A Appendix A Conventions and Units

Conventions and units should be one's friend in physics, in that they should both make an analysis more transparent and they should nudge people away from, rather than towards, common mistakes. They are also often acquired without much thought as one grows up.

This section is meant to explain the ones used in this book. The focus here is to list conventional choices for aficionados, with little effort made explaining the field theories involved. A reader seeking more detailed (though still cursory) background information should try their luck with Appendix B (for the quantum mechanics of scattering) and Appendix C (for quantum field theory).

A.1 Fundamental units

It is common to use specific units adapted to specific problems so that numerical values are not too far from one (such as using the Angstrom - or Rydberg - for atomic electrons, fm for nuclear processes, astronomical units for the solar system or megaparsecs in cosmology). Such choices are mostly not made here, since one of the points of this book is to emphasize the broad utility of EFT methods many different areas in physics.

Instead, this book uses *fundamental units*, for which the fundamental constants $\hbar$, c and k_B (Planck's constant, the speed of light and Boltzmann's constant) equal 1. For instance $c = 1$ is ensured by measuring time and distances both in seconds (where a second of distance means a light-second; the distance light travels in a second). Similarly, $\hbar = 1$ if $(\text{energy})^{-1}$ and time are both measured in seconds — where an inverse-second of energy means the amount $\hbar/(1\text{sec}) = 6.58211 \times 10^{-16}$ eV — and so on.

In this book usually the basic unit is taken to be energy, given in eV or multiples thereof. The utility of this choice is that the proton and neutron rest masses in these units are (respectively) 0.938 GeV and 0.940 GeV (which is to say, the energy tied up in the rest mass of a nucleon is just shy of 1 GeV). This is useful because once told that the mass of the earth is $M_\oplus \simeq 3.35 \times 10^{51}$ GeV you also know roughly how many nucleons are in it, since the biggest contributor to an object's mass usually comes from the mass of each nucleon residing in its constituent nuclei.

Fundamental units have the very useful benefit of boiling equations down to relations between physical quantities without cluttering them up with symbols purely to do with units. This is a particularly good virtue when identifying which scales are relevant to any given problem, as is central to the utility of EFT methods. Electromagnetic units are set by using the proton charge e as the unit of charge rather than the Coulomb.

Ordinary units may always be retrieved by putting in any missing factors of $\hbar$, c or k_B as required by dimensional analysis. Useful rules of thumb for this purpose are:

$$1 \text{ fm} \simeq (0.2 \text{ GeV})^{-1} \simeq 3 \times 10^{-24} \text{ sec} \quad \text{and} \quad 1 \text{ K} \simeq 9 \times 10^{-5} \text{ eV}\,. \tag{A.1}$$

The conversions of other units into powers of eV and to powers of metres are given below.

Length and Time

$1/M_p\ (= G_N/\hbar c)^{\frac{1}{2}}$	=	8.1897×10^{-29}	c^2/eV	=	1.6161×10^{-35}	m$c/\hbar$
$1/m_p$	=	1.0658×10^{-9}	c^2/eV	=	2.1031×10^{-16}	m$c/\hbar$
1 fm	=	5.06773×10^{-9}	$\hbar c$/eV	=	10^{-15}	m
$1/m_e$	=	1.957×10^{-6}	c^2/eV	=	3.8616×10^{-13}	m$c/\hbar$
$a_0\ (= 1/\alpha m_e)$	=	2.6818×10^{-4}	c^2/eV	=	5.2918×10^{-11}	m$c/\hbar$
1 A	=	5.06773×10^{-4}	$\hbar c$/eV	=	10^{-10}	m
1 nm	=	5.06773×10^{-3}	$\hbar c$/eV	=	10^{-9}	m
1 μm	=	5.06773	$\hbar c$/eV	=	10^{-6}	m
1 cm	=	5.06773×10^{4}	$\hbar c$/eV	=	0.01	m
1 m	=	5.06773×10^{6}	$\hbar c$/eV	=	1	m
1 km	=	5.06773×10^{9}	$\hbar c$/eV	=	10^{3}	m
1 sec	=	1.51927×10^{15}	$\hbar$/eV	=	2.99792×10^{8}	m/c
1 min	=	9.11562×10^{16}	$\hbar$/eV	=	1.79875×10^{10}	m/c
1 hr	=	5.46937×10^{18}	$\hbar$/eV	=	1.07925×10^{12}	m/c
1 day	=	1.31265×10^{20}	$\hbar$/eV	=	2.59020×10^{13}	m/c
1 yr	=	4.795×10^{22}	$\hbar$/eV	=	9.461×10^{15}	m/c
1 pc	=	1.564×10^{23}	$\hbar c$/eV	=	3.08568×10^{16}	m
1 kpc	=	1.564×10^{26}	$\hbar c$/eV	=	3.08568×10^{19}	m
1 Mpc	=	1.564×10^{29}	$\hbar c$/eV	=	3.08568×10^{22}	m

Microscopic Energy and Mass

1 eV	=	10^{-9}	GeV	=	5.06773×10^{6}	$\hbar c/\mathrm{m}$
1 keV	=	10^{-6}	GeV	=	5.06773×10^{9}	$\hbar c/\mathrm{m}$
1 MeV	=	10^{-3}	GeV	=	5.06773×10^{12}	$\hbar c/\mathrm{m}$
1 GeV	=	1	GeV	=	5.06773×10^{15}	$\hbar c/\mathrm{m}$
αm_e	=	3.7289×10^{-6}	GeV/c^2	=	1.8897×10^{10}	$\hbar/\mathrm{m}c$
m_e	=	5.10999×10^{-4}	GeV/c^2	=	2.5896×10^{12}	$\hbar/\mathrm{m}c$
	=	9.10939×10^{-28}	g			
m_p	=	0.938272	GeV/c^2	=	4.75491×10^{15}	$\hbar/\mathrm{m}c$
	=	1.67262×10^{-24}	g			
	=	1.83615×10^{3}	m_e			
$M_p = (\hbar c/G_N)^{\frac{1}{2}}$	=	1.22105×10^{19}	GeV/c^2	=	6.1879×10^{34}	$\hbar/\mathrm{m}c$
	=	2.17671×10^{-5}	g			
	=	1.30138×10^{19}	m_p			
$\hat{M}_p = (\hbar c/8\pi G_N)^{\frac{1}{2}}$	=	2.43564×10^{18}	GeV/c^2	=	1.23431×10^{34}	$\hbar/\mathrm{m}c$
	=	4.34191×10^{-6}	g			
	=	2.59588×10^{18}	m_p			

Ordinary Units Expressed Microscopically

1 g	=	5.60959×10^{23}	GeV/c^2	=	2.84279×10^{39}	$\hbar/\mathrm{m}c$
1 kg	=	5.60959×10^{26}	GeV/c^2	=	2.84279×10^{42}	$\hbar/\mathrm{m}c$
1 Joule = 1 kg $\mathrm{m}^2/\mathrm{s}^2$	=	6.24151×10^{9}	GeV	=	3.16303×10^{25}	$\hbar c/\mathrm{m}$
1 erg = 1 g $\mathrm{cm}^2/\mathrm{s}^2$	=	6.24151×10^{2}	GeV	=	3.16303×10^{18}	$\hbar c/\mathrm{m}$
$= 10^{-7}$ J						
1 Newton = 1 kg m/s^2	=	1.23162×10^{-6}	$\mathrm{GeV}^2/\hbar c$	=	3.16303×10^{25}	$\hbar c/\mathrm{m}^2$
	=	1.23162×10^{12}	$\mathrm{eV}^2/\hbar c$			
1 dyne = 1 g cm/s^2	=	1.23162×10^{-11}	$\mathrm{GeV}^2/\hbar c$	=	3.16303×10^{20}	$\hbar c/\mathrm{m}^2$
$= 10^{-5}$ N	=	1.23162×10^{7}	$\mathrm{eV}^2/\hbar c$			
1 Watt = 1 J/s	=	4.10824×10^{-15}	$\mathrm{GeV}^2/\hbar$	=	1.05507×10^{17}	$\hbar c^2/\mathrm{m}^2$
	=	4.10824×10^{3}	$\mathrm{eV}^2/\hbar$			
1 Hz = 1/s	=	6.5821×10^{-25}	$\mathrm{GeV}/\hbar$	=	3.3356×10^{-9}	c/m
1 Kelvin	=	8.61742×10^{-14}	GeV/k_B	=	4.36707×10^{2}	$\hbar c/\mathrm{m}k_B$
	=	8.61742×10^{-5}	eV/k_B	=	1/11604.4	eV/k_B

Electromagnetic Units

1 Coulomb	=	6.24151×10^{18}	e			
1 Volt = 1 J/C	=	1	eV/e	=	5.06773×10^{6}	$\hbar c/\mathrm{m}e$
	=	10^{-9}	GeV/e			
1 Farad = 1 C/V	=	6.24151×10^{18}	e^2/eV	=	1.23162×10^{12}	$\mathrm{m}e^2/\hbar c$
1 Ampere = 1 C/s	=	4.10824×10^{3}	$\mathrm{eV}e/\hbar$	=	2.08194×10^{10}	ec/m
1 Ohm = 1 V/A	=	2.43413×10^{-4}	$\hbar/e^2$			
1 Mho = 1/Ohm	=	4.10824×10^{3}	$e^2/\hbar$			
1 Weber = 1 V s	=	1.51927×10^{15}	$\hbar/e$			
1 Tesla = 1 Weber/m^2	=	59.1572	$\mathrm{eV}^2/\hbar ec^2$	=	1.51927×10^{15}	$\hbar/e\mathrm{m}^2$
1 Gauss = 10^{-4} Tesla	=	5.91572×10^{-3}	$\mathrm{eV}^2/\hbar ec^2$	=	1.51927×10^{11}	$\hbar/e\mathrm{m}^2$
$\phi_0 = 2\pi\hbar/e$	=	6.28319	$\hbar/e$	=	4.13567×10^{-15}	Weber
				=	$1/(2.418 \times 10^{14}\ )$	Weber
$\epsilon_0 = 8.854 \times 10^{-12}$ F/m	=	10.905	$e^2/\hbar c$			
$\mu_0 = 4\pi \times 10^{-7}$ N/A^2	=	0.0917012	$\hbar/ce^2$		$\epsilon_0\mu_0 = 1/c^2$	
$\alpha = e^2/(4\pi\epsilon_0\hbar c)$	=	7.2974×10^{-3}			$1/\alpha = 137.036$	

In these tables m_e denotes the electron mass, m_p is the proton mass, α is the electromagnetic fine-structure constant (evaluated at low energies, $\mu \sim m_e$).

A.2 Conventions

Like religion and politics, conventions are a subject normally avoided in polite company for fear of provoking strong words or fisticuffs. Any practicing physicist should usually adopt a set of conventions and stick to them, and (as is the case for many) the ones used here are largely the ones I learned as a student. (Because of this they usually agree with those used in Steven Weinberg's many textbooks.) This section explains my rationale for the main choices made.

A.2.1 Geometrical conventions

The conventions for vectors are such that Greek indices represent spacetime coordinates in 3+1 dimensions, with $x^\mu = \{x^0, x^1, x^2, x^3\} = \{t, x, y, z\}$ a contravariant vector built from Cartesian coordinates. Spatial indices are denoted by latin letters, such as $x^a = \{x, y, z\}$ or $x^i = \{x, y, z\}$, with letters chosen early or later in the alphabet in a way that distinguishes them from any other indices present (such as those describing internal symmetries, or space-time spinors, *etc.*).

The Einstein summation convention is used throughout the book, unless explicitly stated otherwise. In this convention any repeated appearance of an index represents a summation of that index over its entire range. So $a^\mu a_\mu = a^0 a_0 + a^1 a_1 + a^2 a_2 + a^3 a_3$ while $a^i a_i = a^1 a_1 + a^2 a_2 + a^3 a_3$, and so on.

Metric conventions

The spacetime metric is denoted $g_{\mu\nu}(x)$ and defines the invariant line-element by $\mathrm{d}s^2 = g_{\mu\nu}(x)\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$, that gives the square of the distance $\mathrm{d}s$ between two infinitesimally separated points: x^μ and $x^\mu + \mathrm{d}x^\mu$. The signature of the metric is $(-+++)$, so the Minkowski metric that describes the flat space of special relativity in Cartesian coordinates is given explicitly by

$$\mathrm{d}s^2 = \eta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\mathrm{d}t^2 + \delta_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\,. \tag{A.2}$$

This is one of the choices that generates the most heat when discussed, since half the world learns this choice (often called the 'east-coast' or 'mostly plus' or 'right' metric) while the other half adopts the opposite sign for $\eta_{\mu\nu}$ (called the 'west-coast' or 'mostly minus' or 'wrong' metric). Normally much heat (and not much light) is spent on whether it is more sensible for time intervals or space intervals to be negative. With (A.2) time-like vectors have negative length, while vectors in the three space-like directions have positive length.

A more compelling reason for using the convention (A.2) comes once Wick rotations are made to Euclidean space, such as often is done when discussing thermal systems (for which temperature can often be conveniently regarded as periodicity in imaginary time – see *e.g.* §A.2.2). In this case $\tau = \mathrm{i}t$ and so $\mathrm{d}t^2 = -\mathrm{d}\tau^2$. With the above choice the metric becomes positive definite, as do the lengths of all vectors, like $a^2 := \eta_{\mu\nu}\,a^\mu a^\nu$. With the 'mostly-minus' metric convention all such squares become negative when Euclideanized (and when quantities like a^2 are negative it can be a nightmare finding sign mistakes). As mentioned earlier, your conventions should be your friend, and should nudge you towards making fewer errors rather than more errors.

Since notation is part of language, part of the thinking behind metric conventions is also the practice of the community with which one wishes to communicate. Broadly speaking most relativists, cosmologists and string theorists use the $(-+++)$ metric used here, while particle physicists are more split, though with a majority of phenomenologists using mostly-minus conventions.

With the above metric choice the action for scalars and gauge bosons (see next Appendix) have negative coefficients, since this is required to have positive kinetic energies. That is

$$\mathfrak{L} = -\frac{1}{2}\,\eta^{\mu\nu}\partial_\mu\phi\,\partial_\nu\phi = \frac{1}{2}\left[(\partial_t\phi)^2 - (\nabla\phi)^2\right] \tag{A.3}$$

while

$$\mathfrak{L} = -\frac{1}{4}\,F_{\mu\nu}F^{\mu\nu} = \frac{1}{2}\left(\mathbf{E}^2 - \mathbf{B}^2\right)\,. \tag{A.4}$$

Curvature conventions

A natural convention is to define the curvature so that the same sign also applies to the action for the metric in General relativity, which is given – see, for example, (10.1) – by

$$\frac{\mathfrak{L}}{\sqrt{-g}} = -\frac{R}{16\pi G_N}\,. \tag{A.5}$$

This is ensured if one adopts the curvature conventions $R := g^{\mu\nu}R_{\mu\nu}$ with Ricci tensor defined by $R_{\mu\nu} := R^{\lambda}{}_{\mu\lambda\nu}$ and Riemann curvature tensor given by

$$R^{\mu}{}_{\nu\lambda\rho} = \partial_\rho \Gamma^{\mu}_{\nu\lambda} + \Gamma^{\mu}_{\rho\sigma}\Gamma^{\sigma}_{\nu\lambda} - (\rho \leftrightarrow \lambda)\,. \tag{A.6}$$

Here

$$\Gamma^{\mu}_{\nu\lambda} = \frac{1}{2} g^{\mu\alpha}[\partial_\nu g_{\alpha\lambda} + \partial_\lambda g_{\alpha\nu} - \partial_\alpha g_{\nu\lambda}] \tag{A.7}$$

is the Christoffel symbol (of the second kind) built from derivatives of the metric, and its inverse $g^{\mu\nu}$ defined by $g^{\mu\nu}g_{\nu\lambda} = \delta^{\mu}_{\lambda}$.

The above definitions are the same as used in the well-known book [400], and are also almost the same as a very popular choice (often called the 'MTW' – or 'geometrical' – choice, with MTW representing the authors Misner, Thorne and Wheeler, of an influential relativity textbook [399]). 'Almost the same' here means the only difference relative to MTW conventions is the overall sign of the definition (A.6). The motivation for the MTW choice is that it gives a positive curvature for spheres in euclidean space (and negative curvatures to hyperbolae), though at the expense of introducing an unusual gravity-specific sign in the action.

Levi-Civita conventions

Finally, another useful geometrical tensor (in four spacetime dimensions) is the four-index Levi-Civita completely antisymmetric tensor $\epsilon_{\mu\nu\lambda\rho}$. In flat space this is defined to be completely antisymmetric under the interchange of any two indices (and so to vanish whenever two indices take the same values); to have elements ± 1 when all indices are different. The convention used here takes $\epsilon^{0123} = +1$ and then all other components are dictated by the antisymmetry condition.

On curved space it is worth working with a *vierbein* (or tetrad): defined as a basis of four vector fields, $e^{a}{}_{\mu}(x)$, with $a = 0, 1, 2, 3$. The basis is chosen to be orthonormal and complete in the sense that

$$g^{\mu\nu}e^{a}{}_{\mu}e^{b}{}_{\nu} = \eta^{ab} \quad \text{and} \quad \eta_{ab}\, e^{a}{}_{\mu}e^{b}{}_{\nu} = g_{\mu\nu}\,, \tag{A.8}$$

with the Einstein summation convention in full force, and where η^{ab} is the signature-$(-+++)$ Minkowski tensor, and $g^{\mu\nu}$ is the inverse of the spacetime metric $g_{\mu\nu}$, so $g^{\mu\nu}g_{\nu\lambda} = \delta^{\mu}_{\lambda}$. Evidently the matrix $e^{a}{}_{\mu}$ is morally the square root of the metric $g_{\mu\nu}$.

The above orthogonality and completeness relations allow the definition of the inverse $e_{a}{}^{\mu}$, defined to satisfy $e_{a}{}^{\mu}e^{b}{}_{\mu} = \delta^{b}_{a}$ and $e_{a}{}^{\mu}e^{a}{}_{\nu} = \delta^{\mu}_{\nu}$. Any tensor can then be described by its world-index components, like $T_{\mu\nu}$, or its tangent-frame components, like $T_{ab} = e_{a}{}^{\mu}e_{b}{}^{\nu}T_{\mu\nu}$, and so on. The basis vectors in the vierbein are not unique, with the freedom to do local Lorentz transformations, $e^{a}{}_{\mu} \to \Lambda^{a}{}_{b}e^{b}{}_{\mu}$ where the position-dependent matrices $\Lambda^{a}{}_{b}$ satisfy the Lorentz-group definition: $\eta_{ac}\Lambda^{a}{}_{b}\Lambda^{c}{}_{d} = \eta_{bd}$. These definitions ensure it is consistent to raise and lower indices with either η_{ab} or $g_{\mu\nu}$ (or their inverses) in arbitrary order, so $T^{ab} = e^{a}{}_{\mu}e^{b}{}_{\nu}T^{\mu\nu} = \eta^{ac}\eta^{bd}T_{cd}$ and so on.

