing normalization, we figure out $A = \sqrt{2/L}$. Therefore, the wave function solution inside the infinite well is

$$\psi_n(x) = \sqrt{\frac{2}{L}}\sin\frac{n\pi}{L}x \tag{7.42}$$

From Eq (7.39), you can find the possible energy values are

$$\boxed{E_n = \frac{\hbar^2k^2}{2m} = \frac{n^2\pi^2\hbar^2}{2mL^2}} \tag{7.43}$$

We see that the energies can be "quantized" as discrete values depending on positive integer n. When $n = 1$ it represent the **ground state**, while for the larger n it represent the **excited state(s)**. If you plug the wave function into the general expression Eq (7.35), it will look like

$$\Psi(x,t) = \sum_{n=1}^{\infty} c_n\sqrt{\frac{2}{L}}\sin\Big(\frac{n\pi}{L}x\Big)e^{-i\frac{n^2\pi^2\hbar^2}{2mL^2}t} \tag{7.44}$$

.

7.7 The delta potential barrier

Physicists noticed we can find numerous interesting phenomena when studying different cases of the potential well. The delta potential barrier is a simple model for the finite potential barrier cases. Now, let us consider the potential as a Dirac δ function. According to Eq (1.15), when $x = 0$ the function reaches infinity, while when $x \neq 0$ the function will be $\delta(x) = 0$. Writing down the Schrödinger equation, the delta potential well's expression looks like

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \alpha\delta(x)\psi = E\psi \tag{7.45}$$

To let our analysis be mathematically meaningful, we here consider $V(x) = 0$'s case, so this leads to two different cases we are going to discuss – $x < 0$ and $x > 0$. After setting up the potential, the equation becomes

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2}\psi \tag{7.46}$$

Formally, we have to consider the following two situations:

- $E < V(\pm\infty) \Longleftrightarrow E < 0$ (bound state)
- $E > V(\pm\infty) \Longleftrightarrow E < 0$ (scattering state)

In most of the cases $V(\pm\infty) \to 0$. Some students might be confused: why can the energy be negative, $E < 0$? Well, "negative energy" is not really physical, for sure. In this case the energy we discuss is the "relative" energy (relative to the potential). For example, we say an electron bounded in a nucleus has an energy of -13.6 eV. This doesn't mean the energy; instead, it represents the minimum energy that an electron has to overcome to be released through the photoelectric effect.

7.7.1 Bound state

We first look at the bound states, that is, $E < 0$. In this case, the Schrödinger equation looks like

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2}\psi = \kappa^2\psi \tag{7.47}$$

where

$$\kappa \equiv \frac{\sqrt{-2mE}}{\hbar}. \tag{7.48}$$

Since $E < 0$, κ has to be real and positive. The general solution is

$$\begin{aligned} \psi(x) &= Ae^{\kappa x} + Be^{-\kappa x} \quad (x < 0) \\ \psi(x) &= Ce^{\kappa x} + De^{-\kappa x} \quad (x > 0) \end{aligned} \tag{7.49}$$

As mentioned, we need to discuss two different cases of x:

- When $x \to -\infty$, the term $Be^{-\kappa x}$ diverges so we have to set $B = 0$. Thus,

$$\psi(x) = Ae^{\kappa x} \quad (x < 0)$$

- When $x \to +\infty$, the term $Ce^{\kappa x}$ diverges so we have to set $C = 0$. Thus,

$$\psi(x) = De^{\kappa x} \quad (x > 0)$$

We see that boundary conditions play significant roles for us to solve differential equations. The key is to analyze the "continuity" of a function, that is,

1. $\psi(x)$ has to be always continuous;
2. $d\psi/dx$ is continuous except at points where $V(x) = \pm\infty$.

The first condition implies when $x = 0$ both wave functions have to be the same, so $A = D$,

$$\psi(x) = \begin{cases} Ae^{\kappa x} & (x \leq 0), \\ Ae^{-\kappa x} & (x \geq 0), \end{cases}$$

For the second condition, please bear in mind the following expression:

$$\Delta\left(\frac{d\psi}{dx}\right) = -\frac{2m\alpha}{\hbar^2}\psi(0) \tag{7.50}$$

This comes from the integration over the Schrödinger equation and the massage of Dirac δ function's discontinuity. Calculating the derivatives gives

$$\begin{aligned} d\psi/dx &= +A\kappa \quad (x < 0), \\ d\psi/dx &= -A\kappa \quad (x > 0) \end{aligned}$$

By subtracting with each other, we have

$$\Delta\left(\frac{d\psi}{dx}\right) = 2A\kappa = \frac{2m\alpha}{\hbar^2}A$$

Then we find

$$\kappa = \frac{m\alpha}{\hbar^2}$$

Hence, the allowed energy is

$$E = -\frac{m\alpha^2}{2\hbar^2} \tag{7.51}$$

Looking back to the wave function, we only need to normalize it, and then we get

$$A = \sqrt{\kappa} = \frac{\sqrt{m\alpha}}{\hbar}$$

Thus, the wave function of the delta potential well will be

$$\psi(x) = \frac{\sqrt{m\alpha}}{\hbar}e^{-\frac{m\alpha|x|}{\hbar^2}} \tag{7.52}$$

, and it has exactly one bound state energy Eq (7.51).

7.7.2 Scattering state

Next, let us consider the scattering states, $E > 0$. The procedures are almost the same as the bound state's case. First, we look at the Schrödinger equation:

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2}\psi = -k^2\psi \tag{7.53}$$

where

$$k \equiv \frac{\sqrt{2mE}}{\hbar}. \tag{7.54}$$

Since $E > 0$, the k is naturally positive, as what we have seen in the infinite square well's case. The general solutions, as the previous case, are

$$\begin{aligned} \psi(x) &= Ae^{ikx} + Be^{-ikx} \quad (x < 0) \\ \psi(x) &= Ce^{ikx} + De^{-ikx} \quad (x > 0) \end{aligned} \tag{7.55}$$

(In fact, the exponential term here is the same as the sinusoidal solutions we had for infinite square well.) Remember we have to look at the continuity right after getting the general solutions. From the first condition, at $x = 0$, we have

$$A + B = C + D$$

Moving to the second condition of continuity, we have

$$d\psi/dx = ik(A - B) \quad (x < 0),$$
$$d\psi/dx = ik(C - D) \quad (x > 0)$$

Eq (7.50) tells us that

$$\Delta\left(\frac{d\psi}{dx}\right) = ik(C - D - A + B) = -\frac{2m\alpha}{\hbar^2}(A + B)$$

Almost done! We write it down in a more concise way:

$$C - D = A(1 + 2i\beta) - B(1 - 2i\beta)$$

where

$$\beta \equiv \frac{m\alpha}{\hbar^2 k}$$

For scattering from the left side, we obtain $D = 0$. We defined that A is the amplitude of the incident wave, B is the amplitude of the reflected wave, and C is the amplitude of the transmitted wave. Specifically,

$$B = \frac{i\beta}{1 - i\beta}A, \quad C = \frac{1}{1 - i\beta}A$$

Now, we can determine the reflection coefficient

$$R = \frac{|B|^2}{|A|^2} = \frac{\beta^2}{1 + \beta^2} = \frac{1}{1 + \dfrac{2\hbar^2 E}{m\alpha^2}} \tag{7.56}$$

as well as the transmission coefficient

$$T = \frac{|C|^2}{|A|^2} = \frac{1}{1 + \beta^2} = \frac{1}{1 + \dfrac{m\alpha^2}{2\hbar^2 E}} \tag{7.57}$$

Higher energy leads to the greater probability of transmission. Also, the sum of probabilities is simply 1, $R + T = 1$. Note this is fairly different from classical physics: For a potential barrier the reflection coefficient would be 1 because a wave should entirely bounce back; for a potential well it should be 0 because a wave would pass through without any disturbs. But in quantum mechanics, the transmission coefficient is not equal to 0, even if $E < V_{\max}$, implies a particle can naturally pass through a "wall". This interesting phenomenon is called **quantum tunneling**. Numerous electronic devices have employed this effect, for example, the scanning tunneling microscope.

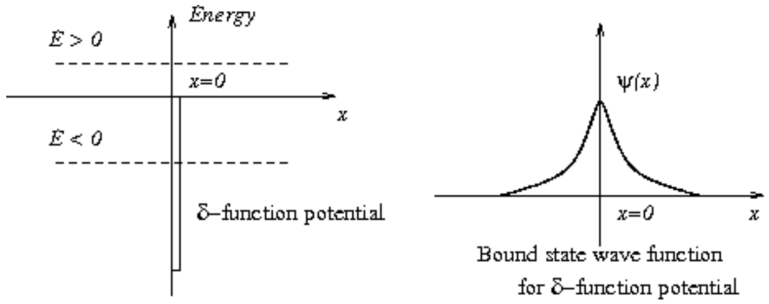

Figure 95: An illustration about Dirac δ potential well and its bound state wave function. *Source*: Lehman College.

