# A Bridge to Modern Physics 

John F. Lindner<br>Physics Department<br>The College of Wooster

2021 January 9


## Contents

List of Tables ..... 7
List of Figures ..... 9
1 Introduction ..... 11
1.1 Two Teasers ..... 11
1.1.1 First Teaser: Chasing a Light Beam ..... 11
1.1.2 Second Teaser: Sorting Photons ..... 13
1.2 Two Jewels ..... 16
1.2.1 First Jewel: Extremal Action \& Newton's 2nd Law ..... 16
1.2.2 Second Jewel: Maxwell's Equations \& Light Waves ..... 17
Introduction Problems ..... 21
2 Relativistic Physics ..... 23
2.1 Flat Spacetime of Special Relativity ..... 23
2.1.1 Time Dilation ..... 23
2.1.2 Length Contraction ..... 25
2.1.3 Muon Decay Example ..... 27
2.1.4 Clock Desynchronization \& Simultaneity Relativity ..... 28
2.1.5 Tethered Rockets Example ..... 29
2.1.6 Lorentz-Einstein Transformation ..... 30
2.1.7 Recovering the Fundamental Effects ..... 31
2.1.8 Velocity Addition ..... 32
2.1.9 Spacetime Rotations ..... 33
2.1.10 Spacetime Interval ..... 34
2.1.11 Twin Paradox ..... 36
2.1.12 Loedel Spacetime Diagrams ..... 38
2.1.13 Recovering the Fundamental Effects (Again) ..... 39
2.1.14 Pole-in-the-Barn Example ..... 39
2.1.15 Faster-Than-Light Implies Backward-In-Time ..... 40
2.1.16 Magnetism is Relativistic Electricity ..... 41
2.2 Relativistic Dynamics ..... 43
2.2.1 Spacetime 4-Vectors ..... 43
2.2.2 Invariant Mass ..... 45
2.2.3 Einstein's Box Derivation of $E=m c^{2}$ ..... 46
2.2.4 Compton Scattering ..... 48
2.2.5 Pair Production ..... 50
2.3 Curved Spacetime of General Relativity ..... 51
2.3.1 Gravity and Electricity ..... 51
2.3.2 The Equivalence Principle ..... 52
2.3.3 Gravitational Time Dilation ..... 53
2.3.4 Weakly Curved Spacetime ..... 55
2.3.5 Newton's Laws from Curved Spacetime ..... 56
Relativity Problems ..... 58
3 Quantum Physics ..... 63
3.1 Interference and Superposition ..... 63
3.1.1 Beam Splitter Probabilities ..... 63
3.1.2 Two Interpretations ..... 65
3.1.3 Mach-Zehnder Interferometer ..... 66
3.1.4 Quantum Eraser ..... 67
3.1.5 Interaction-Free Measurement ..... 68
3.1.6 Quantum Computing ..... 69
3.2 Indeterminacy and Entanglement [Optional] ..... 70
3.2.1 Mach-Zehnder Classical Model ..... 70
3.2.2 Mach-Zehnder Quantum Model 1 ..... 72
3.2.3 Mach-Zehnder Quantum Model 2 ..... 72
3.2.4 Hilbert Space ..... 75
3.2.5 Quantum Evolution ..... 75
3.2.6 The Measurement Problem (and Schrödinger's Cat) ..... 75
3.2.7 Polarization ..... 77
3.2.8 Heisenberg Indeterminacy ..... 78
3.2.9 Crossed Polarizers ..... 79
3.2.10 Entangled States (and Schrödinger's Kittens) ..... 80
3.2.11 EPR-Bohm Experiment ..... 81
3.2.12 Bell's Inequality ..... 82
3.3 Feynman to Schrödinger [Optional] ..... 83
3.3.1 Feynman Postulate ..... 83
3.3.2 Path Integral ..... 85
3.3.3 Recovering Classical Mechanics ..... 86
3.3.4 Schrödinger Equation ..... 87
3.4 Simple Schrödinger Solutions ..... 89
3.4.1 Analogies ..... 89
3.4.2 Schrödinger Wave Equation Simply ..... 91
3.4.3 Probability Conservation ..... 92
3.4.4 Schrödinger Wave Packets ..... 93
3.4.5 Separation of Variables ..... 93
3.4.6 Hamiltonian Eigenfunctions ..... 94
3.4.7 Qualitative Solutions ..... 96
3.4.8 Particle in a Box ..... 97
3.4.9 Particle in a Rigid Box ..... 99
3.4.10 Escape from the Box: Negative Kinetic Energy? ..... 101
3.4.11 Escape from the Box: STM! ..... 101
3.4.12 Quantum Tunneling in $\mathrm{NH}_{3}$ ..... 103
3.4.13 Band Structure of Solids ..... 106
3.5 Quantum Harmonic Oscillator ..... 106
3.5.1 Classical Harmonic Oscillator ..... 107
3.5.2 Heuristic Spectrum ..... 107
3.5.3 Dimensionless Variables ..... 108
3.5.4 Asymptotic Behavior ..... 109
3.5.5 Power Series Solution ..... 110
3.5.6 Power Series Diverges ..... 110
3.5.7 Truncate Series ..... 111
3.5.8 Standard Form Solutions ..... 111
3.5.9 Classical Correspondence ..... 113
Quantum Problems ..... 115
A Mathematics Background ..... 121
A. 1 Complex Numbers ..... 121
A. 2 Hyperbolic Functions ..... 122
A. 3 Spatial Rotations ..... 124
A. 4 Partial Derivatives ..... 125
A. 5 Function Notation ..... 125
Problems ..... 126
Bibliography ..... 127

## List of Tables

2.1 Space vs. spacetime: Einstein is the new Pythagoras. ..... 37
2.2 Angle vs. twin "paradoxes". ..... 37
3.1 Contrasting Newton with Schrödinger. ..... 90
3.2 Contrasting Maxwell with Schrödinger. ..... 90
3.3 First few harmonic oscillator eigenvalues and eigenfunctions. ..... 112

## List of Figures

1.1 Dipole antenna ..... 12
1.2 Light-pulse clock ..... 13
1.3 Calcite crystal ..... 13
1.4 Polarization sorters ..... 14
1.5 Photon and a $\oplus$ sorter ..... 15
1.6 Photon and a $\oplus$ sorter \& antisorter ..... 15
1.7 Real path and nearby imaginary path ..... 17
1.8 Transverse electromagnetic wave ..... 19
2.1 Perpendicular light-pulse clock ..... 23
2.2 Parallel light-pulse clock ..... 25
2.3 Perpendicular lengths don't contract ..... 26
2.4 Photographing a relativistic grid ..... 27
2.5 Cosmic ray muons ..... 28
2.6 Relativity of simultaneity ..... 29
2.7 Tethered rockets ..... 30
2.8 Observers in relative motion ..... 30
2.9 Spacetime rotation ..... 34
2.10 Light cones ..... 36
2.11 Loedel spacetime diagram ..... 38
2.12 Graphical relativity ..... 39
2.13 Pole-in-the-barn spacetime ..... 40
2.14 Instanton communication ..... 41
2.15 Parallel currents attract ..... 42
2.16 Antiparallel currents repel ..... 42
2.17 4-velocity \& 4-momentum ..... 44
2.18 Loedel momentum-energy diagram ..... 45
2.19 Einstein's box ..... 47
2.20 Einstein's box in spacetime ..... 47
2.21 Compton scattering ..... 48
2.22 Pair production ..... 50
2.23 Equivalence principle ..... 52
2.24 Gravity deflects light ..... 53
2.25 Curved spacetime ..... 56
3.1 Beam splitter ..... 64
3.2 Beam splitter \& single photons ..... 64
3.3 Mach-Zehnder interferometer ..... 66
3.4 Mach-Zehnder interferometer \& single photons ..... 67
3.5 Quantum eraser ..... 68
3.6 Interaction-free measurement ..... 69
3.7 Quantum computing ..... 70
3.8 Rotating arrow ..... 72
3.9 Rotating arrow \& single photons ..... 73
3.10 Photon interferometer states ..... 74
3.11 Schrödinger cat states ..... 76
3.12 Linear polarizer ..... 79
3.13 Crossed polarizers \& single photons ..... 80
3.14 Decay of positronium ..... 81
3.15 EPR-Bohm experiment ..... 82
3.16 Double slit interference ..... 84
3.17 From slits to space ..... 86
3.18 Wave packet ..... 93
3.19 Generic potential \& eigenfunction ..... 97
3.20 Electric trap ..... 98
3.21 Finite square well ..... 98
3.22 Infinite square well ..... 99
3.23 Rigid box classical correspondence ..... 100
3.24 STM model potential energy ..... 102
$3.25 \mathrm{NH}_{3}$ inversion ..... 103
$3.26 \mathrm{NH}_{3}$ potential energy model ..... 104
$3.27 \mathrm{NH}_{3}$ spectrum ..... 105
3.28 Band structure ..... 106
3.29 Three Spectra ..... 108
3.30 Oscillator eigenfunctions ..... 112
3.31 Oscillator classical correspondence ..... 113
A. 1 Complex plane ..... 122
A. 2 Hyperbolic functions ..... 123
A. 3 Rotated coordinates ..... 124

## Chapter 1

## Introduction

Two thought experiments whet our appetite for modern physics, while two jewels summarize classical physics.

### 1.1 Two Teasers

This text is an introduction to relativistic (or spacetime) and quantum physics, the twin pillars of modern physics, which profoundly changed the way we think about the world. This section is a "teaser" for the text. Here, we will consider two thought experiments involving light. "Chasing a Light Beam" introduces relativity, and "Sorting Photons" introduces quantum physics.

### 1.1.1 First Teaser: Chasing a Light Beam

The 16-year-old Albert Einstein wondered what riding on a beam of light would be like. Classically, light is a transverse electromagnetic wave, where a changing electric field induces a changing magnetic field and vice versa, like the dipole radiation of Figure 1.1. Can one chase a light beam and thereby cause the oscillating and self-inducing electric and magnetic fields to slow and even stop?

No! If the fields of Figure 1.1 were static, they would violate Maxwell's equations (as, for example, static electric field lines don't form closed loops). Einstein realized that not being able to slow light by chasing it exposes a larger problem: The mechanics of Newton is inconsistent with the electromagnetism of Maxwell.

Newton's laws embody Galileo's principle of relativity, according to which the laws of physics are the same for all observers in uniform (nonaccelerating) motion. Since all such observers can consider themselves at rest, all uniform motion is relative, even if the relativity is often implicit. For example, on a road you drive $100 \mathrm{~km} / \mathrm{hr}$ relative to the pavement, $200 \mathrm{~km} / \mathrm{hr}$ relative to an oncoming car, and $0 \mathrm{~km} / \mathrm{hr}$ relative to a trailing car.


Figure 1.1: Dipole antenna of oscillating current radiates an electromagnetic wave. (Because the net electric field vanishes at large distances, we only draw the $1 / r$ transverse dynamic fields and not the $1 / r^{2}$ radial static fields.)

However, Maxwell's equations imply that light propagates at the speed

$$
\begin{equation*}
c=\frac{1}{\sqrt{\epsilon_{0} \mu_{0}}} \approx 10^{9} \mathrm{~km} / \mathrm{hr} \tag{1.1}
\end{equation*}
$$

where $\epsilon_{0}$ and $\mu_{0}$ are the SI electric and magnetic constants. (For example, $\epsilon_{0}$ appears in Gauss's law while $\mu_{0}$ appears in Ampère's law.) So, light speed $c$ is about a billion $\mathrm{km} / \mathrm{hr}$, but relative to what? It's a billion $\mathrm{km} / \mathrm{hr}$, period. It is our universe's unique invariant speed.

Can the relativity of motion be reconciled with the invariancy of the speed of light? Something must give! Einstein realized that Maxwell's equations had been successfully tested at both low and high speeds, in static and optical experiments, but Newton's laws had only been tested at low speeds (compared to a billion $\mathrm{km} / \mathrm{hr}$ ). Hence, he generalized the latter to conform to the former.

To get an idea of what's involved, consider a light-pulse clock, in which a light pulse is reflected back to its source by a mirror, as in Figure 1.2, at relative rest and in relative motion. Because light speed is the same for all observers, light takes less time to travel the compact, folded path at relative rest than the expanded, kinked path in relative motion, as the hypotenuse of a right triangle is longer than either of its legs. Thus, a clock in relative motion ticks slowly compared to a clock at relative rest.

As Einstein discovered, and as we shall derive, in order that all observers in relative motion measure the same invariant speed of light, their measured values of space and time intervals - and their definitions of "now" - must be relative. They will agree on the speed of light, but disagree about lengths and durations and simultaneity. However, this will have dramatic consequences


Figure 1.2: One tick-tock of a light-pulse clock at relative rest (left) and in relative motion (right). The vertical paths (left) are shorter than the diagonal paths (right).
only for relative speeds approaching the billion $\mathrm{km} / \mathrm{hr}$ speed of light. We will elucidate these consequences later.

### 1.1.2 Second Teaser: Sorting Photons

Optically anisotropic materials can sort light according to its polarization (the oscillation direction of its electric field). For example, because of its crystal structure, calcite is birefringent with different indices of refraction for electric fields perpendicular and parallel to its optic axis, as illustrated in Figure 1.3. We can schematically represent the action of the calcite by a box with one input and two outputs, as in Figure 1.4. We can convert a vertical and horizontal $\oplus$ sorter into a $\pm 45^{\circ}$ diagonal $\otimes$ sorter by rotating the calcite. For bright classical light, if diagonally polarized light is input to a $\oplus$ sorter, then half of the input light intensity will appear in each output channel. Similarly, if vertically polarized light is input to a $\otimes$ sorter, then half of the input light intensity will appear in each output channel.


Figure 1.3: A calcite crystal sorts classical light into vertical (up-down arrows) and horizontal (in-out dots) polarizations. Thus, a black disk seen through calcite appears as two partially overlapping gray disks, which disappear alternately when viewed through a rotating polarizer.


Figure 1.4: Schematic diagrams of diagonally polarized bright light of intensity $I_{0}$ input to a $\oplus$ sorter (left) and vertically polarized bright light input to a $\otimes$ sorter (right).

What happens if we repeat this experiment with very faint light? Near the beginning of the twentieth century, the Einstein photoelectric effect and the Compton scattering experiment demonstrated the "granularity" of faint light. In fact, they suggested that light consists of particles whose energy is proportional to the light's classical temporal frequency

$$
\begin{equation*}
E=\hbar \omega \tag{1.2}
\end{equation*}
$$

whose momentum is proportional to the light's classical spatial frequency

$$
\begin{equation*}
\vec{p}=\hbar \vec{k} \tag{1.3}
\end{equation*}
$$

and whose spin angular momentum corresponds to the light's classical (circular) polarization

$$
\begin{equation*}
\vec{S}= \pm \hbar \hat{k} \tag{1.4}
\end{equation*}
$$

where the common proportionality $\hbar=h / 2 \pi$ (pronounced "h bar") is Planck's reduced constant. These particles are now called photons. Classical wave-like light emerges from a large ensemble of particle-like photons. Electrons and other subatomic particles (and even atoms and molecules ... ) exhibit similar wave-particle duality. Such "wavicles" or "matter-waves" have been called "the dreams that stuff are made of".

We can use neutral density filters or crossed polarizers to reduce the intensity of the light so on average only one photon is in our sorter at any one time. (Alternately, we could use the Section 3.1.1 single photon source to guarantee only one photon in the sorter.) To count the photons in each of the output channels, we can use photomultiplier tubes or avalanche photodiodes, which exploit the photoelectric effect to convert a single photon into a macroscopic cascade of electrons. (Alternately, we could use a frog's eye, which is apparently sensitive to single photons!) Each experimental trial will report " 1 " if it detects a photon and " 0 " if it does not.

So, we input a diagonally polarized photon to a $\oplus$ sorter. What happens? The input photon must emerge in one of the two output channels, if only because it must go somewhere, but why would it emerge in one channel and not the other? In fact, as illustrated in Figure 1.5, the experiment is not repeatable,
which is itself a disaster for classical physics; rather, the photon emerges half the time in each channel, randomly. We cannot predict into which output channel any given input photon will emerge, but we can predict the probability that it will emerge in either channel, and the equal probabilities of $1 / 2$ correspond well with the classical, bright-light result. The apparent indeterminism of the individual trials of this experiment is in striking disagreement with classical physics and is a hallmark of quantum mechanics.


Figure 1.5: Diagonally polarized photons input one-by-one to a $\oplus$ sorter emerge randomly but equally in each output channel (left). Similarly, vertical polarized photons input to a $\otimes$ sorter emerge randomly (right). The experiment is not repeatable, except statistically!

Even more classically strange is what happens if we recombine the two output channels with a reversed $\oplus$ sorter, as in Figure 1.6. Now the experiment is repeatable and determined! If the $\oplus$ sorter randomizes the diagonal polarization, how does the recombination preserve it? Surely, the diagonally polarized photon entering the first $\oplus$ sorter can not "know" it will be recombined by the second, reversed $\oplus$ sorter?


Figure 1.6: One $\oplus$ sorter randomizes the diagonal polarization but adding a second, recombining, reversed $\oplus$ sorter preserves the diagonal polarization.

Classically, the diagonal light can be thought of as a superposition of horizontal and vertical light and constructive interference between the two channels can preserve its polarization. We will return later to investigate such superposition and interference experiments involving single photons.

### 1.2 Two Jewels

Before beginning our study of modern physics, we first review or introduce two jewels of classical physics: we will deduce Newton's second law from the principle of extremal action and infer electromagnetic waves from Maxwell's equations. Later, we will generalize the action principle to curved spacetime and quantum mechanics, and we will compare and contrast the electromagnetic wave equation with the Schrödinger matter-wave equation of quantum mechanics.

### 1.2.1 First Jewel: Extremal Action \& Newton's 2nd Law

Toss a ball, and it goes up and comes down in a certain time, while traveling a certain path (probably some approximation to a parabola). Imagine, instead, that it went via some other path, in the same time. If we calculate the kinetic energy at every moment, subtract the potential energy, and sum over the path, the result will always be larger for the imagined path and least for the real path. Alternately, since the time is the same for all paths, the average kinetic energy less the average potential energy is minimized by the real path.

More formally, in one spatial dimension, the path is some function $x[t]$. As is traditional in this context, we label the kinetic energy $T=\frac{1}{2} m \dot{x}^{2}$ and the potential energy $V$. Their sum is the total energy $E=T+V$ and their difference is the Lagrangian $L=T-V$. The time integral of the Lagrangian is the action,

$$
\begin{equation*}
S[x[t]]=\int_{t_{1}}^{t_{2}} L[x[t], \dot{x}[t]] d t \tag{1.5}
\end{equation*}
$$

The action $S$ is extremal or stationary for the real path $x[t]$. (It can be a maximum, but in classical mechanics, it is usually a minimum and the principal of extremal action is often called the principle of least action.) Finding the real path is then analogous to the ordinary calculus problem of finding the extremum of a function. However, note that the action $S$ is a functional of the path $x[t]$. (A function maps numbers to numbers, while a functional maps functions to numbers.) Hence, we will need to employ variational calculus to determine the real path.

If $x[t]$ is the real path, consider a path $x[t]+\xi[t]$, where the variation $\xi[t]$ vanishes at the end points $\xi\left[t_{1}\right]=0=\xi\left[t_{2}\right]$, as in Figure 1.7. If the action for the real path is

$$
\begin{equation*}
S[x]=\int_{t_{1}}^{t_{2}}\left(\frac{1}{2} m \dot{x}^{2}-V[x]\right) d t \tag{1.6}
\end{equation*}
$$

then the action for the nearby path is

$$
\begin{equation*}
S[x+\xi]=\int_{t_{1}}^{t_{2}}\left(\frac{1}{2} m(\dot{x}+\dot{\xi})^{2}-V[x+\xi]\right) d t \tag{1.7}
\end{equation*}
$$

Expanding the integrand to first order in the small quantities $\xi$ and $\dot{\xi}$ using the Taylor series $V[x+\xi]=V[x]+\xi V^{\prime}[x]+O\left[\xi^{2}\right]$, we find that

$$
\begin{equation*}
S[x+\xi]=S[x]+\delta S, \tag{1.8}
\end{equation*}
$$



Figure 1.7: Real path and nearby imaginary path.
where

$$
\begin{equation*}
\delta S=\int_{t_{1}}^{t_{2}}\left(m \dot{x} \dot{\xi}-\xi V^{\prime}[x]\right) d t . \tag{1.9}
\end{equation*}
$$

In order to factor the perturbation $\xi$ out of the integrand, we integrate its first term $m \dot{x} \dot{\xi}$ by parts. This involves moving the derivative from one factor to the other, incurring a minus sign and a vanishing boundary term,

$$
\begin{equation*}
\int_{t_{1}}^{t_{2}} m \dot{x} \dot{\xi} d t=-\int_{t_{1}}^{t_{2}} m \ddot{x} \xi d t \tag{1.10}
\end{equation*}
$$

For the action $S$ to be an extremum, it must be stationary, and hence its first variation $\delta S$ must vanish. Consequently, to ensure that $S$ is a extremum, we demand

$$
\begin{equation*}
0=\delta S=\int_{t_{1}}^{t_{2}}\left(-m \ddot{x}-V^{\prime}[x]\right) \xi[t] d t \tag{1.11}
\end{equation*}
$$

Since this is true for all variational functions $\xi[t]$, the rest of the integrand must vanish identically. (If the integrand were nonzero near some $t^{*}$, a variation $\xi^{*}$ that vanished everywhere except for a blip near $t^{*}$ would nonzero the integral.) Hence

$$
\begin{equation*}
-V^{\prime}[x]=m \ddot{x} \tag{1.12}
\end{equation*}
$$

which we recognize as Newton's second law of motion, $F_{x}=m a_{x}$. (For example, for a mass $m$ connected to a spring of stretch $x$ and stiffness $k$, the potential energy $V[x]=\frac{1}{2} k x^{2}$ and the force $F_{x}=-V^{\prime}[x]=-k x$, which is Hooke's law.)

Why does it work? Later, we will derive the principle of extremal action from the fundamentals of quantum mechanics!

### 1.2.2 Second Jewel: Maxwell's Equations \& Light Waves

To write Maxwell's equations in integral form, first recall the definitions of flux and circulation. Given a vector field $\vec{v}[\vec{r}]$, define the flux through a surface area
$a$ by

$$
\begin{equation*}
\Phi_{v}=\iint_{a} \vec{v} \cdot d \vec{a} \tag{1.13}
\end{equation*}
$$

and the circulation around a closed loop $l$ by

$$
\begin{equation*}
\Gamma_{v}=\oint_{l} \vec{v} \cdot d \vec{l} \tag{1.14}
\end{equation*}
$$

Flux tells us whether or not the field diverges (or converges), like the electric field in the vicinity of an electric charge. Circulation tells us whether the field circulates, like the magnetic field in the vicinity of an electric current. Physically realistic fields are determined by specifying the flux and circulation everywhere. Maxwell's equations do this for the interrelated electric and magnetic fields.

For a closed surface containing an electric charge $Q$, Gauss's law for electricity is

$$
\begin{equation*}
\epsilon_{0} \Phi_{\mathcal{E}}=Q \tag{1.15}
\end{equation*}
$$

and Gauss's law for magnetism is

$$
\begin{equation*}
\Phi_{\mathcal{B}}=0 \tag{1.16}
\end{equation*}
$$

For an open surface bounded by a closed loop and pierced by an electric current $I=\dot{Q}$, Faraday's law of induction is

$$
\begin{equation*}
\Gamma_{\mathcal{E}}=-\dot{\Phi}_{\mathcal{B}} \tag{1.17}
\end{equation*}
$$

and Ampère's law combined with Maxwell's law of induction is

$$
\begin{equation*}
\Gamma_{\mathcal{B}}=\mu_{0} I+\epsilon_{0} \mu_{0} \dot{\Phi}_{\mathcal{E}} \tag{1.18}
\end{equation*}
$$

In a vacuum, no sources exist, the charges and the currents vanish, and these four Maxwell's equations simplify. Both fluxes vanish everywhere, so that any electric and magnetic fields must form closed loops, and the circulation of one field is induced by the time-varying flux of the other.

Consider a transverse electromagnetic wave polarized in the $\hat{z}$-direction and traveling in $\hat{x}$-direction, as in Figure 1.8, with electric field

$$
\begin{equation*}
\overrightarrow{\mathcal{E}}[x, t]=\hat{z} \mathcal{E}_{m} \sin [k x-\omega t] \tag{1.19}
\end{equation*}
$$

and magnetic field

$$
\begin{equation*}
\overrightarrow{\mathcal{B}}[x, t]=\hat{y} \mathcal{B}_{m} \sin [k x-\omega t] \tag{1.20}
\end{equation*}
$$

where

$$
\begin{equation*}
\frac{\omega}{k}=\frac{2 \pi / T}{2 \pi / \lambda}=\frac{\lambda}{T}=c \tag{1.21}
\end{equation*}
$$

We will show that this is a solution to Maxwell's equation provided $c=1 / \sqrt{\epsilon_{0} \mu_{0}}$ $\left(\right.$ and $\left.\mathcal{E}_{m}=c \mathcal{B}_{m}\right)$.


Figure 1.8: Transverse electromagnetic wave. Time-varying flux of one field through an infinitesimal area induces the circulation of the other field around the boundary of the area.

In order to apply Faraday's law of induction to the $x z$-area element of Figure 1.8 , we choose positive circulation to advance in the direction of the $\mathcal{B}$-field according to the right-hand rule. Then, the electric circulation is

$$
\begin{equation*}
\Gamma_{\mathcal{E}}=\mathcal{E} d z-(\mathcal{E}-d \mathcal{E}) d z=d \mathcal{E} d z \tag{1.22}
\end{equation*}
$$

and the magnetic flux is

$$
\begin{equation*}
\Phi_{\mathcal{B}}=+\mathcal{B} d x d z \tag{1.23}
\end{equation*}
$$

and so

$$
\begin{equation*}
d \mathcal{E} d z=-\dot{\mathcal{B}} d x d z \tag{1.24}
\end{equation*}
$$

or

$$
\begin{equation*}
\partial_{x} \mathcal{E}=\frac{\partial \mathcal{E}}{\partial x}=-\frac{\partial \mathcal{B}}{\partial t}=-\partial_{t} \mathcal{B} \tag{1.25}
\end{equation*}
$$

Similarly, in order to apply Maxwell's law of induction to the $x y$-area element of Figure 1.8, we choose positive circulation to advance in the direction of the $\mathcal{E}$-field according to the right-hand rule. Then, the magnetic circulation is

$$
\begin{equation*}
\Gamma_{\mathcal{B}}=-\mathcal{B} d y+(\mathcal{B}-d \mathcal{B}) d y=-d \mathcal{B} d y \tag{1.26}
\end{equation*}
$$

and the electric flux is

$$
\begin{equation*}
\Phi_{\mathcal{E}}=+\mathcal{E} d x d y \tag{1.27}
\end{equation*}
$$

and so

$$
\begin{equation*}
-d \mathcal{B} d y=\epsilon_{0} \mu_{0} \dot{\mathcal{E}} d x d y \tag{1.28}
\end{equation*}
$$

or

$$
\begin{equation*}
-\partial_{x} \mathcal{B}=-\frac{\partial \mathcal{B}}{\partial x}=\epsilon_{0} \mu_{0} \frac{\partial \mathcal{E}}{\partial t}=\epsilon_{0} \mu_{0} \partial_{t} \mathcal{E} \tag{1.29}
\end{equation*}
$$

We differentiate Equation 1.25 with respect to $x$ and Equation 1.29 with respect to $t$. Because the partial derivatives of physical functions commute, we get the wave equation

$$
\begin{equation*}
\partial_{x}^{2} \mathcal{E}=\epsilon_{0} \mu_{0} \partial_{t}^{2} \mathcal{E} \tag{1.30}
\end{equation*}
$$

which for the electromagnetic wave of Equation 1.19 reduces to

$$
\begin{equation*}
-k^{2} \mathcal{E}=-\epsilon_{0} \mu_{0} \omega^{2} \mathcal{E} \tag{1.31}
\end{equation*}
$$

and hence

$$
\begin{equation*}
c=\frac{\omega}{k}=\frac{1}{\sqrt{\epsilon_{0} \mu_{0}}} \tag{1.32}
\end{equation*}
$$

In three spatial dimensions, for the electric field $\overrightarrow{\mathcal{E}}$, the wave equation generalizes to

$$
\begin{equation*}
\partial_{x}^{2} \overrightarrow{\mathcal{E}}+\partial_{y}^{2} \overrightarrow{\mathcal{E}}+\partial_{z}^{2} \overrightarrow{\mathcal{E}}=\frac{1}{c^{2}} \partial_{t}^{2} \overrightarrow{\mathcal{E}} \tag{1.33}
\end{equation*}
$$

with a similar equation for the magnetic field $\overrightarrow{\mathcal{B}}$. This can be compactly written as

$$
\begin{equation*}
\square^{2} \overrightarrow{\mathcal{E}}=\overrightarrow{0} \tag{1.34}
\end{equation*}
$$

where the d'Alembertian operator

$$
\begin{equation*}
-\square^{2}=\nabla^{2}-\frac{1}{c^{2}} \partial_{t}^{2}=\partial_{x}^{2}+\partial_{y}^{2}+\partial_{z}^{2}-\frac{1}{c^{2}} \partial_{t}^{2} \tag{1.35}
\end{equation*}
$$

generalizes the Laplacian operator $\nabla^{2}$ from space to spacetime. (Note that the pseudo-letters $\partial$ and $\nabla$, which are pronounced "del", are introduced in analogy with the Greek letters $\delta$ and $\Delta$, which are pronounced "delta".)

## Introduction Problems

1. Refraction Index. You may have learned that light "slows" to a speed $v=c / n$ when passing though glass of refraction index $n$. What could this mean? Doesn't light always travel at $c$ ? Explain classically what happens inside the glass as light passes through.
2. Variational Calculus. Parallel our derivation of Newton's second law from the principle of "least" (or stationary) action to find the shortest path $y[x]$ connecting two points.
(a) Show that the length of the path is

$$
\begin{equation*}
\ell[y[x]]=\int_{x_{1}}^{x_{2}} d x \sqrt{1+y^{\prime}[x]^{2}} \tag{1.36}
\end{equation*}
$$

(b) Invoke the generalized binomial formula $(1+\epsilon)^{\rho} \sim 1+\rho \epsilon$ for $\epsilon \ll 1$ to show that $\ell[y+\eta]=\ell[y]+\delta \ell$, where the variation

$$
\begin{equation*}
\delta \ell=\int_{x_{1}}^{x_{2}} d x \frac{y^{\prime}[x] \eta^{\prime}[x]}{\sqrt{1+y^{\prime}[x]^{2}}} \tag{1.37}
\end{equation*}
$$

(c) Integrate the above variation by parts and use the blip argument to conclude the stationary path is a straight line. Can you show that it is a minimum and not a maximum?

## 3. Wave Equation.

(a) Explicitly verify that $\overrightarrow{\mathcal{E}}[z, t]=\hat{x} \mathcal{E}_{0} e^{-(\omega t-k z)^{2}}$ satisfies the wave equation $\square^{2} \overrightarrow{\mathcal{E}}=\overrightarrow{0}$ provided $\omega=k c$ by explicitly evaluating all the partial derivatives.
(b) Sketch the wave at two different times.

## Chapter 2

## Relativistic Physics

Albert Einstein's theories of relativity revolutionized our understanding of space and time.

### 2.1 Flat Spacetime of Special Relativity

We begin by investigating ideas emerging from Einstein's 1905 theory of special relativity.

### 2.1.1 Time Dilation

The light-pulse clock, which was introduced in our first teaser, is an idealized clock whose tick-tock involves a pulse of light bouncing back and forth between parallel mirrors. This famous device illustrates the special relativistic phenomena of time dilation and length contraction and thereby probes the structure of space and time.



Figure 2.1: One tick-tock of a light-pulse clock in relative motion perpendicular to its length. Successive images of the clock show the reflection and return of a pulse. The light path is shown kinked in space.

When at relative rest, as on the left in Figure 1.2, the clock's proper length is $l_{0}$. Its proper time is $\Delta t_{0}=2 l_{0} / c$, the duration of its tick-tock. (In relativity, the adjective "proper" refers to quantities measured at relative rest.)