With these definitions in mind the Minkowski flat-space conventions for the Levi-Civita tensor apply to the tangent-frame components. That is, $\epsilon^{abcd} = +1$ when $a = 0$, $b = 1$,

$c = 2$ and $d = 3$, and so $\epsilon_{abcd} = -1$ for the same choices for a, b, c and d. The value for all other choices of indices is then determined by complete antisymmetry under permutations of any pair of indices. With this choice then the world-index versions are defined by

$$\epsilon^{\mu\nu\lambda\rho} := e_a{}^{\mu} e_b{}^{\nu} e_c{}^{\lambda} e_d{}^{\rho} \epsilon^{abcd} , \tag{A.9}$$

and the also completely antisymmetric

$$\epsilon_{\mu\nu\lambda\rho} := e^a{}_{\mu} e^b{}_{\nu} e^c{}_{\lambda} e^d{}_{\rho} \epsilon_{abcd} = g_{\mu\alpha} g_{\nu\beta} g_{\lambda\sigma} g_{\rho\zeta} \epsilon^{\alpha\beta\sigma\zeta} , \tag{A.10}$$

which satisfy $\epsilon^{\mu\nu\lambda\rho} = \det[e_a{}^{\mu}] = \det^{1/2}[-g^{\mu\nu}] = \det^{-1/2}[-g_{\mu\nu}]$ when $\mu = 0$, $\nu = 1$, $\lambda = 2$ and $\rho = 3$. It is conventional to introduce the notation $g := \det[g_{\mu\nu}]$, which is negative given the Lorentzian signature shared by $g_{\mu\nu}$ and η_{ab}. In terms of this $\epsilon_{\mu\nu\lambda\rho} = -\det[e^a{}_{\mu}] = -\sqrt{-g}$ when $\mu = 0$, $\nu = 1$, $\lambda = 2$ and $\rho = 3$, with all other entries defined by antisymmetry.

As eq. (A.9) makes clear, the quantity $\epsilon^{\mu\nu\lambda\rho}$ transforms as a rank-4 contravariant tensor under coordinate transformations (or diffeomorphisms) and is invariant under local Lorentz transformations, since the tangent-frame quantity ϵ^{abcd} transforms as a scalar under diffeomorphisms (and is a rank-4 contravariant tensor under local Lorentz transformations). One sometimes encounters in the literature (but never elsewhere in this book) a related tensor, $\varepsilon^{\mu\nu\lambda\rho} := \sqrt{-g}\ \epsilon^{\mu\nu\lambda\rho}$ whose components equal ± 1 when all indices are different (or its covariant version $\varepsilon_{\mu\nu\lambda\rho} := \epsilon_{\mu\nu\lambda\rho}/\sqrt{-g}$), whose nonzero components are also ± 1. Although these quantities have simple components they transform differently under diffeomorphisms, transforming as a tensor density (of weight $\pm\frac{1}{2}$) due to the additional factor of $\sqrt{-g}$ that is present.

Useful identities use the fact that two Levi-Civitas make a metric, since both are invariant tensors under proper Lorentz transformations (more about which below) while Levi-Civita changes sign under parity and time-reversal (and the metric does not). More precisely,

$$\begin{aligned}
\epsilon_{\mu\nu\lambda\rho}\epsilon^{\alpha\beta\sigma\zeta} &= -\delta^{\alpha}_{\mu}\,\delta^{\beta}_{\nu}\,\delta^{\sigma}_{\lambda}\,\delta^{\zeta}_{\rho} \pm (23 \text{ other permutations of } \alpha, \beta, \sigma \text{ and } \zeta) \\
\epsilon_{\mu\nu\lambda\rho}\epsilon^{\alpha\beta\sigma\rho} &= -(\delta^{\alpha}_{\mu}\,\delta^{\beta}_{\nu}\,\delta^{\sigma}_{\lambda} + \delta^{\beta}_{\mu}\,\delta^{\sigma}_{\nu}\,\delta^{\alpha}_{\lambda} + \delta^{\sigma}_{\mu}\,\delta^{\alpha}_{\nu}\,\delta^{\beta}_{\lambda} - \delta^{\beta}_{\mu}\,\delta^{\alpha}_{\nu}\,\delta^{\sigma}_{\lambda} - \delta^{\sigma}_{\mu}\,\delta^{\beta}_{\nu}\,\delta^{\alpha}_{\lambda} - \delta^{\alpha}_{\mu}\,\delta^{\sigma}_{\nu}\,\delta^{\beta}_{\lambda}) \\
\epsilon_{\mu\nu\lambda\rho}\epsilon^{\alpha\beta\lambda\rho} &= -2(\delta^{\alpha}_{\mu}\,\delta^{\beta}_{\nu} - \delta^{\beta}_{\mu}\,\delta^{\alpha}_{\nu}) \\
\epsilon_{\mu\nu\lambda\rho}\epsilon^{\alpha\nu\lambda\rho} &= -3!\,\delta^{\alpha}_{\mu} \\
\epsilon_{\mu\nu\lambda\rho}\epsilon^{\mu\nu\lambda\rho} &= -4! \,.
\end{aligned} \tag{A.11}$$

The right-hand sides of these identities are the most general possible tensors built only out of the Kronecker delta and the metric with the same symmetries as the left-hand side. The numerical coefficients are most easily determined by evaluating both sides using explicit values for the open indices.

A.2.2 Finite temperature and Euclidean signature

It is often useful to work with a Euclidean-signature metric, for which all of the eigenvalues of $g_{\mu\nu}$ are positive — also called the $(++++)$ metric. For instance the metric in rectangular coordinates for 4D flat Euclidean space is

$$ds^2 = g_{mn}\, dx^m\, dx^n = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 + (dx^4)^2 \,. \tag{A.12}$$

One of the great virtues of using a $(-+++)$ metric in Lorentzian signature is that the positive Euclidean metric is obtained simply by replacing $x^0 = t \to -\mathrm{i}\tau_E$ where $\tau_E = x^4$ is the corresponding Euclidean coordinate. The choice of sign in this transformation ensures that terms in the action transform as

$$\exp\left\{\frac{\mathrm{i}}{2}\int \mathrm{d}t\,\mathrm{d}^3x\left[(\partial_t\phi)^2-(\nabla\phi)^2\right]\right\} \to \exp\left\{-\frac{1}{2}\int \mathrm{d}\tau_E\mathrm{d}^3x\left[(\partial_{\tau_E}\phi)^2+(\nabla\phi)^2\right]\right\} \quad \text{(A.13)}$$

and so the oscillatory factor $e^{\mathrm{i}S(\phi)}$ in the path integral suppresses large gradients.

Equilibrium calculations at finite temperature provide a concrete situation where Euclidean methods are particularly useful. They are useful because of the resemblance between the thermal density matrix, $\rho \propto e^{-\beta H} = e^{-H/T}$, and the time-evolution operator, $U(t,0) = \exp[-\mathrm{i}Ht]$. This resemblance makes it look as if a thermal density matrix enters into calculations in the same way as would the time-evolution operator for a shift in imaginary time through a distance $\Delta t = -\mathrm{i}\Delta\tau_E = -\mathrm{i}\beta$ and so

$$\exp\left[-\mathrm{i}H\Delta t\right] = \exp\left[-H\Delta\tau_E\right] = \exp\left[-\beta H\right]. \quad \text{(A.14)}$$

Furthermore, in this language thermal averages, like the partition function

$$Z = \mathrm{Tr}\left[e^{-\beta H}\right] = \sum_N \left\langle N\left|e^{-\beta H}\right|N\right\rangle, \quad \text{(A.15)}$$

correspond to an evolution of a state $|N\rangle$ through a time interval $-\mathrm{i}\beta$ and then identifying the state obtained with the initial state (and summing). This makes plausible (and can be turned into a proof) that thermal expectation values can be rewritten in terms of field theories in a Euclidean-signature space for which the Euclidean time direction is a circle with circumference β.

Once time becomes a circular direction boundary conditions must be imposed on the fields in this direction. Standard thermal behaviour is reproduced if integer-spin bosons are chosen to be periodic, $\phi(\tau_E+\beta) = \phi(\tau_E)$, and half-integral spin fermions are chosen to be anti-periodic, $\psi(\tau_E+\beta) = -\psi(\tau_E)$.

A.2.3 Dirac conventions

For fermions the metric conventions drive related conventional choices for the Dirac matrices, γ^μ, since essentially everyone agrees these should be defined to satisfy the algebra $\{\gamma^\mu,\gamma^\nu\} = 2\eta^{\mu\nu}$ (in Minkowski space). In curved spaces one instead demands

$$\{\gamma^a,\gamma^b\} = 2\eta^{ab}, \quad \text{(A.16)}$$

in the tangent frame (defined by the tetrad $e^a{}_\mu$ of the previous section) and then converts to world indices using $\gamma^\mu := e_a{}^\mu\,\gamma^a$. With these definitions the world-index Dirac matrices satisfy the generally covariant Clifford algebra

$$\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu} \quad \text{(A.17)}$$

where $g^{\mu\nu}$ is the inverse metric.

In Minkowski space (or in the tangent frame of a curved space) the use of the $(-+++)$

metric implies $(\gamma^0)^2 = -1$ while $(\gamma^i)^2 = +1$ for $i = x, y, z$. Because this makes γ^0 imaginary (when diagonal) it is useful to define $\beta := \mathrm{i}\gamma^0$ so that $\beta^2 = 1$.

A convenient choice of basis for the Dirac matrices (which diagonalizes $\gamma_5 = -\mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$) which satisfies (A.16) is given by

$$\gamma_0 = -\gamma^0 = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \gamma_k = \begin{pmatrix} 0 & -\mathrm{i}\sigma_k \\ \mathrm{i}\sigma_k & 0 \end{pmatrix} \tag{A.18}$$

where σ_k are the usual 2×2 Pauli matrices for $k = 1, 2, 3$. In this basis

$$\beta = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \quad \text{and} \quad \gamma_5 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{A.19}$$

where I is the 2×2 unit matrix.

The Lorentz generators in this representation are given by $\mathcal{J}_{\mu\nu} = -\frac{\mathrm{i}}{4}[\gamma_\mu, \gamma_\nu]$, and so defining rotations, $\mathcal{J}_k$, and boosts, $\mathcal{K}_k$, by $\mathcal{J}_{0k} = \mathcal{K}_k$ and $\mathcal{J}_{ij} = \epsilon_{ijk}\mathcal{J}_k$ allows these generators to be written explicitly as

$$\mathcal{J}_k = \frac{1}{2}\begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}, \qquad \mathcal{K}_k = \frac{\mathrm{i}}{2}\begin{pmatrix} -\sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}. \tag{A.20}$$

Because these are block-diagonal they show that $[\gamma_5, \mathcal{J}_{\mu\nu}] = 0$, and so the 4-dimensional spinor representation is reducible: the two 2-dimensional eigenspaces of γ_5 each furnish separate representations of the Lorentz group. Furthermore, although these two representations agree on their representation of the $\mathcal{J}_k$ (both are spin-half), their representations of the $\mathcal{K}_k$ are conjugates of one another in the precise sense that the Pauli-matrix identity $\sigma_k = -\sigma_2\sigma_k^*\sigma_2$ implies

$$\mathcal{K}_{k\pm} = \sigma_2\,\mathcal{K}_{k\mp}^*\,\sigma_2\,, \tag{A.21}$$

where the sign in the subscript denotes the eigenvalue of γ_5.

Using this representation for Lorentz boosts allows explicit construction of the spinors $\mathbf{u}(\mathbf{p}, \sigma)$ and $\mathbf{v}(\mathbf{p}, \sigma)$ appearing in the field expansion of (C.30), reproduced again here:

$$\Psi(x) = \sum_{\sigma=\pm\frac{1}{2}} \int \frac{\mathrm{d}^3 p}{\sqrt{(2\pi)^3 2E_p}} \left[\mathbf{u}(\mathbf{p}, \sigma)\, \mathfrak{c}_{\mathbf{p}\sigma}\, e^{\mathrm{i}p\cdot x} + \mathbf{v}(\mathbf{p}, \sigma)\, \overline{\mathfrak{c}}^*_{\mathbf{p}\sigma}\, e^{-\mathrm{i}p\cdot x}\right], \tag{A.22}$$

where $E_p = \sqrt{\mathbf{p}^2 + m^2}$ and m is the particle mass.

These spinors satisfy[1] $(\mathrm{i}\not{p} + m)\mathbf{u} = (\mathrm{i}\not{p} - m)\mathbf{v} = 0$, where as usual the slash denotes contraction with a 4-vector, as in $\not{p} := p_\mu\gamma^\mu$. In the rest frame these become the conditions $\beta\,\mathbf{u} = \mathbf{u}$ and $\beta\,\mathbf{v} = -\mathbf{v}$. Boosting $\beta\mathbf{u} = \mathbf{u}$ to a general frame then gives

$$\mathbf{u}(\mathbf{p}, \sigma) = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{E_p + m} - \boldsymbol{\sigma}\cdot\hat{\mathbf{p}}\sqrt{E_p - m} & 0 \\ 0 & \sqrt{E_p + m} + \boldsymbol{\sigma}\cdot\hat{\mathbf{p}}\sqrt{E_p - m} \end{pmatrix}\begin{pmatrix} \chi(\sigma) \\ \chi(\sigma) \end{pmatrix}, \tag{A.23}$$

[1] This follows purely from the consistency of the Poincaré transformation properties of Ψ and $a_{\mathbf{p}\sigma}$, but can equivalently be regarded as a consequence of the field equation $(\not{\partial} + m)\Psi = 0$.

where $\hat{\mathbf{p}} = \mathbf{p}/|\mathbf{p}|$. Here $\chi(\sigma)$ is a 2-component spinor encoding the spin of the particle in its rest frame. If defined as eigenstates of $\mathcal{J}_3$ these become

$$\chi(\sigma = +1/2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \chi(\sigma = -1/2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{A.24}$$

The spinor $\mathbf{v}$ is found by a similar exercise, or by the action of charge conjugation (see below). A short calculation shows that these spinors satisfy the useful completeness relations

$$\sum_{\sigma=\pm\frac{1}{2}} \mathbf{u}(\mathbf{p},\sigma)\overline{\mathbf{u}}(\mathbf{p},\sigma) = -\mathrm{i}\not{p} + m \quad \text{and} \quad \sum_{\sigma=\pm\frac{1}{2}} \mathbf{v}(\mathbf{p},\sigma)\overline{\mathbf{v}}(\mathbf{p},\sigma) = -\mathrm{i}\not{p} - m\,, \tag{A.25}$$

whose right-hand sides reduce in the rest frame to $m(\beta \pm 1)$, projecting onto the appropriate eigenspace of β.

Weyl and Majorana spinors

There are two natural ways to reduce the 4-dimensional Dirac spinor to two components in a Lorentz-invariant way. Since left-handed spinors satisfy $\gamma_5\psi_L = \psi_L$ and right-handed spinors satisfy $\gamma_5\psi_R = -\psi_R$ a general Dirac spinor can be written in this basis as

$$\Psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \tag{A.26}$$

The conditions $\psi_R = 0$ or $\psi_L = 0$ are clearly Lorentz-invariant. Dirac spinors satisfying one of these conditions are called (left- or right-handed) Weyl spinors.

The other Lorentz-invariant way to constrain a Dirac spinor is to demand that ψ_R be the complex conjugate of ψ_L – up to multiplication by σ_2 as in condition (A.21). A spinor satisfying this type of reality condition is called a Majorana spinor,

$$\Psi_M = \begin{pmatrix} \psi_L \\ -\varepsilon\,\psi_L^* \end{pmatrix} = C\,\Psi_M^*\,, \tag{A.27}$$

where $\varepsilon = \mathrm{i}\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is the real 2×2 antisymmetric matrix. The matrix C is called the charge-conjugation matrix, and is given explicitly in this basis by

$$C = \begin{pmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{pmatrix} = -\gamma^2\,. \tag{A.28}$$

A final convention involves the definition of the Dirac conjugate, which here is given by

$$\overline{\Psi} := \Psi^\dagger\beta = \mathrm{i}\Psi^\dagger\gamma^0\,. \tag{A.29}$$

(Keep in mind the factor of i here when comparing with conventions using the opposite signature for the metric.) When applied to Majorana spinors (A.27), becomes

$$\overline{\Psi}_M := \Psi_M^\dagger\beta = \Psi_M^T C\beta = \Psi_M^T\gamma_5\epsilon\,, \tag{A.30}$$

where the superscript 'T' denotes the transpose in spinor space, in the same way that '†'

denotes hermitian conjugation in this space. The matrix ϵ defined here is called the time-reversal matrix, given explicitly in this basis by

$$\epsilon = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}, \tag{A.31}$$

in terms of which $C = \gamma_5 \epsilon \beta$.

The matrices β, ϵ and γ_5 provide very useful set inasmuch as they characterize a spinor's transformation properties under parity, time-reversal and charge conjugation (more about the definition of these is given in §C.4.3 below). Chasing through the definitions shows that these symmetries get realized on Dirac spinors as follows:

$$\begin{aligned} \mathcal{P}\,\Psi(x)\,\mathcal{P}^{-1} &= \eta_p \beta\,\Psi(x_P) \\ C\,\Psi(x)\,C^{-1} &= \eta_c\, C\,\Psi^*(x) \\ \mathcal{T}\,\Psi(x)\,\mathcal{T}^{-1} &= \eta_t\, \epsilon\,\Psi(x_T)\,, \end{aligned} \tag{A.32}$$

where η_p, η_c and η_t are arbitrary phases while $x_P^\mu := P^\mu{}_\nu x^\nu$ and $x_T^\mu := T^\mu{}_\nu x^\nu$ are the parity and time-reversal transforms of the point x^μ (with the matrices $P^\mu{}_\nu$ and $T^\mu{}_\nu$ defined in eq. (C.64)). In particular, an individual Majorana spinor represents a spin-half particle that is its own antiparticle (in much the same way that a real scalar represents a spin-zero particle that is its own antiparticle).

Spinor bilinears

Since local lagrangian densities are scalars they are built from combinations of fermion bilinears of the form $\overline{\Psi}_1 M \Psi_2$ for two fields Ψ_1 and Ψ_2. It is useful to expand the arbitrary 4×4 matrix M in terms of a standard basis that transforms covariantly under Lorentz transformations. This basis is conveniently chosen to be the sixteen matrices

$$1\,, \quad \gamma_5\,, \quad \gamma^\mu\,, \quad \gamma_5\gamma^\mu \quad \text{and} \quad \gamma^{\mu\nu} = \frac{1}{2}\left[\gamma^\mu, \gamma^\nu\right]. \tag{A.33}$$

These satisfy useful symmetry and hermiticity relations together with the parity and time-reversal matrices β and ϵ. In addition to $\gamma_5 = \gamma_5^T = \gamma_5^\dagger$, $\beta = \beta^T = \beta^\dagger$ and $-\epsilon = \epsilon^\dagger = \epsilon^{-1} = \epsilon^T$, one has the identities

$$M^T = \xi(\epsilon M \epsilon)\,, \quad M^\dagger = \zeta(\beta M \beta) \quad M^* = \xi\zeta(\epsilon\beta M \epsilon\beta) \quad \text{and} \quad M = \eta(\gamma_5 M \gamma_5)\,, \tag{A.34}$$

with the signs ξ, ζ and η given for each member of the basis (A.33) in Table A.1.

These identities are useful in that they dictate the reality and symmetry properties of bilinears built from Majorana fermions. That is, if Ψ_1 and Ψ_2 both satisfy condition (A.27), then

$$\overline{\Psi}_1 M \Psi_2 = \lambda(\overline{\Psi}_2 M \Psi_1) \qquad \text{and} \qquad \overline{\Psi}_1 M \Psi_2 = \chi(\overline{\Psi}_1 M \Psi_2)^*\,, \tag{A.35}$$

with signs λ and χ also given in Table A.1.

Table A.1 The signs appearing in (A.34) and (A.35) for M one of the basis (A.33) of Dirac matrices.

	1	γ^μ	$\gamma^{\mu\nu}$	$\gamma_5\gamma^\mu$	γ_5
ξ	−	−	+	+	−
ζ	+	−	−	−	−
η	+	−	+	−	+
$\xi\zeta$	−	+	−	−	+
λ	+	−	−	+	+
χ	+	+	+	−	−

Standard-Model fermions

It is often useful to use real fields when writing down the most general effective couplings, and for spin-half fields this means using Majorana spinors. This section develops the notation for writing a generation of Standard Model fermions in terms of Majorana spinors, following [194].