7.8 The finite square well

Next, let us consider a finite well with the following potential:

$$V(x) = \begin{cases} -V_0 & -a \le x \le a, \\ 0 & |x| > a \end{cases} \tag{7.58}$$

This is slightly similar to the delta potential well, where we have to consider the bound state ($E < 0$) and the scattering state ($E > 0$).

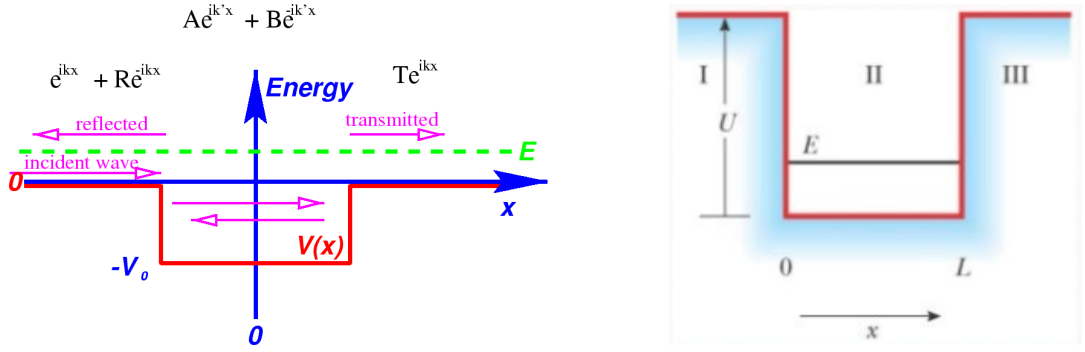

Figure 96: The finite potential well can be divided into three regions (left). Note this is a sketch for scattering state's case. For calculation convenience, we can reset the boundary from $-a$, a to 0, L (right). *Source*: https://www.slideshare.net/Nitians/ph-1019-quantum-machanics (left) coursera.com (right).

7.8.1 Bound state

The mathematical setup is very similar to what we had in delta potential well's case. The first step is always writing down the Schrödinger equation, but now we have to consider three different regions:

$$\begin{cases} \frac{d^2\psi}{dx^2} = \kappa^2\psi, & \kappa \equiv \frac{\sqrt{-2mE}}{\hbar} \quad (x < -a) \\ \frac{d^2\psi}{dx^2} = -\ell^2\psi, & \ell \equiv \frac{\sqrt{2m(E+V_0)}}{\hbar} \quad (-a < x < a) \\ \frac{d^2\psi}{dx^2} = \kappa^2\psi, & \kappa \equiv \frac{\sqrt{-2mE}}{\hbar} \quad (x > a) \end{cases} \tag{7.59}$$

Solve the differential equations and test the divergence at $x = -\infty$ and $x = \infty$, gives:

$$\begin{cases} \psi(x) = Ae^{\kappa x} + Be^{-\kappa x} \xrightarrow{x\to-\infty} \psi(x) = Ae^{\kappa x} & (x < -a) \\ \psi(x) = C\cos\ell x + D\sin\ell x & (-a < x < a) \\ \psi(x) = Fe^{\kappa x} + Ge^{-\kappa x} \xrightarrow{x\to\infty} \psi(x) = Ge^{-\kappa x} & (x > a) \end{cases}$$

Therefore, so far we have the following solutions in the three different regions (as shown in Figure 96):

$$\begin{aligned} \psi(x) &= \begin{cases} Ae^{\kappa x} & (x < -a) \\ C\cos\ell x + D\sin\ell x & (-a < x < a) \\ Ge^{-\kappa x} & (x > a) \end{cases} \\ &= \begin{cases} \psi(-x) & (x < 0) \\ C\cos\ell x & (0 < x < L) \\ Ge^{-\kappa x} & (x > L) \end{cases} \end{aligned} \tag{7.60}$$

where the second line we have followed the graphical interpretation shown in Figure 96. By considering the odd / even function's properties, it would be easier for use to proceed calculations.

As we have seen, once we find the solutions, the next step is to write down the boundary conditions. In this case, we consider the boundary condition at $x = L$ and we will have:

$$\begin{aligned} \psi_{\mathrm{II}}(L) = \psi_{\mathrm{III}}(L) \quad &\Longrightarrow \quad C\cos\ell L = Ge^{-\kappa L} \\ \frac{d\psi_{\mathrm{II}}(L)}{dx} = \frac{d\psi_{\mathrm{III}}(L)}{dx} \quad &\Longrightarrow \quad -\ell C\sin\ell L = -\kappa G e^{-\kappa L} \end{aligned}$$

Solving the two equations gives

$$\kappa = \ell\tan\ell L \tag{7.61}$$

Now we can make the following substitutions:

$$z \equiv \ell L \quad \text{and} \quad z_0 \equiv \frac{a\sqrt{2mV_0}}{\hbar}.$$

Then from Eq.(7.59) we can determine

$$\kappa L = \sqrt{z_0^2 - z^2}$$

Finally, we arrive at the expression:

$$\tan z = \sqrt{\left(\frac{z_0}{z}\right)^2 - 1} \tag{7.62}$$

This is the "transcendental equation" for z. Based on this expression, we may analyze some properties at the limit regions. If the z_0 is small, that means we will have fewer bound states. But no matter how small it is, there is at least one bound state. Conversely, if z_0 is increased, the intersections between two curves will be increased, which means there will be more bound states, followed by

$$E_n + V_0 \approx \frac{n^2\pi^2\hbar^2}{2m(2L)^2}$$

The graphical interpretation is shown as Figure 97.

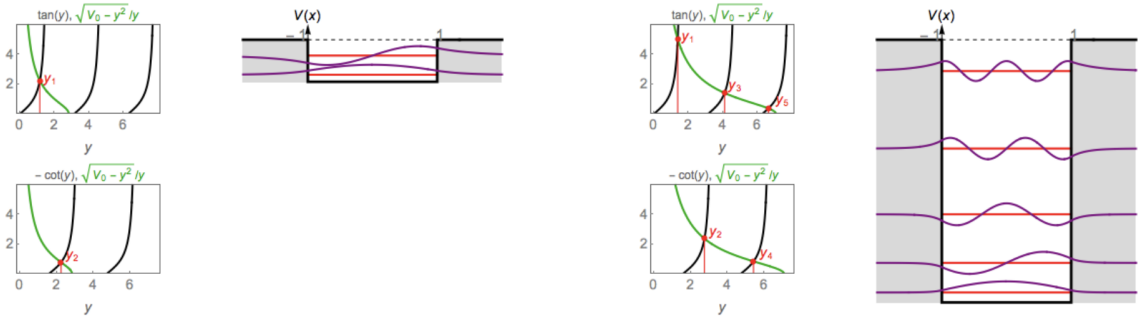

Figure 97: The finite potential well solutions in the shallow well case (left) and the deep well case (right). *Source*: demonstrations.wolfram.com

7.8.2 Scattering state

Now we are familiar with solving the Schrödinger equation, so we can directly write down the solutions of the three different regions in scattering state's case ($E > 0$):

$$\begin{cases} \psi_{\mathrm{I}}(x) = Ae^{ikx} + Be^{-ikx} \quad (x < -a) \\ \psi_{\mathrm{II}}(x) = C\sin\ell x + D\cos\ell x \quad (-a < x < a) \\ \psi_{\mathrm{III}}(x) = Fe^{ikx} \quad (x > a) \end{cases} \tag{7.63}$$

where

$$k \equiv \frac{\sqrt{2mE}}{\hbar}, \quad \ell \equiv \frac{\sqrt{2m(E+V_0)}}{\hbar} \tag{7.64}$$

as we have defined. Note that we did not consider there is no any additional term based on e^{-ikx} for $\psi_{\text{III}}(x)$. The reason is that we assume there is no incident wave in that region. For the graphical definition about the incident wave, the reflected wave, and the transmitted wave, it would be good to refer to Figure 96 (left).