Suppose the clock is in relative motion at speed $v$ perpendicular to its length, as in Figure 2.1. Assuming light travels at the invariant speed $c$ for a time $\Delta t$ along the kinked path, the Pythagorean theorem implies

$$
\begin{equation*}
\left(c \frac{\Delta t}{2}\right)^{2}=\left(v \frac{\Delta t}{2}\right)^{2}+l_{0}^{2} \tag{2.1}
\end{equation*}
$$

where the duration of the tick-tock is now

$$
\begin{equation*}
\Delta t=\frac{2 l_{0}}{\sqrt{c^{2}-v^{2}}}=\frac{2 l_{0} / c}{\sqrt{1-(v / c)^{2}}} \tag{2.2}
\end{equation*}
$$

or

$$
\begin{equation*}
\Delta t=\gamma \Delta t_{0} \geq \Delta t_{0} \tag{2.3}
\end{equation*}
$$

where the relativistic stretch (or Lorentz factor) is

$$
\begin{equation*}
\gamma=\frac{1}{\sqrt{1-(v / c)^{2}}} \geq 1 \tag{2.4}
\end{equation*}
$$

Thus, clocks in relative motion are observed to tick slowly compared to clocks at relative rest, a phenomenon called time dilation. This applies to all clocks, including our heart beats (and physiological aging) because, if another kind of clock went out-of-synch with the light-pulse clocks, we could use that fact to absolutely distinguish motion from rest, in violation of the principle of relativity embodied by Newton's laws.

If the light were instead a massive object bouncing up and down at a speed $v_{m}<c$ when the clock is at relative rest, then when the clock is in relative motion, the vertical component of the ball's velocity would be $v_{m}$, its horizontal component would be $v$, and its total speed would be $\sqrt{v_{m}^{2}+v^{2}}$. As we shall see, this reasoning fails for speeds near $c$, because the rules for adding velocities change.

In 1971, physicists Joe Hafele and Richard Keating flew highly accurate atomic clocks around the world in commercial jets. They found that the airborne clocks were 59 ns slow relative to the ground clocks, as predicted by the theory of relativity. By the 1990s, the Global Positioning Satellite (GPS) system was in wide commercial and military use employing extremely accurate atomic clocks to triangulate positions. GPS doesn't work unless it incorporates time dilation (and other relativistic effects).

While the relativistic stretch $\gamma$ diverges to infinity for speeds $v$ near the billion $\mathrm{km} / \mathrm{hr}$ speed of light $c$, it is nearly unity for everyday speeds, which are very small compared to a billion $\mathrm{km} / \mathrm{hr}$. In fact, using the generalized binomial formula, we can write

$$
\begin{equation*}
\gamma=\left(1-\left(\frac{v}{c}\right)^{2}\right)^{-1 / 2} \sim 1+\frac{1}{2}\left(\frac{v}{c}\right)^{2} \tag{2.5}
\end{equation*}
$$

provided $v \ll c$. Thus, we don't notice time dilation in everyday life. (Only if light speed were much slower, say $c \sim 30 \mathrm{~km} / \mathrm{hr}$, would we have to be careful about our motion relative to family and friends!)

### 2.1.2 Length Contraction



Figure 2.2: One tick-tock of a light-pulse clock in relative motion parallel to its length. Successive images of the clock are expanded in time rightward to prevent overlapping. The light path is shown kinked in spacetime.

Now suppose the light-pulse clock is in relative motion at speed $v$ parallel to its length $l$, which turns out to be different from its proper length $l_{0}$, as in Figure 2.2. Again, assume light travels at the invariant speed $c$ for a time $\Delta t$ along the kinked path. The "tick" time between emission and reflection $\Delta t_{1}$ is longer than the "tock" time between reflection and return $\Delta t_{2}$. In fact, from Figure 2.2, we find the distance relations

$$
\begin{equation*}
l+v \Delta t_{1}=\mathrm{c} \Delta t_{1} \tag{2.6}
\end{equation*}
$$

and

$$
\begin{equation*}
l-v \Delta t_{2}=\mathrm{c} \Delta t_{2} \tag{2.7}
\end{equation*}
$$

so that

$$
\begin{equation*}
\Delta t_{2}=\frac{l}{c+v}<\frac{l}{c-v}=\Delta t_{1} \tag{2.8}
\end{equation*}
$$

Hence, the total tick-tock time is

$$
\begin{equation*}
\Delta t=\Delta t_{1}+\Delta t_{2}=\frac{2 l c}{c^{2}-v^{2}}=\frac{2 l / c}{1-(v / c)^{2}} \tag{2.9}
\end{equation*}
$$

However, time dilation implies

$$
\begin{equation*}
\Delta t=\gamma \Delta t_{0}=\frac{2 l_{0} / c}{\sqrt{1-(v / c)^{2}}} \tag{2.10}
\end{equation*}
$$

Therefore

$$
\begin{equation*}
\frac{2 l / c}{1-(v / c)^{2}}=\frac{2 l_{0} / c}{\sqrt{1-(v / c)^{2}}} \tag{2.11}
\end{equation*}
$$

and

$$
\begin{equation*}
l=l_{0} \sqrt{1-(v / c)^{2}} \tag{2.12}
\end{equation*}
$$

or

$$
\begin{equation*}
l=\frac{l_{0}}{\gamma} \leq l_{0} \tag{2.13}
\end{equation*}
$$

Thus, parallel lengths in relative motion are contracted compared to parallel lengths at relative rest, a phenomenon called length contraction. In contrast, we implicitly assumed in Section 2.1.1 that perpendicular lengths in relative motion do not contract. Suppose they did, as in Figure 2.3, where two similar sticks are in relative motion perpendicular to their lengths. One stick has paint brushes on its ends so that it marks, or does not mark, the other stick when they pass. After the encounter, the presence or absence of the marks determines absolutely which stick was actually in motion, in violation of the principle of relativity. Thus, perpendicular lengths are invariant.


Figure 2.3: If perpendicular lengths contracted in relative motion, we could determine whether the stick with paint brushes on its ends (left) was moving or not by looking for the presence (center) or absence (right) of marks on the other stick after they pass each other.

Suppose two identical rockets L and R are in relative motion. L is shorter relative to $R$, and $R$ is shorter relative to $L$. So, is $R$ shorter than $L$ or isn't it? There is no "is-ness" about it: it's relative!

Can one photograph the length contraction of an object in relative motion? Yes, but the finite speed of light causes significant other distortions to the photograph. For example, imagine photographing a green grid moving left-to-right perpendicular to the line of site at various speeds, as in Figure 2.4. The left edge has a component of velocity toward the camera, and so light from it will be
blue-shifted. Similarly, the right edge has a component of velocity away from the camera, and so light from it will be red-shifted. Light from the top left corner must travel a longer distance to the camera than light from the center, so it must leave earlier to be captured by the camera at the same time as light from the center. Since earlier the grid was translated to the left, vertical lines on the grid will be mapped into hyperbolas on the image. Length contraction is only obvious at extremely high speeds.


$v=0.15 c$

$v=0.3 c$

Figure 2.4: Simulated photographs of a green grid moving left-to-right at increasing speeds displays various aberrations, including Doppler shifting, in addition to length contraction. (Dark gray represents visible light shifted into the infrared and ultraviolet.)

### 2.1.3 Muon Decay Example

By the 1940s, the decay of subatomic particles called muons had provided dramatic evidence of the interrelated special relativistic phenomena of time dilation and length contraction. Muons, which seem to be heavier, "second generation" electrons, are produced when cosmic rays, consisting mainly of energetic protons, strike Earth's atmosphere. In the presence of the nuclei of atmospheric atoms, such as nitrogen, the protons scatter inelastically, so that some of the their kinetic energy converts to sprays of other unstable particles, such as pions, which almost immediately decay to muons and neutrinos. For example,

$$
\begin{equation*}
p^{+} \underset{N}{\longrightarrow} p^{+}+\pi^{0}+\pi^{+}+\pi^{-} \tag{2.14}
\end{equation*}
$$

where almost immediately

$$
\begin{equation*}
\pi^{-} \longrightarrow \mu^{-}+\bar{\nu}_{\mu} \tag{2.15}
\end{equation*}
$$

followed more leisurely by

$$
\begin{equation*}
\mu^{-} \longrightarrow e^{-}+\bar{\nu}_{e}+\nu_{\mu} \tag{2.16}
\end{equation*}
$$

The latter transmutation is an example of radioactive decay. It has an e-folding time $\tau_{0} \sim 2 \mu \mathrm{~s}$ (with a comparable half-life), so that the number of muons remaining after time $t$ is

$$
\begin{equation*}
N[t]=N[0] e^{-t / \tau_{0}} \tag{2.17}
\end{equation*}
$$



Figure 2.5: Cosmic ray shower in the atmosphere produces relativistic muons, as observed by Earth (left) and a muon (right).

Suppose the bulk of the muons are created at an altitude of about $h_{0} \sim$ 9000 m moving at nearly the speed of light $v \sim 0.998 c$, as in Figure 2.5. They travel 300 m in $1 \mu \mathrm{~s}, 600 \mathrm{~m}$ in $2 \mu \mathrm{~s}$, and 9000 m in $30 \mu \mathrm{~s}$. Hence, nonrelativistically, we expect $N[t] / N[0]=e^{-30 \mu \mathrm{~s} / 2 \mu \mathrm{~s}}=e^{-15} \sim 10^{-7}$, or only about one in ten-million muons to survive to Earth's surface. However, the muons' speed corresponds to a relativistic stretch of $\gamma \sim 15$, so that relative to Earth, time dilation stretches their e-folding time to $\gamma \tau_{0} \sim 30 \mu$ s. Hence, relativistically, we expect $N[t] / N[0]=e^{-30 \mu \mathrm{~s} / 30 \mu \mathrm{~s}}=e^{-1} \sim 1 / 3$, or about one third of the muons to survive to Earth's surface! Experiment easily decides in favor of the generous relativistic prediction.

What happens if we adopt the muons' point of view? They are created at relative rest, but due to length contraction, Earth is only $h_{0} / \gamma \sim 600 \mathrm{~m}$ below and rushes up in just $2 \mu \mathrm{~s}$. Hence, we again expect $N[t] / N[0]=e^{-2 \mu \mathrm{~s} / 2 \mu \mathrm{~s}}=$ $e^{-1} \sim 1 / 3$, or about one third of the muons to survive. Earth and muons both compute the same survival fraction, but attribute it to different effects, time dilation or length contraction.

### 2.1.4 Clock Desynchronization \& Simultaneity Relativity

Time dilation and length contraction would not be consistent in every situation without a third interrelated phenomenon, clock desynchronization. Suppose a train is hit by two lightning strikes as it coasts by a train station platform at constant speed. Afterword, scorch marks confirm that the lightning struck the ends of the train when they were coincident with the corresponding ends of the platform, as in Figure 2.6.

The strikes are simultaneous for a platform observer $O$ but not simultaneous for a train observer $O^{\prime}!$ Light from front and rear strikes reach $O$ simultaneously, so $O$ concludes that the strikes were simultaneous. However, light from front strike reaches $O^{\prime}$ first, as $O^{\prime}$ is moving toward the front, so $O^{\prime}$ concludes that the front strike was first. In fact, for $O^{\prime}$, the two strikes were separated by a time

$$
\begin{equation*}
\Delta t^{\prime}=\frac{\Delta l^{\prime}}{v}=\frac{l_{0}-l_{0} / \gamma^{2}}{v}=\frac{l_{0}}{v}\left(1-\frac{1}{\gamma^{2}}\right)=\frac{l_{0} v}{c^{2}} . \tag{2.18}
\end{equation*}
$$



Figure 2.6: For the railway platform observer $O$ (left), the lightning strikes are simultaneous, but for the train observer $O^{\prime}$ (right), they are not. Clocks are labeled assuming they are synchronized in their own frames, and the right clocks of both frames are synchronized at the right lightning strike.

Thus, two clocks synchronized at relative rest are desynchronized in relative motion, and the chasing clock is ahead by the time $l_{0} v / c^{2}$, where $l_{0}$ is the proper separation of the clocks. Many of the so-called "paradoxes" of special relativity can be resolved by correctly incorporating the relativity of simultaneity.

### 2.1.5 Tethered Rockets Example

To illustrate the interconnectedness of the three key kinematical effects of special relativity, time dilation, length contraction, and clock desynchronization, consider the example of the tethered rockets. Earth observes two rockets $L$ and $R$ accelerate identically from relative rest to rendezvous with a uniformly moving space station such that the distance between them always remains constant, as in Figure 2.7. Why then does an inextensible string tethering the midpoints of the rockets break?!

The key to unlocking the solution is the relativity of simultaneity. The rockets are separated in space but accelerate simultaneously relative to Earth. Consequently, they do not accelerate simultaneously relative to the space station. The final proper separation between the rockets, as measured by the space station, is greater than the initial proper separation, as measured by Earth, because according to the space station, $R$ accelerates before $L$ ! According to Earth, the tether must contract in the direction of its motion, and so the rockets' attempt to accelerate the tether rigidly breaks it.

This example thereby illustrates the relativity of rigidity. Relativistically, infinitely rigid objects cannot exists. Tap a steel rod at one end, and the other end moves only after a compression wave travels through the material at the speed of sound, which is always less than the speed of light. If the rod were infinitely rigid, tapping one end would move the other end immediately, thereby enabling us to send messages faster-than-light. As we shall later see, this would


Figure 2.7: Tethered rockets $L-R$ accelerate identically relative to Earth, as observed by Earth $E$ (left) and the space station $S$ (right). If the tether is inextensible, it will break.
allows us to send messages into our past and create causal contradictions.

### 2.1.6 Lorentz-Einstein Transformation

The Lorentz-Einstein [6, 3] transformation provides a formal dictionary relating spacetime observations between uniformly moving observers. It generalizes the corresponding Galilean transformation of Newtonian mechanics, which we first review here. Suppose two observers $O$ and $O^{\prime}$ in relative motion observe a single event, such as a supernova, at coordinates $\{t, x, y, z\}$ and $\left\{t^{\prime}, x^{\prime}, y^{\prime}, z^{\prime}\right\}$, respectively. Assuming they have aligned their coordinate systems so that their relative motion is along the $x$-axis and adjusted their clocks so that they coincide at $t=t^{\prime}=0$, as in Figure 2.8, then their observations are related by

$$
\begin{equation*}
x=x^{\prime}+v t \tag{2.19}
\end{equation*}
$$

where implicitly $t=t^{\prime}$ for all times (and, of course, $y=y^{\prime}$ and $z=z^{\prime}$ ).


Figure 2.8: Two observers in relative motion assign different coordinates to the same event.

Relativistically, we know that due to length contraction, if $O^{\prime}$ measures $x^{\prime}$, then $O$ measures $x^{\prime} / \gamma$, so we generalize the Galilean transformation to $x=$
$x^{\prime} / \gamma+v t$ or

$$
\begin{equation*}
x^{\prime}=\gamma(x-v t) \tag{2.20}
\end{equation*}
$$

A boost to the other coordinate system means replacing $v$ by $-v$ and interchanging primes and unprimes,

$$
\begin{equation*}
x=\gamma\left(x^{\prime}+v t^{\prime}\right) \tag{2.21}
\end{equation*}
$$

Substituting Equation 2.21 into Equation 2.20 and solving for $t$, we find

$$
\begin{equation*}
t=\gamma\left(t^{\prime}+v x^{\prime} / c^{2}\right) \tag{2.22}
\end{equation*}
$$

These equations can be written more symmetrically as

$$
\begin{align*}
c t & =\gamma\left(c t^{\prime}+\frac{v}{c} x^{\prime}\right)  \tag{2.23a}\\
x & =\gamma\left(x^{\prime}+\frac{v}{c} c t^{\prime}\right) \tag{2.23~b}
\end{align*}
$$

Note how the invariant speed converts units of time to units of space. If we measure time in years (for example) and space in light-years, than the speed of light is numerically unity, $c=1 \mathrm{lyr} / \mathrm{yr}$, and the Lorentz-Einstein transformation simplify further to

$$
\begin{align*}
t & =\gamma\left(t^{\prime}+v x^{\prime}\right)  \tag{2.24a}\\
x & =\gamma\left(x^{\prime}+v t^{\prime}\right) \tag{2.24b}
\end{align*}
$$

where the relativistic stretch is $\gamma=1 / \sqrt{1-v^{2}}$. These natural, relativistic units not only emphasize the symmetry between space and time, they also simplify calculations. In matrix notation, being careful to preserve the order of the variables, and in the full $3+1$ dimensions, we have

$$
\left(\begin{array}{c}
t  \tag{2.25}\\
x \\
y \\
z
\end{array}\right)=\left(\begin{array}{cccc}
\gamma & \gamma v & 0 & 0 \\
\gamma v & \gamma & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right)\left(\begin{array}{l}
t^{\prime} \\
x^{\prime} \\
y^{\prime} \\
z^{\prime}
\end{array}\right)
$$

The Lorentz-Einstein transformation formally reduces to the Galilean transformation if $c \rightarrow \infty$, a universe without an invariant speed. The Galilean transformation is a good approximation to the Lorentz-Einstein transformation when speeds are slow and distance small, $v \ll c$ and $x \ll c t$.

### 2.1.7 Recovering the Fundamental Effects

All three of the key kinematical effects of special relativity, time dilation, length contraction, and clock desynchronization, are implicit in the Lorentz-Einstein transformation. Suppose two observers $O$ and $O^{\prime}$ in relative motion observe two
events separated in space and time by $(\Delta t, \Delta x)$ and $\left(\Delta t^{\prime}, \Delta x^{\prime}\right)$. Taking finite differences of Equation 2.24, these separations are related by

$$
\begin{align*}
& \Delta t=\gamma\left(\Delta t^{\prime}+v \Delta x^{\prime}\right)  \tag{2.26a}\\
& \Delta x=\gamma\left(\Delta x^{\prime}+v \Delta t^{\prime}\right) \tag{2.26b}
\end{align*}
$$

or, after a boost,

$$
\begin{align*}
\Delta t^{\prime} & =\gamma(\Delta t-v \Delta x)  \tag{2.27a}\\
\Delta x^{\prime} & =\gamma(\Delta x-v \Delta t) \tag{2.27b}
\end{align*}
$$

Suppose the two events are successive ticks of a clock at rest in $O^{\prime}$. Substituting $\Delta x^{\prime}=0$ into Equation 2.26a gives $\Delta t=\gamma \Delta t^{\prime} \geq \Delta t^{\prime}$, which is time dilation. Substituting $\Delta x^{\prime}=0$ into Equation 2.27b gives $\Delta x=v \Delta t$, and combining this with the first gives $\Delta t^{\prime}=\Delta t / \gamma \leq \Delta t$, which is again time dilation.

Suppose the two events are simultaneous measurements by $O$ of the front and rear of $O^{\prime}$. Substituting $\Delta t=0$ (a "slice of constant $O$-time") into Equation 2.27 b gives $\Delta x^{\prime}=\gamma \Delta x \geq \Delta x$, which is length contraction. Substituting $\Delta t=0$ into Equation 2.27a gives $\Delta t^{\prime}=-v \Delta x^{\prime}=-v \Delta x^{\prime} / c^{2}$, which is clock desynchronization, and combining this with the second gives $\Delta x=\Delta x^{\prime} / \gamma \leq \Delta x^{\prime}$, which is again length contraction.

### 2.1.8 Velocity Addition

Suppose we are in $O^{\prime}$ walking in the direction of relative motion of $O$ and $O^{\prime}$ so that our successive steps are separated by small distances and times $d x^{\prime}$ and $d t^{\prime}$. Taking differentials of Equation 2.24, these separations are related by

$$
\begin{align*}
d t & =\gamma\left(d t^{\prime}+v d x^{\prime}\right)  \tag{2.28a}\\
d x & =\gamma\left(d x^{\prime}+v d t^{\prime}\right) \tag{2.28b}
\end{align*}
$$

Dividing the second of these equations by the first gives

$$
\begin{equation*}
\frac{d x}{d t}=\frac{d x^{\prime}+v d t^{\prime}}{d t^{\prime}+v d x^{\prime}}=\frac{d x^{\prime} / d t^{\prime}+v}{1+v d x^{\prime} / d t^{\prime}} \tag{2.29}
\end{equation*}
$$

or

$$
\begin{equation*}
u=\frac{u^{\prime}+v}{1+u^{\prime} v}=\frac{u^{\prime}+v}{1+u^{\prime} v / c^{2}} \tag{2.30}
\end{equation*}
$$

where $u=d x / d t$ and $u^{\prime}=d x^{\prime} / d t$. This reduces to the familiar Galilean velocity addition formula if $c \rightarrow \infty$, but generalizes it to any speeds, even those close to the speed of light. It implies that the light from a forward laser on a nearlight speed rocket is never observed to be superluminal. In fact, if we use Equation 2.30 to define a relativistic velocity combination operator $\oplus$ by $u=$ $u^{\prime} \oplus v$, then we have $1 \oplus 1=1$ !

### 2.1.9 Spacetime Rotations

We can parameterize relative velocities by the rapidity $\varphi$ using an Appendix A. 2 hyperbolic tangent function, so that

$$
\begin{equation*}
v=\tanh \varphi \tag{2.31}
\end{equation*}
$$

and

$$
\begin{equation*}
\gamma=\frac{1}{\sqrt{1-v^{2}}}=\frac{1}{\sqrt{1-\tanh ^{2} \varphi}}=\cosh \varphi \tag{2.32}
\end{equation*}
$$

While the relativistic velocities combine in an unfamiliar way, rapidities simply add. Using the parameterization of Equation 2.31 as a template, we substitute rapidities into the relativistic velocity addition formula of Equation 2.30 and employ a standard hyperbolic identity to show

$$
\begin{equation*}
\tanh \varphi_{3}=\frac{\tanh \varphi_{1}+\tanh \varphi_{2}}{1+\tanh \varphi_{2} \tanh \varphi_{2}}=\tanh \left[\varphi_{1}+\varphi_{2}\right] \tag{2.33}
\end{equation*}
$$

Hence

$$
\begin{equation*}
\varphi_{3}=\varphi_{1}+\varphi_{2} \tag{2.34}
\end{equation*}
$$

Furthermore, if we substitute rapidities into the Lorentz-Einstein transformation of Equation 2.24 and use $\cosh \varphi \tanh \varphi=\sinh \varphi$, we get

$$
\begin{align*}
t & =t^{\prime} \cosh \varphi+x^{\prime} \sinh \varphi  \tag{2.35a}\\
x & =x^{\prime} \cosh \varphi+t^{\prime} \sinh \varphi \tag{2.35b}
\end{align*}
$$

If we introduce an imaginary angle $\varphi=i \alpha$, where $\alpha$ is real and $i=\sqrt{-1}$, then

$$
\begin{array}{r}
t=t^{\prime} \cos \alpha+i x^{\prime} \sin \alpha \\
x=x^{\prime} \cos \alpha+i t^{\prime} \sin \alpha \tag{2.36b}
\end{array}
$$

Multiplying the first of these equations by $i$, introducing $y=i t=i c t$, and rearranging the equations and terms into standard form, we get

$$
\begin{align*}
& x=x^{\prime} \cos \alpha+y^{\prime} \sin \alpha  \tag{2.37a}\\
& y=-x^{\prime} \sin \alpha+y^{\prime} \cos \alpha \tag{2.37b}
\end{align*}
$$

As a matrix equation, this is

$$
\binom{x}{y}=\left(\begin{array}{rr}
\cos \alpha & \sin \alpha  \tag{2.38}\\
-\sin \alpha & \cos \alpha
\end{array}\right)\binom{x^{\prime}}{y^{\prime}}
$$

which represents a rotation through an angle $\alpha$. Thus, the Lorentz-Einstein transformation can be understood as a rotation through an imaginary angle in a complex space!

This suggests the following deep analogy. Consider the ordinary rotation of an object in space, as in Figure 2.9. The relative slope $s=\Delta x / \Delta y$, the ratio of two space changes, can parameterize the rotation. The most the rotation can


Figure 2.9: When an object rotates in space, its vertical cross section dilates, its horizontal cross section contracts, and its top (and bottom) corners no longer coincide vertically. These are analogues of time dilation, length contraction, and clock desynchronization, for a four-dimensional object rotating in spacetime, as it changes its velocity in space.
achieve is one revolution (beyond which it retraces itself). During the rotation, the object's two-dimensional cross sections change, one contracts and the other dilates. The left and right corners, which are at the same height initially, are at different heights finally.
"Can an instantaneous cube exist?" asked H. G. Well's Time Traveler. Familiar objects, which naturally appear three-dimensional to us, are really extended in time, the fourth dimension. A velocity like $v_{x}=\Delta x / \Delta t$, the ratio of a space change and a time change, corresponds to a spacetime rotation. The most a speed can change from zero is light speed, $v=c=1$ in natural units. Length contraction, time dilation, and clock desynchronization are merely the geometric projection effects of observing three-dimensional cross-sections of four-dimensional objects rotated in four-dimensional spacetime.

Why time dilation? It's the rotation.

### 2.1.10 Spacetime Interval

Spatial rotations preserve distances between points. Consider two observers relatively rotated in 2-dimensional space. If the coordinates of a place, say a flagpole, are $\{x, y\}$ for one and $\left\{x^{\prime}, y^{\prime}\right\}$ for the other, then these are related by the rotation transformation

$$
\begin{align*}
x^{\prime} & =x \cos \theta+y \sin \theta  \tag{2.39a}\\
y^{\prime} & =-x \sin \theta+y \cos \theta \tag{2.39b}
\end{align*}
$$

where $s=\tan \theta$ is the relative slope of their axes. The trigonometric identity $\cos ^{2} \theta+\sin ^{2} \theta=1$ implies that the sum of the squares of the space coordinates are the same,

$$
\begin{equation*}
x^{\prime 2}+y^{\prime 2}=x^{2}+y^{2} \equiv l^{2} \tag{2.40}
\end{equation*}
$$

where $l$ is the invariant distance of the place from the observers' common origin. More generally, in 3-dimensional space,

$$
\begin{equation*}
\Delta x^{\prime 2}+\Delta y^{\prime 2}+\Delta z^{\prime 2}=\Delta x^{2}+\Delta y^{2}+\Delta z^{2}=\Delta l^{2} \tag{2.41}
\end{equation*}
$$

where $\Delta x^{2}=(\Delta x)^{2}$, and so on.
Spacetime rotations, or boosts, preserve the spacetime interval between events. Consider two observers in relative motion in 1+1-dimensional spacetime. If the coordinates of an event, say a supernova, are $\{t, x\}$ for one and $\left\{t^{\prime}, x^{\prime}\right\}$ for the other, then these are related by the Lorentz-Einstein transformation

$$
\begin{align*}
t & =t^{\prime} \cosh \varphi+x^{\prime} \sinh \varphi  \tag{2.42a}\\
x & =x^{\prime} \cosh \varphi+t^{\prime} \sinh \varphi \tag{2.42b}
\end{align*}
$$

where $v=\tanh \varphi$ is their relative velocity. The hyperbolic identity $\cosh ^{2} \varphi-$ $\sinh ^{2} \varphi=1$ implies that the difference in the squares of the space and time coordinates are the same,

$$
\begin{equation*}
x^{\prime 2}-t^{\prime 2}=x^{2}-t^{2} \equiv+\sigma^{2} \equiv-\tau^{2} \tag{2.43}
\end{equation*}
$$

where $\tau^{2}=-\sigma^{2}$ is the square of the spacetime interval of the event from the observers' common origin. More generally, in 3+1-dimensional spacetime,

$$
\begin{equation*}
\Delta{x^{\prime}}^{2}+\Delta y^{\prime 2}+\Delta z^{\prime 2}-\Delta t^{\prime 2}=\Delta x^{2}+\Delta y^{2}+\Delta z^{2}-\Delta t^{2}=+\Delta \sigma^{2}=-\Delta \tau^{2} \tag{2.44}
\end{equation*}
$$

The invariant interval generalizes the Pythagorean theorem to spacetime. The corresponding generalized geometry is sometimes called "pseudo-Euclidean" or "flat", and Einstein has been called the "new Pythagoras" for this profound discovery. In spacetime physics, space and time are on an equal footing, and if it were not for the minus sign in the interval, time would be exactly like space, and we could walk back to yesterday!

The Equation 2.41 Pythagorean theorem provides a notion of distance or metric for Euclidean space. The Equation 2.44 spacetime interval provides a notion of distance for flat spacetime. But due to the minus sign, it is technically a semi-metric, meaning it can be positive or negative. In fact, this distinction leads to an invariant partition of spacetime at every event, as summarized in Figure 2.10. The origin of the coordinates is a fiducial event $E_{0}$, which is separated from other events by the interval $s^{2}-t^{2}=+\sigma^{2}=-\tau^{2}$, where $s^{2}=$ $x^{2}+y^{2}+z^{2}$. Events $E_{0}$ and $E_{1}$ are separated by a timelike interval, where their time separation dominates their space separation, so that $\tau^{2}=t^{2}-s^{2}>0$. They can be causally related because, for example, they can both be events on the worldine of the same massive object. Events $E_{0}$ and $E_{3}$ are separated by a spacelike interval, where their time separation dominates their space separation, so that $\sigma^{2}=s^{2}-t^{2}>0$. They are causally unrelated. Finally, events $E_{0}$ and $E_{2}$ are separated by a lightlike interval, where their time separation equals their space separation, so that $\sigma^{2}=\tau^{2}=0$. They can be joined by a light ray.


Figure 2.10: Invariant light cones partition spacetime at every event, in $1+1$ dimensional (left) and 2+1-dimensional (right) spacetimes.

In 2+1-dimensional spacetime, this partition is obviously a light cone at every event. The worldline of a massive object "enters" every event via the past light cone and "exits" via the future light cone. We draw the future light cones opening to the right (instead of opening up) so that the slopes of the worldlines of massive objects are their velocities (instead of their inverse velocities).

### 2.1.11 Twin Paradox

The interval between two events on the worldline of a massive object has an important interpretation. Suppose two observers $O$ and $O^{\prime}$ are in relative motion along a common $x, x^{\prime}$ direction. Then, for successive ticks of a clock at rest relative to $O^{\prime}$, the square of the spacetime interval is

$$
\begin{equation*}
0^{2}+0^{2}+0^{2}-d t^{\prime 2}=d x^{2}+0^{2}+0^{2}-d t^{2}=-d \tau^{2} \tag{2.45}
\end{equation*}
$$

so that

$$
\begin{equation*}
d \tau=d t^{\prime}=\sqrt{d t^{2}-d x^{2}}=\sqrt{1-\left(\frac{d x}{d t}\right)^{2}} d t=\frac{d t}{\gamma} \tag{2.46}
\end{equation*}
$$

Thus, $d \tau$ is the proper time between the ticks. More generally, the spacetime interval between two events on the worldline of a massive object is the proper time or aging between them. Different observers in relative motion will disagree about space and time intervals, but they will always agree on aging, $\Delta \tau=\int d \tau$, which is the length of the corresponding worldline. Table 2.1 summarizes the analogy between space and spacetime.

The twin "paradox" is one famous consequence of the equivalence of aging and worldline length. Stella and Terra are identical twins. Stella leaves Earth $\oplus$ for a fast voyage to a nearby star $\star$ and back, while Terra stays behind. Upon Stella's return, the twins are distinctly different ages! Why? Because their worldlines have distinctly different lengths. Also, due to the minus sign in the spacetime interval, Stella's kinked worldline is actually shorter than Terra's
unkinked worldline, as summarized in Table 2.2, where the diagrams are necessarily rendered in Euclidean space. In general, the unaccelerated, unkinked worldline between any two events is the one of extremal (longest) proper time.