The types of spin-half 2-component Weyl fermions in a Standard Model generation are

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \quad u_R\,, \quad d_R\,, \quad e_R\,. \tag{A.36}$$

A 4-component Majorana field for each right-handed particle is then defined by

$$\gamma_R U = u_R\,, \quad \gamma_R D = d_R \quad \text{and} \quad \gamma_R E = e_R\,. \tag{A.37}$$

The left-handed components of these spinors are simply given by the conjugate fields, so

$$\gamma_L U = u_L^c = \varepsilon\, u_R^*\,, \quad \gamma_L D = d_L^c = \varepsilon\, d_R^* \quad \text{and} \quad \gamma_L E = e_L^c = \varepsilon\, e_R^*\,, \tag{A.38}$$

with the matrix ε defined below eq. (A.27).

The same construction applies to the $SU_L(2)$ doublets, starting with the definitions

$$\gamma_L Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \quad \text{and} \quad \gamma_L L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \tag{A.39}$$

and so the right-handed components become

$$\gamma_R Q = \begin{pmatrix} u_R^c \\ d_R^c \end{pmatrix} \quad \text{and} \quad \gamma_R L = \begin{pmatrix} \nu_R^c \\ e_R^c \end{pmatrix}, \tag{A.40}$$

where $d_R^c = -\varepsilon\, d_L^*$ and so on.

A.2.4 Dimensional Regularization

Dimensional regularization is usually the regularization of choice for practical calculations, both because of its comparative simplicity and because it treats symmetries relatively be-

nignly. This section collects some of the useful formulae associated with this regularization used in the main text.

The fundamental formula used in the main text involves a single loop integral of the form

$$J(q) := \int \frac{\mathrm{d}^4 p}{(2\pi)^4} \left[\frac{(p^2)^A}{(p^2+q^2)^B}\right], \tag{A.41}$$

where $p^2 = p_\mu p^\mu$ and q^2 is a Lorentz-invariant function of $q_\mu q^\mu$, and possible low-energy masses. The squares of all 4-momenta are taken in Lorentzian signature, so $J(q)$ is a Lorentz-invariant function of q^μ. At face value this integral diverges in the ultraviolet for $2A + 4 \geq 2B$, and the goal is to define the integral so as to be able sensibly to evaluate physical quantities before the divergence is ultimately eliminated by absorbing it into the value of a bare parameter when renormalizing.

The denominator of the integrand usually has an implicit $\mathrm{i}\varepsilon$ factor that tells how to navigate around any poles in the energy integrations, and the result is the same as what is obtained from Wick rotating the energy to imaginary values, using[2] $p^0 = \mathrm{i}p^4$, so $p^\mu p_\mu = -(p^0)^2 + \mathbf{p}^2 = (p^4)^2 + \mathbf{p}^2 \geq 0$, while $\mathrm{d}^4 p = \mathrm{i}\mathrm{d}^4 p_E$, where $\mathrm{d}^4 p_E := \mathrm{d}p^4 \mathrm{d}^3\mathbf{p}$ is the Euclidean integration measure. The rotation occurs because p^4 is integrated through real values rather than imaginary ones. Once this is done the angular integration over the direction of p^m can be done by inspection, leaving only a divergent one-dimensional integral to be regularized.

The idea of dimensional regularization is to consider the same expression in D dimensions,

$$I_D(q) := \int \frac{\mathrm{d}^D p_E}{(2\pi)^D} \left[\frac{p^{2A}}{(p^2+q^2)^B}\right], \tag{A.42}$$

with the desired answer formally obtained by $J = \lim_{D\to 4} I_D$. The virtue of introducing D as a variable is that the integral converges in the ultraviolet for $D < 2(B-A)$, with the finite result obtained by explicit integration being

$$\begin{aligned} I_D(q) &= \left[\frac{S_{D-1}}{2(2\pi)^D}\right] \left(q^2\right)^{A-B+D/2} \int_0^\infty \mathrm{d}x \left[\frac{x^{A+(D-2)/2}}{(x+1)^B}\right] \\ &= \left[\frac{\pi^{D/2}}{(2\pi)^D \Gamma\left(\frac{D}{2}\right)}\right] \frac{\Gamma\left(A+\frac{D}{2}\right)\Gamma\left(B-A-\frac{D}{2}\right)}{\Gamma(B)} \left(q^2\right)^{A-B+D/2}, \end{aligned} \tag{A.43}$$

where S_n is the area of the n-dimensional unit sphere and $\Gamma(z)$ is Euler's Gamma function, defined by analytically continuing the defining relation $\Gamma(z+1) = z\Gamma(z)$ to the complex plane.

The last equality here extends the definition of I_D to any complex D except for the the poles of $\Gamma\left(A+\frac{D}{2}\right)\Gamma\left(B-A-\frac{D}{2}\right)$, which occur whenever the argument of a Γ-function is a non-positive integer (and so includes the case of real interest where $D = 4$). The regularization is performed by evaluating the result at $D = 4 - 2\epsilon$ for $0 < \epsilon \ll 1$, with the limit $\epsilon \to 0$ take after renormalization has removed the divergence. The incipient divergence in

[2] The sign here is chosen by the requirement that the rotation from the real to the imaginary axis avoids the poles at $p^0 = \sqrt{\mathbf{p}^2 + m^2} - \mathrm{i}\varepsilon$.

this limit appears as a pole, arising from asymptotic formulae for the Gamma function like

$$\Gamma[\epsilon] = \frac{1}{\epsilon} - \gamma + O(\epsilon)\,, \tag{A.44}$$

where the Euler-Mascheroni constant, γ, is defined by the limit

$$\gamma := \lim_{n\to\infty}\left[\sum_{k=1}^{n}\frac{1}{k} - \ln n\right] \simeq 0.57721566490153286060651209008240243104 2\ldots\,. \tag{A.45}$$

Poles near negative integers are found by repeatedly using $\Gamma(z+1) = z\,\Gamma(z)$. For instance, choosing $z = -1+\epsilon$ implies

$$\Gamma[-1+\epsilon] = \frac{\Gamma[\epsilon]}{-1+\epsilon} \simeq -\frac{1}{\epsilon} + (\gamma - 1) + O(\epsilon)\,, \tag{A.46}$$

and so on.

Renormalization schemes

The pole in this expression expresses the divergence that the integral possesses when $D = 4$, which (if ultraviolet[3] in origin) is usually absorbed into the renormalization of a bare coupling. This is possible because the full expression for a physical quantity depends on both this bare coupling and the loop integral. Although it is unambiguous to say that the bare coupling cancels the divergent part of a loop integral, there *is* an ambiguity associated with how much of the finite parts of a loop are also subtracted in the same way. A precise statement about how much of the finite part to absorb when cancelling divergences defines what is called the renormalization 'scheme'. There is nothing unique about any scheme, with different choices simply corresponding to different ways for precisely defining the meaning of the coupling in question.

To see how this works in practice, consider the divergences associated with the vacuum polarization of the electromagnetic field. In quantum electrodynamics the Fourier transform of the propagator, $\langle TA_\mu(x)A_\nu(y)\rangle$, for non-interacting photons is (in a Lorentz-covariant gauge)

$$\Delta^{(0)}_{\mu\nu}(p) = \frac{1}{p^2 - i\varepsilon}\left[\eta_{\mu\nu} + C_{\scriptscriptstyle A}\,\frac{p_\mu p_\nu}{p^2}\right], \tag{A.47}$$

where ε is a small positive infinitesimal (*not* the ϵ from $D = 4-2\epsilon$) and $C_{\scriptscriptstyle A}$ is a quantity whose precise value depends on the gauge being used (*e.g.* with $C_{\scriptscriptstyle A} = 0$ in Feynman gauge or $C_{\scriptscriptstyle A} = -1$ in Landau gauge). The precise form of $C_{\scriptscriptstyle A}$ is not important since it does not appear in any physical predictions. This propagator has a pole at $p^2 = 0$ that defines the photon's energy-momentum dispersion relation: $\varepsilon_p = p^0 = |\mathbf{p}|$. The freedom to rescale fields, $A_\mu \to \lambda A_\mu$, is used when writing (A.47) to ensure the $\eta_{\mu\nu}$ term has unit residue at this pole (this is an automatic consequence of canonical normalization).

Once interactions are included the propagator does not remain as simple as in (A.47),

[3] Poles as $D \to 4$ can also arise due to infra-red divergences, and these should *not* be renormalized (as may be seen from the discussion following Eq. (12.19) in the main text).

but it turns out that Lorentz-covariance and gauge invariance require that its most general form must be

$$\begin{aligned}\Delta_{\mu\nu}(p) &= \frac{1}{(p^2-\mathrm{i}\varepsilon)[1-\Pi(p^2)]}\left[\eta_{\mu\nu}-\Pi(p^2)\,\frac{p_\mu p_\nu}{p^2}\right]+C_{\scriptscriptstyle A}\,\frac{p_\mu p_\nu}{p^2}\\ &= \frac{\eta_{\mu\nu}}{(p^2-\mathrm{i}\varepsilon)[1-\Pi(p^2)]}+\widetilde{C}_{\scriptscriptstyle A}\,\frac{p_\mu p_\nu}{p^2}\,,\end{aligned} \tag{A.48}$$

where the function, $\Pi(p^2)$, is known as the vacuum polarization (and the second line defines the quantity $\widetilde{C}_{\scriptscriptstyle A}$, whose gauge-dependent value still does not matter for physical predictions).

Although $\Pi(p^2)$ vanishes for non-interacting photons, it is nonzero once couplings to charged particles are included. If $\Pi(p^2)$ were also to have a pole at $p^2 = 0$, such as if $\Pi(p^2) = A/p^2 + B + \cdots$, then the propagator's pole gets moved to $p^2 = A$ (which – provided A is negative – would imply that the interactions give the photon a nonzero mass). So long as $\Pi(p^2)$ is less singular than this near $p^2 = 0$ the pole in $\Delta_{\mu\nu}$ survives, indicating that no mass gets developed.

In quantum electrodynamics (QED) — the theory of interacting electrons and photons — $\Pi(p^2)$ is obtained by evaluating 1-particle irreducible graphs with two external photon legs (see Fig. 7.4). The absence of reducible photon lines in these graphs precludes them from introducing a pole and this ensures that loops of virtual electrons do not shift the photon mass. The one-loop vacuum polarization graph with an electron in the loop contributes

$$\begin{aligned}\Pi(p^2)_{1-\text{loop}} &= -\frac{8e^2}{(4\pi)^{D/2}}\,\Gamma\!\left(2-\frac{D}{2}\right)\int_0^1 \mathrm{d}u\; u(1-u)\left[\frac{m^2+p^2u(1-u)}{\mu^2}\right]^{(D-4)/2}\\ &= \frac{e^2}{2\pi^2}\int_0^1 \mathrm{d}u\; u(1-u)\left\{\frac{1}{(D/2)-2}+\gamma+\ln\left[\frac{m^2+p^2u(1-u)}{4\pi\mu^2}\right]+O(D-4)\right\}\,,\end{aligned} \tag{A.49}$$

once regularized in dimensional regularization. Here $-e$ is the electron charge and m is the electron mass, and μ is an arbitrary scale introduced by replacing $e^2 \to e^2\,\mu^{4-D}$ so that e remains dimensionless in D spacetime dimensions.

Although this expression has no pole at $p^2 = 0$, neither does it vanish there since

$$\Pi(0)_{1-\text{loop}} = \frac{e^2}{12\pi^2}\left[\frac{1}{(D/2)-2}+\gamma+\ln\left(\frac{m^2}{4\pi\mu^2}\right)+O(D-4)\right]. \tag{A.50}$$

Consequently the propagator's residue at $p^2 = 0$ is no longer unity. To fix this the field must be rescaled once more — that is to say: 'renormalized' — to ensure unit residue, by taking $A_\mu \to \sqrt{Z_3}\,A_\mu$, after which the propagator rescales to $\Delta^{(0)}_{\mu\nu}(p) \to Z_3\,\Delta^{(0)}_{\mu\nu}(p)$. In perturbation theory writing $Z_3 = 1 + \delta Z$, with $\delta Z \sim O(e^2)$, then shows that the renormalized vacuum polarization becomes

$$\Pi(p^2)_{\text{ren}} = \Pi(p^2)_{1-\text{loop}} - (Z_3 - 1)\,. \tag{A.51}$$

Now comes the main point regarding convenient renormalization schemes. The physical renormalization choice (or 'on-shell' scheme) for Z_3 requires $\Pi(0)_{\text{ren}} = 0$ since this

guarantees unit residue at the propagator's pole at $p^2 = 0$. This gives

$$Z_3^{\text{phys}} - 1 \simeq \Pi(0)_{1-\text{loop}} = \frac{e^2}{12\pi^2}\left[\frac{1}{(D/2)-2} + \gamma + \ln\left(\frac{m^2}{4\pi\mu^2}\right) + O(D-4)\right], \tag{A.52}$$

after which the limit $D \to 4$ can be taken to give

$$\Pi(p^2)_{\text{ren}}^{\text{phys}} = \frac{e^2}{2\pi^2}\int_0^1 \mathrm{d}u\, u(1-u)\ln\left[1 + \frac{p^2u(1-u)}{m^2}\right]. \tag{A.53}$$

But if the only goal is to subtract off divergences, the minimalist choice — called the *minimal subtraction* or MS scheme — merely subtracts the pole at $D = 4$, so

$$Z_3^{\text{MS}} - 1 \simeq \frac{e^2}{6\pi^2}\left(\frac{1}{D-4}\right), \tag{A.54}$$

and so (again taking $D \to 4$)

$$\Pi(p^2)_{\text{ren}}^{\text{MS}} = \frac{e^2}{12\pi^2}\left[\gamma + \ln\left(\frac{m^2}{4\pi\mu^2}\right)\right] + \Pi(p^2)_{\text{ren}}^{\text{phys}}. \tag{A.55}$$

A slightly more convenient and equally minimal choice [35, 36, 37], called the *modified minimal subtraction* (or $\overline{\text{MS}}$) scheme, subtracts the universal factors γ and $\ln 4\pi$ as well as the divergent pole, leading to

$$Z_3^{\overline{\text{MS}}} - 1 \simeq \frac{e^2}{6\pi^2}\left[\frac{1}{(D/2)-2} + \gamma - \ln(4\pi)\right], \tag{A.56}$$

and so

$$\Pi(p^2)_{\text{ren}}^{\overline{\text{MS}}} = \frac{e^2}{12\pi^2}\ln\left(\frac{m^2}{\mu^2}\right) + \Pi(p^2)_{\text{ren}}^{\text{phys}}. \tag{A.57}$$

Although these last two renormalization schemes do not use canonically normalized fields, they trade this against the advantage of simplicity for other types of calculations. In particular they allow more simple direct integration of the renormalization-group evolution of couplings with scale and so simplify the resummation of leading logarithms (such as described in the main text in §7.2.1).

B Appendix B Momentum eigenstates and scattering

This appendix collects (often only with telegraphic derivation) some useful relations for computing scattering and decay rates. Since some subtleties of continuum normalization for momenta are dealt with by appealing to discrete normalization, this discussion starts with a summary of conventions regarding momentum eigenstates.

B.1 Momentum eigenstates

There are three different conventions often used: discrete normalization, continuum normalization, and relativistic continuum normalization. All three types arise in this book, so this section furnishes a brief reminder of how to convert from one to another. For simplicity this is done here for one spatial dimension, though identical arguments also work in other choices for the number of dimensions.

Discrete normalization corresponds to situations where momentum takes a denumerably infinite set of values, such as occurs if spatial dimensions have finite length, L say, perhaps satisfying periodic boundary conditions so fields satisfy $\psi(x + L) = \psi(x)$ for any x. For momentum eigenstates, $\psi(x) \propto \exp[\mathrm{i}px]$, this condition implies $p = 2\pi n/L$ for integer n, making p denumerable as required. Normalization and completeness relations for states then take the usual quantum form, such as

$$(p\,|\,q) = \delta_{pq} \qquad \text{and} \qquad \sum_p |\,p)(p\,| = 1\,, \tag{B.1}$$

where the sum over p is really a sum over the integer n. Here the 'rounded ket' notation, $|\,p)$, is used to distinguish states normalized this way from non-denumerable situations normalized in the continuum.

Inserting a complete set of position eigenstates and using the wavefunction $\langle x\,|\,p) \propto e^{\mathrm{i}px}$ shows that the orthonormality condition of (B.1) becomes

$$(p\,|\,q) = \int_0^L \mathrm{d}x\,(p\,|x\rangle\,\langle x\,|\,q) = \delta_{pq}\,, \tag{B.2}$$

and so $\langle x\,|\,p) = L^{-1/2}\,e^{\mathrm{i}px}$. This normalization then implies completeness takes the usual form

$$\langle x\,|\,y\rangle = \sum_p \langle x\,|\,p)(p\,|\,y\rangle = \frac{1}{L}\sum_{n=-\infty}^{\infty} e^{2\mathrm{i}\pi n(x-y)/L} = \delta(x-y)\,. \tag{B.3}$$

Continuum normalization

Continuum-normalized states $|p\rangle$ are obtained from discrete-normalized states in the infinite-volume limit $L \to \infty$. In this limit the spacing, $2\pi/L$, between adjacent levels goes to zero so the denumerable label p goes over to a continuum one. For L very large but still finite there are $\mathrm{d}N = \mathrm{d}p/(2\pi/L)$ states in a small continuous interval $\mathrm{d}p$, and so the density of states is $\mathrm{d}N/\mathrm{d}p = L/(2\pi)$. Therefore any sum over p goes over to an integral according to the rule

$$\sum_p F(p) = \int \mathrm{d}p\, F(p) \frac{\mathrm{d}N}{\mathrm{d}p} = L \int \frac{\mathrm{d}p}{2\pi}\, F(p)\,. \tag{B.4}$$

Using this conversion, for very large L the completeness relation for $|p)$ becomes

$$1 = \sum_p |p)(p\,| = L \int \frac{\mathrm{d}p}{2\pi}\, |p)(p| =: \int \mathrm{d}p\, |p\rangle\langle p\,|\,, \tag{B.5}$$

where the last equality suggests the definition of the continuum-normalized state

$$|p\rangle := \sqrt{\frac{L}{2\pi}}\, |p)\,. \tag{B.6}$$

Multiplying (B.5) through on the right by $|q\rangle$ shows that consistency requires the continuum state must satisfy the normalization condition

$$\langle p\,|\,q\rangle = \delta(p-q)\,, \tag{B.7}$$

which can also be inferred directly from the definitions using

$$\langle p\,|\,q\rangle = \frac{L}{2\pi}\,(p\,|\,q) = \lim_{L\to\infty} \frac{L}{2\pi}\, \delta_{pq}\,. \tag{B.8}$$

The right-hand side of this expression is zero if $p \neq q$ and if $p = q$ it goes to infinity as $L \to \infty$. This suggests it is a Dirac delta function, $\delta(p-q)$, up to normalization. To get the normalization notice that the integral over p of (B.8) in this limit is given by

$$\int \mathrm{d}p\, \langle p\,|\,q\rangle = \frac{2\pi}{L} \sum_p \langle p\,|\,q\rangle = \sum_p (p\,|\,q) = 1\,, \tag{B.9}$$

and so the right-hand side of (B.8) goes to $\delta(p-q)$ as $L \to \infty$, as claimed.