Next, we can set up the boundary conditions. Since in this case we are not going to consider the odd / even function properties, we shall have four boundary conditions (two at $x = -a$ and two at $x = a$):

$$\begin{aligned}
\psi_{\text{I}}(-a) = \psi_{\text{II}}(-a) \quad &\Longrightarrow \quad Ae^{-ika} + Be^{ika} = -C\sin\ell a + D\cos\ell a \\
\frac{d\psi_{\text{I}}(-a)}{dx} = \frac{d\psi_{\text{II}}(-a)}{dx} \quad &\Longrightarrow \quad ik(Ae^{-ika} - Be^{ika}) = \ell(C\cos\ell a + D\sin\ell a) \\
\psi_{\text{II}}(a) = \psi_{\text{III}}(a) \quad &\Longrightarrow \quad C\sin\ell a + D\cos\ell a = Fe^{ika} \\
\frac{d\psi_{\text{II}}(a)}{dx} = \frac{d\psi_{\text{III}}(a)}{dx} \quad &\Longrightarrow \quad \ell(C\cos\ell a - D\sin\ell a) = ikFe^{ika}
\end{aligned} \tag{7.65}$$

After collecting all of the boundary conditions, we are almost done. What we have to do is to eliminate the terms and to calculate the coefficients, as what we have done in delta potential's case. Eventually, we should arrive at the following:

$$\begin{aligned}
\frac{B}{F} &= i(\ell^2 - k^2)\frac{\sin 2\ell a}{2k\ell}, \\
\frac{F}{A} &= \frac{e^{-2ika}}{\cos 2\ell a - i\dfrac{k^2+\ell^2}{2k\ell}\sin 2\ell a}
\end{aligned}$$

Since the coefficient F determines the transmission (that is, the Region III), we can write down the transmission coefficient:

$$T = \frac{|F|^2}{|A|^2} = \left[1 + \frac{V_0^2}{4E(E+V_0)}\sin^2\left(\frac{2a\sqrt{2m(E+V_0)}}{\hbar}\right)\right]^{-1} \tag{7.66}$$

As for the reflection coefficient, it can be simply written down as

$$R = \frac{|B|^2}{|A|^2}$$

where

$$T + R = 1 \tag{7.67}$$

must be satisfied. To sum up, we can see in the scattering state's case, once the particle has enough energy, it can escape the well. As for the bound state's case, we have shown that there are a finite number of energy levels exist.

7.9 Harmonic oscillators

This is one of the most important topic in quantum mechanics. First, we know that the Schrödinger equation has the following form:

$$H\psi = E\psi$$

Classically, the potential of a particle in the 1D harmonic oscillator model is given by

$$V(x) = \frac{1}{2}m\omega^2 x^2 \tag{7.68}$$

In other words,

$$\begin{aligned}
E\psi &= \left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right]\psi \\
&= \frac{1}{2m}\left[\hat{p}^2 + (m\omega\hat{x})^2\right]
\end{aligned}$$

where we have expressed momentum p and position x as the operators. Recall Eq.(7.17).

Armed with the Schrödinger equation, we would typically like to solve that. In the case of quantum harmonic oscillators, we can do either algebraically or analytically. In order to get insight into quantization, I shall show algebraic method here. First, we recognize the Hamiltonian has the form like

$$\hat{H} = \frac{1}{2m}[\hat{p}^2 + (m\omega\hat{x})^2]$$

This is the tricky part to think about the calculations further. What we are going to do is the "factor" the Hamiltonian out. To achieve this, let us introduce the following operators (based on convention, I am not going to write an extra "hat" symbol):

$$\begin{aligned} a &= \sqrt{\frac{m\omega}{2\hbar}}\hat{x} + i\sqrt{\frac{1}{2m\hbar\omega}}\hat{p} \\ a^\dagger &= \sqrt{\frac{m\omega}{2\hbar}}\hat{x} - i\sqrt{\frac{1}{2m\hbar\omega}}\hat{p} \end{aligned} \tag{7.69}$$

These operators are called **ladder operators**. Specifically, a is called the **lowering operator** and $a^\dagger$ is called the **raising operator**. In QFT, $a^\dagger$ and a are called the **creation operator** and the **annihilation operator**, respectively. We will shortly see how they work.

To see how these operators work for factoring the Hamiltonian, let us consider, for example, the following calculation:

$$\begin{aligned} aa^\dagger &= \frac{1}{2m\hbar\omega}(i\hat{p} + m\omega\hat{x})(-i\hat{p} + m\omega\hat{x}) \\ &= \frac{1}{2m\hbar\omega}[\hat{p}^2 + (m\omega\hat{x})^2 - im\omega(\hat{x}\hat{p}) - \hat{p}\hat{x}] \\ &= \frac{1}{2m\hbar\omega}[\hat{p}^2 + (m\omega\hat{x})^2] - \frac{i}{2\hbar}[\hat{x}, \hat{p}] \end{aligned}$$

, where between the third and the bottom line, we employed the **commutator** defined by

$$\boxed{[\hat{\alpha}, \hat{\beta}] = \hat{\alpha}\hat{\beta} - \hat{\beta}\hat{\alpha}} \tag{7.70}$$

Next, let use see what happens if we calculate the commutator of $[\hat{x}, \hat{p}]$. To get insight into how the commutator works, we can introduce a "test function" $f(x)$:

$$\begin{aligned} [\hat{x}, \hat{p}]f(x) &= x(-i\hbar)\frac{df}{dx} - (-i\hbar)\frac{d}{dx}(xf) \\ &= -i\hbar\left(x\frac{df}{dx} - x\frac{df}{dx} - f\right) \\ &= i\hbar f(x) \end{aligned}$$

That is, we obtain the following expression:

$$\boxed{[\hat{x}, \hat{p}] = i\hbar} \tag{7.71}$$

This is called the **canonical commutation relation**. Recall the section that we discussed Hamiltonian mechanics, x and p play the leading roles there because they are canonical conjugate quantities. If the commutator leads to zero that means the quantities are commute. However, in quantum mechanics, we see that $\hat{x}$ and $\hat{p}$ are not commute, this motivates Heisenberg proposed the "uncertainty principle" where we have discussed earlier.

Then, we can employed the results to our calculations about $aa^\dagger$. Repeat the same procedure for $a^\dagger a$, we obtain:

$$\begin{aligned} aa^\dagger &= \frac{1}{\hbar\omega}\hat{H} + \frac{1}{2}, \quad \text{where} \quad \hat{H} = \left(aa^\dagger - \frac{1}{2}\right)\hbar\omega \\ a^\dagger a &= \frac{1}{\hbar\omega}\hat{H} - \frac{1}{2}, \quad \text{where} \quad \hat{H} = \left(a^\dagger a + \frac{1}{2}\right)\hbar\omega \end{aligned} \tag{7.72}$$

Note the condition

$$[a, a^\dagger] = 1 \tag{7.73}$$

is satisfied. Now consider we have an energy ε and see what the ladder operators act on the Schrödinger equation:

$$\begin{aligned} \hat{H}a|\varepsilon\rangle &= (a\hat{H} - [a, \hat{H}])|\varepsilon\rangle \\ &= (a\hat{H} - a)|\varepsilon\rangle \\ &= (\varepsilon - 1)a|\varepsilon\rangle \end{aligned} \tag{7.74}$$

$$\begin{aligned} \hat{H}a^\dagger|\varepsilon\rangle &= (a^\dagger\hat{H} - [a^\dagger, \hat{H}])|\varepsilon\rangle \\ &= (a^\dagger\hat{H} + a^\dagger)|\varepsilon\rangle \\ &= (\varepsilon + 1)a^\dagger|\varepsilon\rangle \end{aligned} \tag{7.75}$$

So far we have noticed that the operator a lowers down the energy while the operator $a^\dagger$ raises the energy level up. That is essentially the reason why they are called the "ladder operators".

It is worthwhile to notice that we cannot use the lowering operator to lower down the energy levels forever because

$$a|\varepsilon_0\rangle = 0 \tag{7.76}$$

This implies

$$a^\dagger a|\varepsilon_0\rangle = 0$$

In other words,

$$\left(\hat{H} - \frac{1}{2}\right)|\varepsilon_0\rangle = 0$$

The Schrödinger equation will be

$$\hat{H}|\varepsilon_0\rangle = \frac{1}{2}|\varepsilon_0\rangle$$

The lowest energy eigenvalue will be $\varepsilon_0 = 1/2$. That means for a given n-th energy level, the eigenvalue will be

$$\varepsilon_n = n + \frac{1}{2}, \quad n = 0, 1, 2, ... \tag{7.77}$$

Therefore, the energies are given by

$$\boxed{E_n = \left(n + \frac{1}{2}\right)\hbar\omega} \tag{7.78}$$

This is the well-known expression for energy levels in quantum mechanics. This indicates that the ground state has a non-zero energy

$$E_0 = \frac{1}{2}\hbar\omega \tag{7.79}$$

In QFT, such an energy is called the **zero-point energy**, which implies even in the empty space of the vacuum, there is still some fluctuation about the energy. This also shows the energy can be "quantized".