Table 2.1: Space vs. spacetime: Einstein is the new Pythagoras.

| Pythagoras | Einstein |
| :--- | :--- |
| 2-dimensional Euclidean space | 1+1-dimensional spacetime |
| Path between places | Worldline between events |
| P | $>y$ |

Table 2.2: Angle vs. twin "paradoxes".

| Angle "paradox" | Twin "paradox" |
| :---: | :---: |
| Direct path shorter than kinked path $d l=\sqrt{d y^{2}+d x^{2}} \geq d y$ <br> Faithfully rendered | Direct worldline ages more than kinked worldline $d \tau=\sqrt{d t^{2}-d x^{2}} \leq d t$ <br> Unfaithfully rendered |

### 2.1.12 Loedel Spacetime Diagrams

We can obtain a helpful and very insightful visualization of a 1+1-dimensional Lorentz-Einstein transformation by beginning with a representation of an orthogonal rotation between two coordinate systems $O$ and $O^{\prime}$, as in Figure 2.11. This can not simultaneously record the observations of two observers in relative motion, as the two sets of orthogonal coordinates are related by the invariance of distance, not interval. However, if we interchange the two time axes, the resulting skew coordinate systems do preserve the interval. Define coordinates by drawing parallels (Loedel convention) rather than perpendiculars (Brehme convention) or mixing parallels and perpendiculars (Minkowski convention). Specifically, the $x$-coordinate of an event is determined by drawing a line parallel to the $t$-axis, and so on. Similarly, a line parallel to the $t$-axis represents a point at rest relative to the $O$ observer.



Figure 2.11: Swapping time coordinates converts a spatial rotation preserving distance into a spacetime rotation preserving interval. The spatial coordinates (left) are orthogonal, while the spacetime coordinates (right) are skew.

Therefore, from the geometry of the Loedel spacetime diagram, the velocity of $O$ relative to $O^{\prime}$ is

$$
\begin{equation*}
v=\frac{\Delta x^{\prime}}{\Delta t^{\prime}}=\sin \theta \tag{2.47}
\end{equation*}
$$

and the relativistic stretch is

$$
\begin{equation*}
\gamma=\frac{1}{\sqrt{1-v^{2}}}=\sec \theta \tag{2.48}
\end{equation*}
$$

Note that if $v=0$, then $\theta=0$, and the two skew coordinate systems collapse into a single orthogonal coordinate system.

### 2.1.13 Recovering the Fundamental Effects (Again)

All three of the key kinematical effects of special relativity, time dilation, length contraction, and clock desynchronization, are implicit in the Loedel spacetime diagram. For example, consider Figure 2.12. The row of dots represents the regular flashes of a clock at rest in $O^{\prime}$. From the geometry of the diagram, $\Delta t^{\prime} / \Delta t=\cos \theta=1 / \gamma$, which is time dilation. The shaded slab without dots represents the worldtube of a stick at rest in $O^{\prime}$. From the geometry of the diagram, $\Delta x^{\prime} / \Delta x=1 / \cos \theta=\gamma$, which is length contraction. The stick's worldtube is the invariant reality, while different observers in relative motion take different slices of constant time to measure different lengths! Note how the diagrams faithfully represent the magnitudes of the time dilation and length contractions (unlike the more popular Minkowski spacetime diagrams). Finally, the shaded slab with dots represents two synchronized and regularly flashing clocks at rest in $O^{\prime}$. From the geometry of the diagram, $-\Delta t^{\prime} / \Delta x^{\prime}=\sin \theta=v$ or $\Delta t^{\prime}=-v \Delta x^{\prime}$, which is clock desynchronization.


Figure 2.12: Spacetime diagrams illustrating time dilation, length contraction, and clock desynchronization.

### 2.1.14 Pole-in-the-Barn Example

To illustrate the utility of the Loedel spacetime diagram, we will use it to analyze the famous pole-in-the-barn "paradox". Can one fit a 20 -meter pole in a 10meter barn by running so fast that it contracts to half its length? (Safety first: Assume both the front and rear barn doors are open!) For the barn observer, the pole does contract to half its length, and it would seem to fit, but for the pole observer, it is the barn that contracts to half its length, and a fit seems impossible. A Loedel spacetime diagram with an angle $\theta=60^{\circ}$, corresponding to relativistic stretch $\gamma=2$ and a velocity $v=\sqrt{3} / 2$, can represent both sets of
observations at once, as in Figure 2.13.


Figure 2.13: In the spacetime diagram, intersecting worldtubes represent the pole and the barn. Each observer sequences the events with different families of parallel slices of constant time.

The pole is at rest for the pole observer, and hence is represented as a worldtube parallel to the $t_{P}$-axis. The barn is at rest for the barn observer, and hence is represented as a worldtube parallel to the $t_{B}$-axis, but half as wide. The crossed worldtubes are the invariant reality, which each observer interprets with a different family of parallel slices of constant time. The pole is in entirely in the barn for one slice of constant barn time, but is not entirely in the barn for any slice of constant pole time. Thus, the consistency of the descriptions depends crucially on the relativity of simultaneity.

So, is the pole ever in the barn or not? There is no is-ness about it! It's relative. Interestingly, one can trap the pole in the barn by quickly and carefully closing the doors, because the rear of the pole can't know that the front of the pole has stopped until a compression wave moving at the speed of sound $v_{S} \ll 1$ travels the length of the pole.

The doubly sliced crossed worldtubes of Figure 2.13 remind us of Wheeler's dictum: Space is different for different observers, time is different for different observers, spacetime is the same for all observers!

### 2.1.15 Faster-Than-Light Implies Backward-In-Time

As a second example of the utility of the Loedel spacetime diagram, consider a spaceship installed with the following foolproof protection system. If an accident occurs, such as a collision with interstellar debris, then the spaceship will send a faster-than-light signal to Earth, which will relay the signal faster-than-light back to the spaceship, so that it arrives before the accident to allow evasive action, as illustrated in Figure 2.14 for the special case of an infinitely fast signal (like the science-fiction ansible introduced by Ursula K. LeGuin and employed by Orson Scott Card).

Obviously, this is logically problematic. If the spaceship takes evasive action to avoid the accident, what sent the signal? Because faster-than-light signalling


Figure 2.14: Rocket and Earth exchange an instanton traveling at $v=\infty$, which returns to the rocket before it left. (The times of reflection and return can be readily calculated from the geometry of the diagram using Equation 2.48.)
can lead to causality violations, most physicists believe that information cannot travel faster than light. (However, things that don't carry information, such as the intersection of the blades of a long scissors quickly closed, can go faster than light.)

### 2.1.16 Magnetism is Relativistic Electricity

Magnetism is a strange, velocity-dependent, deflecting force. While a gravitational field $\vec{g}$ results in a gravitational force $m \vec{g}$, and an electrical field $\overrightarrow{\mathcal{E}}$ results in an electrical force $q \overrightarrow{\mathcal{E}}$, a magnetic field $\overrightarrow{\mathcal{B}}$ results not in a magnetic force $q \overrightarrow{\mathcal{B}}$, but in a magnetic force $q \vec{v} \times \overrightarrow{\mathcal{B}}$. In fact, its strangeness betrays its origin as a tiny, relativistic correction to electricity!

Under boosts, the Lorentz-Einstein transformation alters forces and electrical fields in nontrivial ways. However, electric charge, like the spacetime interval, is a Lorentz-invariant scalar, the same for all observers. (Charge along with mass and spin characterize and classify elementary particles.) In the rest frame of the electric charges, where the relativistic corrections vanish, the electric force $\overrightarrow{\mathcal{E}}$ on a charge $q$ remains $q \overrightarrow{\mathcal{E}}$. By working in such a reference frame, we can show that parallel currents must attract and antiparallel currents must repel - the very definition of the magnetic force - without invoking magnetism at all. We won't need any "new magic" beyond electricity and relativity.

The drift speed of the electrons in copper wire at room temperature is typically very small, with a relativistic stretch of nearly unity. Consequently, the "magnetic" corrections to electricity will be of order $\gamma-1 \sim v^{2} / 2 \ll 1$. They are only significant because the enormous electric force is effectively screened in bulk matter.

We will first show that parallel current-carrying wires attract. We model the wires as superpositions of one-dimensional positive (ionic) and negative (elec-


Figure 2.15: Parallel current-carrying wires as superpositions of positive and negative charges, displaced vertically for clarity, for observers at rest relative to the bottom wire's positive (left) and negative (right) charges. The relativistic stretch $\gamma$ is greatly exaggerated.
tron) charge densities, displaced slightly vertically for clarity in Figure 2.15. We assume the wires are electrically neutral in their (ionic) rest frame. In the rest frame of the bottom wire's positive charges, the top wire is uncharged, and hence the bottom wire's positive charges are neither attracted nor repelled by the top wire. However, in the rest frame of the bottom wire's negative charges, the top wire is positively charged, due to length contraction, and hence the bottom wire's negative charges are attracted upward. The net result is that the bottom wire is attracted to the top wire (and vice versa).


Figure 2.16: Antiparallel current-carrying wires as superpositions of positive and negative charges, displaced vertically for clarity, for observers at rest relative to the bottom wire's positive (left) and negative (right) charges. The compound velocity $v^{\prime}=2 v /\left(1+v^{2}\right)$ and the corresponding relativistic stretch $\gamma^{\prime}=\gamma^{2}(1+$ $v^{2}$ ).

We will next show that antiparallel current-carrying wires repel. Again, we model the wires as superpositions of one-dimensional positive and negative charge densities, displaced slightly vertically for clarity in Figure 2.16. In the rest frame of the bottom wire's positive charges, the top wire is uncharged, and hence the bottom wire's positive charges are neither attracted nor repelled by the top wire. However, in the rest frame of the bottom wire's negative charges, the top wire is negatively charged, due to length contraction, and hence the
bottom wire's negative charges are repelled downward. The net result is that the bottom wire is repelled by the top wire (and vice versa).

### 2.2 Relativistic Dynamics

Changes to our understanding of space and time imply changes to our understanding of energy and momentum.

### 2.2.1 Spacetime 4-Vectors

Motion shrinks rulers, slows and desynchronizes clocks, and consequently alters momentum and energy. The transition from relativistic kinematics to relativistic dynamics is facilitated by the introduction of spacetime vectors or 4-vectors, which are like space or 3-vectors with the addition of a time component and a generalized scalar product. For example, if a place in space is $\vec{r}$, than an event in spacetime is

$$
\vec{r}=\left(\begin{array}{c}
t  \tag{2.49}\\
x \\
y \\
z
\end{array}\right)=\binom{t}{\vec{r}}
$$

where the latter expression is an obvious shorthand. If the space scalar product is $\vec{r} \cdot \vec{r}^{\prime}=x x^{\prime}+y y^{\prime}+z z^{\prime}$, so the magnitude $\operatorname{mag}[\vec{r}]=\sqrt{\vec{r} \cdot \vec{r}}=\sqrt{x^{2}+y^{2}+z^{2}}$, then the spacetime scalar product is

$$
\begin{equation*}
\vec{r} \cdot \vec{r}^{\prime}=t t^{\prime}-\vec{r} \cdot \vec{r}^{\prime}=t t^{\prime}-x x^{\prime}-y y^{\prime}-z z^{\prime} \tag{2.50}
\end{equation*}
$$

so the magnitude

$$
\begin{equation*}
\operatorname{mag}[\overrightarrow{\vec{r}}]=\sqrt{\overrightarrow{\vec{r}} \cdot \overrightarrow{\vec{r}}}=\sqrt{t^{2}-\vec{r}^{2}}=\sqrt{t^{2}-x^{2}-y^{2}-z^{2}} \tag{2.51}
\end{equation*}
$$

where once again the negative signs reflect the distinction between space and time.

Consider the worldine of an object of mass $m$, as in Figure 2.17. In a short time $d t$, the mass $m$ undergoes a 3 -displacement $d \vec{r}$ and a 4-displacement

$$
d \overrightarrow{\vec{r}}=\left(\begin{array}{c}
d t  \tag{2.52}\\
d x \\
d y \\
d z
\end{array}\right)=\binom{d t}{d \vec{r}}
$$

The magnitude of the 4-displacement is the elapsed proper time,

$$
\begin{equation*}
\operatorname{mag}[d \vec{r}]=\sqrt{d t^{2}-d \vec{r}^{2}}=d t \sqrt{1-\left(\frac{d \vec{r}}{d t}\right)^{2}}=\frac{d t}{\gamma}=d \tau \tag{2.53}
\end{equation*}
$$



Figure 2.17: 4-velocity and 4-momentum tangent to the worldine of a massive particle.

Since the 4-displacement $d \overrightarrow{\vec{r}}$ is a 4-vector and the proper time $d \tau$ is a Lorentzinvariant scalar, and because multiplication (and division) of a vector by a scalar is well-defined, we may define the 4 -velocity as

$$
\vec{v}=\frac{d \overrightarrow{\vec{r}}}{d \tau}=\left(\begin{array}{c}
d t / d \tau  \tag{2.54}\\
d x / d \tau \\
d y / d \tau \\
d z / d \tau
\end{array}\right)=\binom{d t / d \tau}{d \vec{r} / d \tau}=\binom{\gamma}{\gamma \vec{v}}
$$

Its time component is the relativistic stretch, $v_{t}=\gamma$, and its magnitude is light speed,

$$
\begin{equation*}
\operatorname{mag}[\vec{v}]=\sqrt{\gamma^{2}-(\gamma v)^{2}}=\gamma \sqrt{1-v^{2}}=1=c \tag{2.55}
\end{equation*}
$$

In space masses have different relative speeds $v$, but in spacetime they have the same invariant speed $c$ (even as they move in space and exist in spacetime).

Finally, since mass $m$ is also a Lorentz-invariant scalar, we may define the 4-momentum as

$$
\begin{equation*}
\vec{p}=m \overrightarrow{\vec{v}}=\binom{\gamma m}{\gamma m \vec{v}} \tag{2.56}
\end{equation*}
$$

To identify its time component, consider the slow motion limit $v \ll 1$, where

$$
\begin{equation*}
p_{t}=\gamma m=m\left(1-v^{2}\right)^{-1 / 2} \sim m\left(1+\frac{1}{2} v^{2}\right)=m+\frac{1}{2} m v^{2} \tag{2.57}
\end{equation*}
$$

which is the nonrelativistic kinetic energy shifted by a constant. Hence, we identify the time component of the 4 -momentum as the total energy

$$
\begin{equation*}
E=\gamma m \tag{2.58}
\end{equation*}
$$

and the relativistic kinetic energy becomes

$$
\begin{equation*}
T=E-m=(\gamma-1) m \tag{2.59}
\end{equation*}
$$

In a rest frame, $v=0, \gamma=1$, and $E=m$, which in nonrelativistic units is $E=m c^{2}$, the most famous equation in modern physics [2]. Meanwhile, the space components of the 4-momentum generalize Newton's 3-momentum to high speeds

$$
\begin{equation*}
\vec{p}=\gamma m \vec{v} . \tag{2.60}
\end{equation*}
$$

The magnitude of the 4 -momentum is the mass,

$$
\begin{equation*}
\operatorname{mag}[\vec{p}]=\sqrt{E^{2}-\vec{p}^{2}}=\sqrt{(\gamma m)^{2}-(\gamma m v)^{2}}=\gamma m \sqrt{1-v^{2}}=m \tag{2.61}
\end{equation*}
$$



Figure 2.18: Loedel spacetime (left) and momentum-energy (right) diagrams for a mass $m$ at rest relative to observer $O^{\prime}$ and in uniform motion relative to observer $O$. Note that the 4-momentum (right) is parallel to the worldine (left).

Since the 4 -momentum is a 4 -vector, it transforms under boosts like events in spacetime. For example, suppose two observers $O$ and $O^{\prime}$ are in relative motion along a common $x$ and $x^{\prime}$ axis. Then, as in Equation 2.24,

$$
\begin{align*}
E & =\gamma\left(E^{\prime}+v p_{x}^{\prime}\right)  \tag{2.62a}\\
p_{x} & =\gamma\left(p_{x}^{\prime}+v E^{\prime}\right) \tag{2.62b}
\end{align*}
$$

Consider a mass $m$ at rest in $O^{\prime}$, as in Figure 2.18. In $O^{\prime}$, its 4-momentum components are

$$
\begin{equation*}
\binom{E^{\prime}}{\vec{p}^{\prime}}=\binom{m}{\overrightarrow{0}} \tag{2.63}
\end{equation*}
$$

while in $O$, using the Lorentz transformation of Equation 2.62, its 4-momentum components are

$$
\begin{equation*}
\binom{E}{\vec{p}}=\binom{\gamma m}{\gamma m \vec{v}} . \tag{2.64}
\end{equation*}
$$

### 2.2.2 Invariant Mass

Mass is neither additive nor conserved, but it is an invariant under boosts, as it is the magnitude of a 4 -vector. Conversely, energy and momentum are
additive and conserved, but they are not invariant under boosts, as they are components of 4 -vectors. For example, if $\vec{p}_{i}+\Delta \vec{p}=\vec{p}_{f}$, then $E_{i}+\Delta E=E_{f}$ and $\vec{p}_{i}+\Delta \vec{p}=\vec{p}_{f}$ but $m_{i}+\Delta m \neq m_{f}$ by the (spacetime) triangle inequality.

A good example of the nonadditivity of mass is a box of many molecules of heated air. In the rest frame of the box, if the momentum of a typical air molecule is $\vec{p}_{i}$, then the total 3-momentum of the air vanishes, $\vec{p}_{a}=\sum \vec{p}_{i}=\overrightarrow{0}$, and so its 4 -momentum is

$$
\begin{equation*}
\vec{p}_{a}=\binom{E_{a}}{\overrightarrow{0}} . \tag{2.65}
\end{equation*}
$$

The mass of the box of air is therefore $m_{a}=\operatorname{mag}\left[\vec{p}_{a}\right]=E_{a}$, but the energy of the air $E_{a}=\sum E_{i}$, where $E_{i}=m_{i}+T_{i}$, and so

$$
\begin{equation*}
m_{a}=\sum_{i} m_{i}+T_{a} \geq \sum_{i} m_{i} \tag{2.66}
\end{equation*}
$$

were $T_{a}=\sum T_{i}$ is the total kinetic energy of the air in the box. (In practice, typically $m_{a}=m_{a} c^{2} \gg T_{a}$, so the mass is nearly additive.)

Particles, like photons, that move at light speed are necessarily massless. Combining $E=\gamma m$ and $p=\gamma m v$, we get

$$
\begin{equation*}
p=E v \tag{2.67}
\end{equation*}
$$

Consequently, if $v=1$, then $p=E$, and the mass $m=\sqrt{E^{2}-p^{2}}=0$.
The concepts of "relativistic" and "rest" mass are misleading. Defining $m_{v}=$ $\gamma m$ allows us to write $\vec{p}=m_{v} \vec{v}$, which resembles the corresponding Newtonian expression. However, $m$ and not $m_{v}$ is a relativistic invariant. Invariants are jewels, and we don't throw away jewels. The mass of a particle doesn't increase with speed; rather, its energy and momentum diverge as its speed approaches light speed.

### 2.2.3 Einstein's Box Derivation of $E=m c^{2}$

Einstein's famous $E=m c^{2}$ almost effortlessly fell out of our 4-vector formalism, but a simple and more physical derivation of this important result would be nice. Einstein provided such a derivation in 1906 [4] involving a closed box that shifts its position due to a burst of radiant energy traveling inside it from one end to the other, as in Figure 2.19.

Switching momentarily to nonrelativistic units, we can choose the mass $M$ of the box sufficiently large so that the recoil speed is arbitrarily slow, $v \ll c$, the relativistic stretch is nearly unity, $\gamma \sim 1$, and the Newtonian momentum is practically conserved. If the energy of the radiation is $E$, then its 3-momentum is $p=E / c$, and

$$
\begin{equation*}
0=M v-\frac{E}{c} \tag{2.68}
\end{equation*}
$$

so that the box's recoil speed is $v=E / M c$. Since the mass is large, the recoil distance will be small, $\Delta x \ll L$, and the recoil time $\Delta t=(L-\Delta x) / c \sim L / c$.


Figure 2.19: A burst of radiation of energy $E$ and 3-momentum $p=E / c$ shifts a box of mass $M$ a distance $\Delta x$ in a time $\Delta t$.


Figure 2.20: Spacetime (left) and momentum-energy (right) diagrams of the shifting box. The sum of the 4-momenta of the recoiling box and the radiation equals the 4-momentum of the stationary box, but the sum of the mass of the recoiling box and the zero mass of the radiation does not equal the mass of the stationary box. (Masses are distorted in the diagram because spacetime is rendered in the Euclidean space of the page.)

Hence, the recoil distance is

$$
\begin{equation*}
\Delta x=v \Delta t=\left(\frac{E}{M c}\right)\left(\frac{L}{c}\right)=\frac{E}{M c^{2}} L \tag{2.69}
\end{equation*}
$$

which is indeed much less than $L$ if $E \ll M c^{2}$.
Since no external force acts on the system, its center of mass cannot move. Hence, the radiation must transfer an effective mass $m \ll M$, and the shifts in $m$ and $M$ must offset each other, so that the center of mass does not move. Computing the location of the center of mass before and after the shift gives

$$
\begin{equation*}
\frac{M(L / 2)+m L}{M+m}=\bar{x}=\frac{M(L / 2+\Delta x)+m \Delta x}{M+m} \tag{2.70}
\end{equation*}
$$

which implies

$$
\begin{equation*}
\Delta x=\frac{m}{M+m} L \sim \frac{m}{M} L \tag{2.71}
\end{equation*}
$$

Combining our two expressions for $\Delta x$ gives

$$
\begin{equation*}
E=m c^{2} \tag{2.72}
\end{equation*}
$$

Note how radiation can transfer mass even though the mass of the radiation itself is zero! This is possible because masses, being the length of 4-momenta, are not additive, as demonstrated in Figure 2.20. Note also that Einstein's box cannot move as a rigid body, because sound speed (and hence vibrations) $v_{s} \ll c$. However, including this complication does not change the result.

### 2.2.4 Compton Scattering

In the 1920s, Arthur Compton scattered gamma rays from electrons and found that the light behaved as particles, now called photons, of energy $E=h \nu$ and momentum $p=h / \lambda$. In natural, spacetime units, $c=1$ and $\nu=1 / \lambda$, and so $E=p=h \nu$.

Consider the photon-electron scattering represented in Figure 2.21. Before the scattering, the photon has 4-momentum

$$
\vec{p}=\left(\begin{array}{c}
E  \tag{2.73}\\
p \\
0 \\
0
\end{array}\right)=h \nu\left(\begin{array}{l}
1 \\
1 \\
0 \\
0
\end{array}\right)
$$

and the stationary electron of mass $m$ has 4-momentum

$$
\vec{p}_{e}=\left(\begin{array}{c}
m  \tag{2.74}\\
0 \\
0 \\
0
\end{array}\right)=m\left(\begin{array}{l}
1 \\
0 \\
0 \\
0
\end{array}\right)
$$



Figure 2.21: Before (left) and after (right) the scattering of a photon and an electron.

After the scattering, the photon has 4-momentum

$$
\vec{p}^{\prime}=\left(\begin{array}{c}
E^{\prime}  \tag{2.75}\\
p^{\prime} \cos \theta \\
p^{\prime} \sin \theta \\
0
\end{array}\right)=h \nu^{\prime}\left(\begin{array}{c}
1 \\
\cos \theta \\
\sin \theta \\
0
\end{array}\right)
$$

and the recoiling and undetected electron has 4 -momentum $\vec{p}_{e}^{\prime}$. By the conservation of 4-momentum, which relativistically conserves both energy and momentum,

$$
\begin{equation*}
\vec{p}+\vec{p}_{e}=\overrightarrow{\vec{p}}^{\prime}+\vec{p}_{e}^{\prime} \tag{2.76}
\end{equation*}
$$

We can eliminate the unknown $\vec{p}_{e}^{\prime}$ by squaring it, so that

$$
\begin{equation*}
{\overrightarrow{\vec{p}_{e}}}_{e}^{2}=\left(\overrightarrow{\vec{p}}+\overrightarrow{\vec{p}}_{e}-\overrightarrow{\vec{p}}^{\prime}\right)^{2}=\overrightarrow{\vec{p}}^{2}+\overrightarrow{\vec{p}}_{e}^{2}+{\vec{p}^{\prime}}^{2}+2\left(\overrightarrow{\vec{p}} \cdot \vec{p}_{e}-\overrightarrow{\vec{p}} \cdot \overrightarrow{\vec{p}}^{\prime}-\overrightarrow{\vec{p}}_{e} \cdot \overrightarrow{\vec{p}}^{\prime}\right) \tag{2.77}
\end{equation*}
$$

Since the magnitude of a 4-momentum is a mass, this implies

$$
\begin{equation*}
m^{2}=0^{2}+m^{2}+0^{2}+2\left(\overrightarrow{\vec{p}} \cdot \vec{p}_{e}-\overrightarrow{\vec{p}} \cdot \overrightarrow{\vec{p}}^{\prime}-\overrightarrow{\vec{p}}_{e} \cdot \overrightarrow{\vec{p}}^{\prime}\right) \tag{2.78}
\end{equation*}
$$

and so

$$
\begin{equation*}
\overrightarrow{\vec{p}} \cdot \overrightarrow{\vec{p}}^{\prime}=\overrightarrow{\vec{p}} \cdot \overrightarrow{\vec{p}}_{e}-\overrightarrow{\vec{p}}_{e} \cdot \overrightarrow{\vec{p}}^{\prime} \tag{2.79}
\end{equation*}
$$

Implementing the spacetime scalar products with Equation 2.50 gives

$$
\begin{equation*}
h^{2} \nu \nu^{\prime}(1-\cos \theta)=m h \nu-m h \nu^{\prime} \tag{2.80}
\end{equation*}
$$

or

$$
\begin{equation*}
1-\cos \theta=\frac{m}{h}\left(\frac{1}{\nu^{\prime}}-\frac{1}{\nu}\right)=\frac{m}{h}\left(\lambda^{\prime}-\lambda\right) \tag{2.81}
\end{equation*}
$$

or

$$
\begin{equation*}
\frac{\Delta \lambda}{\lambda_{c}}=1-\cos \theta \tag{2.82}
\end{equation*}
$$

where the wavelength change $\Delta \lambda=\lambda^{\prime}-\lambda$ and the Compton wavelength of the electron is

$$
\begin{equation*}
\lambda_{c}=\frac{h}{m}=\frac{h}{m c} \sim \frac{1}{40} \AA . \tag{2.83}
\end{equation*}
$$

The largest absolute wavelength change is for Compton backscattering, when $\theta=\pi$ and $\Delta \lambda=2 \lambda_{c}$. To maximize the relative wavelength change, Compton used the smallest feasible wavelengths, those of gamma rays, whose wavelengths are less than $0.1 \AA$.

### 2.2.5 Pair Production

An electron $e^{-}$may combine with an antielectron $e^{+}$to annihilate into two (or three) photons. For example,

$$
\begin{equation*}
e^{-}+e^{+} \longrightarrow 2 \gamma \tag{2.84}
\end{equation*}
$$

where here, as is conventional, $\gamma$ represents a photon (and not the relativistic stretch). In the presence of matter, the inverse process, pair production, is also possible. For example,

$$
\begin{equation*}
\gamma+N \longrightarrow e^{-}+e^{+}+N \tag{2.85}
\end{equation*}
$$

where $N$ might be the nucleus of an atom.


Figure 2.22: Pair production, before and after (left and right), as observed in two different reference frames (top and bottom).

In a minimum energy, or threshold, situation, the particles are created at rest in the center-of-mass, or zero 3-momentum, reference frame, as in Figure 2.22. Let the initial 4 -momentum of the system in the $N$-frame be

$$
\vec{p}_{N}=\left(\begin{array}{c}
E+M  \tag{2.86}\\
E \\
0 \\
0
\end{array}\right)
$$

where $M$ is the mass of the nucleus and $E$ is the energy and 3-momentum of the photon. Let the final 4-momentum of the system in the zero 3-momentum frame be

$$
\vec{p}_{0}^{\prime}=\left(\begin{array}{c}
2 m+M  \tag{2.87}\\
0 \\
0 \\
0
\end{array}\right)
$$

where $m$ is the mass of the electron and antielectron. Conservation of 4momentum implies $\vec{p}_{0}=\vec{p}_{0}^{\prime}$ and invariance of mass implies $\vec{p}_{0}^{2}=\vec{p}_{N}^{2}$, and so

$$
\begin{equation*}
\vec{p}_{N}^{2}=\vec{p}_{0}^{\prime 2} \tag{2.88}
\end{equation*}
$$

Performing the implied spacetime scalar products, again with Equation 2.50, gives

$$
\begin{equation*}
(E+M)^{2}-E^{2}=(2 m+M)^{2} \tag{2.89}
\end{equation*}
$$

or

$$
\begin{equation*}
E=2 m\left(1+\frac{m}{M}\right) \geq 2 m \tag{2.90}
\end{equation*}
$$

If $M=\infty$, then the electron-antielectron pair can be created most economically with the energy $E=2 m=2 m c^{2}$. However, if $M=0$, then the threshold energy is an unattainable $E=\infty$. A mass $M>0$ must "catalyze" the process; without it, the 3-momentum and hence the energy of the photon in the zero 3 -momentum reference frame would vanish.

### 2.3 Curved Spacetime of General Relativity

Here we explore Einstein's 1915 theory of general relativity, which incorporates gravity. Newton's theory of gravity makes a relativistic theory of gravity both necessary and possible: necessary because Newtonian gravity acts instantaneously, with all the faster-than-light difficulties that entails; possible because of the equality of inertial mass and gravitational charge, which we discuss below.

### 2.3.1 Gravity and Electricity

First we review Newton's law of gravity by comparing and contrasting it with Coulomb's law of electricity. A test electrical charge $q$ separated from a fixed electrical charge $Q$ by a distance $r$ experiences a force

$$
\begin{equation*}
\vec{F}=+k \frac{Q q}{r^{2}} \hat{r}=q \overrightarrow{\mathcal{E}}=-q \vec{\nabla} \varphi=-\vec{\nabla} U \tag{2.91}
\end{equation*}
$$

where $\overrightarrow{\mathcal{E}}$ is the electric field, $\varphi$ is the electric potential (with the SI unit of volts), and $U$ is the electric potential energy (with the SI unit of Joules). If the charge $q$ has an inertial mass $m_{I}$, then its acceleration is

$$
\begin{equation*}
\vec{a}=\frac{\vec{F}}{m_{I}}=\frac{q}{m_{I}} \overrightarrow{\mathcal{E}} \tag{2.92}
\end{equation*}
$$

which clearly depends on the charge-to-mass ratio.
A test gravitational charge $m_{G}$ separated from a fixed gravitational charge $M_{G}$ by a distance $r$ experiences a force

$$
\begin{equation*}
\vec{F}=-G \frac{M_{G} m_{G}}{r^{2}} \hat{r}=m_{G} \vec{g}=-m_{G} \vec{\nabla} \varphi=-\vec{\nabla} U \tag{2.93}
\end{equation*}
$$

where $\vec{g}$ is the gravitational field, $\varphi$ is the gravitational potential, and $U$ is the gravitational potential energy. If the gravitational charge $m_{G}$ has an inertial mass $m_{I}$, then its acceleration is

$$
\begin{equation*}
\vec{a}=\frac{\vec{F}}{m_{I}}=\frac{m_{G}}{m_{I}} \vec{g} \tag{2.94}
\end{equation*}
$$

which appears to depend on the charge-to-mass ratio. However, in Newton's theory, by an inexplicable coincidence, the inertial mass and gravitational charge are always the same,

$$
\begin{equation*}
m_{I}=m_{G} \tag{2.95}
\end{equation*}
$$

and hence $\vec{a}=\vec{g}$ always. Thus, in 1971, when Apollo 15 astronaut Dave Scott simultaneously dropped a hammer and a feather in the vacuum of Luna's surface, they hit the ground together.

The equivalence of inertial mass and gravitational charge makes possible a geometric theory of gravity in which the fall of an object depends not on force fields but on geodesics, paths of extremal length, in curved spacetime.

### 2.3.2 The Equivalence Principle

Einstein's "happiest thought" was that gravity disappears for a freely falling observer.


Figure 2.23: The uniform gravitational field of a flat Earth (left) is equivalent to the constant acceleration of a spacecraft (right).

Imagine you are hanging tightly to a spinning merry-go-round. You feel a centrifugal pseudoforce pulling you outward, but it's not real, as you can cancel it completely by simply letting go. You feel it because you are in the "wrong" reference frame. Similarly, Einstein reasoned, gravity disappears in the "right" reference frame, that of a freely falling observer. Astronauts aboard
the International Space Station freely fall - or float - around Earth, effectively cancelling gravity.