A useful relation when converting between discrete and continuum normalizations is

$$\sum_p |p)(p\,| = \frac{2\pi}{L} \sum_p \left[\frac{L}{2\pi}\, |p)(p\,|\right] \to \int \mathrm{d}p\, |p\rangle\langle p\,|\,, \tag{B.10}$$

showing that completeness sums are the same, regardless of whether momenta are normalized discretely or in the continuum. Often these kinds of sums arise weighted by quantities like an initial probability distribution for $P(p)$, and when this is so $P(p)$ goes over in the continuum limit to a phase-space distribution, $f(p)$, as follows. If $P(p)$ is the probability of having any one value for p, and varies slowly enough to be regarded as being constant in a short interval $\mathrm{d}p$, then the density of probability, $\mathrm{d}\mathcal{P}(p)$, for finding the particle in $\mathrm{d}p$ is:

$$\mathrm{d}\mathcal{P}(p) = \frac{\mathrm{d}N}{\mathrm{d}p}\, P(p)\, \mathrm{d}p = \frac{L}{2\pi}\, P(p)\, \mathrm{d}p\,. \tag{B.11}$$

and so the differential probability per-unit-spatial-volume of finding the particle in this momentum region (*i.e.* the phase-space probability density) becomes

$$\frac{f(p)}{2\pi} := \frac{1}{L}\left(\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p}\right) = \frac{P(p)}{2\pi}\,. \tag{B.12}$$

The 2π in the left-hand side's denominator is conventional, and ensures that for a thermal distribution (say) for which the position-space probability density is

$$\mathfrak{p} := \int \frac{\mathrm{d}p}{2\pi}\,\frac{1}{e^{E/T}\pm 1}\,, \tag{B.13}$$

one has $f(p) = (e^{E/T}\pm 1)^{-1}$ with no additional factors of 2π. In the $L\to\infty$ limit one therefore has

$$\sum_p P(p)\,|p)(p\,| \to \int \mathrm{d}p\; f(p)\,|p\rangle\langle p\,|\,. \tag{B.14}$$

For three spatial dimensions identical arguments show that the density of states is $\mathrm{d}N/\mathrm{d}^3\mathbf{p} = \mathcal{V}/(2\pi)^3$ where $\mathcal{V} := L^3$ is the system's large spatial volume. This means discrete sums go over into 3D integrals according to

$$\sum_{\mathbf{p}} \to \mathcal{V}\int \frac{\mathrm{d}^3p}{(2\pi)^3} = \mathcal{V}\int \frac{\mathrm{d}p_x\mathrm{d}p_y\mathrm{d}p_z}{(2\pi)^3}\,, \tag{B.15}$$

so if $|\mathbf{p}\rangle = [\mathcal{V}/(2\pi)^3]^{1/2}|\mathbf{p})$ then as $\mathcal{V}\to\infty$ the completeness formula (B.10) becomes

$$1 = \sum_{\mathbf{p}} |\mathbf{p})(\mathbf{p}| \to \int \mathrm{d}^3p\;|\mathbf{p}\rangle\langle\mathbf{p}|\,, \tag{B.16}$$

while orthogonality goes over to

$$\langle\mathbf{p}|\mathbf{q}\rangle = \delta^3(\mathbf{p}-\mathbf{q}) = \delta(p_x-q_x)\,\delta(p_y-q_y)\,\delta(p_z-q_z)\,. \tag{B.17}$$

A sum weighted by an initial probability distribution similarly goes over to

$$\sum_{\mathbf{p}} P(\mathbf{p})\,|\mathbf{p})(\mathbf{p}\,| \to \int \mathrm{d}^3p\; f(\mathbf{p})\,|\mathbf{p}\rangle\langle\mathbf{p}\,|\,. \tag{B.18}$$

Covariant normalization

An additional normalization change is often made for relativistic theories, since for these it can be inconvenient that $|p\rangle$ satisfies a Lorentz non-invariant condition like (B.7) and (B.9). In particular, (B.9) implies $\langle p|q\rangle$ transform inversely to the way the measure $\mathrm{d}p$ transforms.

It happens, however, that the combination $\mathrm{d}p/E_p$ *is* invariant, if $E_p = \sqrt{p^2+m^2}$ is the energy associated with a given momentum p. This makes it useful to define the covariantly normalized state $|p\,\rangle_r := \sqrt{2E_p}\,|p\rangle$, which satisfies a Lorentz-invariant completeness condition

$$\int \frac{\mathrm{d}p}{2E_p}\,|p\,\rangle_r\,{}_r\langle\,p\,| = \int \mathrm{d}p\;|p\rangle\langle p| = 1\,, \tag{B.19}$$

and orthogonality relation

$$ {}_r\langle\, p \,|\, q \,\rangle_r = \sqrt{4E_pE_q}\,\langle p \,|\, q\rangle = 2E_p\,\delta(p-q) = 2E_q\,\delta(p-q)\,. \tag{B.20} $$

For three spatial dimensions the relativistic normalization is again defined by $|\mathbf{p}\,\rangle_r := \sqrt{2E(\mathbf{p})}\,|\mathbf{p}\rangle$, which satisfies a Lorentz-invariant completeness condition

$$ \int \frac{\mathrm{d}^3p}{2E(\mathbf{p})}\,|\mathbf{p}\,\rangle_r\,{}_r\langle\,\mathbf{p}\,| = \int \mathrm{d}\mathbf{p}\,|\mathbf{p}\rangle\langle\mathbf{p}| = 1\,, \tag{B.21} $$

and orthogonality relation

$$ {}_r\langle\,\mathbf{p}\,|\,\mathbf{q}\,\rangle_r = 2E(\mathbf{p})\,\delta^3(\mathbf{p}-\mathbf{q}) = 2E(\mathbf{q})\,\delta^3(\mathbf{p}-\mathbf{q})\,. \tag{B.22} $$

B.2 Basics of scattering theory

Scattering describes interactions for which the particles involved start off as widely separated wave-packets, then approach one another and interact briefly as their wave-packets overlap, and then separate to great distances again. Theoretical simplicity arises because many details are not required, with only the total change in energy and momentum due to the scattering being measured (rather than, say, their detailed trajectories for all times).

In principle, in quantum mechanics a particle moving in an initial (or final) wave-packet cannot be exact momentum (or energy) eigenstates because the uncertainty principle ensures that such eigenstates are not localized in space or time at all. It is nonetheless often possible to approximate the real states using a class of energy eigenstate, since scattering results are often largely insensitive to the details of the wave packets describing the initial states. The idealized energy eigenstates used for this purpose (described below) are called *scattering states*.

The goal is to compute scattering perturbatively in the interaction that dominates when the scattering wave-packets overlap. To this end suppose the complete Hamiltonian, H, can be written $H = H_0 + V$, where H_0 describes the evolution of the initial and final wave packets before and after the scattering. A key assumption is that the full set of energy eigenvalues for H contains (but need not be identical with) the spectrum of energy eigenstates for H_0. For instance, eigenstates of both H and H_0 can be labelled by their asymptotic momentum in the remote past, or the remote future. States in the spectrum of H but not in H_0 might include bound states whose existence relies on the presence of the interaction V.

Denote the energy eigenstates of H_0 by $|\alpha\rangle$, with α collectively denoting all of the labels required to describe single- and many-particle states and $H_0|\alpha\rangle = E_\alpha|\,\alpha\rangle$. A wave packet of such states can be schematically written

$$ |\,\phi_f\rangle := \int \mathrm{d}\alpha\; f(\alpha)|\alpha\rangle \tag{B.23} $$

where $f(\alpha)$ defines a normalizable packet. The label α is treated as continuous because it contains continuum-normalized momentum states (possibly among other labels). Because the spectrum of H_0 lies within the spectrum of H the same labels, α, and energies, E_α,

also describe some of the eigenstates of the full system, $H|\alpha\rangle\!\rangle = E_\alpha|\alpha\rangle\!\rangle$ (where – in this Appendix only – a double ket $|(\cdots)\rangle\!\rangle$ denotes an eigenstate of H).

In the Schrödinger picture the burden of time evolution is carried by the state of the system, and for scattering the time evolution of states prepared in appropriate wave packets, $|\,\phi_f\rangle$, have essentially the same evolution in the remote past and the remote future when evolved by either H or H_0 (because the particle wave-packets no longer overlap). That is, at very late times there is a state of the full system for which

$$\lim_{t\gg T} \mathrm{e}^{-\mathrm{i}Ht}|\,\phi_f\rangle\!\rangle_{\mathrm{o}} = \lim_{t\gg T} \mathrm{e}^{-\mathrm{i}H_0t}|\,\phi_f\rangle\,, \tag{B.24}$$

and a similar state, $|\,\phi_f\rangle\!\rangle_{\mathrm{i}}$, – in general different than $|\,\phi_f\rangle\!\rangle_{\mathrm{o}}$ if scattering actually occurs – whose evolution under H agrees with the evolution of a packet $|\,\phi_f\rangle$ under H_0 in the remote past:

$$\lim_{t\ll -T} \mathrm{e}^{-\mathrm{i}Ht}|\,\phi_f\rangle\!\rangle_{\mathrm{i}} = \lim_{t\ll -T} \mathrm{e}^{-\mathrm{i}H_0t}|\,\phi_f\rangle\,. \tag{B.25}$$

Taking the limiting case of appropriately peaked wave packets, $f(\alpha)$, allows the definition of idealized 'in' and 'out' scattering eigenstates of the full Hamiltonian, $|\alpha\rangle\!\rangle_{\mathrm{o,i}}$, that satisfy

$$\lim_{t\gg T} \mathrm{e}^{-\mathrm{i}Ht}|\alpha\rangle\!\rangle_{\mathrm{o}} = \lim_{t\gg T} \mathrm{e}^{-\mathrm{i}H_0t}|\alpha\rangle \quad \text{and} \quad \lim_{t\ll -T} \mathrm{e}^{-\mathrm{i}Ht}|\alpha\rangle\!\rangle_{\mathrm{i}} = \lim_{t\ll -T} \mathrm{e}^{-\mathrm{i}H_0t}|\alpha\rangle\,. \tag{B.26}$$

Scattering asks only for transition amplitudes between states that evolve like an eigenstate of H_0 in the remote past to similar states that evolve like eigenstates of H_0 in the remote future. Any such a scattering amplitude can be reconstructed from the matrix of all possible amplitudes between scattering energy eigenstates,

$$S_{\beta\alpha} := {}_{\mathrm{o}}\langle\!\langle\beta|\alpha\rangle\!\rangle_{\mathrm{i}}\,. \tag{B.27}$$

This quantity is called the S-matrix. The scattering operator, $\mathcal{S}$, is defined as that operator whose matrix elements between H_0 eigenstates, $|\alpha\rangle$, reproduce the amplitudes (B.27):

$$\langle\beta|\mathcal{S}|\alpha\rangle := S_{\beta\alpha}\,. \tag{B.28}$$

Formally $\mathcal{S}$ can be computed in terms of the Møller wave operators

$$\Omega(t) := \mathrm{e}^{\mathrm{i}Ht}\,\mathrm{e}^{-\mathrm{i}H_0t}\,, \tag{B.29}$$

because

$$|\,\alpha\rangle\!\rangle_{\mathrm{o}} = \lim_{t\gg T}\Omega(t)|\,\alpha\rangle \qquad \text{and} \qquad |\,\alpha\rangle\!\rangle_{\mathrm{i}} = \lim_{t\ll -T}\Omega(t)|\,\alpha\rangle\,. \tag{B.30}$$

The operators $\Omega^{\pm} = \lim_{t\to\pm\infty}\Omega(t)$ are isometric operators, but strictly speaking are not unitary in the presence of 'bound states' that are contained in the spectrum of H but not in the spectrum of H_0. The S-matrix is then given by

$$\mathcal{S} = \lim_{t\to\infty}\lim_{t'\to-\infty}\Omega^*(t)\Omega(t') = (\Omega^+)^*\Omega^-\,, \tag{B.31}$$

in which the limit $t\to\mp\infty$ must be defined with some care (which is where the appropriately normalized wave packets come in).

B.2.1 Time-dependent perturbation theory

An approximate expression for $\mathcal{S}$ in powers of $V = H - H_0$ can be obtained using standard steps of time-dependent perturbation theory. To this end notice that $\Omega^*(t)\,\Omega(t') = \mathrm{e}^{\mathrm{i}H_0 t}\,\mathrm{e}^{-\mathrm{i}H(t-t')}\,\mathrm{e}^{-\mathrm{i}H_0 t'}$ satisfies the differential equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}[\Omega^*(t)\,\Omega(t')] = \mathrm{e}^{\mathrm{i}H_0 t}(H - H_0)\,\mathrm{e}^{-\mathrm{i}H(t-t')}\,\mathrm{e}^{-\mathrm{i}H_0 t'} = V(t)\,\Omega^*(t)\,\Omega(t')\,, \tag{B.32}$$

where this last equality defines the interaction picture operator $V(t) := \mathrm{e}^{\mathrm{i}H_0 t} V\,\mathrm{e}^{-\mathrm{i}H_0 t}$. This can be solved iteratively, leading to the following expression for $\mathcal{S} = \lim_{\substack{t\to+\infty \\ t'\to-\infty}} \Omega^*(t)\Omega(t')$:

$$\begin{aligned}\mathcal{S} &= \sum_{n=0}^{\infty}(-\mathrm{i})^n \int_{-\infty}^{\infty}\mathrm{d}\tau_1 \int_{-\infty}^{\tau_1}\mathrm{d}\tau_2 \cdots \int_{-\infty}^{\tau_{n-1}}\mathrm{d}\tau_n\, V(\tau_1)V(\tau_2)\cdots V(\tau_n) \\ &= \sum_{n=0}^{\infty}\frac{(-\mathrm{i})^n}{n!}\int_{-\infty}^{\infty}\mathrm{d}^4x_1 \cdots \int_{-\infty}^{\infty}\mathrm{d}^4x_n\; T[\mathfrak{H}(x_1)\cdots\mathfrak{H}(x_n)]\;,\end{aligned} \tag{B.33}$$

which uses $V(t) = \int \mathrm{d}^3x\; \mathfrak{H}(x)$ and introduces the time-ordering operation $T[O_1(t)O_2(t')] = \Theta(t-t')O_1(t)O_2(t') + \Theta(t'-t)O_2(t')O_1(t)$ (where Θ denotes a step function).

Using momentum eigenstates, for which $\langle\beta|O(x)|\alpha\rangle = \mathrm{e}^{\mathrm{i}(p_\alpha - p_\beta)\cdot x}\langle\beta|O(0)|\alpha\rangle$, the S-matrix can therefore be written

$$S_{\beta\alpha} = \delta_{\beta\alpha} - \mathrm{i}\mathcal{M}_{\beta\alpha}(2\pi)^4\delta^4(p_\beta - p_\alpha) \tag{B.34}$$

with

$$\mathcal{M}_{\beta\alpha} = \langle\beta|\mathfrak{H}_I(0)|\alpha\rangle - \frac{\mathrm{i}}{2!}\int \mathrm{d}^4x\langle\beta|T\left[\mathfrak{H}_I(x)\mathfrak{H}_I(0)\right]|\alpha\rangle + \cdots\,. \tag{B.35}$$

B.2.2 Transition rates

The expressions for the S-matrix are proportional to an energy-conserving (and often momentum-conserving) delta function when expressed in terms of energy eigenstates rather than wave packets. This means that the square of S-matrix elements – what should be the transition probabilities – are proportional to $[\delta(E)]^2 = \delta(E)\,\delta(0)$ and so must diverge. Physically, this divergence arises because energy eigenstates are an idealization of the wave packets that scattering really involves. The convenience of using scattering energy eigenstates to compute the S-matrix has as its price the necessity to more carefully sort out the relationship between physical quantities and S-matrix elements.

Going back to the wave-packet description, $|\,\phi_f\rangle\!\rangle_{\mathrm{i}} = \int \mathrm{d}\alpha\, f(\alpha)|\alpha\rangle\!\rangle_{\mathrm{i}}$, the probability of finding the system in the final state labeled by β becomes

$$P_f(\beta) = |{}_{\mathrm{o}}\langle\!\langle\beta|\phi_f\rangle\!\rangle_{\mathrm{i}}|^2 = \int \mathrm{d}\alpha\,\mathrm{d}\alpha'\, f^*(\alpha')f(\alpha)\;{}_{\mathrm{o}}\langle\!\langle\beta|\alpha\rangle\!\rangle_{\mathrm{i}}\;{}_{\mathrm{i}}\langle\!\langle\alpha'|\beta\rangle\!\rangle_{\mathrm{o}}\,. \tag{B.36}$$

In practice, the packet $f(\alpha)$ is peaked about some value $\overline{\alpha}$ with a width about this value that is small compared with experimental resolutions but not so small as to run into trouble from the uncertainty relations. It is also usually true that $S_{\beta\alpha}$ does not have a strong dependence

on α in the regime over which $f(\alpha)$ has its support. Concretely, the energy width of a wave packet is usually small compared with the energy dependence of the scattering cross section. (If this were not true the experiment would do a poor job measuring the S-matrix.)

Under these circumstances – and assuming β is distinguishable from all of the α in the support of $f(\alpha)$, so $S_{\beta\alpha} \simeq -\mathrm{i}\mathcal{M}_{\beta\alpha}(2\pi)^4\delta^4(p_\beta - p_\alpha)$ – then (B.36) factorizes,

$$P_f(\beta) \approx |\mathcal{M}_{\beta\overline{\alpha}}|^2 \int [\mathrm{d}\alpha]\,[\mathrm{d}\alpha']\, f^*(\alpha')f(\alpha) \tag{B.37}$$

where $\mathrm{d}\alpha\,(2\pi)^4\delta^4(p_\alpha - p_\beta) = [\mathrm{d}\alpha]$, and similarly for $[\mathrm{d}\alpha']$. The delta functions enforcing energy-momentum conservation are no longer a problem because they are used to perform part of the integration over α and α'.

The finite-volume/finite-time trick

What is important about (B.37) is its factorization of reaction probabilities into an interaction part, $|\mathcal{M}_{\beta\overline{\alpha}}|^2$, and a part involving precisely how the initial wave-packet is set up. Given this factorization, it would be useful to identify the interaction part as efficiently as possible without having to set up the wave-packets in detail each time.

A trick for doing so is to directly compute $S_{\beta\alpha}$ with the system imagined to be inside a box with large but finite volume $\mathcal{V}$, and allowing the interactions to last only over a large but finite time interval, T. When this is true probabilities can be computed as in ordinary quantum mechanics by squaring the relevant transition amplitude, and the result's dependence on $\mathcal{V}$ and T can then be studied to identify what remains physical in the limit $\mathcal{V}, T \to \infty$. Once such a quantity is identified, the temporary theoretical contrivance of the box and interval can be dropped.

Conventions for discrete and continuum normalizations for momentum eigenstates are summarized in §B.1. Following the discussion there, states using discrete normalization for momenta, $(\mathbf{p}|\mathbf{q}) = \delta_{\mathbf{pq}}$, are denoted $|\,\alpha)$, those using nonrelativistic continuum normalization, $\langle\mathbf{p}|\mathbf{q}\rangle = \delta^3(\mathbf{p}-\mathbf{q})$, are denoted $|\,\alpha\rangle$, while those using relativistic normalization, ${}_r\langle\mathbf{p}|\mathbf{q}\rangle_r = 2E_p\,\delta^3(\mathbf{p}-\mathbf{q})$, are denoted $|\,\alpha\rangle_r$.