Back to the topic about the ladder operators, it is also useful to express the canonical operators in terms of a and $a^\dagger$:

$$\begin{aligned} \hat{x} &= \sqrt{\frac{\hbar}{2m\omega}}(a^\dagger + a) \\ \hat{p} &= i\sqrt{\frac{m\hbar\omega}{2}}(a^\dagger - a) \end{aligned} \tag{7.80}$$

If we calculate the commutator, we will get the same result as shown in Eq.(7.71).

Besides, given any state $|n\rangle$, the ladder operators have the following relations:

$$\begin{aligned} a|n\rangle &= \sqrt{n}|n-1\rangle \\ a^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle \end{aligned} \tag{7.81}$$

In other words,

$$\begin{aligned}\langle n'|a|n\rangle &= \sqrt{n}\langle n'|a|n-1\rangle = \sqrt{n}\delta_{n',n-1}\\ \langle n'|a^\dagger|n\rangle &= \sqrt{n+1}\langle n'|a|n+1\rangle = \sqrt{n+1}\delta_{n',n+1}\end{aligned} \tag{7.82}$$

where δ is the Kronecker delta notation. Finally, we have not actually solved the Schrödinger equation for quantum harmonic oscillator yet, that is,

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \frac{1}{2}m\omega^2 x^2\psi = E\psi \tag{7.83}$$

The direct solution is:

$$\psi(x) = \frac{1}{\sqrt{2^n n!}}\left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} e^{-\frac{m\omega x^2}{2\hbar}} H_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right) \tag{7.84}$$

where $H_n(y)$ is the **Hermite polynomials** and $n = 0, 1, 2, ...$ For instance, $H_0(y) = 1$, $H_1(y) = 2y$, and so forth.

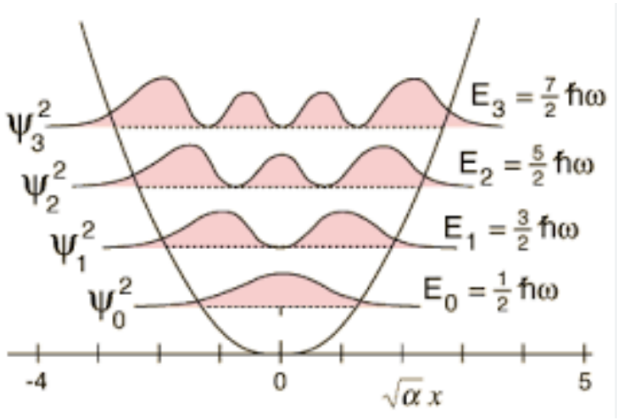

Figure 98: The probability distributions and the energy levels of the quantum harmonic oscillator. *Source*: Hyperphysics.

7.10 3-D Schrödinger equation (spherically symmetric potential)

So far we discussed a series of cases about 1-D Schrödinger equation. Now, we can consider the general case in three dimension. It would be better to consider the spherical symmetric case first, since it will be useful to illustrate the hydrogen atom later. First of all, we need to recall the spherical coordinate system and Eq.(1.7). Basically, what we are going to do is to rewrite the 1-D derivative d^2/dx^2 as the 3-D differential operator ∇^2 (which is, technically, called the "Laplacian"). In this way, the 3-D Schrödinger equation has the following form:

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \tag{7.85}$$

Explicitly, we have

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}\right] + V\psi = E\psi \tag{7.86}$$

where the Laplacian in spherical coordinates is defined by

$$\nabla^2 \equiv \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{7.87}$$

Indeed, Eq.(7.86) looks very tricky to solve, compared to the typical second-order linear differential equation we discussed previously. Nevertheless, to solve this sort of differential equation, we can separate the variables.

In this case, since we have the radial component (r) and the angular components (θ and ϕ), the wavefunction can be written as

$$\psi(r,\theta,\phi) = R(r)Y(\theta,\phi) \tag{7.88}$$

Next, we plug this into Eq.(7.86) and then divide that by $-2mr^2/\hbar^2$, we obtain:

$$\left[\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{2mr^2}{\hbar^2}(V(r)-E)\right] + \frac{1}{Y}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right] = 0$$

Now we notice this equation is consist of the radial part and the angular part (the first and the second bracket on the left-hand side), and adding up both together gives zero. Let us introduce a new constant l and write down a set of two differential equations (we shall explain the reason later):

$$\begin{aligned}
&\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{2mr^2}{\hbar^2}(V(r)-E) = l(l+1) \quad \text{(Radial equation)} \\
&\frac{1}{Y}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right] = -l(l+1) \quad \text{(Angular equation)}
\end{aligned} \tag{7.89}$$

Now we can try to solve the radial and the angular equation respectively.

7.10.1 The radial part

The radial part of the 3-D Schrödinger equation can be written as

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{2mr^2}{\hbar^2}[V(r)-E]R = l(l+1)R \tag{7.90}$$

To solve this, we can make a substitution $u \equiv rR(r)$, and we have

$$\frac{dR}{dr} = \frac{1}{r^2}\left(r\frac{du}{dr} - u\right)$$

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) = r\frac{d^2u}{dr^2}.$$

In this way, our equation can be expressed as

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dr^2} + \left[\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} + V\right]u = Eu \tag{7.91}$$

, where

$$V_{\text{eff}} = V + \underbrace{\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2}}_{\text{centrifugal term}} \tag{7.92}$$

is the effective potential. Note the shape of the potential $V(r)$ only affects the radial part $R(r)$ of the equation.

It would be good to illustrate an example. Let us assume we have an "infinite spherical well",

$$V(r) = \begin{cases} 0 & r \le a \\ \infty & r > a \end{cases} \tag{7.93}$$

Simply speaking, this model is considering quantum physics confined in a sphere. Inside the sphere, the radial part of the Schrödinger equation (7.91) becomes

$$\frac{d^2u}{dr^2} = \frac{l(l+1)}{r^2}u - k^2u \quad \text{where} \quad k \equiv \frac{\sqrt{2mE}}{\hbar} \tag{7.94}$$

This looks a little bit familiar with what we have seen before. However, the $l(l+1)$ term might trouble us to solve that. In fact, this type of differential equation is not that uncommon – we are glad that mathematicians already had the general solution to this type of equation:

$$u(r) = Ar\, j_l(kr) + Br\, n_l(kr) \tag{7.95}$$

Here involves two special functions:

$$j_l(x) = (-x)^l \left(\frac{1}{x}\frac{d}{dx}\right)^l \frac{\sin x}{x} \tag{7.96}$$

is called the **spherical Bessel function**, and

$$n_l(x) = -(-x)^l \left(\frac{1}{x}\frac{d}{dx}\right)^l \frac{\cos x}{x} \tag{7.97}$$

is called the **spherical Neumann function**. For example, if $l = 0$, we have

$$j_0(x) = \frac{\sin x}{x} \quad \text{and} \quad n_0(x) = -\frac{\cos x}{x}$$

Note that at $x = 0$, the spherical Bessel function is finite whereas the spherical Neumann function blows up. Therefore, we can demand $B = 0$, and the solution becomes

$$R(r) = A\, j_l(kr). \tag{7.98}$$

7.10.2 The angular part

Now, let us dig into the angular equation (7.89). The equation can be rewritten as

$$\sin\theta \frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial Y}{\partial\theta}\right) + \frac{\partial^2 Y}{\partial\phi^2} = -l(l+1)\sin^2\theta\, Y \tag{7.99}$$

When solving differential equation, one useful tip is to observe the number of variables. In this case, because we have θ, ϕ, that might be a good motivation for use to solve by separating the variables. As what we have seen at the beginning of this section, we can write $Y(\theta, \phi)$ in terms of two parts:

$$Y(\theta, \phi) \equiv \Theta(\theta)\, \Phi(\phi) \tag{7.100}$$

Plugging into the differential equation and divide that by $\Theta\Phi$ gives a set of two equations:

$$\begin{aligned} &\frac{1}{\Theta}\left[\sin\theta \frac{d}{d\theta}\left(\sin\theta \frac{d\Theta}{d\theta}\right)\right] + l(l+1)\sin^2\theta = m^2 \quad (\Theta\text{-equation}) \\ &\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = -m^2 \quad (\Phi\text{-equation}) \end{aligned} \tag{7.101}$$

, where we have introduced another constant, m. We will see the reason immediately. It is apparent that the second equation is much easier to solve. The solution will be

$$\Phi(\phi) = e^{im\phi} \tag{7.102}$$

Note that the exponent ensures the periodic condition, that is,

$$\Phi(\phi) = \Phi(\phi + 2\pi)$$

Meanwhile, the Euler formula indicates

$$e^{2i\pi m} = 1$$

Hence, the condition of m will be

$$m = 0, \pm 1, \pm 2, ...$$

Then, we can look into the Θ-equation in Eq.(7.101). The solution to this differential equation is

$$\Theta(\theta) = A\, P_l^m(\cos\theta). \tag{7.103}$$

Here A is a constant, $P_l^m(\cos\theta)$ can be defined by the following relation:

$$P_l^m(x) = (-1)^m(1-x^2)^{\frac{m}{2}}\left(\frac{d}{dx}\right)^m P_l(x) \tag{7.104}$$

This is called the **associated Legendre polynomials**, which is a very common solution when solving the Laplace's equation (the equation takes the form such that $\nabla^2 f = 0$) in spherical coordinate system. On the right-hand side, the $P_l(x)$ is called the **Legendre polynomials**, which is defined by

$$P_l(x) = \frac{1}{2^l l!}\left(\frac{d}{dx}\right)^l (x^2-1)^l \tag{7.105}$$

This relation is also called the **Rodrigues' formula**. For instance, $P_0(x) = 1$, $P_1(x) = x$, and so forth.....In fact, this is not the first time we meet these relations. In the section we discussed the electric potential, we also solved the similar equation by separating the variables, and we got the pretty similar result, as shown in Eq.(4.30).