Figure 2.24: A photon's passage through an accelerating glass spaceship in the star frame (left) and spaceship frame (center) and in the equivalent Earth frame (right). The deflection is greatly exaggerated.

Conversely, thanks to the equality of inertial and gravitational mass, constant acceleration is equivalent to a uniform gravitational field, as is demonstrated in Figure 2.23. Sealed inside a closed capsule, no experiment can distinguish being at rest on a flat Earth from being in constant acceleration relative to the stars. In both cases, for example, the distances fallen in successive equal time intervals are proportional to the odd integers, and the cumulative distances fallen are proportional to the squares of the integers, just as Galileo discovered in the early 1600s.

A dramatic consequence of the equivalence principle is that light must fall in a gravitational field with the same acceleration as massive objects, even though no corresponding Newtonian equation of motion exists, as illustrated in Figure 2.24. Arthur Eddington's 1919 expedition to Africa to observe the deflection of starlight near the sun during a solar eclipse verified this prediction and helped make Einstein famous.

### 2.3.3 Gravitational Time Dilation

Another consequence of the equivalence principle is that, near a planet or star, the lower a clock, the slower it ticks. Suppose we freely fall past two clocks fixed at different heights above Earth's surface. Our fall cancels gravity and makes us an inertial observer, as in the special theory of relativity. We observe the clocks rushing upward, and hence they both tick slowly, but the lower clock moves faster - as we have fallen farther and have accelerated relative to Earth - and hence ticks even more slowly than the higher clock.

To make this quantitative, assume we drop from the higher clock to the lower clock. If the clocks are separated by a small height $h$, our speed $v \ll 1$,
then a Newtonian analysis of our motion suffices. Constant acceleration $a=g$ implies linearly increasing velocity $v=g t$ and quadratically increasing position $h=g t^{2} / 2$. Hence, the time to fall is $t=\sqrt{2 h / g}$ and our final speed is $v=\sqrt{2 g h}$. The corresponding relativistic stretch is

$$
\begin{equation*}
\gamma=\frac{1}{\sqrt{1-v^{2}}}=\left(1-v^{2}\right)^{-1 / 2} \sim 1+\frac{1}{2} v^{2}=1+g h>1 \tag{2.96}
\end{equation*}
$$

The clock we carry runs at the same rate as the higher clock and, of course, doesn't vary relative to us. Consequently, as we fall past the lower clock, and observe it moving relative to us, we can directly compare the higher and lower clock rates. Suppose all observers measure the duration of a distant event, perhaps the brightening of a supernova. We measure a duration $\Delta t=\Delta t_{H}$, while the lower clock measures a duration $\Delta t_{L}$. These durations are related by

$$
\begin{equation*}
\Delta t_{H}=\Delta t=\gamma \Delta t_{L}=(1+g h) \Delta t_{L}>\Delta t_{L} \tag{2.97}
\end{equation*}
$$

More time passes for the higher, faster clock, and less time for the lower, slower clock. A lower charged oscillator oscillates slower and emits lower frequency "redder" radiation, a gravitational redshift. People living in a penthouse apartment age more rapidly than those living on the ground floor (but perhaps the view is worth it).

In a Newtonian flat Earth model, the corresponding gravitational potential energy change is $\Delta U=m g h$. Since the gravitational potential is the gravitational potential energy per unit mass, $\varphi=U / m$, we may write

$$
\begin{equation*}
\Delta t_{H}=(1+\Delta \varphi) \Delta t_{L} \tag{2.98}
\end{equation*}
$$

If observing distant light, where wavelength is proportional to period,

$$
\begin{equation*}
\lambda_{H}=(1+\Delta \varphi) \lambda_{L} \tag{2.99}
\end{equation*}
$$

and so the fractional change in wavelength with height is

$$
\begin{equation*}
\frac{\Delta \lambda}{\lambda}=\frac{\lambda_{H}-\lambda_{L}}{\lambda_{L}}=\Delta \varphi=g h=\frac{g h}{c^{2}} \tag{2.100}
\end{equation*}
$$

where we briefly revert to SI units to suggest how small this correction is for everyday heights.

In 1960, Robert Pound and Glen Rebka first observed this redshift in photons traversing an elevator shaft in Harvard University's Jefferson tower. The shift was so small, Pound and Rebka needed to exploit the then recently discovered "recoilless" emission and absorption of gamma rays from $\mathrm{Fe}^{57}$ nuclei "nailed" to crystal lattices (the Mössbauer effect) to precisely define the frequencies. By the late 1990s, the Global Positioning Satellite (GPS) system was already widely used in consumer products like car navigation systems, and it had to incorporate both gravitational and special relativistic time dilation in order to operate correctly.

### 2.3.4 Weakly Curved Spacetime

Recall that the flat space of Euclidean geometry is characterized by the Pythagorean expression

$$
\begin{equation*}
d r^{2}=d x^{2}+d y^{2}+d z^{2} \tag{2.101}
\end{equation*}
$$

for the invariant distance between two nearby places, and the flat spacetime of special relativity is characterized by the pseudo-Euclidean expression

$$
\begin{equation*}
d \tau^{2}=d t^{2}-d r^{2} \tag{2.102}
\end{equation*}
$$

for the invariant interval between two nearby events. We can account for gravitational time dilation by assuming a curved spacetime with interval

$$
\begin{equation*}
d \tau^{2}=(1+2 \varphi) d t^{2}-(1-2 \varphi) d r^{2} \tag{2.103}
\end{equation*}
$$

where $0 \leq \varphi \ll 1$ is the weak Newtonian gravitational potential: $1 \pm 2 \varphi$ correspond to "slow time" and "space warp". This is a good approximation to the curved spacetime geometry of static, weak sources, such as Earth. For simplicity, in what follows, let $x>0$ be the height above the surface, $d y=d z=0$, and $\varphi=g x$. The metric simplifies to

$$
\begin{equation*}
d \tau^{2}=(1+2 g x) d t^{2}-(1-2 g x) d x^{2} \tag{2.104}
\end{equation*}
$$

We can characterize the resulting 1+1-dimensional spacetime by its light cone structure, which we find by demanding $d \tau^{2}=0$ and solving Equation 2.104 for the slopes of the light lines,

$$
\begin{equation*}
\pm \frac{d x}{d t}=\sqrt{\frac{1+2 g x}{1-2 g x}} \sim \sqrt{(1+2 g x)(1+2 g x)} \sim 1+2 g x \tag{2.105}
\end{equation*}
$$

as $2 g x=2 g x / c^{2} \ll 1$. These slopes determine the openings of the light cones and imply exponential light curves, some of which are depicted in Figure 2.25.

Suppose observers at rest at higher and lower heights $x_{H}$ and $x_{L}$ measure the proper times $\Delta \tau_{H}$ and $\Delta \tau_{L}$ between light flashes from the surface separated by the same coordinate time $\Delta t$. Because the spacetime is curved, the proper times will not be the same. In fact, from the metric of Equation 2.104, $\Delta x_{H}=0$ implies

$$
\begin{equation*}
\Delta \tau_{H}^{2}=\left(1+2 g x_{H}\right) \Delta t^{2} \tag{2.106}
\end{equation*}
$$

and hence

$$
\begin{equation*}
\Delta \tau_{H}=\left(1+2 g x_{H}\right)^{1 / 2} \Delta t \sim\left(1+g x_{H}\right) \Delta t \tag{2.107}
\end{equation*}
$$

Similarly, $\Delta x_{L}=0$ implies

$$
\begin{equation*}
\Delta \tau_{L}=\left(1+2 g x_{L}\right)^{1 / 2} \Delta t \sim\left(1+g x_{L}\right) \Delta t \tag{2.108}
\end{equation*}
$$

Therefore,

$$
\begin{equation*}
\frac{\Delta \tau_{H}}{\Delta \tau_{L}}=\frac{1+g x_{H}}{1+g x_{L}} \sim\left(1+g x_{H}\right)\left(1-g x_{L}\right) \sim 1+g\left(x_{H}-x_{L}\right) \tag{2.109}
\end{equation*}
$$

or

$$
\begin{equation*}
\Delta \tau_{H}=(1+g \Delta x) \Delta \tau_{L}=(1+\Delta \varphi) \Delta \tau_{L}>\Delta \tau_{L} \tag{2.110}
\end{equation*}
$$

which is gravitational time dilation.


Figure 2.25: Light cone structure of the static, weakly curved spacetime near Earth's surface (left), and high and low observers measuring the time between flashes of light from the surface (right). Although the coordinate times are the same, the proper times are different. (Proper times are distorted in the diagram as curved spacetime is rendered in the Euclidean space of the page.)

### 2.3.5 Newton's Laws from Curved Spacetime

We know from our discussion of the twin paradox that the unaccelerated, unkinked worldline between any two events is the one of longest proper time. Thus a free particle in flat spacetime follows a path of extremal proper time. By demanding that a free particle in the weakly curved spacetime near Earth also follow a path of extremal proper time, we can recover Newton's laws of motion, including gravity!

The proper time between two events 1 and 2 near Earth is

$$
\begin{equation*}
\Delta \tau=\int_{1}^{2} d \tau \tag{2.111}
\end{equation*}
$$

or using the metric Equation 2.104,

$$
\begin{equation*}
\Delta \tau=\int_{1}^{2} \sqrt{(1+2 \varphi) d t^{2}-(1-2 \varphi) d x^{2}}=\int_{t_{1}}^{t_{2}} \sqrt{1+2 \varphi-(1-2 \varphi) v^{2}} d t \tag{2.112}
\end{equation*}
$$

where $v=d x / d t$. Consistent with our assumption that $\varphi=g x \ll 1$ is the assumption that $v \ll 1$. Hence, neglecting terms that are cubic or higher in small quantities,

$$
\begin{equation*}
\Delta \tau \sim \int_{t_{1}}^{t_{2}}\left(1+2 \varphi-v^{2}\right)^{1 / 2} d t \sim \int_{t_{1}}^{t_{2}}\left(1+\frac{1}{2}\left(2 \varphi-v^{2}\right)\right) d t \tag{2.113}
\end{equation*}
$$

which we can write as

$$
\begin{equation*}
\Delta \tau \sim \int_{t_{1}}^{t_{2}}\left(1-\frac{1}{m}\left(\frac{1}{2} m \dot{x}^{2}-V[x]\right)\right) d t \tag{2.114}
\end{equation*}
$$

where $V[x]=m \varphi=m g x$ is the potential energy and $\dot{x}=v$ is the velocity. The worldline that extremizes $\Delta \tau$ must therefore also extremize

$$
\begin{equation*}
S=\int_{t_{1}}^{t_{2}}\left(\frac{1}{2} m \dot{x}^{2}-V[x]\right) d t \tag{2.115}
\end{equation*}
$$

which we have already extremized, in Section 1.2.1, to get

$$
\begin{equation*}
m a_{x}=m \ddot{x}=-V^{\prime}[x]=-m g \tag{2.116}
\end{equation*}
$$

or

$$
\begin{equation*}
a_{x}=-g \tag{2.117}
\end{equation*}
$$

which is Newton's law of gravity for a flat Earth, in accordance with the equivalence principle. We obtain the same result by neglecting the "space warp" term in Equation 2.103, so objects fall down because time runs slow near Earth!

## Relativity Problems

1. Relativistic Units. How tall are you in seconds? How old are you in meters? (Hint: Use the exact constant $c=299792458 \mathrm{~m} / \mathrm{s}$ to convert from traditional units.)
2. Eris Round Trip. A spaceship leaves Earth and travels to the dwarf planet Eris, about 30 light-hours away, and returns 75 hours later, as measured by clocks on Earth. How much time has elapsed on the spaceship, assuming the spaceship accelerates for only a negligible fraction of its journey? (Hint: Write light-hour as $c \cdot \mathrm{~h}$ and don't convert to SI units.)
3. Nuclear Pancake. The Brookhaven National Laboratory's Relativistic Heavy Ion Collider (RHIC) accelerates lead nuclei to $v=0.99995 c$, where a lead nucleus at rest is a sphere $d=14 \mathrm{fm}$ in diameter.
(a) In the lab frame, what is the shape and size of the nuclei?
(b) In a head-on collision, how much time elapses between the nuclei first touching and completely overlapping?
4. Moving Stick. A stick of length $\ell$ moves past you at speed $v$. As observed in frame $\mathcal{O}$ and calculated in frame $\mathcal{C}$, let $\Delta t_{\mathcal{O C}}$ be the time between the front and rear ends of the stick coinciding with you. Calculate the following times. (Hint: Imagine clocks at both ends of the stick, synchronized in the stick's frame, and imagine you as a clock.)
(a) Time $\Delta t_{\mathcal{Y} Y}$ as observed by you and calculated by you.
(b) Time $\Delta t_{\mathcal{S} \mathcal{Y}}$ as observed by stick and calculated by you.
(c) Time $\Delta t_{\mathcal{Y S}}$ as observed by you and calculated by stick.
(d) Time $\Delta t_{\mathcal{S S}}$ as observed by stick and calculated by stick.
5. Pole \& Barn. Alice says she can fit a 20 -meter pole in a 10 -meter barn by running with the pole so fast that it is contracted to half its length. However, Bob points out that, from the pole's frame of reference, it is the barn that is contracted to half its length! Analyze this classic paradox, as we analyzed the tethered rockets paradox, with synchronized clocks at each end of the pole.
(a) What happens in the pole's frame?
(b) What happens in the barn's frame?
(c) Is the pole every entirely in the barn? Can you trap it in the barn by quickly closing the front and rear doors?
6. Cookie Cutter. Cookie dough lies on a conveyor belt moving with speed $v$. A circular stamp of proper diameter $\ell$ stamps out cookies as the dough rushes by beneath it. Incorporating relativistic effects, quantitatively answer the following.
(a) What is the shape of the cookies (when you later eat them)?
(b) Explain the shape in the factory's frame (where the cutter moves only up and down).
(c) Explain the shape in the dough's frame.
7. Radio Exchange. Two spaceships move in opposite directions at Earth speed $v=3 c / 5$. Assume the spaceships were together at Earth position $x_{E}=0$ when Earth time $t_{E}=0$, and the leftward one sends a radio signal (at the invariant speed $c$ ) toward the rightward one when they are an Earth distance $d$ apart.
(a) In Earth's frame, compute the spacetime coordinates $\left\{t_{E_{1}}, x_{E_{1}}\right\}$ of the emission event, $\left\{t_{E_{2}}, x_{E_{2}}\right\}$ of the reception event, and the duration $\Delta t_{E}=t_{E_{2}}-t_{E_{1}}$. (Hint: Answer in a multiple of $d / c$.)
(b) From these, Lorentz transform to find the corresponding $\left\{t_{L_{1}}, x_{L_{1}}\right\}$, $\left\{t_{L_{2}}, x_{L_{2}}\right\}$, and $\Delta t_{L}=t_{L_{2}}-t_{L_{1}}$ in the leftward frame.
(c) Lorentz transform to find the corresponding $\left\{t_{R_{1}}, x_{R_{1}}\right\}$, $\left\{t_{R_{2}}, x_{R_{2}}\right\}$, and $\Delta t_{R}=t_{R_{2}}-t_{R_{1}}$ in the rightward frame.
8. Staging. The first stage of a multi-stage rocket boosts the rocket to a speed of 0.1 c relative to Earth before being jettisoned. The next stage boosts the rocket to a speed of 0.1 c relative to the final speed of the first stage, and so on. How many stages is needed to boost the payload to a speed in excess of $0.95 c$ relative to Earth?

## 9. Headlight Effect.

(a) A rocket coasts by Earth with velocity $v$. It emits a flash of light at an angle $\varphi_{R}$ from its direction of motion relative to itself. Show that the direction of the flash is $\varphi_{E}$ from the rocket's direction relative to Earth, where in natural units

$$
\begin{equation*}
\cos \varphi_{E}=\frac{\cos \varphi_{R}+v}{1+v \cos \varphi_{R}} \tag{2.118}
\end{equation*}
$$

(b) Suppose the rocket emits the light uniformly in all directions, relative to itself. Using the previous result, show that the light emitted in the forward hemisphere, relative to the rocket, is concentrated in a cone of angle

$$
\begin{equation*}
\varphi_{E}=\sin ^{-1} \frac{1}{\gamma}=\arcsin \left[\frac{1}{\gamma}\right] \tag{2.119}
\end{equation*}
$$

relative to Earth.
(c) What is the angle of this cone for $v=0.99$ ?
10. Spacetime Causality. Consider two events $E_{1}$ and $E_{2}$ in spacetime. Use Loedel spacetime diagrams to illustrate the following statements.
(a) If $E_{1}$ and $E_{2}$ are separated by a timelike interval, and $E_{1}$ precedes $E_{2}$ in one reference frame, then $E_{1}$ precedes $E_{2}$ in all reference frames.
(b) If $E_{1}$ and $E_{2}$ are separated by a spacelike interval, then in some references frames $E_{1}$ precedes $E_{2}$ and in other reference frames $E_{2}$ precedes $E_{1}$.
(c) Which kind of events can be causally related?
11. Ansible. The ansible is a science fiction interstellar instantaneous communications device invented by Ursula K. Le Guin and used by Orson Scott Card. Show how you could use an ansible to communicate with your past and create a causal paradox. Specifically, show how you can send a signal at time $t$ and receive it at time $t / \gamma^{2}<t$.
12. Tethered Rockets. Analyze the Section 2.1.5 tethered rockets paradox with a carefully drawn Loedel spacetime diagram relating the measurements of the uniformly moving Earth and space station observers. (Your diagram should fill most of an $8.5 \times 11$-inch page and be accurate and neat. Use a straight edge and compass or a computer drawing program.)

## 13. Things That Go Faster Than Light.

(a) A very long straight rod, inclined at an angle $\varphi$ to a horizontal rod, moves downward at constant speed $v$ so that the intersection between the two rods moves horizontally at constant speed $v_{H}$. Under what conditions can this latter speed exceed light speed? Can the rods (or scissors) be used to transmit a message horizontally faster than light?
(b) A powerful searchlight sweeps out a circle in time $T$. You and I are a distance $d$ from the searchlight and an angle $\theta$ apart. Under what conditions will the searchlight beam sweep from me to you faster than a light signal could travel between us? Can we use the searchlight to transmit information faster than light?
14. 4-Momentum. A $2.0-\mathrm{kg}$ object moves with speed $1.8 \times 10^{8} \mathrm{~m} / \mathrm{s}$ in the $x$-direction. What are the components of the object's 4 -momentum, in SI units of $\mathrm{kg} \mathrm{m} / \mathrm{s}$ ?
15. Box of Light. Consider photons of energy $E=p$ and spacetime momen$\operatorname{tum} \vec{p}=\{E, \vec{p}\}$.
(a) Calculate the mass of one of the photons.
(b) Calculate the mass of two photons moving parallel.
(c) Calculate the mass of two photons moving perpendicular.
(d) Can the contents of a box of light have nonzero mass?
16. Pion-Muon Race. A neutral pion and a muon each have energy 10 GeV . In a $100-\mathrm{m}$ race, which wins and by how much distance? (Hint: Assume $m_{\pi_{0}} c^{2}=135 \mathrm{MeV}$ and $m_{\mu} c^{2}=106 \mathrm{MeV}$.)
17. Light Bulb. How much mass does a 100 -watt light bulb dissipate (in heat and light) in one year?
18. Energy Costs. Assume electrical energy costs about $\$ 0.05$ per kW-hr.
(a) If you have $\$ 1$ million to buy electrical energy to convert to kinetic energy, about how fast can you make a $1.0-\mathrm{g}$ object travel?
(b) To compete with email, the U.S. Postal Service offers Super Express Mail, where a letter is sent to its destination at $0.99 c$ using a special Letter Accelerator. If a typical letter has a mass of about 25 g , what will be the minimum cost of the letter's stamp?
(c) If the Super Express Mail letter misses its target and hits a nearby building, describe the consequences. (For comparison, a megaton of TNT is about 4 petajoules.)
19. Inelastic Collision. Two identical particles of mass $m$ moving at speed $v$ collide and stick together.
(a) If relativistic energy and momentum are conserved in the center-ofmass frame, what is the mass of the final particle?
(b) Now boost to the frame of reference of the right particle and check that the new 4-momenta are consistent with the relativistic velocity addition formula. (Hint: Show that the relativistic stretch of the left particle is $\gamma^{\prime}=\left(1+v^{2}\right) /\left(1-v^{2}\right)$.)
20. More Sleep. To exploit gravitational time dilation and be better rested for the next day's exam, should you sleep in the attic or the basement? Estimate how much you can adjust your sleep this way.

## Chapter 3

## Quantum Physics

Relativity "completes" classical physics; quantum physics subsumes it.
Richard Feynman wrote, "Things on a very small scale behave like nothing that you have any direct experience about. They do not behave like waves, they do not behave like particles, they do not behave like clouds, or billiard balls, or weights on springs, or like anything that you have ever seen."

You can't learn about atoms by playing with billiard balls, but you can learn about billiard balls by studying atoms. Classical physics follows from quantum physics, not the other way around.

### 3.1 Interference and Superposition

The quantum analogues of the classic wave concepts of interference and superposition reveal deep and surprising features of quantum reality.

### 3.1.1 Beam Splitter Probabilities

A beam splitter is an optical device that transmits half the light incident on it and reflects the other half. It could be a mirror with an unusually thin metal layer or a dielectric slab whose thickness and index of refraction together produce the desired constructive and destructive interference. We will imagine it to be two prisms separated by a small gap, as in Figure 3.1. Varying the thickness of the spacer, a thin film that separates the two prisms, can produce any ratio of transmitted to reflected light, via an exponentially decaying evanescent wave propagating through the spacer, a phenomenon called frustrated total internal reflection.

For simplicity, we imagine that our light source is monochromatic. This could be a laser, which consists of an electrically excited medium bounded by two mirrors, one of which is partially reflecting. De-excitation results in monochromatic, coherent, and directional light escaping the partially reflecting mirror.


Figure 3.1: Beam splitter reduces the reflected and transmitted bright light intensity by $1 / 2$ and amplitude by $1 / \sqrt{2}$.

At sufficiently high intensity, light behaves like an electromagnetic wave. The frequency of visible light is so high ( $\nu=\omega / 2 \pi \sim 100 \mathrm{THz}$ ) that our eyes and cameras cannot follow its oscillations. Instead, we are sensitive to the time averaged square of its electric field, which is called intensity (or irradiance). Intensity is the energy per unit area per unit time transported by the wave. If the electric field varies sinusoidally, $\mathcal{E}=\mathcal{E}_{0} \cos [k x-\omega t]$, then its intensity is proportional to the electric field amplitude squared, $I \propto \overline{\mathcal{E}^{2}} \propto \mathcal{E}_{0}^{2}$. In appropriate units, we will take $I=\mathcal{E}_{0}^{2}$. Thus, in reducing the intensity of the transmitted and reflected waves by $1 / 2$, the beam splitter of Figure 3.1 reduces the electric field amplitude by $1 / \sqrt{2}$.


Figure 3.2: Beam Splitter reflects and transmits photons with probability $1 / 2$. For each trial, each PhotoMultiplier Tube reports " 1 " if it detected a photon and " 0 " otherwise. Single Photon Source produces a pair of opposing photons so that one enters the Beam Splitter and the other announces the trial.

At sufficiently low intensity, the graininess of light becomes apparent, and
light behaves like a stream of particles, called photons. The energy of single visible-light photons is so small $(E=h \nu \sim 1 \mathrm{eV})$ that our eyes are not (quite) able to detect them. Instead, as in our second teaser, we will detect them with photomultiplier tubes (PMTs) or semiconductor avalanche photodiodes (APDs), which exploit the photoelectric effect to initiate an electron cascade that amplifies a single photoelectron to a macroscopic current pulse with near $100 \%$ efficiency.

We might radically dim our light source using neutral density filters (NDFs) or crossed polarizers, so that only one photon is in the beam splitter at any one time. To avoid photon bunching, we instead use a single photon source. For example, we wait for positronium to decay into opposing photons and detect one to herald the other. More practically, we pump a nonlinear birefringent crystal, like $\beta-\mathrm{BaB}_{2} \mathrm{O}_{4}(\mathrm{BBO})$, with a UV laser to produce two opposing IR photons and again detect one to announce the other.

What happens? If the first photon is reflected, shouldn't they all be reflected? If the first is transmitted, shouldn't they all be transmitted? How then could we recover the bright light classical results from the faint light quantum results by gradually increasing the light intensity?

We put nature to the test and find that the experiment is not repeatable. Instead, individual photons are transmitted or reflected with probability $1 / 2$, as in Figure 3.2, where the binary data strings at each PMT indicate whether a photon has been detected (1) or not (0) during each trial. More generally, we find that the probability of detecting a photon is proportional to the square of the amplitude of the electric field of the corresponding classical wave. In this way, faint light quantum experiments correspond to bright light classical experiments.

### 3.1.2 Two Interpretations

The conventional or Copenhagen interpretation (CI) is that quantum probabilities are ontological rather than epistemological. They reflect how things really are, not merely what we can know about them. They are inherent in nature, not merely limitations in our measuring apparatus.

Einstein famously objected, "God does not play dice with the universe".
In the post-Einstein Many Worlds interpretation (MWI), the ontological probabilities are eliminated. Instead, each photon is both transmitted and reflected, and the world splits into two histories, one for each possibility! Epistemological randomness is apparent only to observers, like us, confined to single histories. From a God's eye point of view, the MWI is deterministic and, for the beam splitter, symmetric (both of two equally likely possibilities are realized), but at the ontological expense of invoking an infinity of equally real worlds to explain our single observable world.

Other interpretations exist, but none preserve classical reality.

### 3.1.3 Mach-Zehnder Interferometer

Probabilities alone don't exhaust the novelty of quantum reality.
Suppose we recombine the light from a beam splitter using two mirrors and a second beam splitter, as in Figure 3.3. Such a device is called a Mach-Zehnder interferometer. If "T" and "R" represent "transmitted" and "reflected", then the four paths through the interferometer can be denoted RRR, TRT, RRT, TRR, where the first two paths exit up and the second two paths exit right. All paths have the same length, but each transmission and reflection is accompanied by a phase shift that depends on the details of the optics. Assuming a phase shift of $\pi / 2$ at each reflection, light waves interfere constructively when exiting right (and hence destructively when exiting up), because the corresponding paths involve the same number of transmissions and the same number of reflections. (In practice, if one of the mirrors or beam splitters is slightly canted, than the interference produces a fringe pattern of parallel stripes.)


Figure 3.3: Mirrors (left) and a second beam splitter (right) recombine bright light split by the first beam splitter.

Suppose we radically dim our light source, so that only one photon is in the interferometer at any one time, as in Figure 3.4. What happens? Without the recombining beam splitter, the data strings at the PMTs are perfectly anticorrelated but random. With the recombining beam splitter, the data strings are still perfectly anticorrelated but are now homogeneous, and all photons exit right, in agreement with the high intensity experiment. Apparently, interference happens even with only one photon in the apparatus at a time!

Note how the addition of the recombining beam splitter radically alters the output of the device. If individual photons were somehow "splitting" (or not) at the first beam splitter, how could they know whether (or not) the recombining beams splitter was in place? In fact, since the speed of the photons is $c \sim$ $10^{9} \mathrm{~km} / \mathrm{hr} \sim 0.3 \mathrm{~m} / \mathrm{ns}$, using nanosecond electronics in a table-top experiment, we can decide to remove or introduce the recombining beam splitter after the photon has interacted with the first beam splitter! The results of such delayed choice experiments are exactly the same: in those trials with the recombiner, all photons exit right; in those trials without the recombiner, half the photons


Figure 3.4: Mirrors (left) and a second beam splitter (right) recombine photon paths split by the first beam splitter.
exit right and half exit up.
Can we check the paths taken by the photons? Since each photon carries momentum $p=h / \lambda$, if we float one of the two mirrors, then the mirror's recoil (or not) reveals the photon's path. However, in such which-way experiments, the constructive and destructive interference, which makes all photons exit right and none exit up, is destroyed, and instead half the photons exit right and half exit up. Indeed, which-way information is consistent with the particle nature of light but is inconsistent with the wave nature of light. Particles take definite paths and do not interfere, while waves take all paths and do interfere. Apparently, incompatible experimental arrangements elicit complementary aspects of the wave-particle duality of light: which-way information (no recombiner or floating mirrors) elicits the particle aspect of light, while no which-way information (recombiner and fixed mirrors) elicits the wave aspect of light.

### 3.1.4 Quantum Eraser

A quantum eraser is a measurement that destroys which-way information. Because the eraser can restore an interference pattern, Neils Bohr's classic argument that the interference pattern is lost because it has been randomly disrupted by the measurement process is not applicable. Suppose we insert a $0^{\circ}$ (horizontal) polarizer in one path of our Mach-Zehnder interferometer, a $90^{\circ}$ (vertical) polarizer in the other path, a $45^{\circ}$ (diagonal) polarizer at the input, and final (analyzing) polarizers at the outputs, as in Figure 3.5. The path polarizers encode which-way information in a photon by altering its spin angular momentum, the quantum counterpart to the polarization of the corresponding classical wave, while not changing its linear momentum. (If we float one of the mirrors and it recoils, we obtain which-way information at the expense of changing the photon's linear momentum; although, if we float one of the mirrors and it does not recoil, we obtain which-way information without affecting the photon.)

If we rotate the analyzers to $0^{\circ}$, then we observe light propagating in only one path of the interferometer, and no interference occurs, so half the photons


Figure 3.5: Path polarizers provide which-way information and interference is lost (left diagram), where photons exit equally up and right. Rotating the final, analyzing polarizers $45^{\circ}$ destroys the which-way information and restores the interference (right diagram), where photons exit right but not up.
exit right and half exit up. If we rotate the analyzers to $90^{\circ}$, we obtain similar results. If we remove the analyzers, still no interference occurs. Apparently, actually detecting a particular path is not necessary, as merely encoding whichway information is sufficient to destroy the interference. However, if we rotate the analyzers to $45^{\circ}$, so we can no longer distinguish between the $0^{\circ}$ (horizontal) and $90^{\circ}$ (vertical) polarizations, which-way information is erased, interference is restored, and all photons exit right - and this is so even if the quantum erasure is a delayed choice!

### 3.1.5 Interaction-Free Measurement

When we float one of the two mirrors in the Mach-Zehnder interferometer, we lose the single-photon interference, even if the single photon reflects off the other, stationary mirror. How can the floating mirror affect the photon if the photon doesn't even come near it? Quantum physics allows us to test counterfactuals, things that might have happened but did not!

The bomb testing problem of Avshalom Elitzur and Lev Vaidman dramatically illustrates such interaction-free or null measurements. Consider bombs so sensitive that even the slightest movement of their detonators will explode them. Unfortunately, some fraction of the detonators are jammed and the attached bombs are consequently duds. No classical way exists to identify good bombs without exploding them, but quantum physics provides a way. We can test a bomb by attaching its detonator to one of the mirrors of the Mach-Zehnder interferometer, as in Figure 3.6.

If all photons exit right, the different alternatives are interfering, constructively right and destructively up. The mirror and its attached detonator must be fixed, so no which-way information exists. Hence, the bomb is a dud. However,


Figure 3.6: Interference (left) reveals a jammed detonator (and a fixed mirror), while no interference (right) reveals a working detonator (and a recoilable mirror).
if even one photon exits up, the different alternatives are not interfering. The mirror and the attached detonator can, in principle, recoil and thereby provide which-way information. Hence, the bomb is good, and if the photons have all reflected off the stationary mirror, the bomb is unexploded.

In practice, we can only harvest

$$
\begin{equation*}
\frac{1}{4}+\frac{1}{16}+\frac{1}{64}+\cdots=\frac{1}{3} \tag{3.1}
\end{equation*}
$$

of the working bombs this way. However, a variation of this technique can arbitrarily reduce the fraction of wasted bombs. In the future, null measurements may allow doctors to x-ray patients without exposing them to (potentially harmful) x-rays!

### 3.1.6 Quantum Computing

Classically, to distinguish a real coin, with a head and a tail, from a trick coin, with two heads (or two tails), we would need to look at each side separately and then compare the results. Using David Deutsch's "two-bit" quantum algorithm, we can do it all at once!