For a state involving N_α particles in a large but finite box of volume, $\mathcal{V}$, these states have the relative normalization

$$|\,\alpha\rangle = \left[\frac{\mathcal{V}}{(2\pi)^3}\right]^{N_\alpha/2} |\,\alpha)] \quad \text{and} \quad |\,\alpha\rangle_r = \left[\frac{2\overline{E}\mathcal{V}}{(2\pi)^3}\right]^{N_\alpha/2} |\,\alpha)]\,, \tag{B.38}$$

where $(2\overline{E})^{N_\alpha/2}$ is shorthand for the product $\prod_{i=1}^{N_\alpha}\sqrt{2E_i}$. The corresponding S-matrix elements, $S^{\text{cont}}_{\beta\alpha} = \langle\beta|\mathcal{S}|\,\alpha\rangle$, $S^{\text{disc}}_{\beta\alpha} = (\beta|\mathcal{S}|\,\alpha)$ and $S^{\text{rel}}_{\beta\alpha} = {}_r\langle\beta|\mathcal{S}|\,\alpha\rangle_r$, are therefore related by

$$S^{\text{cont}}_{\beta\alpha} = \left[\frac{\mathcal{V}}{(2\pi)^3}\right]^{(N_\alpha+N_\beta)/2} S^{\text{disc}}_{\beta\alpha} \quad \text{and} \quad S^{\text{rel}}_{\beta\alpha} = \left[\frac{2\overline{E}\mathcal{V}}{(2\pi)^3}\right]^{(N_\alpha+N_\beta)/2} S^{\text{disc}}_{\beta\alpha}\,. \tag{B.39}$$

For translationally invariant theories the S-matrix in each case differs from unity by

$$S_{\beta\alpha} = \delta_{\beta\alpha} - \mathrm{i}\mathcal{M}_{\beta\alpha}(2\pi)^4\delta^4_{\mathcal{V}T}(\mathbf{p}_\beta - \mathbf{p}_\alpha)\,, \tag{B.40}$$

where the finite-volume delta functions arise when evaluating in the form,

$$(2\pi)^4\delta^4_{\mathcal{V}T}(p_\alpha - p_\beta) := \int_{\mathcal{V}T} \mathrm{d}^4x\, \mathrm{e}^{\mathrm{i}(p_\alpha - p_\beta)\cdot x}\,, \tag{B.41}$$

with the spatial integration over the volume, $\mathcal{V}$, and the time integration runs from $-T/2$ to $+T/2$. As $\mathcal{V}T \to \infty$ the quantity $\delta_{\mathcal{V}T}$ goes to the standard delta-function that enforces energy-momentum conservation.

Now comes the main point. Transition probabilities involve $|S_{\beta\alpha}|^2$ and so necessarily also contain a factor of $\delta^4_{\mathcal{V}T}(0)$. But direct evaluation of (B.41) shows that for finite T and $\mathcal{V}$ this factor is $(2\pi)^4\delta_{\mathcal{V}T}(0) = \mathcal{V}T$. The $\delta_{\mathcal{V}T}(0)$ divergence in $|S_{\beta\alpha}|^2$ is thereby seen to have a physical origin. In a system invariant under translations in space and time (*i.e.* precisely those that generate energy-momentum conserving delta functions), the instantaneous transition probability is the same everywhere and at every instant, making the total rate scale with the volume and time interval over which the initial wave-packets overlap. But this volume and time interval are infinite if the states involved are chosen for convenience to be momentum and energy eigenstates.

This diagnosis suggests a remedy:[1] it is the transition rate per unit volume per unit time that is well-behaved as $\mathcal{V}, T \to \infty$, so this is what should be computed in the continuum limit. Suppose the initial particles have some probability, $P(\alpha)$, to be in a particular region of phase space containing $\Delta\alpha$ states and small enough that $P(\alpha)$ is approximately constant within it. Using discrete normalization the total transition probability for scattering from $\Delta\alpha$ to a similar small region of final-state particles, $\Delta\beta$, is then simply $\mathcal{P}(\alpha \to \beta) = |S^{\mathrm{disc}}_{\beta\alpha}|^2 P(\alpha)\Delta\alpha\Delta\beta$. Since the density of states in momentum space is $\mathcal{V}/(2\pi)^3$, the total number of states in an interval $\mathrm{d}\beta = \prod_{i=1}^{N_\beta} \mathrm{d}^3p_i$ for an N_β-particle state is $\Delta\beta = [\mathcal{V}/(2\pi)^3]^{N_\beta}\,\mathrm{d}\beta$, and similarly for $\Delta\alpha$. The differential transition rate per-unit-time then becomes

$$\begin{aligned}
\mathrm{d}\Gamma(\alpha \to \beta) := \frac{\mathrm{d}\mathcal{P}(\alpha \to \beta)}{T} &= \frac{|S^{\mathrm{disc}}_{\beta\alpha}|^2}{T}\; P(\alpha)\,\Delta\alpha\Delta\beta \\
&= \frac{f(\alpha)}{T}\left[|S^{\mathrm{cont}}_{\beta\alpha}|^2\left(\frac{\mathcal{V}}{(2\pi)^3}\right)^{-(N_\alpha+N_\beta)}\right]\left(\frac{\mathcal{V}}{(2\pi)^3}\right)^{(N_\beta+N_\alpha)} \mathrm{d}\alpha\,\mathrm{d}\beta \\
&= \mathcal{V}f(\alpha)\,|\mathcal{M}^{\mathrm{cont}}_{\beta\alpha}|^2\,(2\pi)^4\delta^4_{\mathcal{V}T}(p_\beta - p_\alpha)\,\mathrm{d}\alpha\,\mathrm{d}\beta \\
&= \mathcal{V}f(\alpha)\,|\widehat{\mathcal{M}}^{\mathrm{cont}}_{\beta\alpha}|^2\,(2\pi)^4\delta^4_{\mathcal{V}T}(p_\beta - p_\alpha)\,\mathrm{d}\hat\alpha\,\mathrm{d}\hat\beta \\
&= \mathcal{V}f(\alpha)\,|\widehat{\mathcal{M}}^{\mathrm{rel}}_{\beta\alpha}|^2\,(2\pi)^4\delta^4_{\mathcal{V}T}(p_\beta - p_\alpha)\,\mathrm{d}\hat\alpha_r\,\mathrm{d}\hat\beta_r\,,
\end{aligned} \tag{B.42}$$

where the replacement $P(\alpha) \to f(\alpha)$ follows the argument leading to (B.14) — where

$$f(\alpha) = \prod_{j\in\alpha} f_j(\alpha)\,, \tag{B.43}$$

is the joint phase-space probability density of initial particles, assumed to be independent

[1] §16.4 provides a more sophisticated version of this same remedy, at least for the growth with T. See also the discussion leading to (16.120).

of one another. The new notation is

$$\mathrm{d}\hat{\alpha} := \prod_{j\in\alpha} \frac{\mathrm{d}^3 p_j}{(2\pi)^3}\,, \quad \mathrm{d}\hat{\alpha}_r := \prod_{j\in\alpha} \frac{\mathrm{d}^3 p_j}{(2\pi)^3 2E_j}\,, \quad \mathrm{d}\hat{\beta} := \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3}\,, \quad \mathrm{d}\hat{\beta}_r := \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3 2E_j}\,, \tag{B.44}$$

and $\widehat{\mathcal{M}}_{\beta\alpha} := (2\pi)^{3(N_\alpha+N_\beta)/2}\mathcal{M}_{\beta\alpha}$. $\widehat{\mathcal{M}}_{\beta\alpha}$ is a natural quantity to define because the factor of $(2\pi)^{3/2}$ for each particle cancels the $(2\pi)^{-3/2}$ appearing in the expansion of all fields in terms of creation and annihilation operators – see for instance eqs. (C.28), (C.30), (C.33) or (C.39) – since these appear systematically when evaluating matrix elements like those appearing in (B.35).

Fermi's Golden Rule

The special case where only the leading term of the transition amplitude of (B.35) dominates the rate of (B.42) leads to the very useful formula

$$\mathrm{d}\Gamma(\alpha\to\beta) = \mathcal{V} f(\alpha)\, |\langle\beta|\mathfrak{H}_I(0)|\alpha\rangle|^2\, (2\pi)^4 \delta^4_{\mathcal{V}T}(p_\beta - p_\alpha)\, \mathrm{d}\alpha\, \mathrm{d}\beta\,, \tag{B.45}$$

known as Fermi's Golden Rule.

Decays: $N_\alpha = 1$

Expression (B.42) can be directly used – right out of the box, so to speak – in the special case $N_\alpha = 1$, since in this case $\mathcal{V} f(\alpha)\,\mathrm{d}\hat{\alpha} = \mathcal{V} f(\mathbf{p})\,\mathrm{d}^3 p/(2\pi)^3$ is the probability of the initial particle being in the initial momentum region $\mathrm{d}^3 p$, making the limit $\mathcal{V}, T \to \infty$ a trivial one. The resulting expression for the decay rate of a single particle then is

$$\mathrm{d}\Gamma(\alpha\to\beta) = \frac{1}{2E_\alpha} |\widehat{\mathcal{M}}^{\mathrm{rel}}_{\beta\alpha}|^2 (2\pi)^4 \delta^4(p_\alpha - p_\beta)\, \mathrm{d}\hat{\beta}_r\,. \tag{B.46}$$

This result is not quite Lorentz-invariant, because of the $1/(2E_\alpha)$ in front. This factor is just what is needed to provide the time-dilation of a fast-moving particle's decay lifetime.

Scattering: $N_\alpha \geq 2$

More generally, for $N_\alpha \geq 2$ the scattering rate per-unit-volume in the large-volume, late-time limit for initially uncorrelated particles is

$$\begin{aligned}\frac{\mathrm{d}\Gamma(\alpha\to\beta)}{\mathcal{V}} &= \left[\prod_{k\in\alpha} f_k(\mathbf{p}_k)\, \frac{\mathrm{d}^3 p_k}{(2\pi)^3}\right] |\widehat{\mathcal{M}}^{\mathrm{cont}}_{\beta\alpha}|^2\, (2\pi)^4 \delta^4(p_\beta - p_\alpha) \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3} \\ &= \left[\prod_{k\in\alpha} f_k(\mathbf{p}_k)\, \frac{\mathrm{d}^3 p_k}{(2\pi)^3 2E_k}\right] |\widehat{\mathcal{M}}^{\mathrm{rel}}_{\beta\alpha}|^2\, (2\pi)^4 \delta^4(p_\beta - p_\alpha) \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3 2E_j}\,,\end{aligned} \tag{B.47}$$

which (reasonably) is proportional to the phase-space density of each particle in the initial state. The second line makes the Lorentz-transformation properties of this rate manifest.

Using again that $\mathcal{V} f(\alpha)\,\mathrm{d}\hat{\alpha} = \mathcal{V} f(\mathbf{p})\,\mathrm{d}^3 p/(2\pi)^3$ is the probability of an initial particle

being in the initial momentum region d^3p, expression (B.47) states that the interaction rate per particle, for particle 'B' (for 'beam'), is

$$\begin{aligned} \mathrm{d}\Gamma_B &= (2\pi)^3 \frac{\mathrm{d}\Gamma(\alpha \to \beta)}{\mathcal{V} f_B \, \mathrm{d}^3 p_B} \\ &= \frac{1}{2E_B} \left[\prod_{\substack{k\in\alpha \\ k\neq B}} f_k(\mathbf{p}_k) \frac{\mathrm{d}^3 p_k}{(2\pi)^3 2E_k} \right] |\widehat{\mathcal{M}}^{\mathrm{rel}}_{\beta\alpha}|^2 \, (2\pi)^4 \delta^4(p_\beta - p_\alpha) \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3 2E_j} \,, \end{aligned} \tag{B.48}$$

Two-body scattering: $N_\alpha = 2$

In the special case of two-particle scattering, $N_\alpha = 2$, the product over initial particles just involves the 'other' (non-beam) particle, denoted 'T' (for 'target'):

$$\mathrm{d}\Gamma_B = \frac{f_T(\mathbf{p}_T)}{(2\pi)^3} \left[\frac{\mathrm{d}^3 p_T}{4E_B E_T} \right] |\widehat{\mathcal{M}}^{\mathrm{rel}}_{\beta\alpha}|^2 \, (2\pi)^4 \delta^4(p_\beta - p_\alpha) \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3 2E_j} \,, \tag{B.49}$$

and so depends explicitly on the density of target particles. It is conventional to normalize out the target-dependent factors and define the cross section by $\mathrm{d}\sigma = \mathrm{d}\Gamma_B / F$ where F satisfies two conditions. First, the Lorentz-transformation properties of F are fixed by requiring $\mathrm{d}\sigma$ to be Lorentz invariant. Inspection of (B.49) shows this implies

$$F = \left(\frac{f_T(\mathbf{p}_T)}{(2\pi)^3} \right) \frac{\mathcal{F} \mathrm{d}^3 p_T}{4E_T E_B} \,, \tag{B.50}$$

where $\mathcal{F}$ is a Lorentz invariant quantity. Second, F should evaluate to the particle flux (as seen by the beam particle), $\mathrm{d}n_T v_{\mathrm{rel}}$, when evaluated in the target-particle rest frame. Here $\mathrm{d}n_T = f_T \, \mathrm{d}^3 p_T / (2\pi)^3$ is the target's ordinary-space particle density and v_{rel} is the (Lorentz-invariant) relative velocity of the initial two particles,

$$v_{\mathrm{rel}} = \sqrt{1 - \frac{m_B^2 \, m_T^2}{(p_B \cdot p_T)^2}} \,, \tag{B.51}$$

where p_B^μ and p_T^μ are the 4-momenta of the two initial particles. The resulting two-body cross section becomes

$$\begin{aligned} \mathrm{d}\sigma(\alpha \to \beta) &= \frac{|\widehat{\mathcal{M}}^{\mathrm{rel}}_{\beta\alpha}|^2}{\mathcal{F}} (2\pi)^4 \delta^4(p_\alpha - p_\beta) \prod_{j\in\beta} \frac{\mathrm{d}^3 p_j}{(2\pi)^3 2E_j} \\ \text{with} \quad \mathcal{F} &:= (-4 p_B \cdot p_T) v_{\mathrm{rel}} = 4 \sqrt{(p_B \cdot p_T)^2 - m_B^2 \, m_T^2} \,. \end{aligned} \tag{B.52}$$

C Appendix C Quantum Field Theory: a Cartoon

Quantum field theory (QFT) plays a central role in most areas of theoretical physics, but this is not really a deep statement. At one level quantum field theory is merely ordinary quantum mechanics applied to processes that change the total number of particles present, and this makes it particularly useful for relativistic systems since fundamental principles (the consistency of quantum mechanics and special relativity) forbid relativistic interactions from ever leaving the total number of particles unchanged.

But the utility of QFT methods are not restricted to processes that change the total number of particles. It is also useful when framing quantum systems with a fixed number of particles, largely because its main feature — the language of creation and annihilation operators — lends itself to efficiently expressing natural laws in such a way that ensures the validity of a few fundamental principles right from the start. The principles that get baked in in this way include 'unitarity' (which is to say, conservation of probability in quantum evolution) and 'cluster decomposition' (which means the factorization of probabilities for independent events when these events are causally separated from one another in space-time).

Although this book is not meant as a textbook on quantum field theory, QFT tools are nonetheless often used within these pages. Consequently this Appendix is offered, both as a quick refresher on some elements of quantum theory of fields, as well as a way to collect together some of the main useful formulae used elsewhere. Although this summary possibly provides a useful reminder for those already with some QFT background, it likely has insufficient detail to teach the subject to a complete newbie.

C.1 Creation and annihilation operators

The goal in quantum field theory is to set up a quantum mechanical framework in which the total number of particles can change. The first step when doing so is to identify the Hilbert space within which quantum operators act.

In elementary single-particle quantum mechanics a basis of states, $|i\rangle$, for the single-particle Hilbert space, $\mathcal{H}_1 = \{|i\rangle\}$, can be chosen consisting of eigenstates of the complete set of commuting observables that label single-particle states. In concrete examples these labels are often chosen to be momentum and any internal quantum numbers, like total spin, s, and its third component, σ: so $|i\rangle = |\mathbf{p}, s, \sigma, \cdots\rangle$.

Ordinary quantum mechanics involving N particles similarly involves a Hilbert space, $\mathcal{H}_N = \{|i_1, i_2, \cdots i_N\rangle\}$, built as products of N copies of the single-particle basis states. For

bosons these states are defined to be completely symmetric in the interchange of any two pairs of labels for identical particles, while the states are antisymmetric under this interchange for fermions. For example, for a two-particle state

$$|i_1, i_2\rangle = \pm|i_2, i_1\rangle \tag{C.1}$$

where the upper (lower) sign corresponds to the particles being bosons (fermions).

The Hilbert space for quantum field theory, $\mathcal{H}$, combines the Hilbert spaces $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$ and so on, up to $\mathcal{H}_N$ and beyond, with N arbitrarily large. Here $\mathcal{H}_0 = \{|0\rangle\}$ is the one-dimensional space spanned by the zero-particle state, $|0\rangle$, while $\mathcal{H}_N$ for $N \geq 1$ is defined as above. A Hilbert space constructed in this way is called a Fock space.

When dealing with different kinds of particles it is useful to label states using the *occupation-number representation*. Instead of listing the single-particle labels for all particles in the state, this representation lists the number of *independent* particle labels, together with the number of particles present in the state that carry these labels. For instance, in the occupation-number representation a state containing two particles with momentum $\mathbf{p}$, is denoted $|\mathbf{p}^{(2)}\rangle$ rather than $|\mathbf{p}, \mathbf{p}\rangle$. For general labels '$i$', the occupation-number representation for a five-particle state contining two particles having single-particle quantum number i_1 and three particles with quantum number i_2 is similarly $|i_1^{(2)}, i_2^{(3)}\rangle$ rather than $|i_1, i_1, i_2, i_2, i_2\rangle$.

A very convenient basis of operators acting within the Fock space of quantum field theory is given by *creation* and *annihilation* operators in the following way. The annihilation operator, $\mathfrak{a}_i$ is defined as the operator that removes a particle with quantum number i from a given state. If the state on which $\mathfrak{a}_i$ acts does not contain the particle in question then the operator is defined to give zero. That is,

$$\mathfrak{a}_i|0\rangle = 0\,, \qquad \mathfrak{a}_i|j\rangle = \delta_{ij}|0\rangle\,, \qquad \mathfrak{a}_i|j,k\rangle = \delta_{ij}|k\rangle + (-)^{ij}\delta_{ik}|jl\rangle\,, \tag{C.2}$$

and so on, where the sign $(-)^{ij}$ is -1 if both particles 'i' and 'j' are fermions and is $+1$ otherwise.

This definition implies that the Hermitian conjugate, $\mathfrak{a}_i^*$, is a *creation operator* for the same particle type, that satisfies

$$\mathfrak{a}_i^*|0\rangle = |i\rangle\,, \qquad \mathfrak{a}_i^*|j\rangle = |i,j\rangle\,, \tag{C.3}$$

and so on. Together with the normalization convention $\langle i|j\rangle = \delta_{ij}$, these definitions imply the following properties. For bosons $|i,j\rangle = |j,i\rangle$, and so

$$\left[\mathfrak{a}_i, \mathfrak{a}_j\right] = \left[\mathfrak{a}_i^*, \mathfrak{a}_j^*\right] = 0 \quad \text{and} \quad \left[\mathfrak{a}_i, \mathfrak{a}_j^*\right] = \delta_{ij}\,. \tag{C.4}$$

For fermions $|i,j\rangle = -|j,i\rangle$, and so

$$\{\mathfrak{a}_i, \mathfrak{a}_j\} = \{\mathfrak{a}_i^*, \mathfrak{a}_j^*\} = 0 \quad \text{and} \quad \{\mathfrak{a}_i, \mathfrak{a}_j^*\} = \delta_{ij}\,, \tag{C.5}$$

in which $[A,B] = AB - BA$ and $\{A,B\} = AB + BA$, as usual.