Note l has to be a non-negative integer. Also, it is required to have $m \leq l$. If $m > l$, $P_l^m = 0$ does not make sense. Therefore, we have

$$\begin{cases} l = 0, 1, 2, ... \\ m = -l, -l+1, ..., -1, 0, 1, ..., l \end{cases} \tag{7.106}$$

In short, we notice that there are $(2l+1)$ possible values of m. Formally, the l is called the **azimuthal quantum number** and m is called the **magnetic quantum number** (specifically, it is called the orbital magnetic quantum number, often denoted by m_l, in case of getting confused with the spin magnetic quantum number). In general, the l decides the shape of the orbitals, and the m determines the orientation of atomic orbitals. We sill visit these again in the next section.

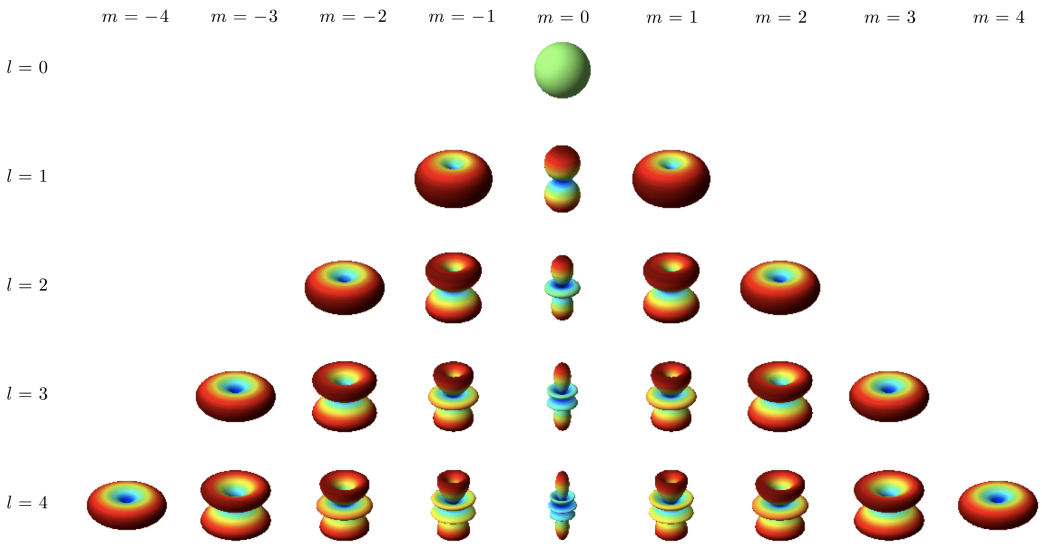

Figure 99: Examples of the first few spherical harmonics from $l = 0$ to $l = 4$. Note the number of m is determined by the value of l. *Source*: http://opticaltweezers.org/

Finally, we can normalize the constant and get an expression for $Y(\theta, \phi)$ in terms of l, m:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\, P_l^m(\cos\theta)\, e^{im\phi} \tag{7.107}$$

This is called the **spherical harmonics**, which is the solution to the angular part of the 3-D Schrödinger equation. All of these functions describe the shapes of the atomic orbitals, as shown in Figure 99.

7.11 The hydrogen atom

One practical example is the hydrogen atom. Physicists figured out the atomic orbitals have different shells, subshells for electrons, and each of them has different shapes. To consider the hydrogen atom's case, we need to consider the Coulomb potential

$$V(r) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r} \tag{7.108}$$

Since the potential affects the radial equation only, we can plug this into the 3-D Schrödinger equation (7.91):

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dr^2} + \left[\frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} - \frac{e^2}{4\pi\epsilon_0}\frac{1}{r}\right]u = Eu. \tag{7.109}$$

Note here e denotes the charge and m is the electron's mass. Rewriting the radial equation gives

$$\frac{d^2u}{dr^2} = \left[\frac{l(l+1)}{(\kappa r)^2} - \frac{me^2}{2\pi\epsilon_0\hbar^2\kappa}\frac{1}{\kappa r} + 1\right]\kappa^2 u \quad \text{where} \quad \kappa \equiv \frac{\sqrt{-2mE}}{\hbar}. \tag{7.110}$$

Again, this equation looks tricky for use to solve. The best way to solve this is to rewrite the component in terms of r, as what we did in the previous section, so we can introduce

$$\rho \equiv \kappa r \quad \text{and} \quad \rho_0 \equiv \frac{me^2}{2\pi\epsilon_0\hbar^2\kappa}$$

Thus, the equation can be rearranged as

$$\frac{d^2u}{d\rho^2} = \left[\frac{l(l+1)}{\rho^2} - \frac{\rho_0}{\rho} + 1\right]u \tag{7.111}$$

Still, this equation looks not that simple for us to tackle. But one way to guide us to the right track is to test the conditions under the limits. We can test the conditions when ρ goes to infinity and when ρ goes to zero. Under these limits, we can solve the equation much simply and put certain constraints on the general solutions. Here is the systematics:

1. $\rho \to \infty$:,

$$\begin{aligned}
&\frac{d^2u}{d\rho^2} = u \\
&\Longrightarrow u(\rho) = Ae^{-\rho} + Be^{\rho}, \quad B = 0 \quad \text{since} \quad e^{\rho} \to \infty \quad \text{when} \quad \rho \to \infty \\
&\Longrightarrow u(\rho) = Ae^{-\rho}
\end{aligned}$$

2. $\rho \to 0$:,

$$\begin{aligned}
&\frac{d^2u}{d\rho^2} = \frac{l(l+1)}{\rho^2}u \\
&\Longrightarrow u(\rho) = C\rho^{l+1} + D\rho^{-l}, \quad D = 0 \quad \text{since} \quad \rho^{-l} \to \infty \quad \text{when} \quad \rho \to 0 \\
&\Longrightarrow u(\rho) = C\rho^{l+1}
\end{aligned}$$

Then, we can combine two results together and make a connection by a new variable, $\xi(\rho)$. Moreover, we can take the second derivative and try to reproduce what we have had in Eq.(7.111):

$$\begin{aligned}
&u(\rho) = \rho^{l+1}e^{-\rho}\,\xi(\rho) \\
&\longrightarrow \frac{du}{d\rho} = \rho^l e^{-\rho}\left[(l-\rho+1)\xi + \rho\frac{d\xi}{d\rho}\right] \\
&\longrightarrow \frac{d^2u}{d\rho^2} = \rho^l e^{-\rho}\left[\left(-2l+\rho-2+\frac{l(l+1)}{\rho}\right)\xi + 2(l-\rho+1)\frac{d\xi}{d\rho} + \rho\frac{d^2\xi}{d\rho^2}\right]
\end{aligned} \tag{7.112}$$

Compare with Eq.(7.111), we have the following equation:

$$\rho\frac{d^2\xi}{d\rho^2} + 2(l-\rho+1)\frac{d\xi}{d\rho} + [\rho_0 - 2(l+1)]\xi = 0 \tag{7.113}$$

One tip to solve the differential equations is using the power series, and then we can take the first and second derivatives:

$$\begin{aligned}
\xi(\rho) &= \sum_{j=0}^{\infty} c_j \rho^j \\
\frac{d\xi}{d\rho} &= \sum_{j=0}^{\infty} c_{j+1}(j+1)\rho^j \\
\frac{d^2\xi}{d\rho^2} &= \sum_{j=0}^{\infty} c_{j+1} j(j+1)\rho^{j-1}
\end{aligned} \tag{7.114}$$