As a slightly simplified quantum version of the problem, suppose that a piece of $\pi$-phase-shifting dielectric may or may not be in one or both paths of a Mach-Zehnder interferometer, as in Figure 3.7. Classically, the presence of the dielectric in one path but not the other converts, at the exit, constructive interference to destructive interference, and vice versa. Quantumly, a single photon explores both paths in parallel. If it exits right instead of up, we know that both paths are the same; with one photon, we have obtained two bits of information.


Figure 3.7: An "inverting" ( $\pi$-phase shifting) dielectric is in one or both paths of the interferometer. If the photon exits right (top row), the paths are identical. If it exits up (bottom row), the paths are nonidentical.

Deutsch's 1985 two-bit scheme was the first quantum computing algorithm. In 1994, Peter Shor discovered a quantum computing algorithm to factor numbers in polynomial time, so that factoring an $N$-bit number requires time $O\left[N^{k}\right]$, for constant $k$. This is something no classical computer can do. Shor's algorithm would revolutionize cryptography, if implemented. In 1996, Lov Grover discovered a quantum computing algorithm to search a database of $N$ elements in time $O[\sqrt{N}]$, again faster than any classical computer, which requires time $O[N]$. In 2001, an early quantum computer ran Shor's algorithm and successfully factored 15 . (That's not a 15 -digit number; that's the number $15=3 \times 5$.)

The MWI provides an easy heuristic for understanding the source of the advantage of these quantum algorithms: they distribute the calculations among many parallel universes!

### 3.2 Indeterminacy and Entanglement [Optional]

Quantum reality includes the additional nonclassical surprises of indeterminacy and entanglement.

### 3.2.1 Mach-Zehnder Classical Model

Prior to creating a more explicit quantum model of the Mach-Zehnder interferometer, let's first create a more quantitative classical model. At high intensity,
light is split into two wave trains at the first beam splitter, which are recombined at the second beam splitter and exit up and right. Let the electric field magnitude at the entrance be

$$
\begin{equation*}
\mathcal{E}[0, t]=\mathcal{E}_{0} \cos [\omega t] \tag{3.2}
\end{equation*}
$$

where $t$ is the time elapsed, and $\omega=2 \pi / T$ is the temporal frequency of the wave train. Then, the electric field magnitude at the exit due to the wave train reflected by mirror $n$ is

$$
\begin{equation*}
\mathcal{E}_{n}[x, t]=\frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \mathcal{E}_{0} \cos \left[\omega t-k z+\delta_{n}\right] \tag{3.3}
\end{equation*}
$$

where $z$ is the distance traveled, $k=2 \pi / \lambda$ is the spatial frequency, and $\delta_{n}$ is the extra phase shift due to reflection. Since $\omega / k=\lambda / T=c$, the spacetime phase $\varphi=\omega t-k z=k(c t-z)$ is zero at $z=c t$, and hence represents a wave traveling in the $\hat{z}$ direction at speed $c$. The factors of $1 / \sqrt{2}$ are due to the beam splitters. The total electric field magnitude at the exit is the superposition

$$
\begin{equation*}
\mathcal{E}=\mathcal{E}_{1}+\mathcal{E}_{2}=\frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \mathcal{E}_{0}\left(\cos \left[\varphi+\delta_{1}\right]+\cos \left[\varphi+\delta_{2}\right]\right) \tag{3.4}
\end{equation*}
$$

Our eyes and cameras are sensitive to the time-averaged square of this electric field, which is the intensity

$$
\begin{equation*}
I=\overline{\mathcal{E}^{2}}=\frac{1}{2} \frac{1}{2} \mathcal{E}_{0}^{2} \overline{\left(\cos ^{2}\left[\varphi+\delta_{1}\right]+2 \cos \left[\varphi+\delta_{1}\right] \cos \left[\varphi+\delta_{2}\right]+\cos ^{2}\left[\varphi+\delta_{2}\right]\right)} \tag{3.5}
\end{equation*}
$$

Using the trigonometric identity $2 \cos u \cos v=\cos [u+v]+\cos [u-v]$, this becomes

$$
\begin{equation*}
I=\frac{1}{2} \frac{1}{2} \mathcal{E}_{0}^{2}\left(\overline{\cos ^{2}\left[\varphi+\delta_{1}\right]}+\overline{\cos \left[2 \varphi+\delta_{1}+\delta_{2}\right]}+\overline{\cos \left[\delta_{1}-\delta_{2}\right]}+\overline{\cos ^{2}\left[\varphi+\delta_{2}\right]}\right) \tag{3.6}
\end{equation*}
$$

Since the time average of a sinusoid (over an integer number of periods) vanishes, and the time average of the square of a sinusoid is $1 / 2$, we have

$$
\begin{equation*}
I=I_{0} \frac{1+\cos \delta}{2} \tag{3.7}
\end{equation*}
$$

where $I_{0}=\mathcal{E}_{0}^{2} / 2$ is the entrance intensity and $\delta=\delta_{1}-\delta_{2}$ is the difference between the reflection phase shifts of the two paths.

We now assume a phase shift of $\pi / 2$ radians at each reflection. (The actual phase shifts depend on the detailed characteristics of the optical elements, but can always be adjusted by inserting dielectric slabs in one or both paths of the interferometer). At the up exit, the difference in phase shifts $\delta_{u}=3(\pi / 2)-$ $(\pi / 2)=\pi$, and so the intensity $I_{u}=0$. At the right exit, the difference in phase shifts $\delta_{r}=2(\pi / 2)-2(\pi / 2)=0$, and so the intensity $I_{r}=I_{0}$.

### 3.2.2 Mach-Zehnder Quantum Model 1

Now, how can we create a more quantitative quantum model of the MachZehnder interferometer, one that works for faint light, when only single photons are in the interferometer? In particular, how can we reproduce wave interference with particles? We will adopt a model due to Richard Feynman.


Figure 3.8: Vertical projection (left) of a rotating vector (right) varies sinusoidally.

The projection of a rotating arrow in a fixed direction varies sinusoidally, like a wave train, as in Figure 3.8. This suggests the following scheme. Along each path through the interferometer, we imagine that a photon carries an arrow that rotates at the frequency of the corresponding classical light. For the purposes of the illustration in Figure 3.9, we assume that each arrow rotates $\pi / 4$ radians per step, plus an extra $\pi / 2$ radians per reflection, and shortens by $1 / \sqrt{2}$ at each beam splitter. At the mirrors and beam splitters, we draw the arrow just before in gray and just after in black. Adding the arrows for both paths at the exit and squaring correctly gives the probability of detecting the photon. For bright light, this corresponds to squaring the electric field amplitude to obtain the intensity.

### 3.2.3 Mach-Zehnder Quantum Model 2

We can conveniently and compactly represent the rotating arrows by complex numbers $\rho e^{i \varphi}$ of modulus $\rho$ and argument $\varphi=k z-\omega t$, where $z$ and $t$ are the (real) propagation distance and time, and $\omega / k=c$. We can label the states of a photon in each segment of the interferometer using the conventional quantum vector notation $|\bullet\rangle$, called a ket (from the word bracket), as in Figure 3.10. These kets can represent the photon's non-spin degrees of freedom, such as position and linear momentum. At the first beam splitter, the initial photon state $|A\rangle$ evolves to a quantum superposition of a transmitted photon state $|B\rangle$ and a reflected photon state $|C\rangle$. If $l$ is the length and width of the interferometer, and if the first beam splitter is at $z=0$, then at time $t$ the initial photon evolves into the superposition

$$
\begin{equation*}
e^{-i \omega t}|A\rangle \xrightarrow{S} \frac{1}{\sqrt{2}} e^{i(k l / 2-\omega t)}|B\rangle+\frac{1}{\sqrt{2}} e^{i(k l / 2+\pi / 2-\omega t)}|C\rangle, \tag{3.8}
\end{equation*}
$$



Figure 3.9: A photon carries an imaginary arrow that rotates at the frequency of the corresponding classical light. Adding the arrows at the exit for both paths and squaring gives the probability of detecting the photon: unity for exiting right (top row) and zero for exiting up (bottom row).
where the complex numbers multiplying each state record the amplitude and phase of the rotating arrows: the moduli $1 / \sqrt{2}$ account for the passage through the beam splitter, while the $\pi / 2$ in the argument of the second complex number accounts for the reflection phase shift.

According to the CI, if the experiment were stopped here, and we observed whether the photon were transmitted or reflected, the square of the moduli of these complex numbers would be the corresponding probabilities,

$$
\begin{align*}
& \mathcal{P}[B]=\left|\frac{1}{\sqrt{2}} e^{i(k l / 2-\omega t)}\right|^{2}=\frac{1}{2}  \tag{3.9a}\\
& \mathcal{P}[C]=\left|\frac{1}{\sqrt{2}} e^{i(k l / 2+\pi / 2-\omega t)}\right|^{2}=\frac{1}{2} \tag{3.9b}
\end{align*}
$$

According the MWI, the quotient of these two numbers $\mathcal{P}[B] / \mathcal{P}[C]=1$ is the branching ratio for the two different histories.

In practice, to calculate the interference, we need only record the difference in the phases of the two paths. Consequently, we can abbreviate the effect of the first beam splitter by the evolution

$$
\begin{equation*}
|A\rangle \xrightarrow{S} \frac{1}{\sqrt{2}}(|B\rangle+i|C\rangle), \tag{3.10}
\end{equation*}
$$



Figure 3.10: Photon states in the interferometer.
where $i=e^{i \pi / 2}$ accounts for the reflection phase shift. Similarly, the mirrors induce

$$
\begin{align*}
& |B\rangle \xrightarrow{S} i|D\rangle,  \tag{3.11a}\\
& |C\rangle \xrightarrow{S} i|E\rangle, \tag{3.11b}
\end{align*}
$$

while the second beam splitter induces

$$
\begin{align*}
& |D\rangle \xrightarrow{S} \frac{1}{\sqrt{2}}(i|F\rangle+|G\rangle),  \tag{3.12a}\\
& |E\rangle \xrightarrow{S} \frac{1}{\sqrt{2}}(|F\rangle+i|G\rangle) . \tag{3.12b}
\end{align*}
$$

The complete evolution is

$$
\begin{align*}
&|A\rangle \xrightarrow{S} \\
& \quad \xrightarrow{S}(|B\rangle+i|C\rangle) \\
& \frac{1}{\sqrt{2}}(i|D\rangle-|E\rangle) \\
& \xrightarrow{S} \frac{1}{2}(-|F\rangle+i|G\rangle-|F\rangle-i|G\rangle)  \tag{3.13}\\
& \quad \xrightarrow{S}-|F\rangle
\end{align*}
$$

or more explicitly

$$
\begin{equation*}
|A\rangle \xrightarrow{S}-1|F\rangle+0|G\rangle . \tag{3.14}
\end{equation*}
$$

Hence, the probabilities

$$
\begin{equation*}
\mathcal{P}[F]=|-1|^{2}=1 \tag{3.15}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathcal{P}[G]=|0|^{2}=0, \tag{3.16}
\end{equation*}
$$

as expected. The certainty of $|F\rangle$ (exiting right) and the impossibility of $|G\rangle$ (exiting up) is an example of quantum interference.

### 3.2.4 Hilbert Space

In general, if $|A\rangle$ and $|B\rangle$ are quantum states, than any linear combination $a|A\rangle+b|B\rangle$, with complex coefficients $a$ and $b$, is also a quantum state. In fact, such states form a Hilbert space: a linear vector space with a complex scalar product. For example, the calcite crystal of Section 1.1.2 can induce a photon to evolve to a state $|\psi\rangle$ that is a superposition of horizontal $|h\rangle$ and vertical $|v\rangle$ polarization, namely

$$
\begin{equation*}
|\psi\rangle=a|h\rangle+b|v\rangle \tag{3.17}
\end{equation*}
$$

where $|a|^{2}+|b|^{2}=1$ to conserve probability. (Measurement will certainly find the photon in one of the two states.) The set of all such states form a quantum bit or qubit, which is of fundamental importance in quantum computing: while a classical bit can be in one of two states, a qubit can be in an infinite number of superpositions of states.

A quantum superposition is a kind of complex-number-weighted coexistence of possibilities (or potentialities). According to the CI, the absolute square of the weights correspond to the probabilities of measuring the alternatives. According to the MWI, the quotient of the weights is the branching ratio for the two different histories. (The branching ratio must be a rational number, but rationals can approximate real numbers arbitrarily well.)

### 3.2.5 Quantum Evolution

As we shall show, superpositions evolve continuously and deterministically under the Schrödinger differential equation, in both the CI and the MWI. For example,

$$
\begin{equation*}
|\psi\rangle \xrightarrow{S}\left|\psi^{\prime}\right\rangle=a^{\prime}|h\rangle+b^{\prime}|v\rangle . \tag{3.18}
\end{equation*}
$$

The CI, but not the MWI, also includes a discontinuous and probabilistic collapse of a superposition to classical probability-weighted alternatives when the system is measured (or observed or registered). For example,

$$
\left|\psi^{\prime}\right\rangle \xrightarrow{M}\left\{\begin{array}{ll}
|h\rangle, & \mathcal{P}[h]=\left|a^{\prime}\right|^{2}  \tag{3.19}\\
|v\rangle, & \mathcal{P}[v]=\left|b^{\prime}\right|^{2}
\end{array}\right\} .
$$

While the $S$-evolution is uncontroversial, the same cannot be said about the $M$-evolution.

### 3.2.6 The Measurement Problem (and Schrödinger's Cat)

Consider a variation of the (in)famous Schrödinger cat experiment, wherein a (working) bomb amplifies a microscopic superposition to macroscopic proportions, as in Figure 3.11, where a single photon interacts with a beam splitter.

In the absence of a measurement, the system $|\psi\rangle$ evolves into a superposition of reflected and transmitted photons

$$
\begin{equation*}
|\psi\rangle \xrightarrow{S}\left|\psi^{\prime}\right\rangle=\frac{1}{\sqrt{2}}|\uparrow\rangle+\frac{1}{\sqrt{2}}|\rightarrow\rangle, \tag{3.20}
\end{equation*}
$$

and unexploded and exploded bombs

$$
\begin{equation*}
\left|\psi^{\prime}\right\rangle \xrightarrow{S}\left|\psi^{\prime \prime}\right\rangle=\frac{1}{\sqrt{2}}|\uparrow, \bullet\rangle+\frac{1}{\sqrt{2}}|\rightarrow, \star\rangle \tag{3.21}
\end{equation*}
$$

and calm and distressed observers

$$
\begin{equation*}
\left.\left.\left.\left|\psi^{\prime \prime}\right\rangle \xrightarrow{S}\left|\psi^{\prime \prime \prime}\right\rangle=\frac{1}{\sqrt{2}} \right\rvert\, \uparrow, \bullet, \cdot\right)\right\rangle+\frac{1}{\sqrt{2}}|\rightarrow, \star, \odot\rangle . \tag{3.22}
\end{equation*}
$$

Such macroscopic superpositions are called Schrödinger cat states. (In the original thought experiment, the observer was a cat.) However, we do not observe superpositions of unexploded and exploded bombs, nor of calm and distressed people, whatever that might mean. According to the CI, to collapse the superposition

$$
\left|\psi^{\prime \prime \prime}\right\rangle \xrightarrow{M}\left\{\begin{array}{ll}
|\uparrow, \bullet, \odot\rangle, & \mathcal{P}=|1 / \sqrt{2}|^{2}=1 / 2  \tag{3.23}\\
|\rightarrow, \star, \odot\rangle, & \mathcal{P}=|1 / \sqrt{2}|^{2}=1 / 2
\end{array}\right\}
$$

a measurement must occur at the beam splitter, or at the bomb, or at the observer, or ... .


Figure 3.11: Single photon incident on a beam splitter is reflected and detected by a PMT, calming the observer (left), or is transmitted and detonates a bomb, distressing the observer (right). The $S$-evolution places the photon, the bomb, and the observer in a macroscopic quantum superposition, a Schrödinger cat state.

But exactly where and when does the superposition collapse? Are not the beam splitter, the bomb, and the observer all ultimately quantum systems?

Where is the threshold between microscopic and macroscopic, between experiment and experimenter, between phenomenon and observer, between quantum and classical? Physicist Eugene Wigner argued that the threshold is human consciousness. The chief architect of the CI, Neils Bohr, argued that the threshold is relative; it depends on one's point of view, on how one chooses to analyze the experiment. So no one right answer exists as to where and when the superposition's complex-number-weighted coexistence of multiple possibilities collapses into a probability-weighted single reality.

The MWI dispenses with this so-called "measurement" problem by entirely eliminating the discontinuous, probabilistic $M$-evolution. According to the MWI, two histories continuously and deterministically emerge from the experiment, one including a calm observer, an unexploded bomb, and a reflected photon, the other including a distressed observer, an exploded bomb, and a transmitted photon. The apparent probabilities and discontinuities are merely artifacts of individual observers being confined to single histories.

### 3.2.7 Polarization

Beam splitters and mirrors control the direction of classical light and the linear momenta of photons. Calcite crystals, quarter wave plates, and polarizers control the polarization of classical light and the angular momenta (or spin) of photons. This latter capability facilitates investigation of additional aspects of quantum reality.

In classical optics, polarization refers to the oscillation of the electric field of light. For example, if light is traveling in the $z$-direction at speed $c=\omega / k$, then

$$
\begin{align*}
\overrightarrow{\mathcal{E}}_{H}[\delta] & =\hat{x} \mathcal{E}_{0} \cos [k z-\omega t+\delta]  \tag{3.24a}\\
\overrightarrow{\mathcal{E}}_{V}[\delta] & =\hat{y} \mathcal{E}_{0} \cos [k z-\omega t+\delta] \tag{3.24b}
\end{align*}
$$

represent horizontal and vertical linearly polarized light, because the electric field is oscillating in a line. We can superpose this light with different relative phases $\delta_{H}-\delta_{V}$ to create differently polarized light. For example, if the relative phase shift is zero, then

$$
\begin{equation*}
\overrightarrow{\mathcal{E}}_{D}=\overrightarrow{\mathcal{E}}_{H}[0]+\overrightarrow{\mathcal{E}}_{V}[0]=(\hat{x}+\hat{y}) \mathcal{E}_{0} \cos [k z-\omega t] \tag{3.25}
\end{equation*}
$$

represents diagonally polarized light, which is just linearly polarized light in a different direction. If the relative phase shift is $\pm \pi / 2$, then

$$
\begin{align*}
& \overrightarrow{\mathcal{E}}_{R}=\overrightarrow{\mathcal{E}}_{H}[0]+\overrightarrow{\mathcal{E}}_{V}[+\pi / 2]=(\hat{x} \cos [k z-\omega t]-\hat{y} \sin [k z-\omega t]) \mathcal{E}_{0}  \tag{3.26a}\\
& \overrightarrow{\mathcal{E}}_{L}=\overrightarrow{\mathcal{E}}_{H}[0]+\overrightarrow{\mathcal{E}}_{V}[-\pi / 2]=(\hat{x} \cos [k z-\omega t]+\hat{y} \sin [k z-\omega t]) \mathcal{E}_{0} \tag{3.26b}
\end{align*}
$$

represent right hand and left hand circularly polarized light, because the electric field appears to be rotate in a circle when viewed along the direction of propagation. (We employ the particle physics convention, which is opposite to the
optics convention, and call the light right-handed when the rotation is similar to that of a right-handed screw.)

The corresponding relations for photons correspond to the classical relations for light waves. A "diagonal" photon is a superposition

$$
\begin{equation*}
|D\rangle=\frac{1}{\sqrt{2}}|H\rangle+\frac{1}{\sqrt{2}}|V\rangle \tag{3.27}
\end{equation*}
$$

A right or left "circular" or natural photon is in one of the superpositions

$$
\begin{align*}
& |R\rangle=\frac{1}{\sqrt{2}}|H\rangle+\frac{i}{\sqrt{2}}|V\rangle  \tag{3.28a}\\
& |L\rangle=\frac{1}{\sqrt{2}}|H\rangle-\frac{i}{\sqrt{2}}|V\rangle \tag{3.28b}
\end{align*}
$$

where the $\pm i=e^{ \pm i \pi / 2}$ account for the relative phase shifts. Since photons are naturally circular, appropriately invert these relations and write

$$
\begin{align*}
|H\rangle & =\frac{1}{\sqrt{2}}(|R\rangle+|L\rangle)  \tag{3.29a}\\
|V\rangle & =\frac{-i}{\sqrt{2}}(|R\rangle-|L\rangle) \tag{3.29b}
\end{align*}
$$

In a measurement of the circular polarization of $|V\rangle,|R\rangle$ and $|L\rangle$ are equally likely,

$$
|V\rangle \xrightarrow{M}\left\{\begin{array}{ll}
|R\rangle, & \mathcal{P}=|-i / \sqrt{2}|^{2}=1 / 2  \tag{3.30}\\
|L\rangle, & \mathcal{P}=|+i / \sqrt{2}|^{2}=1 / 2
\end{array}\right\}
$$

but the complex numbers $\pm i$ are crucial to recovering $|R\rangle$ when superposing $|H\rangle$ and $|V\rangle$, as in Equation 3.28.

### 3.2.8 Heisenberg Indeterminacy

Measuring the linear polarization of a photon places it in a superposition of right and left circular polarizations, while measuring the circular polarization places the photon in a superposition of linear polarizations. In fact, a photon cannot have both linear and circular polarization simultaneously; knowing one type of polarization leaves the other type indeterminate, a special case of the Heisenberg indeterminacy principle.

Optically anisotropic materials with different indices of refraction in different directions can transform light from one polarization to another. A calcite crystal can convert a diagonal light beam into parallel beams of horizontal and vertical light. A quarter wave plate can convert diagonal light into circular light (by retarding one component by a distance $\lambda / 4$ ).

### 3.2.9 Crossed Polarizers

An ideal polarizer converts unpolarized light into linearly polarized light by selectively transmitting only one polarization. Consider light traveling in the $z$-direction, and linearly polarized in the $x$-direction, incident on a polarizer with transmission axis an angle $\theta$ from the $x$-direction, as in Figure 3.12. If the transmission direction is $x^{\prime}$ and the perpendicular direction is $y^{\prime}$, then we can decompose the incident electric field amplitude as the superposition

$$
\begin{equation*}
\overrightarrow{\mathcal{E}}_{0}=\hat{x}^{\prime} \mathcal{E}_{0} \cos \theta+\hat{y}^{\prime} \mathcal{E}_{0} \sin \theta \tag{3.31}
\end{equation*}
$$

Therefore, the transmitted amplitude is

$$
\begin{equation*}
\mathcal{E}_{0}^{\prime}=\mathcal{E}_{0} \cos \theta \tag{3.32}
\end{equation*}
$$

and, since intensity is proportional to the amplitude squared, the transmitted intensity is

$$
\begin{equation*}
I^{\prime}=I \cos ^{2} \theta \tag{3.33}
\end{equation*}
$$

which is Malus's Law.


Figure 3.12: A polarizer transmits the component of light parallel to its transmission axis.

Similarly, a photon polarized in the $x$-direction is a superposition of a photon polarized in the parallel and perpendicular directions,

$$
\begin{equation*}
|x\rangle=\cos \theta\left|x^{\prime}\right\rangle+\sin \theta\left|y^{\prime}\right\rangle \tag{3.34}
\end{equation*}
$$

Therefore

$$
|x\rangle \xrightarrow{M}\left\{\begin{array}{ll}
\left|x^{\prime}\right\rangle, & \mathcal{P}=|\cos \theta|^{2}=\cos ^{2} \theta  \tag{3.35}\\
\left|y^{\prime}\right\rangle, & \mathcal{P}=|\sin \theta|^{2}=\sin ^{2} \theta
\end{array}\right\}
$$

and hence the probability of transmission is $\cos ^{2} \theta$, which corresponds to Malus's law.


Figure 3.13: Single photon incident on crossed polarizers. Transmission probabilities correspond to Malus's law.

Consider next a single photon incident on crossed polarizers, as in Figure 3.13. If the probability of transmission at the first polarizer is $1 / 2$ and the probability of transmission at the second polarizer is $\cos ^{2} \theta$, then the probability of transmission through both polarizers is $(1 / 2) \cos ^{2} \theta$. If the relative angle between the two transmission axes is $\theta=\pi / 2$, then no photons get through. However, if we insert a third polarizer between the previous two with transmission axis at $\theta=\pi / 4$, one in eight photons gets through - adding an intermediate polarizer has increased the probability of transmission! These faint light, quantum experiments correspond well to the analogous bright light, classical experiments.

### 3.2.10 Entangled States (and Schrödinger's Kittens)

Pairs of quantum particles can be entangled so that a property of one, such as its polarization (spin), is linked intimately with that of the other. Such entangled or "twinned" pairs of particles are superpositions of states. Entangled states are sometimes called Schrödinger's kittens.

Consider positronium, a bound state of an electron $\mathrm{e}^{-}$and its antiparticle, the positron $\mathrm{e}^{+}$. Its ground state has zero angular momentum and odd (negative) parity. It is unstable and decays after about $10^{-10} \mathrm{~s}$ into a pair of entangled photons, as in Figure 3.14. To conserve linear momentum, the photons must have equal but opposite momenta. To conserve angular momentum, the spin of the photons must also be equal but opposite, implying identical circular polarization. To conserve parity, these two indistinguishable alternatives must superpose with a minus sign to form the entangled state

$$
\begin{equation*}
|\psi\rangle=\frac{1}{\sqrt{2}}|R\rangle|R\rangle-\frac{1}{\sqrt{2}}|L\rangle|L\rangle=\frac{1}{\sqrt{2}}(|R R\rangle-|L L\rangle), \tag{3.36}
\end{equation*}
$$

where the composite state $|a b\rangle=|a\rangle|b\rangle=|a\rangle \otimes|b\rangle$ is a bilinear "tensor" product. (Parity refers to the behavior of a system under coordinate inversion. If $P$ is the
parity operator, then $P|R R\rangle=|L L\rangle$ and $P|L L\rangle=|R R\rangle$, and so $P|\psi\rangle=-|\psi\rangle$.)
RHC

$-\bigcap_{v} \rightarrow$ RHC
LHC




Figure 3.14: Positronium (top) annihilates into a pair of right circular photons (middle) or a pair of left circular photons (bottom). Both possibilities superpose to form an entangled state. The photon emission is isotropic.

Although photons correspond to circular rather than linear light, they can be analyzed into completely anticorrelated plane polarizations. Using Equation 3.28 to express circular polarizations as superpositions of linear polarizations, the entangled state becomes

$$
\begin{equation*}
|\psi\rangle=\frac{i}{\sqrt{2}}|H\rangle|V\rangle+\frac{i}{\sqrt{2}}|V\rangle|H\rangle=\frac{i}{\sqrt{2}}(|H V\rangle+|V H\rangle) . \tag{3.37}
\end{equation*}
$$

Any linear polarization measurement induces a nonlocal collapse of the superposition

$$
|\psi\rangle \xrightarrow{M}\left\{\begin{array}{ll}
|H V\rangle, & \mathcal{P}=|i / \sqrt{2}|^{2}=1 / 2  \tag{3.38}\\
|V H\rangle, & \mathcal{P}=|i / \sqrt{2}|^{2}=1 / 2
\end{array}\right\}
$$

at least in the CI. After the measurement, one photon is horizontally polarized and the other is vertically polarized.

### 3.2.11 EPR-Bohm Experiment

Consider an experiment first proposed in the 1930s by Albert Einstein, Boris Podolsky, and Nathan Rosen (EPR) and modernized in the 1950s by David Bohm. Suppose two observers, Alice and Bob, intercept entangled photons with linear polarizers at a relative angle of $\theta$, as in Figure 3.15. For each photon pair, the two polarization measurements can be separated by a spacelike interval, so far apart that not even light can join them. Each measurement can be reduced to a binary digit, 1 or 0 , indicating a photon transmitted or not. Given Alice and Bob's binary data for many measurements, we can calculate the correlation function

$$
\begin{equation*}
C[\theta]=\frac{\# \text { matches }}{\# \text { trials }}=\mathcal{P}[\text { match }] \tag{3.39}
\end{equation*}
$$

For $\theta=0$, Alice and Bob's data are sequences of random digits, with 0 and 1 equally likely, but perfectly anticorrelated, in agreement with Equation 3.38, so that $C[0]=0$. For $\theta=\pi / 2$, Alice and Bob's data are other sequences of random digits, but now perfectly correlated, again in agreement with Equation


Figure 3.15: Polarization cross correlations of entangled photon pairs.
3.38 , so that $C[\pi / 2]=1$. To calculate the quantum prediction for the correlation function at an arbitrary angle, we begin with basic probability theory. If the conventional symbols $\wedge, \vee, \mid$, denote "and", "or", "given", then

$$
\begin{align*}
\mathcal{P}[\text { match }] & =\mathcal{P}[(A=0 \wedge B=0) \vee(A=1 \wedge B=1)] \\
& =\mathcal{P}[A=0] \mathcal{P}[B=0 \mid A=0]+\mathcal{P}[A=1] \mathcal{P}[B=1 \mid A=1] \\
& =\mathcal{P}[A=0](1-\mathcal{P}[B=1 \mid A=0])+\mathcal{P}[A=1] \mathcal{P}[B=1 \mid A=1] \tag{3.40}
\end{align*}
$$

Hence by the Malus's law results of Section 3.2.9,

$$
\begin{equation*}
C[\theta]=\mathcal{P}[\text { match }]=\underbrace{\frac{1}{2}}_{\text {no }} \underbrace{\left(1-\cos ^{2} \theta\right)}_{\text {no }}+\underbrace{\frac{1}{2}}_{\text {yes }} \underbrace{\cos ^{2}\left[\frac{\pi}{2}-\theta\right]}_{\text {yes }}=\sin ^{2} \theta \tag{3.41}
\end{equation*}
$$

which agrees with the extreme cases $C[0]=0$ and $C[\pi / 2]=1$. For small angles, $\sin \theta \sim \theta \ll 1$ and $C[\theta] \sim \theta^{2}$, so $C[2 \theta] \sim 4 \theta^{2}>2 \theta^{2}=2 C[\theta]$, or

$$
\begin{equation*}
C[2 \theta]>2 C[\theta] \tag{3.42}
\end{equation*}
$$

### 3.2.12 Bell's Inequality

In 1964, John Bell demonstrated [1] that any classical (local realistic) explanation for an EPR-Bohm-type experiment must produce weaker correlations (as the angle $\theta$ increases), as we now show. If Alice and Bob's polarizer are aligned, so that their relative angle $\theta=0$, then their binary data are completely anticorrelated, $C[0]=0$. If Bob now rotates his polarizer through an angle $\theta>0$, the misalignment introduces some matches into his data (by, say, flipping a 1 to a 0 ), so $C[\theta]>0$. If Alice next rotates her polarizer through the same angle,
the realignment removes the matches from her data (by flipping a 1 to a 0 ), so again $C[0]=0$. If Bob next rotates his polarizer through an additional angle $\theta>0$, the second misalignment once more introduces some matches into his data, so again $C[\theta]>0$. However, if Alice had not rotated her polarizer, the successive misalignments of Bob's polarizer might have cancelled some matches (by flipping a 0 to a 1 and then back to a 0 again). Hence

$$
\begin{equation*}
C[2 \theta] \leq 2 C[\theta] \tag{3.43}
\end{equation*}
$$

which is Bell's inequality.
The quantum prediction of Equation 3.42 contradicts the classical prediction of Equation 3.43, and we must put nature to the test. By the 1980s, in a culmination of a series of increasingly better experiments by many research groups, Alain Aspect and colleagues convincingly demonstrated that Bell's inequality is decisively violated in these kind of experiments. Consequently, something must be wrong with Bell's argument, as Bell himself anticipated.

The argument seems to rest on two assumptions: locality and reality. Locality means no superluminal connections, so what happens here and now doesn't depend on what happens then and there. For example, we implicitly assume locality when we reason that, when Bob rotates his polarizer, he alters his data but not Alice's, and vice versa. Reality means counterfactual definiteness, the ability to consistently discuss what might have happened but did not. For example, we reason that if Bob had rotated his polarizer through $\theta$, then he would have introduced some matches, and if he had then rotated through an additional $\theta$, then some of the matches might have cancelled. One of these two classically reasonable assumptions must be wrong.