In the occupation-number representation the above rules are captured by

$$\mathfrak{a}_i|i_1^{(n_1)}, \cdots, i_r^{(n_r)}\rangle = \sum_{j=1}^{r} s_{ij}\delta_{ii_j}\sqrt{n_j}\,|i_1^{(n_1)}, \cdots, i_j^{(n_j-1)}, \cdots, i_r^{(n_r)}\rangle\,. \tag{C.6}$$

where $s_{ij} = (-)^{i(i_1+\cdots+i_{j-1})}$, and

$$\begin{aligned} \mathfrak{a}_i^*|i_1^{(n_1)},\cdots,i_r^{(n_r)}\rangle &= |i^{(1)},i_1^{(n_1)},\cdots,i^{(n_i+1)},\cdots,i_r^{(n_r)}\rangle \qquad \text{if } i\neq i_j \text{ for any } j \\ &= \sqrt{n_{i_j}+1}\; s_{ij}\,|i_1^{(n_1)},\cdots,i_j^{(n_i+1)},\cdots,i_r^{(n_r)}\rangle \qquad \text{if } i=i_j\,. \end{aligned} \tag{C.7}$$

In particular $\mathfrak{a}_i^*\mathfrak{a}_i$ counts the number of particles with quantum number 'i', because the previous two formulae imply

$$\mathfrak{a}_i^*\mathfrak{a}_i|i_1^{(n_1)},\cdots,i_r^{(n_r)}\rangle = \left(\sum_{j=1}^{r}\delta_{ii_j}n_j\right)|i_1^{(n_1)},\cdots,i_r^{(n_r)}\rangle\,. \tag{C.8}$$

What is important about the creation and annihilation operators is that they make a very convenient basis, in terms of which *any* operator acting on $\mathcal{H}$, can be expanded:[1]

$$\begin{aligned} O = A_{0,0} &+ \sum_i \left[A_{0,1}(i)\;\mathfrak{a}_i + A_{1,0}(i)\;\mathfrak{a}_i^*\right] \\ &+ \sum_{ij}\left[A_{0,2}(i,j)\;\mathfrak{a}_i\mathfrak{a}_j + A_{1,1}(i,j)\mathfrak{a}_i^*\mathfrak{a}_j + A_{2,0}(i,j)\;\mathfrak{a}_i^*\mathfrak{a}_j^*\right] + \cdots\,. \end{aligned} \tag{C.9}$$

To see that this is so it suffices to argue that the coefficient functions $\{A_{0,0}, A_{1,0}(i), A_{0,1}(i), \ldots\}$ can be solved for in terms of the matrix elements: $\langle\psi|O|\,\phi\rangle$, for all choices for $\langle\psi|$ and $|\,\phi\rangle$. (This can be shown by induction, starting with $\langle 0|O|0\rangle = A_{0,0}$ (because $\mathfrak{a}_i|0\rangle = \langle 0|\mathfrak{a}_i^* = 0$) and continuing with $\langle j|O|0\rangle = A_{1,0}(j)$, $\langle 0|O|j\rangle = A_{0,1}(j)$ and so on.)

A system's hamiltonian is an important special case of an operator that can be expanded in this way, and part of the reason creation and annihilation operators are so useful is that this particular expansion is usually an efficient one. For instance, non-interacting particles are ones for which the energy cost of adding N particles is just N times the energy of adding one particle (*i.e.* there is no interaction energy). With (C.8) in mind the hamiltonian for a collection of non-interacting particles is therefore

$$H_{\text{free}} = E_0 + \sum_i \varepsilon_i\,\mathfrak{a}_i^*\mathfrak{a}_i\,, \tag{C.10}$$

where the single-particle labels 'i' appearing in the sum are a complete set of mutually commuting labels for single-particle energy eigenstates.

As is easily verified by explicit evaluation, the hamiltonian H_{free} has the occupation-number states as eigenfunctions, with eigenvalues given by

$$H_{\text{free}}|i_1^{(n_1)},\cdots,i_r^{(n_r)}\rangle = \left(E_0 + \sum_j n_j\varepsilon_j\right)|i_1^{(n_1)},\cdots,i_r^{(n_r)}\rangle\,. \tag{C.11}$$

This reveals E_0 to be the energy of the zero-particle state (or vacuum), $|0\rangle$, while ε_i is the energy associated with the addition of a single particle having quantum number 'i'. For momentum eigenstates (using standard non-relativistic normalization) this expression for

[1] Notice that all instances of $\mathfrak{a}^*$ here stand to the left of all instances of $\mathfrak{a}$ in this expression; something that can be arranged without loss of generality by changing, if needed, the order of operators using the commutation relations (C.4) or (C.5) (a process called 'normal ordering').

$H_{\rm free}$ becomes

$$H_{\rm free} = E_0 + \sum_k \int \mathrm{d}^3 p\, \varepsilon_k(\mathbf{p})\, \mathfrak{a}^*_{\mathbf{p},k} \mathfrak{a}_{\mathbf{p},k}\,, \tag{C.12}$$

where k represents all non-momentum single-particle labels (like spin or particle type) and $\varepsilon_k(\mathbf{p})$ is the single-particle dispersion relation (*i.e.* single-particle energy as a function of momentum). In the special case of relativistic systems Lorentz covariance requires $\mathbf{p}$ and $p^0 = \varepsilon(\mathbf{p})$ must be components of a single 4-momentum vector, p^μ, so $\varepsilon_k(\mathbf{p}) = \sqrt{\mathbf{p}^2 + m_k^2}$ where m_k is the corresponding particle's rest mass.

Interactions have similar representations. For instance a term in $H_{\rm int}$ describing the emission or absorption of a photon by a charged particle, $f(\mathbf{p},\sigma) + \gamma(\mathbf{k},\lambda) \to f(\mathbf{q},\zeta)$, could be written

$$H_{\rm int} \ni \sum_{\lambda\sigma\zeta} \int \mathrm{d}^3 p\, \mathrm{d}^3 q\, \mathrm{d}^3 k \left[h_{\lambda\sigma\zeta}(\mathbf{p},\mathbf{q},\mathbf{k})\, \mathfrak{c}^*_{\mathbf{p}\sigma}\, \mathfrak{c}_{\mathbf{q}\zeta}\, \mathfrak{a}_{\mathbf{k}\lambda} + \text{h.c.} \right] \delta^3(\mathbf{p}-\mathbf{q}-\mathbf{k})\,, \tag{C.13}$$

where 'h.c.' denotes the hermitian conjugate, while $\mathfrak{c}_{\mathbf{p}\sigma}$ denotes the annihilation operator for a charged particle with momentum $\mathbf{p}$ and spin component σ, while $\mathfrak{a}_{\mathbf{k}\lambda}$ is the same for the photon of momentum $\mathbf{k}$ and helicity λ.

The above expression can mediate processes like photon absorption (or, from the 'h.c.' term, emission) because it gives a nonzero matrix element

$$\langle \mathbf{p},\sigma | H_{\rm int} | \mathbf{q},\zeta;\mathbf{k},\lambda\rangle = h_{\lambda\sigma\zeta}(\mathbf{p},\mathbf{q},\mathbf{p})\, \delta^3(\mathbf{p}-\mathbf{q}-\mathbf{k})\,, \tag{C.14}$$

which, when used in expressions like Fermi's golden rule (B.45), can contribute a nonzero transition rate. Eq. (C.14) evaluates the matrix element using (continuum-normalized version of) expressions (C.6) and (C.7) defining the action of creation and annihilation operators.

C.2 Nonrelativistic free fields

An important condition satisfied by most physical systems is cluster decomposition: the factorization of probabilities for events that are widely separated in space at a given time.[2] A physical property demanded of H is that this clustering property should be preserved by time evolution.

What is convenient about the creation- and annihilation-operator formalism is that the requirements of cluster decomposition are automatically satisfied if the (possibly complex) coefficient functions in H – like $h_{\lambda\sigma\zeta}(\mathbf{p},\mathbf{q},\mathbf{k})$ in (C.13), for example – are sufficiently smooth functions of their momentum arguments (*e.g.* admitting a Taylor expansion in powers of momenta) once the delta-function is extracted that enforces momentum conservation

[2] This property assumes the system to have started without initial correlations between the particles involved in these events. For relativistic systems the requirement of 'large' spatial separation means events that are sufficiently outside each other's light cones (large *spacelike* separations). For thermal fluids cluster decomposition is included in the condition of local equilibrium.

(if this is conserved). These are the momentum-space ways of saying that in position space the hamiltonian is local:

$$H = \int \mathrm{d}^3x\, \mathfrak{H}(\mathbf{x})\,, \tag{C.15}$$

for some hamiltonian density $\mathfrak{H}(\mathbf{x})$. Locality is related to cluster decomposition because the time-evolution operator, $U(t, t_0) = \exp[-\mathrm{i}H(t - t_0)]$, should schematically be a product over positions (and so should preserve factorization of amplitudes for spatially separated events) when H is a sum over positions.

This suggests that interactions should often simplify when expressed in position space, using the Fourier-transforms of the creation and annihilation operators. For spinless particles this leads to defining position-space fields, $\Phi(\mathbf{x})$, of the form

$$\Phi(\mathbf{x}) := \int \frac{\mathrm{d}^3p}{(2\pi)^{3/2}}\; \mathfrak{a}_{\mathbf{p}}\, e^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}}\,, \tag{C.16}$$

Such a position-space quantum field can similarly be defined for the destruction operator of each separate type of particle. This last equation is written in Schrödinger representation (for which operators do not evolve in time), but it is often more usefully written in the Interaction representation (for which the operators evolve in time using the free field equations). In interaction picture (C.16) becomes

$$\Phi(\mathbf{x}, t) := \int \frac{\mathrm{d}^3p}{(2\pi)^{3/2}}\; \mathfrak{a}_{\mathbf{p}}\, e^{-\mathrm{i}E_p t + \mathrm{i}\mathbf{p}\cdot\mathbf{x}}\,, \tag{C.17}$$

where E_p is the single-particle energy.

The point of expressing H in terms of $\Phi(\mathbf{x})$ and $\Phi^*(\mathbf{x})$ rather than $\mathfrak{a}_{\mathbf{p}\sigma}$ and $\mathfrak{a}^*_{\mathbf{p}\sigma}$ is that the condition of cluster-decomposition — *i.e.* smoothness of coefficients like $h_{\sigma\lambda\zeta}(\mathbf{p}, \mathbf{q}, \mathbf{k})$ in (C.13) — is ensured by the locality requirement, (C.15), where $\mathfrak{H}(\mathbf{x})$ is built from sums of local monomials of $\Phi(\mathbf{x})$, $\Phi^*(\mathbf{x})$ and their derivatives, all evaluated at the same spatial point.

For example, rather than using (C.12) for a system of non-interacting spinless particles with dispersion relation $\varepsilon(\mathbf{p}) = \mathbf{p}^2/2m$, its Hamiltonian could equally well be written

$$H = \frac{1}{2m} \int \mathrm{d}^3x\, \nabla\Phi^*(\mathbf{x}) \cdot \nabla\Phi(\mathbf{x})\,. \tag{C.18}$$

Because this hamiltonian is invariant under rephasing, $\Phi(\mathbf{x}) \to e^{\mathrm{i}\omega}\Phi(\mathbf{x})$, Noether's theorem implies that it necessarily commutes with total particle number, measured by the operator

$$N = \int \mathrm{d}^3x\, \Phi^*(\mathbf{x})\, \Phi(\mathbf{x}) = \int \mathrm{d}^3p\; \mathfrak{a}^*_{\mathbf{p}}\mathfrak{a}_{\mathbf{p}}\,. \tag{C.19}$$

Sometimes momentum is not a good single-particle label for energy eigenstates, such as for particles that interact with an external potential, $V(\mathbf{x})$. For example, a system of such particles interacting only with the potential (which do not mutually interact with one another) can be written

$$H = \int \mathrm{d}^3x \left[\frac{1}{2m}\, \nabla\Phi^*(\mathbf{x}) \cdot \nabla\Phi(\mathbf{x}) + V(\mathbf{x})\, \Phi^*(\mathbf{x})\Phi(\mathbf{x})\right]. \tag{C.20}$$

As is easy to verify the hamiltonian of eq. (C.20) can be written in the diagonal form (C.11) by generalizing (C.16) to

$$\Phi(\mathbf{x}) := \sum_i \mathfrak{a}_i \, \varphi_i(\mathbf{x}) \,, \tag{C.21}$$

where the mode-function $\varphi_i(\mathbf{x})$ is defined as an eigenstate satisfying the single-particle Schrödinger equation,

$$\left[-\frac{1}{2m} \nabla^2 + V(\mathbf{x}) \right] \varphi_i(\mathbf{x}) = \varepsilon_i \varphi_i(\mathbf{x}) \,. \tag{C.22}$$

Using (C.21) and (C.22) in (C.20) then puts it into diagonal form

$$H = \sum_i \mathfrak{a}_i^* \mathfrak{a}_i \, \varepsilon_i \,. \tag{C.23}$$

An example of a system for which interactions are *not* local, and so do not cluster, is the case where particles interact through long-range forces, such as in

$$H = \int \mathrm{d}^3 x \left[\frac{1}{2m} \nabla\Phi^*(\mathbf{x}) \cdot \nabla\Phi(\mathbf{x}) + \int \mathrm{d}^3 y \; U(\mathbf{x}-\mathbf{y}) \; \Phi^*(\mathbf{x}) \, \Phi(\mathbf{x}) \; \Phi^*(\mathbf{y}) \, \Phi(\mathbf{y}) \right] . \tag{C.24}$$

Examples like this are not normally regarded as counter-examples of the requirement that physical systems cluster, since the long-range interaction $U(\mathbf{x}-\mathbf{y})$ usually arises once a massless (or very light) degree of freedom is integrated out. A famous example is the generation of the Coulomb potential once electromagnetic interactions are integrated out. In all such cases, however, the Wilson action for the relevant effective theory (from which massless or light particles are *not* integrated out) is local, and so does cluster.

C.2.1 Nonrelativistic fields with spin

Position-space fields can also incorporate spin, by requiring them to transform under rotations, such as by using a spinor field, a vector field or some other finite-dimensional representation of the rotation group. Generalizing (C.16) to a field Ψ_a transforming in one such a representation leads to

$$\Psi_a(\mathbf{x}) := \sum_{\sigma=-s}^{s} \int \frac{\mathrm{d}^3 p}{(2\pi)^{3/2}} \; u_a(\sigma) \, \mathfrak{a}_{\mathbf{p}\sigma} \, e^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}} \,, \tag{C.25}$$

where the sum is over the 3rd-component-of-spin quantum number, σ, of the particle state (assumed to have spin s). Here the 'polarization tensor' $u_a(\sigma)$ is the Clebsch-Gordan coefficient required to ensure consistency between the assumed transformation properties under rotations of the field and of the particle states.

For spin-half particles $s = \frac{1}{2}$ and Ψ_a (with $a = 1, 2$) can be a two-component spinor, which under rotations with infinitesimal parameter ω^k transform as $\delta\Psi = \frac{\mathrm{i}}{2} \omega^k \sigma_k \Psi$ (in addition to its action on $\mathbf{x}$), where σ_k are the usual 2×2 Pauli matrices. In this case the spin sum runs over $\sigma = \pm\frac{1}{2}$, with

$$u(+1/2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad u(-1/2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \tag{C.26}$$

A spin-one particle has $s = 1$ and can similarly be represented by a vector field, V_k, with $k = x, y, z$. In this case rotations act as $\delta V_k = \epsilon_{klm}\omega^l V^m$ (in addition to their action on $\mathbf{x}$), and $\sigma = 0, \pm 1$. In this case $u_k(\sigma)$ becomes the polarization vector appropriate for each of the three choices for σ. For instance for a particle with momentum parallel to the z axis, $\mathbf{p} = \mathbf{e}_z$, one finds $\mathbf{u}(0) = \mathbf{e}_z$ and $\mathbf{u}(\pm 1) = \frac{1}{\sqrt{2}}(\mathbf{e}_x \pm \mathrm{i}\mathbf{e}_y)$, so

$$\mathbf{V}(\mathbf{x}) = \sum_{\sigma=-1}^{1} \int \frac{\mathrm{d}^3 p}{(2\pi)^{3/2}}\, \mathbf{u}(\sigma)\, \mathfrak{a}_{\mathbf{p}\sigma}\, e^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}}\,. \tag{C.27}$$

C.3 Relativistic free fields

A similar story goes through for position-space fields in relativistic theories, but with two important differences. The simplest difference simply recognizes that the fields must transform in finite-dimensional representations of the Lorentz group rather than just rotations. The more subtle difference is that all position-space fields must come with both destruction and creation parts, in a way that is elaborated below.

It is this second condition that underlies many of the profound consequences — like the existence of antiparticles, the spin-statistics theorem, the CPT theorem and crossing symmetry — of combining quantum mechanics with special relativity. Although it goes beyond this summary to derive these consequences in detail, both types of differences are illustrated in the low-spin examples below.

C.3.1 Relativistic spin-0 fields

Scalar fields can be used for spin-zero particles[3] and in this case the expansion in terms of creation and annihilation operators generalizes the nonrelativistic result (C.17) to

$$\phi(x) = \int \frac{\mathrm{d}^3 p}{\sqrt{(2\pi)^3 2E_p}} \left[\mathfrak{a}_{\mathbf{p}}\, e^{\mathrm{i}p\cdot x} + \bar{\mathfrak{a}}^*_{\mathbf{p}}\, e^{-\mathrm{i}p\cdot x}\right], \tag{C.28}$$

where $p \cdot x$ is short for $p_\mu x^\mu = -E_p t + \mathbf{p}\cdot\mathbf{x}$ where $E_p = \sqrt{\mathbf{p}^2 + m^2}$ is the relativistic particle energy and m its rest mass. This expression normalizes momentum eigenstates in the same way as does (C.17) – *i.e.* using the non-relativistic normalization $\langle \mathbf{p}|\mathbf{q}\rangle = \delta^3(\mathbf{p} - \mathbf{q})$ where $|\mathbf{p}\rangle = \mathfrak{a}^*_{\mathbf{p}}|0\rangle$ – and so the factor of $\sqrt{E_p}$ in the denominator is precisely what is required to make the left-hand side transform as a Lorentz scalar. (If $\mathfrak{a}^*_{\mathbf{p}}|0\rangle = |\mathbf{p}\rangle_r$ were instead normalized covariantly, as in (B.20), then the measure appearing in (C.28) would be the Lorentz-invariant combination $\mathrm{d}^3 p/E_p$, as expected.)

It is the term involving $\bar{\mathfrak{a}}^*_{\mathbf{p}}$ that is the new 'creation part' of the field alluded to above. Here $\bar{\mathfrak{a}}_{\mathbf{p}}$ is the destruction operator for the *antiparticle* for the particle destroyed by $\mathfrak{a}_{\mathbf{p}}$,

[3] The choice of field representation is not unique for any given spin, with the general condition known since the 1960s [472, 51, 473, 14]. Different choices of representation typically do not define physically different theories. For instance a 4-vector field – instead of a scalar field – can represent a spinless particle, but the 4-vector in this case is simply the gradient of the scalar: $V_\mu = \partial_\mu \phi$.

whose properties are dictated by the requirement that the commutator $[\phi(x), \phi(y)]$ vanish for all spacelike separations, $(x-y)^2 > 0$. If this commutator would not vanish, then neither would the same commutator built using the hamiltonian density, $[\mathfrak{H}(x), \mathfrak{H}(y)]$, and if this does not vanish then the time-orderings of H appearing in the S-matrix (see §B.2) become problematic given that different observers can disagree on the time-ordering of x^0 and y^0 for spacelike-separated points.