,where we have used the dummy index $j \to j+1$ from the second line. Then, we can insert these three terms back to Eq.(7.113) and rearrange, which gives

$$j(j+1)c_{j+1} + 2(l+1)(j+1)c_{j+1} - 2jc_j[\rho_0 - 2(l+1)]c_j = 0$$

Then, we obtain the recursion relation:

$$\frac{c_{j+1}}{c_j} = \frac{2(j+l+1)-\rho_0}{(j+1)(j+2l+2)} \tag{7.115}$$

We can test the upper limit of the relation, that is, let $j \to \infty$,

$$\frac{c_{j+1}}{c_j} \xrightarrow{j\to\infty} \frac{2}{j+1}$$

, which implies

$$c_j \approx \frac{2^j}{j!}c_0$$

Consequently, our function $\xi(\rho)$ can be expressed as

$$\xi(\rho) = c_0 \sum_{j=0}^{\infty} \frac{2^j}{j!}\rho^j = c_0 e^{2\rho} \tag{7.116}$$

Thus, combined with the first line of Eq.(7.112), our solution takes the form such like

$$u(\rho) = c_0\, \rho^{l+1} e^{\rho} \tag{7.117}$$

Now, we are going to bring our main topic from mathematics back to physics. Consider when the series Eq.(7.115) terminates, we can intuitively assume $2(j+l+1) - \rho_0 = 0$. Let us define

$$n \equiv j + l + 1$$

so that

$$\rho_0 = 2n$$

In the meantime, the energy can be written as

$$E = -\frac{\hbar^2\kappa^2}{2m} = -\frac{me^4}{8\pi^2\epsilon_0^2\hbar^2\rho_0^2} \tag{7.118}$$

This is the energy for an electron orbiting around the hydrogen atom's shell. If we consider $\rho_0 = 2n$ and calculate the constant, we obtain the expression for energy given by a certain energy level:

$$\boxed{E_n = -13.6\,\text{eV}\,\frac{1}{n^2}} \quad \text{where} \quad n = 1, 2, ... \tag{7.119}$$

This is called the **Bohr's model** of the hydrogen atom. Here eV represents an electronvolt. This formula can be used to calculate the energy an electron processes on a certain energy level. Another empirical extension of this expression is called the **Rydberg formula** (recall $E = hc/\lambda$):

$$\boxed{\frac{1}{\lambda} = R\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)} \tag{7.120}$$

, where the **Rydberg constant** R for hydrogen is about 1.096×10^7 m^{-1}. This expression describes the wavelength corresponding to the different transitions between energy levels of an atom. In other words, the spectral lines essentially depend on the transition between the initial state n_i and the final state n_f.

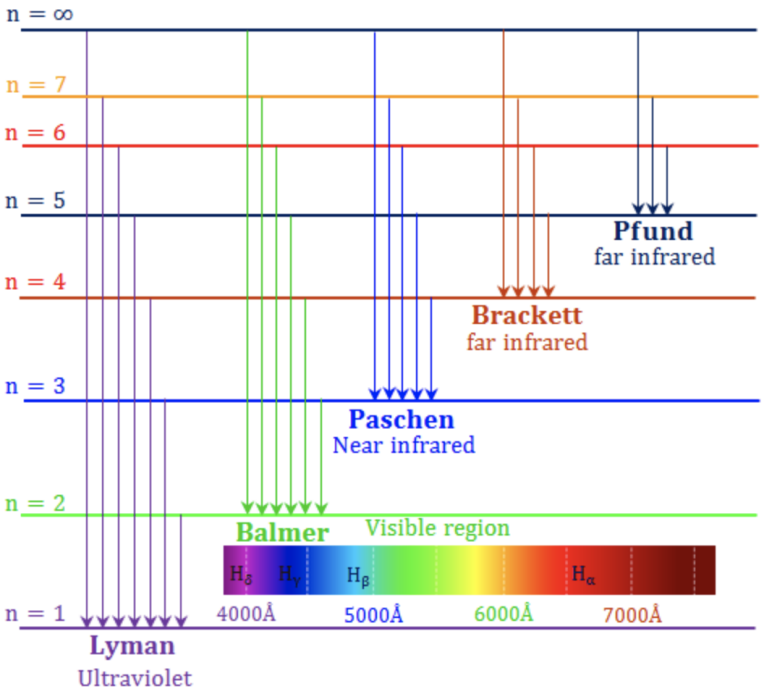

Figure 100: The hydrogen spectral lines: The emission series correspond to the electron's transitions from the different energy levels. The visible region is basically contained in the **Balmer series**. *Source*: https://www.priyamstudycentre.com/2019/02/hydrogen-spectrum.html

Another implication of the Bohr's hydrogen model can be discussed from Eq.(7.118). We can rewrite the differential equation's constant κ as

$$\kappa = \left(\frac{me^2}{4\pi\epsilon_0\hbar^2}\right)\frac{1}{n} \equiv \frac{1}{a_0 n}$$

, where

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} \approx 5.29 \times 10^{-11} \quad \text{m} \tag{7.121}$$

is the **Bohr radius**. The Bohr radius represents the most probable distance between a ground-state electron and a nucleus in a hydrogen atom, and this can be roughly understood as the radius of an atom (the scale is around an angstrom 1 Å$= 10^{-10}$ m).

The last topic is to write down the complete form of our hydrogen wavefunctions. By normalizing the factor, we obtain:

$$\begin{aligned}\psi_{nlm} &= R_{nl}(r)\, Y_l^m(\theta,\phi) \\ &= \sqrt{\left(\frac{2}{a_0 n}\right)^3 \frac{(n-l-1)!}{2n(n+l)!}} e^{-\frac{r}{a_0 n}} \left(\frac{2r}{a_0 n}\right)^l \left[L_{n-l-1}^{2l+1}\left(\frac{2r}{a_0 n}\right)\right] Y_l^m(\theta,\phi)\end{aligned} \tag{7.122}$$

where $Y_l^m(\theta,\phi)$ is given by Eq.(7.107), and L_q^p is the **associated Laguerre polynomials**:

$$L_q^p(x) = (-1)^p \left(\frac{d}{dx}\right)^p L_{p+q}(x) \tag{7.123}$$

, which contains the Laguerre polynomials

$$L_q(x) = \frac{e^x}{q!}\left(\frac{d}{dx}\right)^q (e^{-x}x^q) \tag{7.124}$$

The first few hydrogen wavefunctions are shown as Table 9.

Table 9: Hydrogen wavefunctions from $n = 1$ to 2.

$\boldsymbol{n}$	$\boldsymbol{l}$	$\boldsymbol{m}$	$\boldsymbol{\psi_{nlm}(r,\theta,\phi)}$
1	0	0	$\frac{1}{\sqrt{\pi a_0^3}} e^{-r/a_0}$
2	0	0	$\frac{1}{4\sqrt{2\pi a_0^3}}\left(2-\frac{r}{a_0}\right)e^{-r/2a_0}$
2	1	0	$\frac{1}{4\sqrt{2\pi a_0^3}}\left(\frac{r}{a_0}\right)e^{-r/2a_0}\cos\theta$
2	1	±1	$\frac{1}{8\sqrt{3\pi a_0^3}}\left(\frac{r}{a_0}\right)e^{-r/2a_0}\sin\theta e^{\pm i\phi}$

Another fact we have to notice is about the indices n, l, and m. Previously we mentioned that l represents the azimuthal quantum number and m_l represents the magnetic quantum number. As for the n, it represents the **principal quantum number**, which tells the energy and the average distance between an electron and a nucleus. In summary:

1. **The principal quantum number**: $n = 1, 2, 3, ...$, describes the atomic energy levels (equivalently, the average distance between the electron and the nucleus).
2. **The azimuthal quantum number**: $l = 0, 1, ..., n-1$, describes the shapes of electron's orbitals.
3. **The magnetic quantum number**: $m_l = -l, -l+1, ..., -1, 0, 1, ..., l-1, l$, describes the orientation of electron's orbitals.
4. **The spin quantum number**: $s = \pm 1/2$ for electrons (fermions), describes the spin (will be discussed in the next section).

The orbitals have different names, depending on different l values. For example, $s, p, d, f, ...$ correspond to $l = 0, 1, 2, 3$ and so forth. All of these can be found in Table 10. I will give you a tip to remember all of these at the end of this chapter.