A popular nonlocal interpretation of the EPR-Bohm experiment is that forcing a two-particle interpretation on an entangled particle pair is impossible. While this may violate the spirit of special relativity, it does not violate the letter of special relativity. In the CI, quantum randomness prevents using entangled states for superluminal telegraphs, because any message introduced by rotating one of the polarizers is found only in the correlations between possibly remote and spacelike experiments. In the MWI, measurements don't collapse superpositions, nonlocally or otherwise, and locality is restored.

### 3.3 Feynman to Schrödinger [Optional]

The Schrödinger differential equation governs the continuous evolution of quantum states. We derive the Schrödinger equation from a remarkable postulate and sum-over-paths integral by Feynman.

### 3.3.1 Feynman Postulate

A photon surely doesn't understand partial differential equations, like the wave equation. An electron doesn't understand the Schrödinger equation. How then do they get from point to point?


Figure 3.16: Classical light interferes when passing through a double slit interferometer (left). A photon carries an imaginary arrow that rotates at the frequency of the corresponding classical light (right). Adding the arrows at the bottom for both paths $(a \rightarrow l \rightarrow b$ and $a \rightarrow r \rightarrow b)$ and squaring gives the relative probability of detecting the photon.

Bright light passing through a double slit interferes constructively and destructively to form a pattern of light and dark stripes on a projection screen, as in Figure 3.16. Faint light builds up the same pattern, photon by photon. The double slit is a kind of interferometer, not unlike the Mach-Zehnder interferometer, and we can model it in the same way. The left and right slits, $l$ and $r$, provide two paths through the device. Imagine that, along each path (through each slit), a photon carries an arrow that rotates at the frequency of the corresponding classical light. At the projection screen, vector addition of the arrows for both possibilities, followed by the squaring of the resulting length, gives the relative probability of finding a photon there. Representing the rotating arrows by complex numbers of unit modulus whose arguments record rotations, we can write the probability

$$
\begin{equation*}
\mathcal{P}[a \rightarrow b]=|\mathcal{E}[a \rightarrow b]|^{2} \tag{3.44}
\end{equation*}
$$

where the (normalized) electric field amplitude

$$
\begin{equation*}
\mathcal{E}[a \rightarrow b]=e^{i \varphi[l]}+e^{i \varphi[r]} \tag{3.45}
\end{equation*}
$$

where the phases increase at the rate

$$
\begin{equation*}
\frac{d \varphi}{d t}=\omega=\frac{E}{\hbar} \tag{3.46}
\end{equation*}
$$

where the photon energy $E=h \nu=\hbar \omega$.

In another manifestation of wave-particle duality, if we replace the (massless) photons in the double slit experiment with (massive) electrons, a similar interference pattern develops. We can analyze single electron interference using the rotating arrows of the photon experiment, but at what rate do they rotate? To match experiment, Feynman assumed the probability

$$
\begin{equation*}
\mathcal{P}[a \rightarrow b]=|A[a \rightarrow b]|^{2} \tag{3.47}
\end{equation*}
$$

where the probability amplitude

$$
\begin{equation*}
A[a \rightarrow b]=e^{i \varphi[l]}+e^{i \varphi[r]} \tag{3.48}
\end{equation*}
$$

where the phases increase at the rate

$$
\begin{equation*}
\frac{d \varphi}{d t}=\omega=\frac{L}{\hbar} \tag{3.49}
\end{equation*}
$$

where the classical Lagrangian is difference between the kinetic and potential energies, $L=T-V$. (For a free particle, the potential energy $V=0$, and $\omega=T / \hbar$.) Thus, the phase accumulated by the electron as it travels along a path,

$$
\begin{equation*}
\varphi=\int \omega d t=\int \frac{L}{\hbar} d t=\frac{S}{\hbar} \tag{3.50}
\end{equation*}
$$

is the classical action for that path in units of the quantum of action, $\hbar=h / 2 \pi$. Hence, the electron double slit probability amplitude can be written

$$
\begin{equation*}
A[a \rightarrow b]=e^{i S[l] / \hbar}+e^{i S[r] / \hbar} \tag{3.51}
\end{equation*}
$$

### 3.3.2 Path Integral

If a slit has two holes, then the probability amplitude for an electron to get from point $a$ on one side to point $b$ on the other is

$$
\begin{equation*}
A=A_{1}+A_{2} \tag{3.52}
\end{equation*}
$$

where $A_{1}$ and $A_{2}$ are, respectively, the amplitudes to travel via hole 1 and hole 2. If we drill many more holes in the screen, as in Figure 3.17, the amplitude is the sum

$$
\begin{equation*}
A=\sum_{i} A_{i} \tag{3.53}
\end{equation*}
$$

If we then add many more screens with holes, the amplitude is the double sum

$$
\begin{equation*}
A=\sum_{i, j} A_{i, j} \tag{3.54}
\end{equation*}
$$

If we increase the number of holes and screens until infinitely many screens have infinitely many holes, so the screens aren't there any more, the amplitude is

$$
\begin{equation*}
A=\sum_{\text {all paths }} A_{\text {path }} \tag{3.55}
\end{equation*}
$$



Figure 3.17: Infinitely many screens with infinitely many holes implies infinitely many paths.
where

$$
\begin{equation*}
A_{\mathrm{path}}=e^{i S_{\mathrm{path}} / \hbar} \tag{3.56}
\end{equation*}
$$

More formally, the amplitude to go from point $a$ to point $b$ (in time $t$ ) is called the quantum propagator, which we will write as

$$
\begin{equation*}
A[a \rightarrow b]=\int_{a}^{b} \mathcal{D} x[t] e^{i S[x[t]] / \hbar} \tag{3.57}
\end{equation*}
$$

where the notation $\mathcal{D} x[t]$ reminds us that this is a sum over all paths, a Feynman path integral. (Just as real numbers are more numerous than countable numbers, paths in a plane are more numerous than real numbers. In this sense, path integrals are to real integrals what real integrals are to countable sums.)

### 3.3.3 Recovering Classical Mechanics

Since the action $S=\hbar \varphi$ is an extremum for the classical path $x_{c}[t]$, it is stationary with respect to small variations from this path. Thus, the rotating arrows for paths nearby the classical path will add constructively. Conversely, for every distant path, we can find another distant path, by (say) adding an extra wiggle to the path, with which it will interfere destructively. Hence, the main contribution to the Feynman path integral is from paths near the classical path.

In fact, for a free (nonrelativistic) particle of a given speed $v$,

$$
\begin{equation*}
\omega=\frac{L}{\hbar}=\frac{T}{\hbar}=\frac{(1 / 2) m v^{2}}{\hbar} \propto m \tag{3.58}
\end{equation*}
$$

Hence, the greater the particle's mass, the faster its arrow rotates, and the closer it must be to the classical path to constructively interfere and make a significant contribution to the total amplitude. For a classical object like a billiard ball, the classical, extremal path is overwhelmingly preferred. In this way, Feynman's postulate implies the principle of extremal action in classical mechanics.

### 3.3.4 Schrödinger Equation

Feynman's sum-over-paths are not easy to do - or even to rigorously define. However, like Feynman, we can sidestep this difficulty [5]. First, write the probability amplitude to go from point $a$ to point $b$ as

$$
\begin{equation*}
A[a \rightarrow b]=\int_{a}^{b} \mathcal{D} x[t] e^{i S[a \rightarrow b] / \hbar} \tag{3.59}
\end{equation*}
$$

Next, choose a nonrelativistic Lagrangian, so the action over a path $x[t]$ is

$$
\begin{equation*}
S[a \rightarrow b]=\int_{t_{a}}^{t_{b}} d t\left(\frac{1}{2} m\left(\frac{d x}{d t}\right)^{2}-V[x]\right) \tag{3.60}
\end{equation*}
$$

Now introduce an intermediate point $c$. Since the action $S$ is the time-integral of the Lagrangian $L$, the action is additive

$$
\begin{equation*}
S[a \rightarrow b]=S[a \rightarrow c]+S[c \rightarrow b] \tag{3.61}
\end{equation*}
$$

and so

$$
\begin{equation*}
A[a \rightarrow b]=\int_{a}^{b} \mathcal{D} x[t] e^{i S[a \rightarrow c] / \hbar} e^{i S[c \rightarrow b] / \hbar} \tag{3.62}
\end{equation*}
$$

This is equivalent to the amplitude to go from point $a$ to point $c$ times the amplitude to go from point $c$ to point $b$ summed over all possible intermediate positions $x_{c}$, which can be expressed as the conventional integral

$$
\begin{equation*}
A[a \rightarrow b]=\int_{-\infty}^{\infty} d x_{c} A[a \rightarrow c] A[c \rightarrow b] \tag{3.63}
\end{equation*}
$$

Define the wave function $\Psi[f]=A[a \rightarrow f]$ as the amplitude to be at point $f$. Then

$$
\begin{equation*}
\Psi[b]=\int_{-\infty}^{\infty} d x_{c} \Psi[c] A[c \rightarrow b] \tag{3.64}
\end{equation*}
$$

or more explicitly

$$
\begin{equation*}
\Psi\left[x_{b}, t_{b}\right]=\int_{-\infty}^{\infty} d x_{c} \Psi\left[x_{c}, t_{c}\right] A\left[x_{c}, t_{c} ; x_{b}, t_{b}\right] \tag{3.65}
\end{equation*}
$$

If the points $b$ and $c$ are separated by an infinitesimal interval, this integral equation reduces to a famous differential equation. Take $\left\{x_{b}, t_{b}\right\}=\{x, t+\tau\}$ and $\left\{x_{c}, t_{c}\right\}=\{y, t\}$ and let $\tau \downarrow 0$. Then the future wave function

$$
\begin{equation*}
\Psi[x, t+\tau]=\int_{-\infty}^{\infty} d y \Psi[y, t] A[y, t ; x, t+\tau] \tag{3.66}
\end{equation*}
$$

Since we need only consider a single path between points whose separation is vanishing, the transition amplitude

$$
\begin{equation*}
A[y, t ; x, t+\tau]=N \exp \left[\frac{i}{\hbar} \int_{t}^{t+\tau} d t\left(\frac{1}{2} m\left(\frac{x-y}{\tau}\right)^{2}-V\left[\frac{x+y}{2}\right]\right)\right] \tag{3.67}
\end{equation*}
$$

where $N$ is a normalization constant to be defined below. Since $\int_{t}^{t+\tau} d t=\tau$, the future wave function

$$
\begin{equation*}
\Psi[x, t+\tau]=N \int_{-\infty}^{\infty} d y \Psi[y, t] \exp \left[\frac{i m}{2 \hbar \tau}(x-y)^{2}\right] \exp \left[-\frac{i \tau}{\hbar} V\left[\frac{x+y}{2}\right]\right] \tag{3.68}
\end{equation*}
$$

If we define $\alpha=-i m / 2 \hbar \tau$, then by Euler's theorem, the first exponential

$$
\begin{equation*}
\exp \left[-i|\alpha|(x-y)^{2}\right]=\cos \left[|\alpha|(x-y)^{2}\right]-i \sin \left[|\alpha|(x-y)^{2}\right] \tag{3.69}
\end{equation*}
$$

oscillates rapidly when the integration variable $y$ is far from the independent variable $x$. Since the positive and negative oscillations tend to cancel when integrating, appreciable contributions to the integral occur only when $y$ is near $x$. To exploit this fact, we introduce the change of variable $y=x+\xi$, so that $d y=d \xi$ and

$$
\begin{equation*}
\Psi[x, t+\tau]=N \int_{-\infty}^{\infty} d \xi \Psi[x+\xi, t] \exp \left[\frac{i m}{2 \hbar \tau} \xi^{2}\right] \exp \left[-\frac{i \tau}{\hbar} V\left[x+\frac{\xi}{2}\right]\right] \tag{3.70}
\end{equation*}
$$

If $y$ is near $x$, then $\xi$ is small. More specifically, if we limit the argument of the first exponential to one radian, so that $m \xi^{2} / 2 \hbar \tau \lesssim 1$, then $\xi^{2} \lesssim 2 \hbar \tau^{1} / m$ vanishes as $\tau \downarrow 0$. Hence, in powers of the small quantities, we expand to $O\left[\xi^{2} \tau^{1}\right]$ to get

$$
\begin{align*}
\Psi[x, t]+\tau \partial_{t} \Psi[x, t] & \sim \\
N \int_{-\infty}^{\infty} d \xi\left(\Psi[x, t]+\xi \partial_{x} \Psi[x, t]+\frac{\xi^{2}}{2} \partial_{x}^{2} \Psi[x, t]\right) & \exp \left[\frac{i m}{2 \hbar \tau} \xi^{2}\right]\left(1-\frac{i \tau}{\hbar} V[x]\right) \tag{3.71}
\end{align*}
$$

(We can contract the limits of integration to $\pm \sqrt{2 \hbar \tau / m}= \pm 1 / \sqrt{|\alpha|}$ when expanding in powers of $\xi$ and reset them with impunity afterward thanks to the rapid cancelling oscillations.) Since all the wave functions $\Psi$ are now evaluated at $\{x, t\}$, we can drop these arguments. If we define the Gaussian integrals

$$
\begin{equation*}
G_{n}=\int_{-\infty}^{\infty} d \xi \xi^{n} e^{-\alpha \xi^{2}}=\int_{-\infty}^{\infty} \xi^{n} e^{-\alpha \xi^{2}} d \xi \tag{3.72}
\end{equation*}
$$

with an implicit convergence factor $\delta$, such that

$$
\begin{equation*}
e^{-\alpha \xi^{2}}=e^{-i|\alpha| \xi^{2}} \leftrightarrow e^{-(i+\delta)|\alpha| \xi^{2}}=e^{-i|\alpha| \xi^{2}} e^{-\delta \xi^{2}} \tag{3.73}
\end{equation*}
$$

where $\delta \downarrow 0$ after the integration, then we can expand further to get

$$
\begin{equation*}
\Psi+\tau \partial_{t} \Psi \sim N G_{0} \Psi+N G_{1} \partial_{x} \Psi+\frac{1}{2} N G_{2} \partial_{x}^{2} \Psi-\frac{i \tau}{\hbar} V N G_{0} \Psi \tag{3.74}
\end{equation*}
$$

plus vanishing terms. Comparing both sides of this equation, we must have $N G_{0}=1$, so the normalization constant $N=1 / G_{0}$. Since $G_{0}=\sqrt{\pi / \alpha}, G_{1}=0$, and $G_{2}=G_{0} / 2 \alpha$, we get

$$
\begin{equation*}
\tau \partial_{t} \Psi \sim \frac{1}{2}\left(\frac{\hbar \tau}{-i m}\right) \partial_{x}^{2} \Psi-\frac{i \tau}{\hbar} V \Psi \tag{3.75}
\end{equation*}
$$

or, multiplying both sides by $i \hbar / \tau$ and taking the limit $\tau \downarrow 0$,

$$
\begin{equation*}
i \hbar \partial_{t} \Psi=-\frac{\hbar^{2}}{2 m} \partial_{x}^{2} \Psi+V \Psi \tag{3.76}
\end{equation*}
$$

which is the Schrödinger equation [7].
In $3+1$ dimensions, the Schrödinger equation readily generalizes to

$$
\begin{equation*}
i \hbar \partial_{t} \Psi[\vec{r}, t]=-\frac{\hbar^{2}}{2 m} \nabla^{2} \Psi[\vec{r}, t]+V[\vec{r}] \Psi[\vec{r}, t] \tag{3.77}
\end{equation*}
$$

where the Laplacian $\nabla^{2}=\partial_{x}^{2}+\partial_{y}^{2}+\partial_{z}^{2}$. However, we will focus mainly on $1+1$ dimensions.

### 3.4 Simple Schrödinger Solutions

We explore the nature and implications of the Schrödinger equation and its simple solutions.

### 3.4.1 Analogies

Roughly speaking, Newton's second law is to classical mechanics as Schrödinger's equation is to quantum mechanics. In $1+1$ dimensions, classically, the acceleration of a particle is proportional to a force $F_{x}$ and inversely proportional to its mass $m, a_{x}=F_{x} / m$. The force is often derived from the negative gradient of a potential energy function, $F_{x}=-V^{\prime}[x]$. The result is the worldline $x[t]$. This familiar algorithm is contrasted with the Schrödinger equation in Table 3.1

Contrasting Maxwell's electromagnetic waves with Schrödinger's matter waves is also instructive. Much is sometimes made of the fact that the Schrödinger equation is explicitly complex. However, the complex Schrödinger equation can be written as two coupled real equations. Furthermore, the real Maxwell equations, for the electric and magnetic fields in an electromagnetic wave, can be combined into a single complex equation, as in demonstrated in Table 3.2.

Table 3.1: Contrasting Newton with Schrödinger.

| Newtonian Mechanics | Quantum Mechanics |
| :--- | :--- |
| $\partial_{t}^{2} x=-\frac{1}{m} \partial_{x} V$ | $i \hbar \partial_{t} \Psi=-\frac{\hbar^{2}}{2 m} \partial_{x}^{2} \Psi+V \Psi$ |
| $x[0]=x_{0}$ | $\Psi[x, 0]=\psi_{0}[x]$ |
| $\left(\partial_{t} x\right)[0]=v_{0}$ | $\int d x\|\Psi[x, 0]\|^{2}=1$ |
| $x[t]$ | $\Psi[x, t]$ |

Table 3.2: Contrasting Maxwell with Schrödinger.

| Electromagnetic Waves | Matter Waves |
| :--- | :--- |
| $\overrightarrow{\mathcal{F}}=\overrightarrow{\mathcal{E}}+i c \overrightarrow{\mathcal{B}} \in \mathbb{C}^{3}$ | $\Psi=\Psi_{R}+i \Psi_{I} \in \mathbb{C}$ |
| source free | potential free |
| $\partial_{x}^{2} \overrightarrow{\mathcal{F}}=\frac{1}{c^{2}} \partial_{t}^{2} \overrightarrow{\mathcal{F}}$ | $i \hbar \partial_{t} \Psi=-\frac{\hbar^{2}}{2 m} \partial_{x}^{2} \Psi$ |
| real \& uncoupled | real \& coupled |
| $\partial_{x}^{2} \overrightarrow{\mathcal{E}}=\frac{1}{c^{2}} \partial_{t}^{2} \overrightarrow{\mathcal{E}}$ | $-\hbar \partial_{t} \Psi_{I}=-\frac{\hbar^{2}}{2 m} \partial_{x}^{2} \Psi_{R}$ |
| $\partial_{x}^{2} \overrightarrow{\mathcal{B}}=\frac{1}{c^{2}} \partial_{t}^{2} \overrightarrow{\mathcal{B}}$ | $+\hbar \partial_{t} \Psi_{R}=-\frac{\hbar^{2}}{2 m} \partial_{x}^{2} \Psi_{I}$ |
| energy density | probability density |
| $\|\mathcal{F}\|^{2}=\mathcal{E}^{2}+c^{2} \mathcal{B}^{2}$ | $\|\Psi\|^{2}=\Psi_{R}{ }^{2}+\Psi_{I}{ }^{2}$ |
| dispersion-less solutions | dispersion-full solutions |
| $\overrightarrow{\mathcal{F}}[x, t]=(\hat{y}+i \hat{z}) A \sin [k x-\omega t]$ | $\Psi[x, t]=A \exp [i(k x-\omega t)]$ |
| $\omega=k c$ | $\hbar \omega=\frac{(\hbar k)^{2}}{2 m}$ |

### 3.4.2 Schrödinger Wave Equation Simply

Schrödinger captured wave-particle duality in a probability wave equation. In $1+1$ dimensions $\{x, t\}$, recall that the energy and momentum of photons are proportional to their temporal and spatial frequencies,

$$
\begin{align*}
E & =\hbar \omega  \tag{3.78a}\\
p & =\hbar k \tag{3.78b}
\end{align*}
$$

and assume these also hold for matter waves. Further assume a complex sinusoidal wave

$$
\begin{equation*}
\Psi[x, t]=N e^{i(k x-\omega t)} \tag{3.79}
\end{equation*}
$$

where $e^{i \theta}=\cos \theta+i \sin \theta$ and $i=\sqrt{-1}$. The rates of change of the wavefunction with time and space are

$$
\begin{align*}
\partial_{t} \Psi & =-i \omega \Psi  \tag{3.80a}\\
\partial_{x} \Psi & =+i k \Psi \tag{3.80b}
\end{align*}
$$

or by Eq. 3.78,

$$
\begin{align*}
+i \hbar \partial_{t} \Psi & =E \Psi  \tag{3.81a}\\
-i \hbar \partial_{x} \Psi & =p \Psi \tag{3.81b}
\end{align*}
$$

so the space and time derivatives act like multiplication by energy and momentum.

If a particle of mass $m$ moves in a potential energy $V[x]$ at speed $v \ll c$, then its energy

$$
\begin{equation*}
E=\frac{p^{2}}{2 m}+V \tag{3.82}
\end{equation*}
$$

Multiply by the wavefunction on the right to get

$$
\begin{equation*}
E \Psi=\frac{p^{2}}{2 m} \Psi+V \Psi \tag{3.83}
\end{equation*}
$$

or using Eq. 3.81,

$$
\begin{equation*}
+i \hbar \partial_{t} \Psi=\frac{\left(-i \hbar \partial_{x}\right)^{2}}{2 m} \Psi+V \Psi \tag{3.84}
\end{equation*}
$$

which expands to

$$
\begin{equation*}
i \hbar \partial_{t} \Psi=-\frac{\hbar^{2}}{2 m} \partial_{x}^{2} \Psi+V \Psi \tag{3.85}
\end{equation*}
$$

In Leibniz notation

$$
\begin{equation*}
i \hbar \frac{\partial \Psi}{\partial t}=-\frac{\hbar^{2}}{2 m} \frac{\partial^{2} \Psi}{\partial x^{2}}+V[x] \Psi \tag{3.86}
\end{equation*}
$$

and in $3+1$ dimensions $\{x, y, z, t\}$

$$
\begin{equation*}
i \hbar \partial_{t} \Psi=-\frac{\hbar^{2}}{2 m}\left(\partial_{x}^{2} \Psi+\partial_{y}^{2} \Psi+\partial_{z}^{2} \Psi\right)+V[x, y, z] \Psi \tag{3.87}
\end{equation*}
$$

### 3.4.3 Probability Conservation

Technically, the wave function $\Psi[x, t]$ is the probability density amplitude for a particle to be at a position $x$ at a time $t$, and the absolute square of the wave function $|\Psi|^{2}$ is the corresponding probability density. Thus, in $1+1$ dimensions, $|\Psi|^{2}$ is a probability per unit length, which means that the probability of finding the particle in the interval $d x$ about $x$ at time $t$ is

$$
\begin{equation*}
\mathrm{d} \mathcal{P}=d x|\Psi|^{2} \tag{3.88}
\end{equation*}
$$

and the probability of finding the particle between $x_{1}$ and $x_{2}$ at time $t$ is

$$
\begin{equation*}
\mathcal{P}\left[x_{1}<x<x_{2}\right]=\int_{x_{1}}^{x_{2}} d x|\Psi|^{2} \tag{3.89}
\end{equation*}
$$

provided that we normalize $\Psi[x, t]$ by requiring

$$
\begin{equation*}
1=\int_{-\infty}^{\infty} d x \Psi^{*} \Psi=\int_{-\infty}^{\infty} d x|\Psi|^{2} \tag{3.90}
\end{equation*}
$$

bcause the particle must certainly be found somewhere!
The Schrödinger equation conserves probability, a property called unitarity. For example, if

$$
\begin{equation*}
\mathcal{I}[t]=\int_{-\infty}^{\infty} d x \Psi[x, t]^{*} \Psi[x, t] \tag{3.91}
\end{equation*}
$$

then the time derivative

$$
\begin{equation*}
\dot{\mathcal{I}}=\int_{-\infty}^{\infty} d x\left(\Psi \partial_{t} \Psi^{*}+\Psi^{*} \partial_{t} \Psi\right) \tag{3.92}
\end{equation*}
$$

But from the Schrödinger equation

$$
\begin{equation*}
\partial_{t} \Psi=-\frac{\hbar}{2 m i} \partial_{x}^{2} \Psi+\frac{V}{i \hbar} \Psi \tag{3.93}
\end{equation*}
$$

and by its complex conjugate

$$
\begin{equation*}
\partial_{t} \Psi^{*}=+\frac{\hbar}{2 m i} \partial_{x}^{2} \Psi^{*}-\frac{V}{i \hbar} \Psi^{*} \tag{3.94}
\end{equation*}
$$

Hence, by substitution,

$$
\begin{equation*}
\dot{\mathcal{I}}=\frac{\hbar}{2 m i} \int_{-\infty}^{\infty} d x\left(\Psi \partial_{x}^{2} \Psi^{*}-\Psi^{*} \partial_{x}^{2} \Psi\right) \tag{3.95}
\end{equation*}
$$

Adding and subtracting $\left(\partial_{x} \Psi\right)\left(\partial_{x} \Psi^{*}\right)$ in the integrand yields

$$
\begin{equation*}
\dot{\mathcal{I}}=\frac{\hbar}{2 m i} \int_{-\infty}^{\infty} d x \partial_{x}\left(\Psi \partial_{x} \Psi^{*}-\Psi^{*} \partial_{x} \Psi\right) \tag{3.96}
\end{equation*}
$$

Since integration and differentiation are inverse operations,

$$
\begin{equation*}
\dot{\mathcal{I}}=\left.\frac{\hbar}{2 m i}\left(\Psi \partial_{x} \Psi^{*}-\Psi^{*} \partial_{x} \Psi\right)\right|_{-\infty} ^{\infty}=0 \tag{3.97}
\end{equation*}
$$

because if $\Psi$ did not vanish at infinity, it could not be square normalized. Hence, if $\Psi[x, 0]$ is normalized so that $\mathcal{I}[0]=1$ initially, then $\mathcal{I}[t]=1$ always.

### 3.4.4 Schrödinger Wave Packets

Because the Schrödinger equation is linear in $\Psi$, any superposition of plane wave solutions is also a solution. Hence, we can form the physical (and normalizable) wave packet

$$
\begin{equation*}
\Psi[x, t]=\int_{-\infty}^{\infty} d p \varphi[p] e^{i(p x-E[p] t) / \hbar} \tag{3.98}
\end{equation*}
$$

where the plane wave coefficients $\varphi[p]$ are essentially the Fourier transform, or momentum space representation of the initial wave packet

$$
\begin{equation*}
\Psi[x, 0]=\int_{-\infty}^{\infty} d p \varphi[p] e^{i p x / \hbar} \tag{3.99}
\end{equation*}
$$



Figure 3.18: A wave packet that is static in momentum space (left) but evolving in position space (right).

If $\varphi[p]$ is a Gaussian function of width $\Delta p$ centered on $p_{0}$, then $\Psi[x, t]$ is another Gaussian with width $\Delta x$ centered on $p_{0} t / m$, as in Figure 3.18. By Fourier's theorem, large spatial frequencies $k=p / \hbar$ are required to synthesize a spatially narrow peak, and hence the widths of the wave packet in momentum and position space are inversely related, $\Delta p \Delta x>\hbar / 2$, which in quantum physics is known historically as the Heisenberg uncertainty principle but in reality is another example of quantum indeterminacy. Recall that knowing the circular polarization of a photon renders its linear polarization indeterminate, and vice versa. Similarly, knowing the position $(\Delta x=0)$ of a particle renders its momentum $(\Delta p=\infty)$ indeterminant, and vice versa. Circular and linear polarization are incompatible observables, as are position and momentum.

### 3.4.5 Separation of Variables

Notice that the plane wave solutions can be written as a product of a function of $x$ times a function of $t, e^{i(k x-\omega t)}=e^{i k x} e^{-i \omega t}$. We seek the most general such separable solution,

$$
\begin{equation*}
\Psi[x, t]=X[x] T[t] \tag{3.100}
\end{equation*}
$$

Substituting into the Schrödinger equation, partial derivatives simplify to total derivatives, $\partial_{t} \Psi=X \dot{T}$ and $\partial_{x}^{2} \Psi=X^{\prime \prime} T$, so that

$$
\begin{equation*}
i \hbar X \dot{T}=-\frac{\hbar^{2}}{2 m} X^{\prime \prime} T+V X T \tag{3.101}
\end{equation*}
$$

Dividing by the product $X T$ yields

$$
\begin{equation*}
i \hbar \frac{\dot{T}}{T}=-\frac{\hbar^{2}}{2 m} \frac{X^{\prime \prime}}{X}+V[x] \tag{3.102}
\end{equation*}
$$

the left side of which is a function of $t$ alone, and the right side is a function of $x$ alone. This is only possible if both sides are constant, else by varying $t$, for example, we could change the left side without changing the right side, thereby breaking the equality. The common separation constant has the dimensions of energy, and we denote it by $E$. In summary,

$$
\begin{equation*}
f[t]=g[x]=\text { constant }=E \tag{3.103}
\end{equation*}
$$

Thus, the original partial differential equation separates into two ordinary differential equations. The $t$-equation

$$
\begin{equation*}
i \hbar \frac{\dot{T}}{T}=E \tag{3.104}
\end{equation*}
$$

is exactly integrable

$$
\begin{equation*}
\int_{T[0]}^{T[t]} i \hbar \frac{d T}{T}=\int_{0}^{t} E d t \tag{3.105}
\end{equation*}
$$

and yields

$$
\begin{equation*}
T[t]=T[0] e^{-i E t / \hbar} \tag{3.106}
\end{equation*}
$$

However, the $x$-equation

$$
\begin{equation*}
-\frac{\hbar^{2}}{2 m} \frac{X^{\prime \prime}}{X}+V[x]=E \tag{3.107}
\end{equation*}
$$

or

$$
\begin{equation*}
-\frac{\hbar^{2}}{2 m} X^{\prime \prime}+V[x] X=E X \tag{3.108}
\end{equation*}
$$

involves the generic potential energy function $V[x]$, which must be specified before the equation can be solved.

### 3.4.6 Hamiltonian Eigenfunctions

Sometimes Equation 3.76 is called the time-dependent Schrödinger equation and Equation 3.107 is called the time-independent Schrödinger equation. The relationship between the two equations can be elucidated by defining the following differential operators. The Hamiltonian is

$$
\begin{equation*}
H_{\mathrm{op}}=\frac{p_{\mathrm{op}}^{2}}{2 m}+V \tag{3.109}
\end{equation*}
$$

where the momentum operator is

$$
\begin{equation*}
p_{\mathrm{op}}=-i \hbar \partial_{x}=-i \hbar \frac{\partial}{\partial x} \tag{3.110}
\end{equation*}
$$

The energy operator is

$$
\begin{equation*}
E_{\mathrm{op}}=+i \hbar \partial_{t}=+i \hbar \frac{\partial}{\partial t} \tag{3.111}
\end{equation*}
$$

In $3+1$ dimensions, where

$$
\vec{p}_{\mathrm{op}}=\left(\begin{array}{c}
E_{\mathrm{op}}  \tag{3.112}\\
p_{x, \mathrm{op}} \\
p_{y, \mathrm{op}} \\
p_{z, \mathrm{op}}
\end{array}\right)=\left(\begin{array}{c}
+i \hbar \partial_{t} \\
-i \hbar \partial_{x} \\
-i \hbar \partial_{y} \\
-i \hbar \partial_{z}
\end{array}\right)=i \hbar\left(\begin{array}{c}
+\partial_{t} \\
-\partial_{x} \\
-\partial_{y} \\
-\partial_{z}
\end{array}\right)
$$

is a hint of relativity. However, unlike the classical wave equation, for example, the Schrödinger equation is manifestly nonrelativistic, as it treats space and time asymmetrically, involving as it does a first derivative in time $\partial_{t}$ but a second derivative in space $\partial_{x}^{2}$.