In particular, this condition requires that the antiparticle have precisely the same mass as does the particle, and it must carry precisely the opposite charge for any symmetry that multiplies the field $\phi(x)$ by a phase: $\phi(x) \to e^{i\omega}\phi(x)$. A particle can be its own antiparticle if $\mathfrak{a}_{\mathbf{p}}$ is used instead of $\bar{\mathfrak{a}}_{\mathbf{p}}$ in (C.28). But $\phi(x)$ is then real, and so particle and antiparticle can only be the same in they carry no additive conserved charges. (The photon is an example of a particle of this type that is the same as its antiparticle.)

Cluster decomposition is ensured if local interactions are built as before from powers of fields and their derivatives at a single point, and for relativistic systems this is more conveniently done using the action, $S = \int \mathrm{d}^4x\, \mathfrak{L}$, than with the Hamiltonian, $H = \int \mathrm{d}^3x\, \mathfrak{H}$, because the action is Lorentz invariant while the energy is not. For free charged fields this action is quadratic and the fields can always be defined in such a way that the action becomes

$$S^{\text{free}}_{\text{spin }0} = -\int \mathrm{d}^4x \left[\partial_\mu\phi^*\,\partial^\mu\phi + m^2\phi^*\phi\right], \tag{C.29}$$

where m is the rest mass appearing in the dispersion relation: $\varepsilon(\mathbf{p}) = E_p = \sqrt{\mathbf{p}^2+m^2}$. The fields in this action — and in (C.28) — are chosen to be 'canonical' inasmuch as the canonical equal-time commutation relation, $[\Pi(\mathbf{x},t), \phi(\mathbf{y},t)] = -i\delta^3(\mathbf{x}-\mathbf{y})$, agrees with the creation/annihilation algebra $[a_{\mathbf{p}}, a^*_{\mathbf{q}}] = \delta^3(\mathbf{p}-\mathbf{q})$ when $\Pi = \delta S/\delta(\partial_t\phi)$ is the canonical momentum. (The numerical factors in (C.28) are also chosen to ensure that canonical commutation relations agree with the $a_{\mathbf{p}}, a^*_{\mathbf{p}}$ algebra.) Unlike in non-relativistic systems, relativity makes it necessary to quantize spinless particles using bose statistics — a consequence of the spin-statistics theorem: all integer-spin particles must be bosons and all half-odd spin particles must be fermions.

Interactions are similarly built using non-quadratic (but local) terms in ϕ and its derivatives. It is the requirement that all interactions be built from $\phi(x) \propto \mathfrak{a} + \bar{\mathfrak{a}}^*$ that implies that relativistic interactions *never* preserve particle number, in contrast with the interactions written above for the non-relativistic case.

C.3.2 Relativistic spin-1/2 fields

Spin-half particles are represented using Lorentz-spinor fields, $\psi_a(x)$, which are taken to be distinct from their antiparticles and so represented by 4-component Dirac spinors ($a = 1, \cdots, 4$; see the discussion above eq. (A.27) (or Appendix §A.2.3) for the distinction between these and Majorana or Weyl spinors).

$$\psi_a(x) = \sum_{\sigma=\pm\frac{1}{2}} \int \frac{\mathrm{d}^3p}{\sqrt{(2\pi)^3 2E_p}} \left[u_a(\mathbf{p},\sigma)\, \mathfrak{c}_{\mathbf{p}\sigma}\, e^{ip\cdot x} + v_a(\mathbf{p},\sigma)\, \bar{\mathfrak{c}}^*_{\mathbf{p}\sigma}\, e^{-ip\cdot x}\right], \tag{C.30}$$

where the destruction operator is labelled $\mathfrak{c}$ to distinguish it from the spinless destruction operator described above. For spin-half particles consistency *requires* these operators to anti-commute,

$$\{\mathfrak{c}^*_{\mathbf{p}\sigma}, \mathfrak{c}_{\mathbf{q}\zeta}\} = \delta^3(\mathbf{p}-\mathbf{q})\,\delta_{\sigma\zeta}\,, \tag{C.31}$$

to ensure that the hamiltonian density, $\mathfrak{H}(x)$, can continue to commute with itself when evaluated at spacelike-separated points in spacetime. $u(\mathbf{p},\sigma)$ and $v(\mathbf{p},\sigma)$ are spinors that are chosen to ensure that both sides of (C.30) transform the same way under Lorentz transformations. This implies they satisfy the Dirac conditions $(\mathrm{i}\not{p}+m)u(\mathbf{p},\sigma) = (\mathrm{i}\not{q}-m)v(\mathbf{p},\sigma) = 0$. Here the 'slash' notation denotes $\not{p} = p_\mu\gamma^\mu$, where the Dirac conventions used for the gamma matrices γ^μ are outlined in §A.2.3. Notice that in the particle rest frame these conditions become the projections $\mathrm{i}\gamma^0 u = +u$ for particles and $\mathrm{i}\gamma^0 v = -v$ for antiparticles, and in any other frame they are found by applying the appropriate Lorentz boost to these conditions.

The action that captures these conditions for free fields is

$$S^{\text{free}}_{\text{spin }1/2} = -\int \mathrm{d}^4x\,\overline{\psi}(\not{\partial}+m)\psi\,, \tag{C.32}$$

where again fields are chosen to be canonically normalized, and m is the particle rest mass that enters its dispersion relation $\varepsilon(\mathbf{p}) = E_p = \sqrt{\mathbf{p}^2+m^2}$.

C.3.3 Relativistic spin-1 fields

For spin one and higher the field content needed to treat massive and massless states differs. This is because a massive spin-s state contains $2s+1$ spin components, $\sigma = -s, -s+1, \cdots, s-1, s$, while a minimal massless spin-s state usually contains only two helicities: $\lambda = \pm s$. Although these two options have the same number of states for $s = 0$ and $s = \frac{1}{2}$, they differ from one another for $s \geq 1$.

Massive spin-1 fields

Consider first the massive case. The smallest fields that can be used for massive spin-one particles are vector fields, $V_\mu(x)$, and consistency of the 4-vector Lorentz-transformation rule with the transformations of creation and annihilation operators for massive spin-one particles implies

$$V_\mu(x) = \sum_{\lambda=-1}^{1}\int \frac{\mathrm{d}^3p}{\sqrt{(2\pi)^3 2E_p}}\left[\varepsilon_\mu(\mathbf{p},\lambda)\,\mathfrak{a}_{\mathbf{p}\lambda}\,e^{\mathrm{i}p\cdot x} + \varepsilon^*_\mu(\mathbf{p},\lambda)\,\bar{\mathfrak{a}}^*_{\mathbf{p}\lambda}\,e^{-\mathrm{i}p\cdot x}\right], \tag{C.33}$$

where as before $E_p = \sqrt{\mathbf{p}^2+m^2}$ with m the particle's rest mass. This form is also only consistent with $[\mathfrak{H}(x),\mathfrak{H}(y)] = 0$ for spacelike separations if the particles are bosons, so $[\mathfrak{a}_{\mathbf{p}}, \mathfrak{a}^*_{\mathbf{q}}] = [\bar{\mathfrak{a}}_{\mathbf{p}}, \bar{\mathfrak{a}}^*_{\mathbf{q}}] = \delta^3(\mathbf{p}-\mathbf{q})$.

The polarization vector $\varepsilon_\mu(\mathbf{p},\lambda)$ satisfies $p^\mu\varepsilon_\mu = 0$, and so for momentum pointing up

the z axis, the polarization vector's spatial part can be chosen to be

$$\varepsilon^\mu(\lambda=\pm1)=\frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \mathbf{e}_x\pm\mathrm{i}\,\mathbf{e}_y\end{pmatrix},\ \varepsilon^\mu(\lambda=0)=\begin{pmatrix}p\\ E_p\,\mathbf{e}_z\end{pmatrix}\quad\text{when}\quad p^\mu=\begin{pmatrix}E_p\\ p\,\mathbf{e}_z\end{pmatrix}. \tag{C.34}$$

One rationale for the condition $p^\mu\varepsilon_\mu(\mathbf{p},\lambda)=0$ is that if $V_\mu=\partial_\mu\phi$ were a gradient, it would actually represent a spin-0 particle, and this option must be projected out (as the condition $\partial^\mu V_\mu=0$ indeed does).

A free lagrangian for this type of particle is

$$S^{\text{free}}_{\text{spin 1}}=-\int \mathrm{d}^4x\,\left[\frac{1}{2}\,F^*_{\mu\nu}F^{\mu\nu}+m^2V^\mu V^*_\mu\right], \tag{C.35}$$

where $F_{\mu\nu}:=\partial_\mu V_\nu-\partial_\nu V_\mu$. Notice that the field equation obtained from this by varying V^*_μ is

$$\partial_\mu F^{\mu\nu}-m^2V^\nu=(\Box-m^2)V^\nu-\partial^\nu(\partial_\mu V^\mu)=0\,, \tag{C.36}$$

which, when acted on again by ∂_ν implies $m^2\partial_\mu V^\mu=0$. When $m\neq0$ these field equations both project out the spinless part (*i.e.* ensure $\partial\cdot V=0$) and – once this is used in (C.36) – ensure $p^\mu p_\mu+m^2=-E_p^2+\mathbf{p}^2+m^2=0$, thereby identifying m as the particle's rest mass.

Massless spin-1 fields

Next consider the massless case. It happens that the absence of the longitudinal mode, $\lambda=0$, precludes also using a 4-vector field like $V_\mu(x)$ to represent a massless spin-one particle. In this case the smallest finite-dimension representation of the Lorentz group that can be used to represent the two helicity states of a massless spin-one field turns out to be an antisymmetric tensor, $F_{\mu\nu}=-F_{\nu\mu}$. The two separate helicities are represented by the self-dual and anti-self-dual parts, $F^\pm_{\mu\nu}=F_{\mu\nu}\pm\mathrm{i}\,\widetilde{F}_{\mu\nu}$, where the 'dual' field strength is defined by

$$\widetilde{F}_{\mu\nu}=\frac{1}{2}\,\epsilon_{\mu\nu\lambda\rho}\,F^{\lambda\rho}\,, \tag{C.37}$$

for $\epsilon_{\mu\nu\lambda\rho}$ the completely antisymmetric Levi-Civita tensor of eq. (A.10).

The mode functions that ensure the consistency of the transformation rule for $F_{\mu\nu}$ and for a massless spin-one particle turn out to imply [51, 54]

$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu\,, \tag{C.38}$$

with

$$A_\mu(x)=\sum_{\lambda=\pm1}\int\frac{\mathrm{d}^3p}{\sqrt{(2\pi)^3 2E_p}}\,\left[\,\varepsilon_\mu(\mathbf{p},\lambda)\,\mathfrak{a}_{\mathbf{p}}\,e^{\mathrm{i}p\cdot x}+\ \varepsilon^*_\mu(\mathbf{p},\lambda)\,\mathfrak{a}^*_{\mathbf{p}}\,e^{-\mathrm{i}p\cdot x}\right], \tag{C.39}$$

where $E_p=|\mathbf{p}|$ is the relativistic particle energy for massless particles (and, with photons in mind, the antiparticle is chosen to be identical to the particle). This looks a lot like (C.33), but with the important omission of the $\lambda=0$ polarization.

At this point the astute reader asks: 'How can (C.38) and (C.39) be consistent with the earlier statement that $F_{\mu\nu}$ is the smallest field whose Lorentz-transformation properties

are consistent with representing a massless spin-one particle? Why not simply use the 4-vector A_μ instead?" From this point of view what is important about eqs. (C.38) and (C.39) is this: performing a Lorentz transformation on the creation and annihilation operators to accomplish a Lorentz transformation that takes $p^\mu \to \Lambda^\mu{}_\nu p^\nu$ indeed implies the field $F_{\mu\nu}$ transforms into $\Lambda_\mu{}^\lambda \Lambda_\nu{}^\rho F_{\lambda\rho}$, as should a rank-two tensor.

But the field A_μ does *not* transform as a 4-vector, because of the omission of the $\lambda = 0$ mode. It instead satisfies

$$A_\mu \to \Lambda^\nu{}_\mu A_\nu + \partial_\mu \Omega \tag{C.40}$$

for some scalar function Ω. Although the first term on the right-hand side corresponds to the transformation of a covariant 4-vector, the second term does not. Lorentz-invariant actions for massless spin-one particles can only be built using $A_\mu(x)$ instead of $F_{\mu\nu}(x)$ if they are also chosen to be *gauge invariant*, that is, invariant under the shift $A_\mu \to A_\mu + \partial_\mu \Omega$, for general Ω. It is the freedom to do just such a transformation that allows the $\lambda = 0$ spin-state to be removed, as required for a massless spin-one state.

C.3.4 Massless spin-2 fields

A similar story holds for massless spin-two particles for which there are only two spin states, with helicities $\lambda = \pm 2$. A field that can represent this kind of particle is a tensor $C_{\mu\nu\lambda\rho}$ with the same symmetries as the Riemann tensor: $C_{\mu\nu\lambda\rho} = C_{\lambda\rho\mu\nu} = -C_{\nu\mu\lambda\rho} = -C_{\mu\nu\rho\lambda}$, and in addition a trace-free condition, $\eta^{\lambda\rho} C_{\lambda\mu\rho\nu} = 0$.

Requiring the transformation properties of this field to be consistent with what is obtained once expanded in terms of creation and annihilation operators for a massless spin-two field, implies [51] $C_{\mu\nu\lambda\rho}$ is obtained as two derivatives of a field

$$h_{\mu\nu}(x) = \sum_{\lambda=\pm 2} \int \frac{\mathrm{d}^3 p}{\sqrt{(2\pi)^3 2E_p}} \left[\varepsilon_{\mu\nu}(\mathbf{p}, \lambda)\, \mathfrak{a}_{\mathbf{p}}\, e^{\mathrm{i}p\cdot x} + \varepsilon^*_{\mu\nu}(\mathbf{p}, \lambda)\, \mathfrak{a}^\dagger_{\mathbf{p}}\, e^{-\mathrm{i}p\cdot x} \right], \tag{C.41}$$

where $E_p = |\mathbf{p}|$ is the relativistic particle energy for massless particles and (as for photons) antiparticle is identified with the particle.

When a Lorentz transformation is performed on the particle creation and annihilation operators such that their 4-momentum transforms as $p^\mu \to \Lambda^\mu{}_\nu p^\nu$, then $h_{\mu\nu}$ does not transform as a tensor. It only does so up to a gauge transformation of the form

$$h_{\mu\nu} \to \partial_\mu \Omega_\nu + \partial_\nu \Omega_\mu \,, \tag{C.42}$$

for some field Ω_μ. This gauge symmetry must be preserved if interactions are to be built directly using $h_{\mu\nu}$ rather than $C_{\mu\nu\lambda\rho}$.

Precisely this kind of structure is obtained when the field equations of General Relativity are expanded about a flat background spacetime. In this case the metric is written $g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa\, h_{\mu\nu}$ with $\kappa^2 = 8\pi G_N$ related to Newton's constant, and under coordinate transformations $\delta x^\mu = \xi^\mu(x)$ the fluctuation field $h_{\mu\nu}$ transforms as $\delta h_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$, as above. The field $C_{\mu\nu\lambda\rho}$ is then the linearized Weyl tensor, defined as the completely trace-free part of the Riemann tensor built from $g_{\mu\nu}$.

C.4 Global symmetries

Symmetries play an important role in quantum mechanics, just as they do in classical mechanics. Symmetries are special because their existence allows exact statements to be made about transition probabilities and about energy eigenstates.

In quantum mechanics it is a theorem [41] that transformations that do not change any transition probabilities can always be represented in terms of unitary operators.[4] Being represented by a unitary operator means that the action of the symmetry on any state can be written as

$$|\psi\rangle \to |\tilde{\psi}\rangle = U\,|\psi\rangle\,, \tag{C.43}$$

where $U^*U = UU^* = I$, with I being the identity operator in the Hilbert space. Probabilities remain unchanged, because $\langle\tilde{\psi}_1|\tilde{\psi}_2\rangle \to \langle\psi_1|U^*U|\psi_2\rangle = \langle\psi_1|\psi_2\rangle$. Matrix elements of operators, A, remain unchanged under such transformations because $A \to \tilde{A} = UAU^*$, and so its matrix elements become: $\langle\tilde{\psi}_1|\tilde{A}|\tilde{\psi}_2\rangle = \langle\psi_1|U^*UAU^*U|\psi_2\rangle = \langle\psi_1|A|\psi_2\rangle$.

The complete set of transformations, $\{g_a\}$, that preserve all matrix elements in this way forms a group, where group multiplication, g_1g_2, consists of convolution (*i.e.* successive performance of the two individual transformations). Furthermore, there can be (but need not be) an independent operator $U(g)$ for each such transformation. These operators form a unitary representation inasmuch as $U(g_1)U(g_2) = U(g_1g_2)$.

A *symmetry* is defined to be any transformation of this type that does not change the system's hamiltonian:

$$\widetilde{H} = UHU^* = H\,, \tag{C.44}$$

and so $[U, H] = 0$. These form a subgroup of the group of matrix-element-preserving transformations. Two classic consequences follow immediately from eq. (C.44).

- *Conservation:* Each unitary symmetry, U, defines a hermitian charge, Q, whose quantum numbers for any state are conserved in time. This follows because $[U, H] = 0$ implies $[Q, H] = 0$ and so — because the time-evolution operator is $U(t, t_0) = \exp[-\mathrm{i}H(t-t_0)]$ — then $[Q, U(t, t_0)] = 0$ as well. Therefore, if $Q|\psi(t = t_0)\rangle = q|\psi(t_0)\rangle$ then (in Schrödinger picture)

$$Q\,|\psi(t)\rangle = QU(t, t_0)|\psi(t_0)\rangle = U(t, t_0)Q\,|\psi(t_0)\rangle = q\,|\psi(t)\rangle\,. \tag{C.45}$$

  This expresses conservation in the usual sense that once a state is prepared to have a particular value of Q then it has this same value for all later times.
- *Spectrum degeneracy:* It also follows from (C.44) that if two energy eigenstates are related by a symmetry then they must have the same energy. That is, because $[U, H] = 0$ if $|\psi_2\rangle = U|\psi_2\rangle$ and $H|\psi_i\rangle = E_i\,|\psi_i\rangle$, then

$$E_2|\psi_2\rangle = H|\psi_2\rangle = HU|\psi_1\rangle = UH|\psi_1\rangle = E_1(U|\psi_1\rangle) = E_1|\psi_2\rangle\,, \tag{C.46}$$

  and because $|\psi_2\rangle \neq 0$ this means $E_1 = E_2$. Physically this says that if a transformation

[4] Or anti-unitary operators, as is the case for time-reversal (more about which in §C.4.3).

is a symmetry it should not affect energies; in a rotationally invariant world a ruler has the same total energy regardless of whether it is laid along the x, y or z axes.

For relativistic systems, it is often true that symmetries commute with momentum as well as energy, $[U, P^\mu] = 0$. If so, then because $E_i^2(\mathbf{p}) = \mathbf{p}^2 + m_i^2$ it follows that any two states related by U must have the same rest mass.

All of these statements go over for quantum field theories as well, although for some conclusions (like spectral degeneracy) it matters whether the system's ground state is invariant under the symmetry or not (that is, whether the symmetry is spontaneously broken).

Just as for ordinary quantum mechanics the action of symmetries (besides time-reversal) on states and operators is given by a unitary operator U, with $|\tilde{\psi}\rangle = U|\psi\rangle$ and $\widetilde{O} = UOU^*$. In particular, for creation and annihilation operators $\tilde{\mathfrak{a}}_{\mathbf{p}} \to U\,\mathfrak{a}_{\mathbf{p}}U^*$ and $\tilde{\mathfrak{a}}^*_{\mathbf{p}} \to U\,\mathfrak{a}^*_{\mathbf{p}}U^*$ and so field operators transform as $\tilde{\phi} = U\phi U^*$ and so on.