Table 10: Quantum Numbers

n	l	m_l	m_s	Number of orbitals	Orbital name	Number of electrons	Total electrons
1	0	0	$\pm\frac{1}{2}$	1	$1s$	2	2
2	0	0	$\pm\frac{1}{2}$	1	$2s$	2	8
	1	$-1, 0, 1$	$\pm\frac{1}{2}$	3	$2p$	6	
3	0	0	$\pm\frac{1}{2}$	1	$3s$	2	18
	1	$-1, 0, 1$	$\pm\frac{1}{2}$	3	$3p$	6	
	2	$-2, -1, 0, 1, 2$	$\pm\frac{1}{2}$	5	$3d$	10	

7.12 Angular momentum and spin

We have seen that in quantum mechanics, position and momentum will be expressed as the operators, $\hat{x}$ and $\hat{p}$. Certainly, the angular momentum has the same fashion (note: in the text below I am not going to use the "hat" symbol, but the reader should recongize that the angular momentum is an operator). Classically, the angular momentum is defined by

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}$$

Based on the definition of the cross product, we can figure out each component:

$$\begin{aligned} L_x &= yp_z - zp_y \qquad p_x = -i\hbar\partial_x \\ L_y &= zp_x - xp_z \qquad p_y = -i\hbar\partial_y \\ L_z &= xp_y - yp_x \qquad p_z = -i\hbar\partial_z \end{aligned}$$

If we employ all of these into the commutators, we will finally verify:

$$\begin{aligned} [L_x, L_y] &= i\hbar L_z \\ [L_y, L_z] &= i\hbar L_x \\ [L_z, L_x] &= i\hbar L_y \end{aligned} \tag{7.125}$$

But if we now consider the square of vector magnitude, $L^2 = L_x^2 + L_y^2 + L_z^2$, we will find that

$$[L^2, L_x] = [L^2, L_y] = [L^2, L_z] = 0$$

or,

$$[L^2, \mathbf{L}] = 0. \tag{7.126}$$

That is, L^2 commute with the components of $\mathbf{L}$.

Sometimes, it would be useful to write down the angular momentum operators as the ladder operators,

$$L_\pm = L_x \pm iL_y \tag{7.127}$$

One can verify the following commutator relations:

$$[L^2, L_\pm] = 0 \quad \text{and} \quad [L_z, L_\pm] = \pm i\hbar L_\pm \tag{7.128}$$

In addition, when we apply the ladder operators on the state, we should get something "lower" or "higher". In this case, we have the following expressions:

$$\boxed{L_\pm |l, m_l\rangle = \sqrt{(l \mp m_l)(l \pm m_l + 1)}\hbar\, |l, m_l \pm 1\rangle} \tag{7.129}$$

Using the ladder form of the angular momentum operators can derive the quantization relations:

$$\boxed{\begin{aligned} L^2|l,m_l\rangle &= l(l+1)\hbar^2\,|l,m_l\rangle \\ L_z|l,m_l\rangle &= m_l\hbar\,|l,m_l\rangle \end{aligned}} \tag{7.130}$$

Generally speaking, these mean that if you measure the LHS, you will get the results such as the RHS. Moreover, these show the angular momentum is quantized.

So far we discussed the operators in Cartesian coordinates. If we consider the spherical coordinate system, the angular momentum vector can be expressed as

$$\begin{aligned} \mathbf{L} &= -i\hbar(\mathbf{r}\times\nabla) \\ &= -i\hbar\left[r\underbrace{(\hat{r}\times\hat{r})}_{=0}\frac{\partial}{\partial r} + \underbrace{(\hat{r}\times\hat{\theta})}_{=\hat{\phi}}\frac{\partial}{\partial\theta} + \underbrace{(\hat{r}\times\hat{\phi})}_{=-\hat{\theta}}\frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right] \\ &= -i\hbar\left(\hat{\phi}\frac{\partial}{\partial\theta} - \hat{\theta}\frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right) \end{aligned}$$

Here, we employed the unit vectors' properties in the spherical coordinates. Rewrite the unit vectors in terms of Cartesian components, we will find that:

$$\begin{cases} L_x &= -i\hbar\left(-\sin\phi\frac{\partial}{\partial\theta} - \cos\phi\cot\theta\frac{\partial}{\partial\phi}\right) \\ L_y &= -i\hbar\left(\cos\phi\frac{\partial}{\partial\theta} - \sin\phi\cot\theta\frac{\partial}{\partial\phi}\right) \\ L_z &= -i\hbar\frac{\partial}{\partial\phi} \end{cases} \tag{7.131}$$

The last relation, L_z, is the most useful one. As for the square of angular momentum, we have

$$L^2 = -\hbar^2\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) \tag{7.132}$$

Now, let us dig into another important topic – **spin**. Technically, "spin" is an intrinsic property about the angular momentum of a particle itself. Spin quantum number is typically denoted by s, whereas the spin magnetic quantum number is denoted by m_s, which determines the spin along a specified axis. It is somewhat easy to be thought of as the rotation of the particle itself because it is associated with angular momentum, while this is not exact. We should remind ourselves that spin is an "intrinsic" property such as mass or charge; we can simply refer to "spin" as "a kind of angular momentum". One way to get insight into spin is to categorize it in particle physics:

1. **Fermions**: $s = \frac{1}{2}, \frac{3}{2}, ...$ such as leptons (e.g. electrons, neutrinos) and baryons (e.g. protons, neutrons).
2. **Bosons**: $s = 0, 1, 2, ...$ such as photons, gluons, mesons (e.g. pions, kaons), etc.

Therefore, fermions have half-integer spins whereas bosons have integer spins. As for the spin magnetic number, it has $m_s = -s, -s+1, ..., s-1, s$.

Since spin is directly associated with the angular momentum operator, we can write down the form that it acts on a certain state in terms of ladder operators:

$$\boxed{S_\pm|s,m_s\rangle = \sqrt{(s\mp m_s)(s\pm m_s+1)}\hbar\,|s,m_s\pm1\rangle} \tag{7.133}$$

Similarly, we have the following quantization quantity:

$$\boxed{\begin{aligned} S^2|s,m_s\rangle &= s(s+1)\hbar^2\,|s,m_s\rangle \\ S_z|s,m_s\rangle &= m_s\hbar\,|s,m_s\rangle \end{aligned}} \tag{7.134}$$

In quantum mechanics, an electron is one of the most useful cases to explore the properties related to spin, which has $s = 1/2$. We can write down the general state representation in the form of a **spinor**:

$$\chi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha|\uparrow\rangle + \beta|\downarrow\rangle \tag{7.135}$$

where

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

represent spin-up and spin-down, respectively. We can employ to get the electron's z-component based on Eq.(7.134). Since an electron has $s = 1/2$, we have

$$S_z|\uparrow\rangle = \frac{\hbar}{2}|\uparrow\rangle \quad \text{and} \quad S_z|\downarrow\rangle = -\frac{\hbar}{2}|\downarrow\rangle$$

Or equivalently, we can write it in matrix representation:

$$S_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7.136}$$

Now, we can consider the case for the x and y component. Meanwhile, Eq.(7.133) tells us that

$$S_+|\downarrow\rangle = \hbar|\uparrow\rangle \quad \text{and} \quad S_-|\uparrow\rangle = \hbar|\downarrow\rangle.$$

Also, note that $S_+|\uparrow\rangle = S_-|\downarrow\rangle = 0$. Then we have

$$S_+ = \hbar\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad S_- = \hbar\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Because $S_\pm = S_x \pm iS_y$, we can rearrange and get the following quantities:

$$S_x = \frac{1}{2}(S_+ + S_-) \quad \text{and} \quad S_y = \frac{1}{2i}(S_+ - S_-)$$

Thus, we obtain

$$S_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad S_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{7.137}$$

Eventually, for a spin-1/2 particle, we have a compact form for the spin operator:

$$\boxed{\mathbf{S} = \frac{\hbar}{2}\boldsymbol{\sigma}} \tag{7.138}$$

, where

$$\boxed{\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}} \tag{7.139}$$

are the **Pauli matrices**, which are a set of a Hermitian, unitary 2×2 complex matrices.

The last concept is the **total angular momentum**, **J**, defined by

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \tag{7.140}$$

This can be simply viewed as the addition of two vectors – the angular momentum and the spin (Figure 101). Similarly, $j = l + s$, and it takes a certain range of the allowed values

$$|l - s| \leq j \leq |l + s|$$

It has the same properties as what we had so far:

$$\boxed{J_\pm|j, m\rangle = \sqrt{(j \mp m)(j \pm m + 1)}\hbar\,|j, m \pm 1\rangle} \tag{7.141}$$

and

$$\boxed{\begin{aligned} J^2|j,m\rangle &= j(j+1)\hbar^2\,|j,m\rangle \\ J_z|j,m\rangle &= m\hbar\,|j,m\rangle \end{aligned}} \tag{7.142}$$

, where $m = -j, -j+1, ..., j-1, j$.