If we introduce the more common notation $\psi_{E}[x]=X[x]$ and employ the differential operators, then the time-dependent Schrödinger becomes

$$
\begin{equation*}
H_{\mathrm{op}} \Psi=E_{\mathrm{op}} \Psi \tag{3.113}
\end{equation*}
$$

and the time-independent equation becomes

$$
\begin{equation*}
H_{\mathrm{op}} \psi_{E}=E \psi_{E} \tag{3.114}
\end{equation*}
$$

where

$$
\begin{equation*}
\Psi[x, t]=\psi_{E}[x] e^{-i E t / \hbar} \tag{3.115}
\end{equation*}
$$

Equation 3.114 implies that $\psi_{E}[x]$ is an eigenfunction of the Hamiltonian $H_{\mathrm{op}}$ with eigenvalue $E$. Both $\psi_{E}[x]$ and $E$ may be chosen to be real (rather than complex). The $\psi_{E}[x]$ are called stationary states because the corresponding probabilty density is independent of time,

$$
\begin{equation*}
|\Psi[x, t]|^{2}=\psi_{E}[x] e^{-i E t / \hbar} \psi_{E}[x] e^{+i E t / \hbar}=\psi_{E}[x]^{2} \tag{3.116}
\end{equation*}
$$

Nothing happens in a stationary state. Furthermore, the $\psi_{E}[x]$ form a complete set of solutions, as the most general solution can be written as a linear superposition

$$
\begin{equation*}
\tilde{\Psi}[x, t]=\sum_{E} c_{E} \Psi[x, t]=\sum_{E} c_{E} \psi_{E}[x] e^{-i E t / \hbar} \tag{3.117}
\end{equation*}
$$

where the complex coefficients $c_{E}$ are determined by the initial superposition

$$
\begin{equation*}
\tilde{\Psi}[x, 0]=\sum_{E} c_{E} \psi_{E}[x] \tag{3.118}
\end{equation*}
$$

(In a similar way, any vibration of a guitar string can be written as a linear superposition of normal mode motions.)

In summary, the Hamiltonian eigenfunctions $\psi_{E}[x]$ are stationary states of definite energy forming a complete set of solutions to the Schrödinger equation.

### 3.4.7 Qualitative Solutions

Developing heuristic rules for sketching the Hamiltonian eigenfunctions $\psi_{E}[x]$ for various potential energy functions $V[x]$ is instructive. We begin by rewriting the eigenfunction-eigenvalue Equation 3.114 as

$$
\begin{equation*}
\psi_{E}^{\prime \prime}=-\frac{2 m}{\hbar^{2}}(E-V[x]) \psi_{E}=-\frac{2 m}{\hbar^{2}} T[x] \psi_{E} \tag{3.119}
\end{equation*}
$$

If the kinetic energy $T[x]=E-V[x]>0$, then

$$
\begin{equation*}
\psi_{E}[x] \sim e^{ \pm i k[x] x} \tag{3.120}
\end{equation*}
$$

at least for small ranges of $x$, where the spatial frequency

$$
\begin{equation*}
\frac{2 \pi}{\lambda[x]}=k[x]=\sqrt{+\frac{2 m}{\hbar^{2}} T[x]} . \tag{3.121}
\end{equation*}
$$

Conversely, if the kinetic energy $T[x]<0$, then

$$
\begin{equation*}
\psi_{E}[x] \sim e^{ \pm \kappa[x] x} \tag{3.122}
\end{equation*}
$$

at least for small ranges of $x$, where the decay constant

$$
\begin{equation*}
\kappa[x]=\sqrt{-\frac{2 m}{\hbar^{2}} T[x]} . \tag{3.123}
\end{equation*}
$$

Thus, in the classically allowed regions of positive kinetic energy, $T>0$, the eigenfunctions are sinusoidal with spatial frequency $k \propto \sqrt{+T}$. In the classically forbidden regions of negative kinetic energy, $T<0$, the eigenfunctions are sinusoidal with decay constant $\kappa \propto \sqrt{-T}$. Eigenfunctions bend toward axis in allowed regions and away from axis in forbidden regions. In addition, large positive kinetic energy corresponds to fast classical motion and hence low probability of being there, which suggests (but doesn't guarantee) small eigenfunction amplitude. Conversely, small positive kinetic energy corresponds to slow classical motion and hence large probability of being there, which suggests large eigenfunction amplitude. These heuristics are illustrated in Figure 3.19.

If the potential energy function $V[x]$ is symmetric, then the eigenfunctions $\psi_{E}[x]$ are either symmetric or antisymmetric. To prove this, suppose $V[-x]=$ $V[x]$. This implies $H_{\mathrm{op}}[-x]=H_{\mathrm{op}}[x]$, and so both $H_{\mathrm{op}}[x] \psi_{E}[x]=E \psi_{E}[x]$ and $H_{\mathrm{op}}[x] \psi_{E}[-x]=E \psi_{E}[-x]$. Assuming the eigenvalues are nondegenerate, this means that $\psi_{E}[x]$ and $\psi_{E}[-x]$ must be proportional to each other. If the proportionality constant is $K$, then $\psi_{E}[-x]=K \psi_{E}[x]=K^{2} \psi_{E}[-x]$, and so $K^{2}=1$. Further, since the eigenfunctions and eigenvalues may be assumed real, $K= \pm 1$. Therefore $\psi_{E}[-x]= \pm \psi_{E}[x]$.

If the potential energy function $V[x]$ is continuous - or has only finite discontinuities - both the eigenfunction $\psi_{E}[x]$ and its derivative $\psi_{E}{ }^{\prime}[x]$ are continuous. To prove this, suppose the $V[x]$ has a finite discontinuity at $x_{0}$. Let $V_{c}[x]$ be a continuous version of $V[x]$, which is the same everywhere, except in the interval


Figure 3.19: Generic potential energy function $V[x]$ and one of its energy eigenfunctions $\psi_{E}[x]$. Eigenfunctions bend toward axis in classically allowed regions and away from axis in classically forbidden regions. Large amplitude and large wavelength (small spatial frequency) tend to occur together.
$\left\{x_{0}-\delta, x_{0}+\delta\right\}$, where it linearly interpolates across the discontinuity. The corresponding eigenfunction $\psi_{c}[x]$ obeys

$$
\begin{equation*}
\psi_{c}{ }^{\prime \prime}=-\frac{2 m}{\hbar^{2}}\left(E-V_{c}[x]\right) \psi_{c} \tag{3.124}
\end{equation*}
$$

Integrating both sides, we get

$$
\begin{equation*}
\psi_{c}^{\prime}\left[x_{0}+\delta\right]-\psi_{c}^{\prime}\left[x_{0}-\delta\right]=-\frac{2 m}{\hbar^{2}} \int_{x_{0}-\delta}^{x_{0}+\delta} d x\left(E-V_{c}[x]\right) \psi_{c}[x] \tag{3.125}
\end{equation*}
$$

In the limit $\delta x \rightarrow 0$, both $V_{c} \rightarrow V$ and $\psi_{c} \rightarrow \psi$, and so, using the mean value theorem for integrals,

$$
\begin{equation*}
\psi^{\prime}\left[x_{0}+\delta\right]-\psi^{\prime}\left[x_{0}-\delta\right]=-\frac{2 m}{\hbar^{2}}(2 \delta)\left(E-\frac{V\left[x_{0}+\delta\right]+V\left[x_{0}-\delta\right]}{2}\right) \psi\left[x_{0}\right] \rightarrow 0 \tag{3.126}
\end{equation*}
$$

provided $V[x]$ is bounded. Hence, $\psi^{\prime}[x]$ (and all the more $\psi[x]$ ) is continuous at $x_{0}$.

### 3.4.8 Particle in a Box

Consider the canonical example of a particle confined to a semi-rigid box, perhaps an electric charge confined to a vacuum tube by electric fields, as in Figure
3.20. Potential energy $V[x]$ (in Joules) is proportional to the electric potential $\varphi[x]$ (in volts), $V=q \varphi$. Force $F_{x}[x]$ (in Newtons) is the negative gradient of the potential energy, $F_{x}=-d V / d x$.


Figure 3.20: A positive charge trapped in a vacuum tube by electric fields.


Figure 3.21: Finite square well with energy eigenvalues $E$ and corresponding eigenfunctions $\psi_{E}$. The nondiverging, normalizable, physical solutions $\psi_{n}$ are alternately symmetric or antisymmetric.

Idealize the confining potential energy function $V[x]$ by a finite square well, as in Figure 3.21. Imagine sweeping the energy $E$ from the bottom to the top of the square well and, for each energy, numerically integrating the timeindependent Schrödinger equation from left to right. Beginning with a very small wave function of very small slope on the far left, the wavefunction increases exponentially in the classically forbidden region on the left, joins continuously and smoothly to sinusoidal oscillations in the classically allowed region in the
center, and joins continuously and smoothly to some superposition of increasing and decreasing exponentials in the classical forbidden region on the right. As the energy $E=T[x]+V[x]$ increases, the exponential decay constants decrease and the sinusoidal spatial frequency increases. Only for certain, discrete energies $E_{n}$ does the wave function vanish exponentially both left and right and allow nondiverging, normalizable, physical solutions. In this way, a continuous differential equation gives rise to discrete or quantized energies.

### 3.4.9 Particle in a Rigid Box

The rigid box is a limiting case of the particle-in-the-box model as the depth of the finite square well increases until it becomes an infinite square well. In this instructive special case, our qualitative methods become quantitative!


Figure 3.22: Infinite square well with normalized eigenfunctions superimposed on corresponding energy eigenvalues.

As the depth of the well increases, the exponential decay constants in the classically forbidden region increase until, in the limit, the eigenfunction decays immediately, as in Figure 3.22. Thus, the physical, normalizable eigenfunctions are sinusoids with nodes at each wall of the box and an integer quantum number of half wavelengths in the box,

$$
\begin{equation*}
n \frac{\lambda_{n}}{2}=L \tag{3.127}
\end{equation*}
$$

where $L$ is the length of the box. By the de Broglie relation $p_{n}=h / \lambda_{n}$ and the energy relation $E_{n}=p_{n}^{2} / 2 m$, the set of allowed energies or energy spectrum of
a particle inside the rigid box is

$$
\begin{equation*}
E_{n}=n^{2} E_{1} \tag{3.128}
\end{equation*}
$$

where the lowest or ground state energy is

$$
\begin{equation*}
E_{1}=\frac{h^{2}}{8 m L^{2}}=\frac{\pi^{2} \hbar^{2}}{2 m L^{2}} \tag{3.129}
\end{equation*}
$$

(The term "spectrum" comes from spectroscopy, where the spectral lines of an element arise from transitions among different allowed energy levels.) The corresponding eigenfunctions are, by inspection, the boxed sinusoids

$$
\psi_{n}[x]=\left\{\begin{array}{cc}
N \sin [n \pi x / L], & x \in[0, L]  \tag{3.130}\\
0, & x \notin[0, L]
\end{array}\right\}
$$

where $N=\sqrt{2 / L}$ is a normalization constant determined by the requirement

$$
\begin{equation*}
1=\int_{-\infty}^{\infty} d x\left|\psi_{n}\right|^{2}=\int_{0}^{L} d x \psi_{n}^{2} \tag{3.131}
\end{equation*}
$$

Note how the infinite discontinuities in the potential energy function at the walls of the box kink the eigenfunctions, so that while $\psi_{n}[x]$ is continuous, $\psi_{n}{ }^{\prime}[x]$ is not.


Figure 3.23: For large quantum numbers, the quantum probability distribution of a particle in a rigid box corresponds to the classical probability distribution.

Particle-in-a-box states of small quantum number $n$ do not have classical analogues. In fact, such eigenfunctions are very wave-like, dominated by nodes near which the probability of finding the particle is near zero. However, we can recover a classical correspondence by considering states of large quantum number, say $n=20$, as in Figure 3.23. Classically, a particle in the box with positive kinetic energy would bounce back and forth between the walls at constant speed. It would equally likely be found anywhere, and its position would have a uniform detection probability distribution $\rho_{c}[x]$. The corresponding squared-sinuosidal
quantum probability distribution $\left|\psi_{20}\right|^{2}$ oscillates rapidly. However, any macroscopic detector would average over these oscillations to obtain an effectively uniform distribution.

Note that the particle could not be in the box at rest, because then its (zero) momentum and position would both be known simultaneously, in violation of the Heisenberg uncertainty principle.

### 3.4.10 Escape from the Box: Negative Kinetic Energy?

As we have seen, if a particle is confined to a semi-rigid box, in a stationary state of definite energy, its eigenfunction "leaks" into the classically forbidden region. Thus, when in its ground state, a nonzero probability exists to observe the particle just outside the box, say to the left,

$$
\begin{equation*}
\mathcal{P}[x \leq 0]=\int_{-\infty}^{0} d x\left|\psi_{1}[x]\right|^{2}>0 \tag{3.132}
\end{equation*}
$$

But in this region, classically, the particle would have negative kinetic energy whatever that might mean. Can we observe its negative kinetic energy?

No! Localizing the particle's position to just outside the box delocalizes its momentum, and hence also its kinetic energy, thereby obscuring the observation, as we now argue. Suppose we localize the particle just outside the box, so that it is within the "e-folding" distance of the exponential tail of $\left|\psi_{1}[x]\right|^{2}$, a distance $\Delta x \sim 1 / \kappa$ from the box. The uncertainty in its momentum must therefore be

$$
\begin{equation*}
\Delta p \geq \frac{\hbar}{2 \Delta x} \sim \frac{\hbar}{2} \kappa=\frac{1}{2} \sqrt{-2 m T} \tag{3.133}
\end{equation*}
$$

But if $-T=p^{2} / 2 m$, then the consequent uncertainty in the kinetic energy is

$$
\begin{equation*}
-\Delta T=\frac{p}{m} \Delta p \gtrsim \frac{\sqrt{-2 m T}}{m} \frac{1}{2} \sqrt{-2 m T}=-T \tag{3.134}
\end{equation*}
$$

which is at least as large as the negative kinetic energy we had hoped to measure in the first place.

### 3.4.11 Escape from the Box: STM!

Does the eigenfunction's exponential tail into the classically forbidden region have practical significance? Yes! It is the basis of the exquisitely sensitive Scanning Tunneling Microscope (STM). An electron in a solid is like a particle in a semi-rigid box. It moves in a finite-square-well-type potential, as indicated in Figure 3.24. An electron in the tip of an STM moves in a similar potential. When the two potentials are brought near one another, the exponential tail of a joint eigenfunction can allow an electron to "leak" or "tunnel" from sample to probe across the classically forbidden region. (This is the quantum analogue of frustrated internal reflection's evanescent wave, which was the basis of our beam splitter in Section 3.1.1.)


Figure 3.24: Electron confined to sample (top) moves in finite square well potential. However, when an STM probe is near (bottom), it can "tunnel" across the classically forbidden region.

If the eigenfunction $\psi[x]=\psi[0] e^{-\kappa x}$ decays exponentially, then the probability density $\psi[x]^{2}=\psi[0]^{2} e^{-2 \kappa x}$ does so also. Consequently, the tunnelling current

$$
\begin{equation*}
I[x]=I[0] e^{-2 \kappa x} \sim I[0](1-2 \kappa x) \tag{3.135}
\end{equation*}
$$

is exponentially sensitive to the tiny gap distance $x \lesssim 1 / \kappa$ and, as the probe scans the surface, then the relative change in current

$$
\begin{equation*}
\frac{\Delta I}{I[0]}=\frac{I[0]-I[x]}{I[0]} \sim 2 \kappa x \tag{3.136}
\end{equation*}
$$

is proportional to the gap distance.
For typical energies of $|T| \sim 4 \mathrm{eV}$, from Equation 3.123, the exponential decay constant

$$
\begin{equation*}
\kappa=\sqrt{\frac{2 m|T|}{\hbar^{2}}}=\sqrt{\frac{2\left(m c^{2}\right)|T|}{(\hbar c)^{2}}} \sim \sqrt{\frac{2(0.5 \mathrm{MeV}) 4 \mathrm{eV}}{(2 \mathrm{keV} \cdot \AA)^{2}}}=\frac{1}{1 \AA} . \tag{3.137}
\end{equation*}
$$

Thus, if the gap distance is $x \sim 1 \AA$, then the relative change in current is of the order $\Delta I / I[0] \sim 2(1 / \AA)(1 \AA)=2$, which is easily detected. As the STM scans an atomic surface in a raster pattern, the tunneling current as a function of position reflects the topography of the surface. Thanks to the exponential decay of the electron wavefunctions in the gap, fine STM tips are readily fabricated, as almost all current flows through the single atom in the tip nearest the surface.

### 3.4.12 Quantum Tunneling in $\mathrm{NH}_{3}$

The ammonia molecule $\mathrm{NH}_{3}$ provides another fascinating example of quantum tunneling. $\mathrm{NH}_{3}$ is shaped like a pyramid, with a large N molecule at the apex and a triangle of small H atoms at the base, as depicted in Figure 3.25.


Figure 3.25: By quantum tunneling, the pyramidal ammonia molecule can spontaneously invert, like an umbrella catching a gust of wind.

In addition to electronic, translational, vibrational, and rotational degrees of freedom, $\mathrm{NH}_{3}$ has an additional degree of freedom: the base of H atoms can be on one side of the N atom or the other. Classically, a potential energy barrier prevents such an "inverting umbrella" transition, which we can simply model with a finite square barrier inside an infinite square well, as in Figure 3.26. The potential barrier reflects the repulsion between the N and the H atoms; the potential side walls reflect the chemical bonding, which insures the molecule's cohesion; the two minima represent the two stable configurations. Quantumly, the molecule can tunnel between these two configurations, and it does so spontaneously, in the absence of any forcing.

We can infer the double-well energy eigenvalues, or spectrum, and the corresponding energy eigenfunctions, by beginning with a broad infinite square well, which is exactly solvable, and growing a central barrier to infinity, where it breaks the broad infinite square well into a pair of narrow infinite square wells, which are similarly exactly solvable, as in Figure 3.27. As the barrier height grows from zero, pairs of symmetric and antisymmetric eigenvalues of the broad infinite square well converge to a single degenerate eigenvalue. Conversely, as the barrier height shrinks from infinity, each eigenvalue of the pair of infinite square wells splits into a pair of eigenvalues.

When the barrier height is nonzero, we will denote the first two eigenfunctions, which are necessarily symmetric and antisymmetric, by $\psi_{s}$ and $\psi_{a}$ and the corresponding eigenvalues by $E_{s}$ and $E_{a}$. When the barrier height is high, as in the top right of Figure 3.27, the sum of these eigenfunctions

$$
\begin{equation*}
\psi_{L}[x]=\frac{1}{\sqrt{2}}\left(\psi_{s}[x]+\psi_{a}[x]\right) \tag{3.138}
\end{equation*}
$$

can represent the ammonia molecule with the $H$ base on the left of the $N$, as


Figure 3.26: We idealize the actual potential energy function (top) by a finite square barrier inside an infinite square well (bottom). The coordinate $x_{H}$ locates the base plane containing the hydrogens.
the sum nearly vanishes on the right, while the difference

$$
\begin{equation*}
\psi_{R}[x]=\frac{1}{\sqrt{2}}\left(\psi_{s}[x]-\psi_{a}[x]\right) \tag{3.139}
\end{equation*}
$$

can represent ammonia with the H base on the right, as the difference nearly vanishes on the left.

Suppose, at $t=0$, the H base is on the left, so that the wavefunction for our ammonia model (neglecting its other degrees of freedom) is

$$
\begin{equation*}
\Psi[x, 0]=\psi_{L}[x] \tag{3.140}
\end{equation*}
$$

According to the Equation 3.117 general Schrödinger solution, to find the wave function at later times we must expand the initial state as a linear combination of stationary states of definite energy

$$
\begin{equation*}
\Psi[x, 0]=\frac{1}{\sqrt{2}}\left(\psi_{s}[x]+\psi_{a}[x]\right) \tag{3.141}
\end{equation*}
$$

using Equation 3.138, and thereafter the complex phase of each such state rotates at a frequency proportional to the corresponding energy

$$
\begin{equation*}
\Psi[x, t]=\frac{1}{\sqrt{2}}\left(\psi_{s}[x] e^{-i E_{s} t / \hbar}+\psi_{a}[x] e^{-i E_{a} t / \hbar}\right) \tag{3.142}
\end{equation*}
$$



Figure 3.27: Tunneling eigenvalues and eigenfunctions interpolated from extreme infinite square wells. The barrier height $V_{b}$-axis is nonlinear.
(This is an especially simple case of the the wave packet preparation and evolution recipe of Equations 3.98 and 3.99.) It can be written as

$$
\begin{equation*}
\Psi[x, t]=\frac{1}{\sqrt{2}} e^{-i \bar{E} t / \hbar}\left(\psi_{s}[x] e^{+i \omega t / 2}+\psi_{a}[x] e^{-i \omega t / 2}\right) \tag{3.143}
\end{equation*}
$$

where the average energy $\bar{E}=\left(E_{s}+E_{a}\right) / 2$ and the energy splitting $\hbar \omega=E_{a}-E_{s}$. Consequently, the probability density

$$
\begin{equation*}
|\Psi[x, t]|^{2}=\frac{1}{2}\left(\psi_{s}[x]^{2}+2 \psi_{s}[x] \psi_{a}[x] \cos [\omega t]+\psi_{a}[x]^{2}\right) \tag{3.144}
\end{equation*}
$$

sloshes back

$$
\begin{equation*}
\left|\Psi\left[x, \frac{2 \pi}{\omega}\right]\right|^{2}=\frac{1}{2}\left(\psi_{s}[x]^{2}+2 \psi_{s}[x] \psi_{a}[x]+\psi_{a}[x]^{2}\right)=\psi_{L}[x]^{2} \tag{3.145}
\end{equation*}
$$

and forth

$$
\begin{equation*}
\left|\Psi\left[x, \frac{\pi}{\omega}\right]\right|^{2}=\frac{1}{2}\left(\psi_{s}[x]^{2}-2 \psi_{s}[x] \psi_{a}[x]+\psi_{a}[x]^{2}\right)=\psi_{R}[x]^{2} \tag{3.146}
\end{equation*}
$$

sinusoidally at the frequency $\omega$.
$\mathrm{NH}_{3}$ is a polar molecule, as the N attracts the H electrons creating an electric dipole moment pointing from the apex perpendicular to the base. Consequently, due to quantum tunneling, $\mathrm{NH}_{3}$ can emit or absorb electromagnetic radiation at a frequency ( $\sim 24 \mathrm{GHz}$ ) proportional to the energy splitting $\left(\sim 10^{-4} \mathrm{eV}\right)$ and in the microwave ( $\sim 1.25 \mathrm{~cm}$ ) region of the electromagnetic spectrum. This is the basis of the ammonia maser. (The word MASER was originally an acronym for "Microwave Amplification by the Stimulated Emission of Radiation").

### 3.4.13 Band Structure of Solids

What if we take three infinite square wells and drop the barriers between them? Each energy eigenvalues splits into three, as in Figure 3.28. Thus, the spectrum of three semi-rigid boxes close together consists of triplets of allowed energies separated by bands of forbidden energies.

What if we have Avogadro's number $N_{A} \sim 10^{24}$ such boxes? The spectrum then consists of effectively continuous bands of allowed energies separated by bands of forbidden energies. This band structure is critical to understanding the physics of solids, including insulators, conductors, and semiconductors.


Figure 3.28: Interpolating the energy eigenvalues and corresponding eigenfunctions for three boxes close together.

### 3.5 Quantum Harmonic Oscillator

We quantitatively solve the Schrödinger equation for a simple harmonic oscillator, the most important single example, which is at the core of quantum field
theory, the union of special relativity and quantum mechanics.

### 3.5.1 Classical Harmonic Oscillator

Consider a simple (or ideal) harmonic oscillator, a mass $m$ connected to a Hooke's law spring of stiffness $k$. If the displacement is $x$, then the linear restoring force is $F_{x}=-k x$, and the quadratic (or parabolic) potential energy function is

$$
\begin{equation*}
V[x]=\frac{1}{2} k x^{2} \tag{3.147}
\end{equation*}
$$

The equation of motion follows from Newton's second law $a_{x}=F_{x} / m$, namely

$$
\begin{equation*}
\partial_{t}^{2} x=-\frac{1}{m} \partial_{x} V \tag{3.148}
\end{equation*}
$$

or

$$
\begin{equation*}
\ddot{x}=-\frac{k}{m} x \tag{3.149}
\end{equation*}
$$

This has the well-known sinusoidal solution

$$
\begin{equation*}
x[t]=A \sin [\omega t+\varphi] \tag{3.150}
\end{equation*}
$$

provided the angular frequency $\omega=\sqrt{k / m}$. The constants $A$ and $\varphi$ depend on the initial conditions.

Real springs, of course, aren't so simple. If you stretch them too far, for example, they break. However, almost any potential energy function is approximately parabolic near a local minimum, as we now show. If the potential $V[x]$ has a minimum at $x_{0}$, expand in a Taylor series to get

$$
\begin{equation*}
V[x]=V\left[x_{0}\right]+V^{\prime}[x]\left(x-x_{0}\right)+\frac{1}{2} V^{\prime \prime}\left[x_{0}\right]\left(x-x_{0}\right)^{2}+\cdots \tag{3.151}
\end{equation*}
$$

Since $V^{\prime}\left[x_{0}\right]=0$, near $x_{0}$

$$
\begin{equation*}
V[x]-V\left[x_{0}\right] \sim \frac{1}{2} V^{\prime \prime}\left[x_{0}\right]\left(x-x_{0}\right)^{2} \tag{3.152}
\end{equation*}
$$

or

$$
\begin{equation*}
\delta V \sim \frac{1}{2} k(\delta x)^{2} \tag{3.153}
\end{equation*}
$$

where $k=V^{\prime \prime}\left[x_{0}\right]$. Thus, the simple harmonic oscillator is a canonical system of widespread importance.

### 3.5.2 Heuristic Spectrum

Heuristically, the Eq. 3.128 and Eq. 3.129 energy spectrum for a rigid box follows from kinetic energy $E=p^{2} / 2 m$, the de Broglie relation $p=h / \lambda$, and quantized waves $n \lambda / 2=L$ to give $E \propto n^{2} / L^{2}$, where the length $L$ is the distance between the classical turning points. If the latter increases with energy, the quadratically increasing excited energy levels will fall until they bunch up instead of spread out, as in Fig. 3.29. The energy levels are equally spaced only for the intermediate case of the simple harmonic oscillator.


Figure 3.29: Energy spectra for rigid box, simple harmonic oscillator, and "1D hydrogen atom" have increasing, constant, and decreasing energy level separations.

### 3.5.3 Dimensionless Variables

The parallel quantum problem is to find square normalizable eigenfunctions $\psi[x]$ and the corresponding eigenvalues $E$ for the quantum harmonic oscillator Hamiltonian. This means solving the time independent Schrödinger equation

$$
\begin{equation*}
-\frac{\hbar^{2}}{2 m} \psi^{\prime \prime}+\frac{1}{2} k x^{2} \psi=E \psi \tag{3.154}
\end{equation*}
$$

subject to the constraint

$$
\begin{equation*}
1=\int_{-\infty}^{\infty} d x \psi^{2} \tag{3.155}
\end{equation*}
$$

This famous problem is nontrivial. However, we will follow in the footsteps of those who have solved such problems before us.

A good first step is to introduce dimensionless variables. For the position scale, let $x_{0}$ be the classical turning point for a harmonic oscillator with energy $E_{0}=\hbar \omega / 2$. Thus

$$
\begin{equation*}
\frac{1}{2} \hbar \omega=E_{0}=T+V=\frac{1}{2} m v^{2}+\frac{1}{2} k x^{2}=0+\frac{1}{2} k x_{0}^{2} \tag{3.156}
\end{equation*}
$$

which implies

$$
\begin{equation*}
x_{0}=\sqrt{\frac{\hbar \omega}{k}}=\sqrt{\frac{\hbar}{m \omega}} . \tag{3.157}
\end{equation*}
$$

Use this scale to define a dimensionless position

$$
\begin{equation*}
\xi=\frac{x}{x_{0}} \tag{3.158}
\end{equation*}
$$

and a dimensionless eigenfunction

$$
\begin{equation*}
\varphi[\xi]=\sqrt{x_{0}} \psi[x] \tag{3.159}
\end{equation*}
$$

Then the derivatives transform like

$$
\begin{equation*}
\psi^{\prime}=\frac{d \psi}{d x}=\frac{1}{\sqrt{x_{0}}} \frac{d \varphi}{d x}=\frac{1}{\sqrt{x_{0}}} \frac{d \xi}{d x} \frac{d \varphi}{d \xi}=\frac{1}{x_{0}^{3 / 2}} \varphi^{\prime} \tag{3.160}
\end{equation*}
$$

and

$$
\begin{equation*}
\psi^{\prime \prime}=\frac{d \psi^{\prime}}{d x}=\frac{1}{x_{0}^{3 / 2}} \frac{d \varphi^{\prime}}{d x}=\frac{1}{x_{0}^{3 / 2}} \frac{d \xi}{d x} \frac{d \varphi^{\prime}}{d \xi}=\frac{1}{x_{0}^{5 / 2}} \varphi^{\prime \prime} \tag{3.161}
\end{equation*}
$$

where, as usual, the prime denotes derivative with respect to the argument ( $x$ or $\xi$, as appropriate). Putting this altogether, our problem transforms to solving

$$
\begin{equation*}
\varphi^{\prime \prime}=\left(\xi^{2}-\epsilon\right) \varphi \tag{3.162}
\end{equation*}
$$

subject to the constraint

$$
\begin{equation*}
1=\int_{-\infty}^{\infty} d \xi \varphi^{2} \tag{3.163}
\end{equation*}
$$

where the dimensionless energy

$$
\begin{equation*}
\epsilon=\frac{E}{E_{0}} \tag{3.164}
\end{equation*}
$$

### 3.5.4 Asymptotic Behavior

When $\xi$ is large, we can neglect $\epsilon$ and write

$$
\begin{equation*}
\varphi^{\prime \prime} \sim \xi^{2} \varphi \tag{3.165}
\end{equation*}
$$

which has the approximate exponential solutions

$$
\begin{equation*}
\varphi \sim \pm e^{ \pm \frac{1}{2} \xi^{2}} \tag{3.166}
\end{equation*}
$$

To verify this, note that

$$
\begin{equation*}
\varphi^{\prime} \sim \xi e^{ \pm \frac{1}{2} \xi^{2}} \tag{3.167}
\end{equation*}
$$

and

$$
\begin{equation*}
\varphi^{\prime \prime} \sim\left( \pm 1+\xi^{2}\right) e^{ \pm \frac{1}{2} \xi^{2}} \sim \xi^{2} \varphi \tag{3.168}
\end{equation*}
$$

Since only the decaying exponential is square normalizable, strip off the asymptotic behavior and assume solutions of the form

$$
\begin{equation*}
\varphi[\xi]=h[\xi] e^{-\frac{1}{2} \xi^{2}} \tag{3.169}
\end{equation*}
$$

where we expect the functions $h[\xi]$ to be polynomials. Then

$$
\begin{equation*}
\varphi^{\prime}=\left(h^{\prime}-\xi h\right) e^{-\frac{1}{2} \xi^{2}} \tag{3.170}
\end{equation*}
$$

and

$$
\begin{equation*}
\varphi^{\prime \prime}=\left(h^{\prime \prime}-2 \xi h^{\prime}+\left(\xi^{2}-1\right) h\right) e^{-\frac{1}{2} \xi^{2}} \tag{3.171}
\end{equation*}
$$

With these substitutions, the exponentials cancel, and our differential equation becomes

$$
\begin{equation*}
h^{\prime \prime}-2 \xi h^{\prime}+(\epsilon-1) h=0 \tag{3.172}
\end{equation*}
$$

### 3.5.5 Power Series Solution

We write our solution as a power series

$$
\begin{equation*}
h[\xi]=a_{0}+a_{1} \xi+a_{2} \xi^{2}+a_{3} \xi^{3}+\cdots=\sum_{m=0}^{\infty} a_{m} \xi^{m}, \tag{3.173}
\end{equation*}
$$

and its first derivative

$$
\begin{equation*}
h^{\prime}[\xi]=0+a_{1}+2 a_{2} \xi+3 a_{3} \xi^{2}+\cdots=\sum_{m=0}^{\infty} m a_{m} \xi^{m-1} \tag{3.174}
\end{equation*}
$$

and its second derivative

$$
\begin{equation*}
h^{\prime \prime}[\xi]=0+0+2 a_{2}+3 \cdot 2 a_{3} \xi+\cdots=\sum_{m=0}^{\infty} m(m-1) a_{m} \xi^{m-2} \tag{3.175}
\end{equation*}
$$

Substituting these power series into Equation 3.172 gives

$$
\begin{equation*}
\sum_{m=0}^{\infty} m(m-1) a_{m} \xi^{m-2}-2 \xi \sum_{m=0}^{\infty} m a_{m} \xi^{m-1}+(\epsilon-1) \sum_{m=0}^{\infty} a_{m} \xi^{m}=0 \tag{3.176}
\end{equation*}
$$

By shifting the dummy index $m \rightarrow m+2$ in first summation, we can consolidate this as

$$
\begin{equation*}
\sum_{m=0}^{\infty}\left((m+2)(m+1) a_{m+2}-2 m a_{m}+(\epsilon-2) a_{m}\right) \xi^{m}=0 \tag{3.177}
\end{equation*}
$$

The only way this can be true for all $\xi$ is if the coefficients are all zero, which means

$$
\begin{equation*}
a_{m+2}=\frac{2 m+1-\epsilon}{(m+1)(m+2)} a_{m} \tag{3.178}
\end{equation*}
$$

This recursion relation separately links coefficients of odd and even indices. It thereby specifies two independent solutions, corresponding to the two arbitrary constants determined by the initial conditions of our second-order differential equation. The constant $a_{0}$ specifies symmetric solutions $h[-\xi]=h[\xi]$ in even powers of $\xi$, while the constant $a_{1}$ specifies antisymmetric solutions $h[-\xi]=-h[\xi]$ in odd powers of $\xi$. This is consistent with our expectation that symmetric potentials $V[-x]=V[x]$ imply eigenfunctions of definite symmetry $\psi[-x]= \pm \psi[x]$.