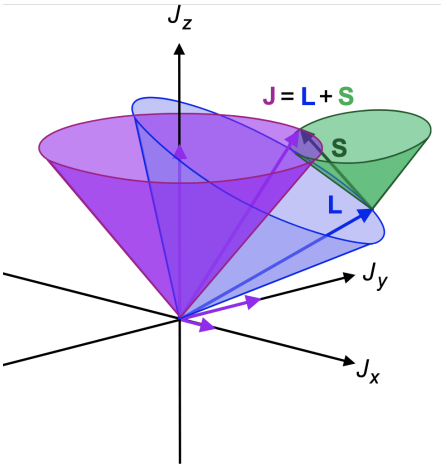

Figure 101: The graphical interpretation of the total angular momentum **J**, the momentum **L**, and the spin **S**. *Source*: Wikipedia

In the section we discussed quantum statistics, we also mentioned the **Pauli exclusion principle**, which indicates that two identical fermions (e.g. electrons) cannot occupy the same quantum state. Physically, "identical" means two particles have the same quantum numbers. In other words, two electrons occupied in the same atomic orbital must have opposite spins. This principle stems from the intrinsic properties about the wavefunction's symmetry and antisymmetry about bosons and fermions. The result helps physicists to explain the stability of the atomic structure.

When studying quantum mechanics, some people might be confused with those quantum numbers. I personally prefer to use the following way to describe: In order to find where the electron locates in an atom, we can think of the principal quantum number n as the number of a "building", l represents which "floor" it locates, m_l represents which "room" it lives, and finally, s represents the "bed" inside the room – Pauli exclusion principle tells us that the spin-up and spin-down electrons cannot occupy the same quantum state, and this can be imagined as that a boy and a girl are not allowed to share the same bed – well, nature is sometimes more orthodox than what you think!

In conclusion, quantum physics explains several phenomena in our daily lives. It also predicted that everything is composed of elementary particles (almost everything can be "quantized"). All of our electronic devices such as computers, smartphones, LEDs, microscopes, and so forth can be attributed to the invention of quantum mechanics.

8 Epilogue

Through these seven chapters so far, we have covered most of the topics related to undergraduate physics courses. Almost all of them will be revisited in graduate school (if you decided to learn physics after college).

Even if you are not going to explore physics in the future, you might have already been aware that physics is a fundamental subject about everything. From the past to today, science have played a more and more crucial role for people.

If you refer to Figure 102, you will find how useful physics is. So far, we have covered the topics about classical mechanics, thermodynamics and statistical mechanics, electromagnetism and optics, special relativity, and quantum mechanics. These are fundamental physics subjects. Based on that diagram, you will find that almost everything about STEM (science, technology, engineering, and mathematics) is related to these subjects. If we go down, we will see that sociology is at the bottom, and that does not mean this is not important but mean the subject is much closer to human interactions and our daily lives. I will probably revise the diagram and leave "philosophy" at the bottom because it is the most fundamental subject to enable humans to think about how our world works.

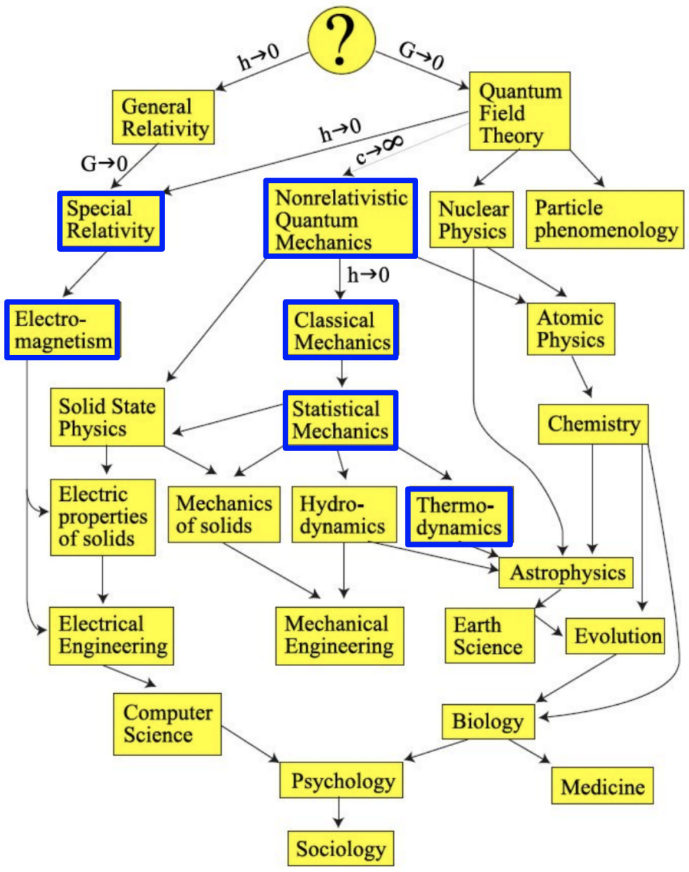

Figure 102: The tree of knowledge: So far we have covered the topics with the blue box in detail. *Source*: https://www.pinterest.com/pin/540502392771519299/

If we continue digging into physics more deeply, we will realize that all of those subjects can be described in the framework of general relativity (GR) and quantum field theory (QFT). The first one pictures the spacetime we live macroscopically, while the second one illustrates the elementary particles consist of everything microscopically. One of the ultimate objective for ambitious physicists is to "unify" GR and QFT. Hence, the top of this diagram is usually called "Theory of Everything" (ToE), string theory is one of the most popular candidates (particularly in the period between 1970s and 1990s). However, this might be very challenging to achieve, since GR and QFT seems to be incompatible at this point. Today's physics focuses

much more on technology (e.g. quantum computing) or experiment aspects (e.g. particle experiments with machine learning).

Yet, everything will change, as shown in Figure 103. The evolution of knowledge shows that science has become more and more significant especially after 16th century, and the complexity of knowledge has been increasing exponentially, owing to the appearance of the Internet.

Learning physics enables us to think about the most fundamental aspect of our world, and mathematics will be the tool for us to write down the logic behind nature. In fact, if we look into all of the subjects – mechanical engineering, biology, computer science, or economics – under the most fundamental scale, we will find things are somewhat random but obey the logic of nature. For example, a computer is composed of the transistors where the signals are controlled by electric current, our neural system is worked by the mechanism of the ion channel and the electric potential difference, the Black-Scholes model implies the financial market can be described by a set of partial differential equations derived from the Brownian motion of particles, and so forth. In other words, physics might not be everything, but everything can be related to physics.

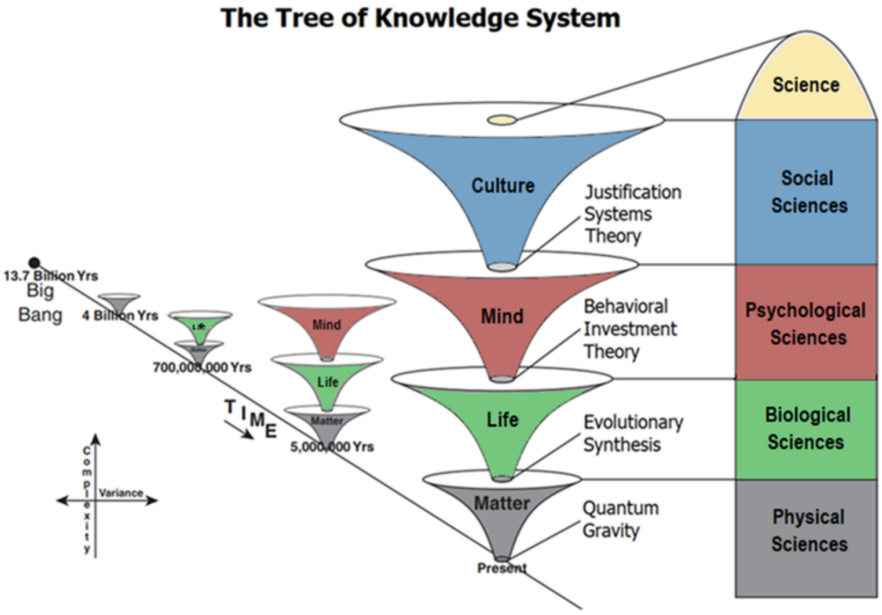

Figure 103: The "tree of knowledge system", which shows the evolution of knowledge. *Source*: Wikipedia

References

[1] D. Halliday, J. Walker, and R. Resnick, *Principles of Physics*. Wiley, 2020.

[2] J. R. Taylor, *Classical Mechanics*. University Science Books, 2005.

[3] D. V. Schroeder, *An Introduction to Thermal Physics*. Oxford University Press, 2021.

[4] D. J. Griffiths, *Introduction to Electrodynamics*. Cambridge University Press, 2017.

[5] D. J. Griffiths, *Introduction to Quantum Mechanics*. Cambridge University Press, 2018.