### 3.5.6 Power Series Diverges

For large $m \gg 1$, the recursion relation simplifies to

$$
\begin{equation*}
a_{m+2} \sim \frac{2 m}{m \cdot m} a_{m}=\frac{a_{m}}{m / 2} \tag{3.179}
\end{equation*}
$$

This has the approximate solution

$$
\begin{equation*}
a_{m} \sim \frac{K}{(m / 2)!} \tag{3.180}
\end{equation*}
$$

for some constant $K$, because it implies

$$
\begin{equation*}
a_{m+2} \sim \frac{K}{(m / 2+1)!}=\frac{K}{(m / 2+1)(m / 2)!} \sim \frac{a_{m}}{m / 2+1} \sim \frac{a_{m}}{m / 2} \tag{3.181}
\end{equation*}
$$

However, this means

$$
\begin{equation*}
h[\xi] \sim \sum_{m \gg 1} \frac{K}{(m / 2)!} \xi^{m}=K \sum_{m \gg 1} \frac{\left(\xi^{2}\right)^{m / 2}}{(m / 2)!} \sim K \sum_{l=0}^{\infty} \frac{\left(\xi^{2}\right)^{l}}{l!}=K e^{\xi^{2}} \tag{3.182}
\end{equation*}
$$

Thus

$$
\begin{equation*}
\varphi[\xi]=h[\xi] e^{-\frac{1}{2} \xi^{2}} \sim \tilde{K} e^{+\frac{1}{2} \xi^{2}} \tag{3.183}
\end{equation*}
$$

for some constant $\tilde{K}$, which is precisely the divergent, unnormalizable behavior we don't want.

### 3.5.7 Truncate Series

The only way to avoid nonphysical solutions is for the infinite power series to terminate. This can happen if the numerator of the recursion relation vanishes for some $m=n<\infty$, in which case $a_{n+2}=0$ and hence $a_{m \geq n+2}=0$. The only way for the numerator to vanish is if the dimensionless energy $\epsilon$ is quantized according to

$$
\begin{equation*}
\epsilon_{n}=2 n+1 \tag{3.184}
\end{equation*}
$$

which implies that the dimensional energy $E$ is quantized according to

$$
\begin{equation*}
E_{n}=\epsilon_{n} E_{0}=(2 n+1) \frac{\hbar \omega}{2}=\left(n+\frac{1}{2}\right) \hbar \omega \tag{3.185}
\end{equation*}
$$

for $n=0,1,2, \ldots$ Thus, the physically relevant recursion relation is

$$
\begin{equation*}
a_{m+2}=\frac{2 m+1-\epsilon_{n}}{(m+1)(m+2)} a_{m}=-\frac{2(n-m)}{(m+1)(m+2)} a_{m} \tag{3.186}
\end{equation*}
$$

where $m=0,1,2, \ldots, n$. This defines a symmetric or antisymmetric $n$th order hermite polynomial $H_{n}[\xi]$.

### 3.5.8 Standard Form Solutions

We can write the harmonic oscillator eigenfunctions in standard form as

$$
\begin{equation*}
\psi_{n}[x]=N_{n} H\left[x / x_{0}\right] e^{-\frac{1}{2}\left(x / x_{0}\right)^{2}} \tag{3.187}
\end{equation*}
$$

where $x_{0}$ is the classical turning point of Equation 3.157 and the normalization constant

$$
\begin{equation*}
N_{n}=\frac{1}{\sqrt{x_{0} 2^{n} n!\sqrt{\pi}}} \tag{3.188}
\end{equation*}
$$

is fixed by the constraint Equation 3.155. The first few eigenfunctions are listed in Table 3.3 and graphed in Figure 3.30.

Table 3.3: First few harmonic oscillator eigenvalues and eigenfunctions.

| $n$ | $E_{n}$ | $\psi_{n}[x]$ |
| ---: | :---: | ---: |
| 0 | $\frac{1}{2} E_{0}$ | $N_{0} e^{-\frac{1}{2}\left(x / x_{0}\right)^{2}}$ |
| 1 | $\frac{3}{2} E_{0}$ | $N_{1} 2\left(x / x_{0}\right) e^{-\frac{1}{2}\left(x / x_{0}\right)^{2}}$ |
| 2 | $\frac{5}{2} E_{0}$ | $N_{2}\left(-2+4\left(x / x_{0}\right)^{2}\right) e^{-\frac{1}{2}\left(x / x_{0}\right)^{2}}$ |
| 3 | $\frac{7}{2} E_{0}$ | $N_{3}\left(-12\left(x / x_{0}\right)+8\left(x / x_{0}\right)^{3}\right) e^{-\frac{1}{2}\left(x / x_{0}\right)^{2}}$ |



Figure 3.30: First few harmonic oscillator eigenfunctions superimposed on the corresponding energy eigenvalues of the quadratic potential. The dots denote concavity changes, the smooth joining of sinusoids and exponentials, at the classical turning points.

The ground state or zero-point energy $E_{0}=\hbar \omega / 2$ is nonzero due to the Heisenberg uncertainty principle. If it were zero, the oscillator's position and momentum would be both be exactly zero, but as we have seen, if one of the two is exact, the other must be indeterminate. The zero-point energy of the
quantum vacuum may be related to the Dark Energy (or Clear Tension!) that seems to be accelerating the expansion of the universe.

The regular energy spacing $\Delta E=E_{n+1}-E_{n}=\hbar \omega$ makes possible the photon model of light. Transitions between adjacent energy levels are accompanied by the emission or absorption of photons of energy $\hbar \omega$, corresponding to classical light of temporal frequency $\omega$.

### 3.5.9 Classical Correspondence

Quantum harmonic oscillator states of small quantum number $n$ do not have classical analogues. In fact, such eigenfunctions are very wave-like, dominated by nodes near which the probability of finding the particle is near zero. However, we can recover a classical correspondence by considering states of large quantum number.


Figure 3.31: At even the modest quantum number $n=20$, the quantum probability density corresponds well to the classical probability density.

For comparison, we must first compute the probability distribution for a classical harmonic oscillator. The sinusoidally oscillating position of Equation 3.150,

$$
\begin{equation*}
x=A \sin [\omega t+\varphi], \tag{3.189}
\end{equation*}
$$

implies a sinusoidally oscillating velocity

$$
\begin{equation*}
\dot{x}=v_{x}=\omega A \cos [\omega t+\varphi] . \tag{3.190}
\end{equation*}
$$

Together, these imply an elliptical phase space $\left\{x[t], v_{x}[t]\right\}$ trajectory

$$
\begin{equation*}
x^{2}+\left(\frac{v_{x}}{\omega}\right)^{2}=A^{2} \tag{3.191}
\end{equation*}
$$

and a speed

$$
\begin{equation*}
\left|v_{x}\right|=\omega \sqrt{A^{2}-x^{2}} \tag{3.192}
\end{equation*}
$$

Suppose the oscillator mass $m$ spends a time $d t$ in distance $d x$ about position $x$. The probability of finding it there is inversely proportional to its speed, so

$$
\begin{equation*}
d \mathcal{P}=\rho_{c}[x] d x=N d t=N \frac{d x}{\left|v_{x}\right|}=N \frac{d x}{\omega \sqrt{A^{2}-x^{2}}} \tag{3.193}
\end{equation*}
$$

where the normalization constant $N$ is determined by the constraint

$$
\begin{equation*}
1=\int_{-A}^{A} \rho_{c}[x] d x=\int_{t_{0}}^{t_{0}+T / 2} N d t=N \frac{T}{2} \tag{3.194}
\end{equation*}
$$

Thus, $N=2 / T=\omega / \pi$, and the classical probability density is

$$
\begin{equation*}
\rho_{c}[x]=\frac{1}{\pi \sqrt{A^{2}-x^{2}}} \tag{3.195}
\end{equation*}
$$

The classical turning point coordinate $x_{n}$ corresponding to the energy $E_{n}$ is defined by

$$
\begin{equation*}
(2 n+1) E_{0}=E_{n}=V\left[x_{n}\right]=E_{0}\left(\frac{x_{n}}{x_{0}}\right)^{2} \tag{3.196}
\end{equation*}
$$

or

$$
\begin{equation*}
x_{n}=x_{0} \sqrt{2 n+1} \tag{3.197}
\end{equation*}
$$

The quantum probability density follows the classical probability density with $A=x_{n}$, as in Figure 3.31, for $n=20$.

## Quantum Problems

1. String Wave. Transverse waves on a string obey

$$
\begin{equation*}
y[x, t]=0.3 \mathrm{~m} \cos \left[1.57 \mathrm{~s}^{-1} t-6.28 \mathrm{~m}^{-1} x\right] \tag{3.198}
\end{equation*}
$$

Find the wave amplitude, period, wavelength, and speed. Which way does the wave propagate?
2. Wave Equation. By direct substitution, show that $\mathcal{E}_{x}=\mathcal{E}_{0} \sin [k z-\omega t]$ obeys the classical wave equation

$$
\begin{equation*}
\frac{\partial^{2} \mathcal{E}_{x}}{\partial z^{2}}=\frac{1}{c^{2}} \frac{\partial^{2} \mathcal{E}_{x}}{\partial t^{2}} \tag{3.199}
\end{equation*}
$$

(Hint: Section A. 4 is a quick review of partial differentiation.)
3. Null Measurements. Justify the Eq. 3.1 interaction-free measurement probability. (Hint: Add probabilities of mutually exclusive events and multiply probabilities of independent events.)
4. Inverse Quantum Zeno Effect. Consider a sequence of $n+1$ polarizers each rotated at an angle $(\pi / 2) / n$ with respect to its neighbors. Suppose a photon passes through the first polarizer.
(a) What is the probability that it passes through all the rest of the polarizers?
(b) Show that the probability of transmission increases to unity as the number of polarizers increases to infinity. Thus, a dense set of "measurements" can rotate the plane of polarization of the photon through a right angle!
5. Classical Polarization. First consider classical light propagating in the $z$-direction.
(a) Show that

$$
\begin{align*}
\overrightarrow{\mathcal{E}}_{r} & =A \frac{1}{\sqrt{2}}(\hat{x}+i \hat{y}) e^{i(k z-\omega t)}  \tag{3.200a}\\
\overrightarrow{\mathcal{E}}_{\ell} & =A \frac{1}{\sqrt{2}}(\hat{x}-i \hat{y}) e^{i(k z-\omega t)} \tag{3.200b}
\end{align*}
$$

represent circularly polarized light. (In classical optics, the real part of these expressions typically represents the light.)
(b) Show that

$$
\begin{align*}
& \overrightarrow{\mathcal{E}}_{x}=\frac{1}{\sqrt{2}}\left(\overrightarrow{\mathcal{E}}_{r}+\overrightarrow{\mathcal{E}}_{\ell}\right)  \tag{3.201a}\\
& \overrightarrow{\mathcal{E}}_{y}=\frac{-i}{\sqrt{2}}\left(\overrightarrow{\mathcal{E}}_{r}-\overrightarrow{\mathcal{E}}_{\ell}\right) \tag{3.201b}
\end{align*}
$$

represent linearly polarized light.
(c) Show that the two kinds of polarizations are also related by

$$
\begin{align*}
\overrightarrow{\mathcal{E}}_{r} & =\frac{1}{\sqrt{2}}\left(\overrightarrow{\mathcal{E}}_{x}+i \overrightarrow{\mathcal{E}}_{y}\right),  \tag{3.202a}\\
\overrightarrow{\mathcal{E}}_{\ell} & =\frac{1}{\sqrt{2}}\left(\overrightarrow{\mathcal{E}}_{x}-i \overrightarrow{\mathcal{E}}_{y}\right) . \tag{3.202b}
\end{align*}
$$

6. Photon Polarization. In order to correspond with classical light, assume that the circularly polarized photons are superpositions of linearly polarized photons and that their states related by

$$
\begin{align*}
|r\rangle & =\frac{1}{\sqrt{2}}(|x\rangle+i|y\rangle)  \tag{3.203a}\\
|\ell\rangle & =\frac{1}{\sqrt{2}}(|x\rangle-i|y\rangle) \tag{3.203b}
\end{align*}
$$

(a) Show that

$$
\begin{align*}
|x\rangle & =\frac{1}{\sqrt{2}}(|r\rangle+|\ell\rangle)  \tag{3.204a}\\
|y\rangle & =\frac{-i}{\sqrt{2}}(|r\rangle-|\ell\rangle) \tag{3.204b}
\end{align*}
$$

(b) Rotate the linear coordinate system through an angle $\theta$ and justifiy

$$
\begin{align*}
\left|x^{\prime}\right\rangle & =+\cos \theta|x\rangle+\sin \theta|y\rangle  \tag{3.205a}\\
\left|y^{\prime}\right\rangle & =-\sin \theta|x\rangle+\cos \theta|y\rangle \tag{3.205b}
\end{align*}
$$

(c) Show that the rotated circular polarizations satisfy

$$
\begin{align*}
\left|r^{\prime}\right\rangle & =e^{-i \theta}|r\rangle  \tag{3.206a}\\
\left|\ell^{\prime}\right\rangle & =e^{+i \theta}|\ell\rangle \tag{3.206b}
\end{align*}
$$

(d) Show then that the probability of measuring a circularly polarized photon to have a particular linear polarization is the same at any angle. (Hint: Expand $\left|r^{\prime}\right\rangle$ and $\left|\ell^{\prime}\right\rangle$ in terms of $|x\rangle$ and $|y\rangle$.)
7. Gaussian Integrals. Gaussian integrals occur frequently in quantum physics (and probability and statistics).
(a) Show that

$$
\begin{equation*}
I_{0}[1]=\int_{-\infty}^{\infty} d x e^{-x^{2}}=\sqrt{\pi} \tag{3.207}
\end{equation*}
$$

using the the method first employed by Laplace in 1778: First compute its square

$$
\begin{equation*}
I_{0}[1]^{2}=\int_{-\infty}^{\infty} d x e^{-x^{2}} \int_{-\infty}^{\infty} d y e^{-y^{2}}=\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d x d y e^{-\left(x^{2}+y^{2}\right)} \tag{3.208}
\end{equation*}
$$

by converting the two-dimensional integral to polar coordinates. Why should we replace the area element $d x d y$ by $r d r d \theta$ ?
(b) Show that

$$
\begin{equation*}
I_{0}[a]=\int_{-\infty}^{\infty} d x e^{-a x^{2}}=\sqrt{\frac{\pi}{a}} \tag{3.209}
\end{equation*}
$$

where $a>0$, by scaling $x$ and using $I_{0}[1]$.
(c) Show that

$$
\begin{equation*}
I_{1}[a]=\int_{-\infty}^{\infty} d x x e^{-a x^{2}}=0 \tag{3.210}
\end{equation*}
$$

using symmetry.
(d) Show that

$$
\begin{equation*}
I_{2}[a]=\int_{-\infty}^{\infty} d x x^{2} e^{-a x^{2}}=\frac{1}{2 a} \sqrt{\frac{\pi}{a}} \tag{3.211}
\end{equation*}
$$

by differentiating $I_{0}[a]$ with respect to the parameter $a$.
(e) If the parameter $a=i|a|$ is imaginary, the Gaussian integrals do not converge. However, in quantum physics, one often defines such integrals by replacing $i|a|$ with $(i-\epsilon)|a|$ and letting the small positive convergence factor $\epsilon \rightarrow 0$ after the integration. Use a convergence factor to evaluate

$$
\begin{equation*}
\int_{-\infty}^{\infty} d x e^{b x^{2}} \tag{3.212}
\end{equation*}
$$

where $b=i|b|$ is imaginary.
8. Qualitative Stationary States. For each of the below 9 potential energy functions $V[x]$, carefully sketch the requested stationary state wave functions $\psi_{E}[x]$. The horizontal dashed lines suggest the appropriate energy levels. $\psi_{n}[x]$ corresponds to the $n$th energy level $E_{n}$, and $n=1$ corresponds to the lowest (ground) state. Check the number of nodes, relative amplitudes, local wavelengths, and decay constants. Relate each wave function to the potential function by using a common $x$-axis. (Hint: For the last two cases, imagine pushing the boxes together.)



Bouncing Ball

$\mathrm{NH}_{3}$ Molecule




Nonrigid Box


Adjacent Boxes


Small \& Large

9. Rigid Box Stationary States. Consider a particle confined to a (onedimensional) box by an infinite square potential energy well of width $L$.
(a) Show that the $n$th definite-energy stationary-state wave function can be written as $\psi_{n}[x]=N \sin k_{n} x$ inside the box. What is $k_{n}$ ? What is $\psi_{n}[x]$ outside the box?
(b) Find the normalization constant $N$ by requiring that $\left|\psi_{n}\right|^{2}$ bound a unit area.
(c) If the particle is in its ground state, calculate the probability that it will be found in the middle half of the well. Compare this with the corresponding classical probability.
(d) Show that the probability for a particle in the $n$th state to be in the middle half of the well approaches $1 / 2$ in the limit as $n \rightarrow \infty$.
10. Rigid Box Dynamic State. The initial state of a particle in a rigid box box is a linear superposition of first and second eigenfunctions

$$
\begin{equation*}
\psi[x]=\sqrt{\frac{1}{3}} \psi_{1}[x]-\sqrt{\frac{2}{3}} \psi_{2}[x] \tag{3.213}
\end{equation*}
$$

with probability $\mathcal{P}_{n}=\int d x \psi^{*} \psi_{n}$ to be in state $\psi_{n}$ with energy $E_{n}$.
(a) Is $\psi[x]$ normalized? Don't assume that the the eigenfunctions are orthonormal.
(b) What is the state $\Psi[x, t]$ at a later time?
(c) What is the probability amplitude to again observe the initial state at a later time, assuming the eigenfunctions are orthonormal? (Hint: Compute the projection $\langle\Psi \mid \psi\rangle=\int d x \Psi^{*} \psi$ using $\langle n \mid n\rangle=\int d x \psi_{n}^{*} \psi_{n}=$ 1 and $\left.\langle 1 \mid 2\rangle=\int d x \psi_{1}^{*} \psi_{2}=0.\right)$
(d) What is the corresponding probability? (Hint: Express $\mathcal{P}=|\langle\Psi \mid \psi\rangle|^{2}$ using a cosine function, where $\mathcal{P}=1$ at $t=0$.)
11. Spring Dynamic State. The initial state of a simple harmonic oscillator is a linear superposition of first and second eigenfunctions

$$
\begin{equation*}
\psi[x]=\sqrt{\frac{1}{5}} \psi_{0}[x]+\sqrt{\frac{4}{5}} \psi_{1}[x] \tag{3.214}
\end{equation*}
$$

with probability $\mathcal{P}_{n}=\int d x \psi^{*} \psi_{n}$ to be in state $\psi_{n}$ with energy $E_{n}$.
(a) Check that $\psi[x]$ is normalized without assuming the eigenfunctions are orthonormal. (Hint: For the Gaussian integrals, check Problem 7.)
(b) Write the state $\Psi[x, t]$ at a later time.
(c) What is the probability amplitude to now observe the initial state, assuming the eigenfunctions are orthonormal?
(d) What is the corresponding probability?
12. Big Mass \& Spring. A spring of stiffness $k=0.1 \mathrm{~N} / \mathrm{m}$ confines a (macroscopic) particle of mass $m=0.01 \mathrm{~kg}$.
(a) What is the spacing $\Delta E$ between its quantized energy levels? Would this be easy to detect experimentally?
(b) Displace the mass 1 cm from equilibrium and release it from rest. What is the (approximate) quantum number $n$ of this state?

## Appendix A

## Mathematics Background

## A. 1 Complex Numbers

Complex numbers are used extensively in quantum mechanics. They also enable beautiful theorems in mathematics, like the Fundamental Theorem of Algebra, which says that an $n$th degree polynomial has exactly $n$ complex roots

$$
\begin{equation*}
z=x+i y \tag{A.1}
\end{equation*}
$$

where $x$ and $y$ are real numbers and the imaginary unit

$$
\begin{equation*}
i=\sqrt{-1} \tag{A.2}
\end{equation*}
$$

A common operation is complex conjugation

$$
\begin{equation*}
z^{*}=x-i y=\bar{z} . \tag{A.3}
\end{equation*}
$$

The real and imaginary parts of a complex number,

$$
\begin{equation*}
\operatorname{Re} z=\frac{z+z^{*}}{2}=x \tag{A.4}
\end{equation*}
$$

and

$$
\begin{equation*}
\operatorname{Im} z=\frac{z-z^{*}}{2 i}=y \tag{A.5}
\end{equation*}
$$

are both real. The modulus

$$
\begin{equation*}
\bmod z=|z|=\sqrt{z^{*} z}=\sqrt{z z^{*}}=\sqrt{x^{2}+y^{2}} \tag{A.6}
\end{equation*}
$$

and argument

$$
\begin{equation*}
\arg z=\operatorname{atan}\left[\frac{y}{x}\right] \tag{A.7}
\end{equation*}
$$

offer an alternate way of specifying the complex number, as in Figure A.1.
Euler's theorem

$$
\begin{equation*}
e^{i \theta}=\cos \theta+i \sin \theta, \tag{A.8}
\end{equation*}
$$



Figure A.1: The complex plane.
which can be proved by expanding each term in a Taylor series, allows us to interconvert the polar and rectangular representations of a complex number by

$$
\begin{equation*}
r e^{i \theta}=r \cos \theta+i r \sin \theta=x+i y \tag{A.9}
\end{equation*}
$$

where $r=\bmod z$ and $\theta=\arg z$. A special case of Euler's theorem, $\theta=\pi$, generates the remarkable formula (purportedly engraved on Euler's tombstone)

$$
\begin{equation*}
e^{i \pi}+1=0 \tag{A.10}
\end{equation*}
$$

which elegantly and surprisingly interconnects the base of the natural logarithms, the imaginary unit, the ratio of a circle's circumference to its diameter, unity, and zero!

## A. 2 Hyperbolic Functions

Hyperbolic functions are intimately related to trigonometric functions. Recall Euler's theorem

$$
\begin{equation*}
e^{i \theta}=\cos \theta+i \sin \theta \tag{A.11}
\end{equation*}
$$

and its complex conjugate

$$
\begin{equation*}
e^{-i \theta}=\cos \theta-i \sin \theta \tag{A.12}
\end{equation*}
$$

Adding and subtracting implies

$$
\begin{equation*}
\cos \theta=\frac{e^{i \theta}+e^{-i \theta}}{2} \tag{A.13}
\end{equation*}
$$

and

$$
\begin{equation*}
\sin \theta=\frac{e^{i \theta}-e^{-i \theta}}{2 i} \tag{A.14}
\end{equation*}
$$

The substitution $\theta \rightarrow i \theta$ replaces a real angle with an imaginary angle and generates hyperbolic functions from trigonometric functions. For example,

$$
\begin{equation*}
\cos [i \theta]=\frac{e^{-\theta}+e^{\theta}}{2}=\cosh \theta \tag{A.15}
\end{equation*}
$$

and

$$
\begin{equation*}
\sin [i \theta]=\frac{e^{-\theta}-e^{\theta}}{2 i}=i \frac{e^{\theta}-e^{-\theta}}{2}=i \sinh \theta \tag{A.16}
\end{equation*}
$$

Hence,

$$
\begin{equation*}
\cosh [i \theta]=\frac{e^{-i \theta}+e^{i \theta}}{2}=\cos \theta \tag{A.17}
\end{equation*}
$$

and

$$
\begin{equation*}
\sinh [i \theta]=\frac{e^{-i \theta}-e^{i \theta}}{2 i}=i \frac{e^{i \theta}-e^{-i \theta}}{2}=i \sin \theta \tag{A.18}
\end{equation*}
$$

Notice how the cosine "swallows" the $i$ when becoming a hyperbolic cosine, while the sine "spits out" the $i$ when becoming a hyperbolic sine. (Similarly, the cosine swallows a minus sign, $\cos [-\theta]=\cos \theta$, while the sine spits out a minus sign, $\sin [-\theta]=-\sin [\theta]$.)


Figure A.2: Graphs of hyperbolic functions, with $\tanh \theta=\sinh \theta / \cosh \theta$.
Every trigonometric identity corresponds to a hyperbolic identity. For example, take $(\cos \theta)^{2}+(\sin \theta)^{2}=1$ and substitute $\theta \rightarrow i \theta$ to get $(\cos [i \theta])^{2}+$ $(\sin [i \theta])^{2}=1$ or

$$
\begin{equation*}
(\cosh \theta)^{2}-(\sinh \theta)^{2}=1 \tag{A.19}
\end{equation*}
$$

The hyperbolic functions are real, exponential, and nonrepeating functions, as depicted in Figure A.2.

## A. 3 Spatial Rotations

Spatial rotations are analogues for Lorentz-Einstein transformations. Suppose an $\left(x^{\prime}, y^{\prime}\right)$ coordinate system is rotated counterclockwise through an angle $\theta$ relative to an $(x, y)$ coordinate system, as in Figure A.3.



Figure A.3: Two coordinate systems with a common origin but rotated through an angle $\theta$ relative to each other.

From the trigonometry, the $y$-coordinate can be expressed as

$$
\begin{equation*}
y=\frac{y^{\prime}}{\cos \theta}+x \tan \theta \tag{A.20}
\end{equation*}
$$

which implies

$$
\begin{equation*}
y^{\prime}=-x \sin \theta+y \cos \theta \tag{A.21}
\end{equation*}
$$

Similarly, the $x^{\prime}$-coordinate can be expressed as

$$
\begin{equation*}
x^{\prime}=\frac{x}{\cos \theta}+y^{\prime} \tan \theta \tag{A.22}
\end{equation*}
$$

which implies

$$
\begin{equation*}
x=x^{\prime} \cos \theta-y^{\prime} \sin \theta \tag{A.23}
\end{equation*}
$$

or, simultaneously negating $\theta$ and interchanging primes and un-primes,

$$
\begin{equation*}
x^{\prime}=x \cos \theta+y \sin \theta \tag{A.24}
\end{equation*}
$$

We can summarize the rotation transformation of Equation A. 24 and Equation A. 21 in the matrix equation

$$
\begin{array}{|c|}
\hline x^{\prime}  \tag{A.25}\\
y^{\prime} \\
\hline
\end{array}=\begin{array}{|cc|}
\hline \cos \theta & \sin \theta \\
-\sin \theta & \cos \theta \\
\hline
\end{array} \begin{aligned}
& x \\
& y
\end{aligned}
$$

which readily checks for $\theta=0$ and $\theta=\pi / 2$.
In terms of the relative slope $s=\tan \theta, \cos \theta=1 / \sqrt{1+s^{2}}=\Gamma$ and $\sin \theta=$ $s / \sqrt{1+s^{2}}=s \Gamma$, and hence

$$
\left.\left.\begin{array}{|c|}
\hline x^{\prime}  \tag{A.26}\\
y^{\prime}
\end{array}=\begin{array}{|cc|}
\hline \Gamma & s \Gamma \\
-s \Gamma & \Gamma
\end{array} \begin{array}{l}
x \\
y
\end{array}\right]=\Gamma \begin{array}{|cc|}
\hline 1 & s \\
-s & 1
\end{array}\right] \begin{aligned}
& x \\
& y
\end{aligned} .
$$

## A. 4 Partial Derivatives

In one-dimension, the function

$$
\begin{equation*}
f[x]=3 x^{2}+1 \tag{A.27}
\end{equation*}
$$

has the derivative

$$
\begin{equation*}
\frac{d f}{d x}=6 x+0=6 x \tag{A.28}
\end{equation*}
$$

In two dimensions, the function

$$
\begin{equation*}
f[x, y]=3 x y^{2}+2 x+3 y+2 \tag{A.29}
\end{equation*}
$$

has the partial derivatives

$$
\begin{align*}
& \frac{\partial f}{\partial x}=3 y^{2}+2+0+0=3 y^{2}+2  \tag{A.30a}\\
& \frac{\partial f}{\partial y}=6 x y+0+3+2=6 x y+5 \tag{A.30b}
\end{align*}
$$

which are just like ordinary derivatives, but with other variables held constant.

## A. 5 Function Notation

Standard mathematics notation suffers a serious ambiguity involving parentheses. In particular, parentheses can be used to denote multiplication, as in $a(b+c)=a b+a c$ and $f(g)=f g$, or they can be used to denote a function evaluated at a point, as in $f(t)$ and $g(b+c)$. One must sometimes struggle to determine the intended meaning from context.

In these notes, to avoid ambiguity, round parentheses ( $\bullet$ ) always denote multiplication, while square brackets $[\bullet]$ always denote function evaluation. Thus, $f[x]$ denotes a function evaluated at a point, while $a(b)=a b$ denotes the product of two quantities. The Wolfram Language and Mathematica employ the same convention.

## Problems

1. Complex Plotting. Plot the following numbers and their complex conjugates in the complex $z=x+i y=\{x, y\}$ plane.
(a) $1+i$
(b) $1-i \sqrt{3}$
(c) $\sqrt{2} e^{-i \pi / 4}$
2. Complex Simplification. Simplify the following numbers to the form $x+i y$.
(a) $\frac{1}{1+i}$
(b) $25 e^{2 i}$
(c) $\frac{3 i-7}{i+4}$
(d) $\left(\frac{1+i}{1-i}\right)^{137}$ (Hint: Don't use a calculator!)
(e) $i^{i}$ (Hint: Find the principal value.)
3. Complex Identities. Derive the following equations.
(a) $e^{i \theta}=\cos \theta+i \sin \theta$ (Hint: Try infinite power series expansion.)
(b) $e^{i \pi}+1=0$
(c) $\cos \theta=\frac{e^{i \theta}+e^{-i \theta}}{2}$
(d) $\sin \theta=\frac{e^{i \theta}-e^{-i \theta}}{2 i}$

## Bibliography

[1] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. Physics Physique Fizika, 1:195-200, Nov 1964.
[2] A. Einstein. Ist die trgheit eines krpers von seinem energieinhalt abhngig? [Does the inertia of a body depend on its energy content?]. Annalen der Physik, 323(13):639-641, 1905.
[3] A. Einstein. Zur Elektrodynamik bewegter Körper [On the Electrodynamics of Moving Bodies]. Annalen der Physik, 322(10):891-921, January 1905.
[4] A. Einstein. Das prinzip von der erhaltung der schwerpunktsbewegung und die trgheit der energie [The principle of the conservation of the center of gravity movement and the inertia of energy]. Annalen der Physik, 20:627633, 1906.
[5] R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. Rev. Mod. Phys., 20:367-387, Apr 1948.
[6] H. A. Lorentz. Electromagnetic phenomena in a system moving with any velocity smaller than that of light. Proceedings Acad. Sci. Amsterdam, 6:809831, 1904.
[7] E. Schrödinger. Quantisierung als eigenwertproblem [Quantization as an Eigenvalue Problem]. Annalen der Physik, 384(4):361-376, 1926.