If $U|0\rangle = |0\rangle$ (which multiplying through by U^* shows also implies $U^*|0\rangle = |0\rangle$) then these transformations amongst creation operators also imply that the corresponding particle states – call them $|\psi(\mathbf{p})\rangle = \mathfrak{a}^*_{\mathbf{p}}|0\rangle$ and $|\tilde{\psi}(\mathbf{p})\rangle = \tilde{\mathfrak{a}}^*_{\mathbf{p}}|0\rangle$ – are related by the action of U, since

$$|\tilde{\psi}(\mathbf{p})\rangle = \tilde{\mathfrak{a}}^*_{\mathbf{p}}|0\rangle = U\mathfrak{a}^*_{\mathbf{p}}U^*|0\rangle = U\mathfrak{a}^*_{\mathbf{p}}|0\rangle = U|\psi(\mathbf{p})\rangle\,. \tag{C.47}$$

When this is true (C.46) implies these particles share the same energy. The above argument shows why this implication also generally fails when $U|0\rangle \neq |0\rangle$ (*i.e.* when the symmetry is spontaneously broken). In this case particles need not align into degenerate multiples for a spontaneously broken symmetry, and the symmetry instead partially acts to shift the vacuum itself (which by assumption is not invariant).[5]

When a symmetry is not spontaneously broken then it can be linearly realized on the fields themselves, as described in the main text in §4.2.1, with

$$\phi^i \to \tilde{\phi}^i := U(g)\,\phi^i U^*(g) = \phi^j\,\mathcal{G}_j{}^i\,. \tag{C.48}$$

Applying two transformations in succession and using $U(g_1g_2) = U(g_2)U(g_2)$ then shows that the matrices $\mathcal{G}_i{}^j$ satisfy $\mathcal{G}_i{}^j(g_1g_2) = \mathcal{G}_i{}^k(g_1)\,\mathcal{G}_k{}^j(g_2)$.

It is not always true that symmetries can be realized linearly in this linear way, with the general case being a nonlinear realization

$$\phi^i \to \tilde{\phi}^i = U(g)\,\phi^i U^*(g) = \xi^i(\phi, g)\,, \tag{C.49}$$

where $\xi^i(\phi, g)$ is potentially a nonlinear function of the ϕ^i. This is the case to which one is led when a symmetry is spontaneously broken, as described in §4.2.2 (with more details given in Appendix C.6).

C.4.1 Lie algebra summary

Many of the symmetries of practical interest are enumerated using continuous parameters (like translations, rotations, chiral symmetries and isospin or gauge transformations), making them Lie groups from the mathematical point of view. This section steps back from the

[5] For translation-invariant ground states the very definitions of Q and U become delicate for spontaneously broken symmetries, at least for field theories in the infinite-volume limit – *c.f.* eq. (C.82).

main line of development to summarize a few facts about these groups, together with their related Lie algebras.

Of particular interest in physical applications are often explicit representations of Lie groups and algebras in terms of matrices (that are often important in specific physical applications). For the present purposes representations are simply examples of matrices or operators in a physical problem whose matrix multiplication rules furnish examples of the underlying group multiplication rule. Eq. (C.48) provides an example of this, with $U(g_1 g_2) = U(g_1)U(g_2)$ providing a unitary representation of the group in the quantum Hilbert space, while the $N \times N$ matrices $\mathcal{G}(g_1 g_2) = \mathcal{G}(g_1)\mathcal{G}(g_2)$ provide a finite-dimensional representation of the group on the space of N fields ϕ^i. Although (C.48) leaves ambiguous where the fields ϕ^i and $\tilde{\phi}^i$ are evaluated, for simplicity in what follows they are taken to be evaluated at the same spacetime point (making the symmetry an 'internal' symmetry (as opposed to a 'spacetime' symmetry – see §C.4.2 below).

For Lie groups the abstract group elements and their explicit realizations are labelled by continuous real parameters, ω^a, with $a = 1, \cdots, N_g$, and both $U(\omega)$ and $\mathcal{G}(\omega)$ are infinitely differentiable. (For instance, for rotations in 3 dimensions the ω_a could correspond to the angles of rotation about three orthogonal axes.)

Continuous symmetries are often efficiently characterized by their generators, t_a, defined by examining transformations arbitrarily close to the identity element: $g(\omega) = I + \mathrm{i}\omega^a t_a + \mathcal{O}(\omega^2)$. It can be shown that the parameters ω^a can be defined in such a way that any group element that is continuously deformable to the identity element can be written as an exponential of the generators: $g(\omega) = \exp[\mathrm{i}\omega^a t_a]$. The statement that a group is closed under multiplication implies that these generators satisfy a set of commutation relations of the form

$$[t_a, t_b] = \mathrm{i}c^d{}_{ab}\, t_d \,. \tag{C.50}$$

The span of all linear combinations of such generators is called the Lie algebra associated with the Lie group. The coefficients $c^d{}_{ab} = -c^d{}_{ba}$ appearing here are called 'structure constants', whose form encodes the multiplication law that defines the underlying group.

Explicit representations of the Lie group also provide representations for the corresponding Lie algebra. For infinitesimal transformations, $g = 1 + \mathrm{i}\omega^a t_a$, the unitary operator in the quantum Hilbert space becomes $U(g) = I + \mathrm{i}\omega^a T_a$ and the representation matrices for the fields become $\mathcal{G} = I + \mathrm{i}\omega^a \mathcal{T}_a$. These act on the fields so that $\delta\phi^i = \tilde{\phi}^i - \phi^i$ has the form

$$\delta\phi^i(x) = \mathrm{i}\omega^a [T_a\,, \phi^i(x)] = \mathrm{i}\omega^a (\mathcal{T}_a)_j{}^i\, \phi^j(x) \,. \tag{C.51}$$

Because the operators U and the matrices $\mathcal{G}$ satisfy the same group multiplication rule as do the group elements g, the operators T_a and the matrices $\mathcal{T}_a$ satisfy the same commutation relations as do the generators t_a: that is, $[\mathcal{T}_a, \mathcal{T}_b] = \mathrm{i}c^d{}_{ab}\mathcal{T}_d$ and $[\mathcal{T}_a, \mathcal{T}_b] = \mathrm{i}c^d{}_{ab}\mathcal{T}_d$, with the same structure constants as in (C.50).

Conjugate and adjoint representations

Any explicit representation of a Lie algebra, $\{\mathcal{T}_a\}$ say, satisfying

$$[\mathcal{T}_a, \mathcal{T}_b] = \mathrm{i}c^d{}_{ab}\, \mathcal{T}_d \,, \tag{C.52}$$

can be used to define two other representations. The first of these is found by taking the transpose of (C.52), which shows that the operators $\mathcal{S}_a = -\mathcal{T}_a^T$ (where the superscript 'T' denotes taking the transpose) also satisfy (C.52). For unitary representations (those for which the matrices $\mathcal{G}$ are unitary) the $\mathcal{T}_a$ are Hermitian matrices and so $\mathcal{S}_a = -\mathcal{T}_a^T = -\mathcal{T}_a^*$ are also related by complex conjugation (not Hermitian conjugation) to the $\mathcal{T}_a$'s.

A second related representation can be built from the ${c^d}_{ab}$'s themselves (and so is more an intrinsic property of the group than of its specific representation in terms of the $\mathcal{T}_a$. To see why, notice that the associative property of matrix multiplication ensures the quantity $[A,[B,C]]+[B,[C,A]]+[C,[A,B]]$ identically vanishes for any three matrices A, B and C. Applying this Jacobi identity to three generators of the Lie algebra then implies

$$\begin{aligned} 0 &= [\mathcal{T}_a,[\mathcal{T}_b,\mathcal{T}_c]] + [\mathcal{T}_b,[\mathcal{T}_c,\mathcal{T}_a]] + [\mathcal{T}_c,[\mathcal{T}_a,\mathcal{T}_b]] \\ &= -\Big({c^d}_{bc}{c^e}_{ad} + {c^d}_{ca}{c^e}_{bd} + {c^d}_{ab}{c^e}_{cd}\Big)\mathcal{T}_e\,. \end{aligned} \tag{C.53}$$

The bracket on the right-hand side of this equation therefore vanishes for any set of structure constants. One way to read this identity is to say that the matrices $\mathcal{A}_a$ with components $(\mathcal{A}_a)^b{}_c := \mathrm{i}{c^b}_{ac}$ satisfy the commutation relation $[\mathcal{A}_a,\mathcal{A}_b] = \mathrm{i}{c^d}_{ab}\mathcal{A}_d$, with precisely the same structure constants as in (C.52), and so therefore furnish another representation – called the *adjoint representation* – of the same Lie algebra.

Finite-dimensional unitary representations

In physical situations continuous symmetry groups often arise as explicit finite-dimensional unitary matrices, such as for the 3×3 orthogonal matrices – *i.e.* $O(3)$ transformations – describing rotations in space, or more generally the internal $N\times N$ unitary matrices – *i.e.* $U(N)$ transformations – amongst N complex fields, ψ^i. This turns out to mean that a special role is often played in physics by *compact groups*, for which the parameter space of the group is a compact set.

Compact groups play a special role because it is a theorem that only compact groups have finite-dimensional, unitary and faithful matrix representations.[6] (The Lorentz group, for instance, is *not* compact although its subgroup of spatial rotations is. Consequently, as found explicitly in §A.2.3 say, although rotations can be represented using finite-dimensional unitary transformations, any finite-dimensional representations of boosts cannot be unitary.)

This section summarizes some useful properties satisfied by representations built from finite-dimensional unitary matrices, for which $\mathcal{G}^\dagger = \mathcal{G}^{-1}$ and so $\mathcal{T}_a^\dagger = \mathcal{T}_a$ if $\mathcal{G}(\omega) = \exp[\mathrm{i}\omega^a\mathcal{T}_a]$. Because the generators are finite-dimensional and hermitian the quantity

$$\gamma_{ab} = \mathrm{Tr}\,(\mathcal{T}_a\mathcal{T}_b)\,, \tag{C.54}$$

is both symmetric and positive definite (and so can be regarded as a metric, called the group's *Killing metric*). Linear combinations of the generators can always be chosen to

[6] A representation is faithful if there is a one-to-one correspondence between the group elements and the matrices which represent them. If the group of interest is *defined* by a finite-dimensional and unitary representation, this representation is by definition faithful.

ensure that

$$\gamma_{ab} = \delta_{ab}\,, \tag{C.55}$$

(such as is true for the standard Pauli-matrix representation of $SU(2)$ – *i.e.* 2×2 unitary matrices with unit determinant – for which $\mathcal{T}_a = \frac{1}{2}\,\sigma_a$ for $a = 1, 2, 3$). This convention for the generators of compact groups is usually assumed to have been made throughout this book.

The metric γ_{ab} can be used to build a completely covariant version of the structure constants:

$$c_{bca} := c^{d}{}_{bc}\,\gamma_{da}\,. \tag{C.56}$$

Its definition automatically implies $c_{abd} = -c_{bad}$, but when the generators are chosen so that $\gamma_{ab} = \delta_{ab}$ it turns out that c_{abd} is completely antisymmetric under the interchange of *any* two of its indices.

Real unitary representations

There is also no loss of generality in assuming representation matrices to be real: $g = g^*$, because any complex representation can always be decomposed into its real and imaginary parts. Although this can always be done, the resulting representation need *not* be irreducible. For reducible representations there is a basis in which all group elements can be written in a block-diagonal form:

$$\mathcal{G} = \begin{pmatrix} \mathcal{G}_{(1)} & & \\ & \ddots & \\ & & \mathcal{G}_{(n)} \end{pmatrix}. \tag{C.57}$$

The unitarity and reality of the group elements, $\mathcal{G}$, then imply the matrices $\mathcal{T}_a$ are antisymmetric and imaginary:

$$\mathcal{T}_a = \mathcal{T}_a^{\dagger} = -\mathcal{T}_a^{*} = -\mathcal{T}_a^{T}\,. \tag{C.58}$$

Subgroups and subalgebras

When describing a symmetry breaking pattern where G breaks to $H \subset G$ it is convenient to choose a basis of generators, T_a, for G that includes the generators, t_i, of H as a subset. To this end decompose the generators T_a, $a = 1, \cdots, N_G$ into the subset t_i, $i = 1, \cdots, N_H$ and X_α, $\alpha = N_H + 1, \cdots, N_G$, so the X_α's constitute a basis of generators not included in the unbroken subalgebra. Here $N_G = \dim G$ is the number of linearly independent generators of the Lie algebra of G and $N_H = \dim H$ is its counterpart for H. Since H is itself a group, its closure under multiplication — *i.e.* the statement that $h_1, h_2 \in H$ implies $h_1 h_2 \in H$ — ensures that

$$t_i\,t_j - t_j\,t_i = i\,c_{ijk}\;t_k\,, \tag{C.59}$$

with no X_α's on the right-hand-side, or (schematically) $c_{ij\alpha} = 0$.

The X_α are not contained in the algebra of H and do not themselves generate a group.

Instead they are said to generate the space, G/H, of *cosets*. A coset is an equivalence class defined to contain all of the elements of G that are related by the multiplication by an element of H. In the applications of §C.6, the X_α's represent those generators of the symmetry group, G, that are spontaneously broken, and (in relativistic applications for internal symmetries) a Goldstone mode is expected for each independent choice of α.

When the generators of G are chosen to ensure the complete antisymmetry of the c_{abd}'s then the group property of H (summarized above as $c_{ij\alpha} = 0$) also implies $c_{i\alpha j} = 0$. This says

$$[t_i, X_\alpha] = i\, c_{i\alpha\beta} X_\beta \,, \tag{C.60}$$

with no t_j's on the right-hand-side. Equivalently this states that the X_α's form a (possibly reducible) representation of H. Once exponentiated into a statement about group multiplication, the condition $[t, X] \sim X$ implies that for any $h \in H$

$$h X_\alpha h^{-1} = L^\beta{}_\alpha X_\beta \tag{C.61}$$

where the coefficients, $L^\beta{}_\alpha(h)$, form a representation of H.

By contrast, the commutator $[X_\alpha, X_\beta]$ need not have a particularly simple form, and can be proportional to both X_γ's and t_i's. (The special case of a coset G/H for which $[X_\alpha, X_\beta]$ does not contain any X_γ's is called a *symmetric* space.)

C.4.2 Internal vs spacetime symmetries

Notice that the above discussion distinguishes unitary transformations (those that preserve matrix elements) from symmetries (those unitary transformations that commute with the hamiltonian). This notion of symmetry is adequate for internal symmetries — *i.e.* those that do not act on spatial position or time, so $U\phi(x)U^* = \tilde{\phi}(x)$ with both sides evaluated at the same position.

A broader definition is needed for spacetime symmetries, for which the transformations act both on the fields and the spacetime point: $U\phi(x)U^* = \tilde{\phi}(\tilde{x})$, with $\tilde{x}^\mu \neq x^\mu$. Lorentz transformations are simple examples where this matters, since for these H generally is not invariant, since it is part of a 4-vector: $UP_\mu U^* = \Lambda^\nu{}_\mu P_\nu$, with $P^0 = H$. In this case a symmetry is defined instead by the invariance of the action, $S = \int \mathrm{d}t\, L$, rather than of H. For scattering problems transformations that are symmetries in this sense also commute with the S-matrix.

The Coleman-Mandula theorem [316] provides an important constraint on the kinds of continuous spacetime symmetries that can be present within interacting relativistic quantum field theories. The Coleman-Mandula theorem states that the most general possible non-Grassman[7] transformations that commute with a (nontrivial – *i.e.* $\mathcal{S} \neq I$) S-matrix are:

$$U = \exp\left[\frac{\mathrm{i}}{2}\,\omega^{\mu\nu} J_{\mu\nu} + \mathrm{i}a^\mu P_\mu + \mathrm{i}\omega^a Q_a\right] \tag{C.62}$$

with generators P_μ, $J_{\mu\nu} = -J_{\nu\mu}$, and Q_a.

Ten of these are no surprise in a relativistic theory: the six generators $J_{\mu\nu}$ satisfy the

[7] This is the assumption that supersymmetric theories violate; see [474] for the generalization to this case.

commutation relations appropriate to the Lorentz group and the four P_μ generate spacetime translations and so mutually commute (and fill out the usual 4-momentum operator). The commutation relations between $J_{\mu\nu}$ and P_μ fill out the algebra of the Poincaré group of Lorentz transformations and translations, making these the defining symmetries of special relativity.[8]

The power of the Coleman-Mandula theorem is what it says about the remaining generators: the Q_a's. These must be internal symmetries, and although they can fail to commute quietly amongst themselves, $[Q_a, Q_b] = \mathrm{i}c^d{}_{ab}T_d$, the theorem states that they must always commute with the spacetime symmetries (*i.e.* the Poincaré generators): $[P_\mu, Q_a] = 0$ and $[J_{\mu\nu}, Q_a] = 0$. The theorem is proven by assuming it to be false, and then showing that the additional conservation laws for the spacetime symmetries are so strong that they generically force the scattering matrix to be trivial: $\mathcal{S} = I$.

C.4.3 Discrete symmetries

Discrete symmetries (those that are not described by continuous parameters) are also important for physics. Some of these can be internal symmetries, such as an example like $\phi(x) \to -\phi(x)$, which defines a discrete Z_2 symmetry in field space. Such symmetries constrain the kinds of interactions that can arise (forbidding, in the Z_2 example, terms involving odd powers of fields). Their representations can also be used to classify states (in the Z_2 case states can be chosen to either change sign or be invariant under the group's action).

Spacetime discrete symmetries are also important. These are defined to be those Poincaré transformations that cannot be continuously deformed to the unit element. There are two such discrete transformations within the Lorentz group. To see why, recall that the general Lorentz transformation is defined by the condition

$$\Lambda^\mu{}_\nu \Lambda^\rho{}_\lambda \eta_{\mu\rho} = \eta_{\nu\lambda} \,. \tag{C.63}$$

This condition throws up two obstructions to being able to deform $\Lambda^\mu{}_\nu$ to the identity transformation. One of these arises because (C.63) implies the determinant of the matrix $\Lambda^\mu{}_\nu$ must be ± 1, but only those with determinant $+1$ can be continuously connected to the identity matrix. Similarly, (C.63) requires $|\Lambda^0{}_0| \geq 1$ and so any matrix with $\Lambda^0{}_0 < -1$ also cannot be continuously related to the identity matrix.

A general solution $\Lambda^\mu{}_\nu$ to (C.63) can be written as a combination of a matrix continuously connected to the identity, $\Lambda^\mu{}_\nu = (e^\omega)^\mu{}_\nu$ (called a 'proper' Lorentz transformation) times a product of one or both of the two specific matrices

$$P^\mu{}_\nu := \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \quad \text{and} \quad T^\mu{}_\nu := \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \tag{C.64}$$

where P (parity) acts to reflect all spatial coordinates while T (time-reversal) flips the sign of time.

[8] For theories involving only massless particles this symmetry group is sometimes a bit larger; comprising the conformal group that also includes rescalings, $x^\mu \to s x^\mu$, and conformal boosts.

These matrices show why time reversal is the lone symmetry that cannot be represented by a unitary operator. Like for any symmetry, for Poincaré transformations the unitary operators must furnish a representation. Denoting by $U(\Lambda, a)$ the representation of the Lorentz transformation $\Lambda^\mu{}_\nu$ and spacetime translation a^μ, this representation in particular means

$$U(\Lambda_1, a_1)U(\Lambda_2, a_2) = U(\Lambda_1\Lambda_2, \Lambda_1 a_2 + a_1)\,. \tag{C.65}$$

Denoting the operator that represents time-reversal by $\mathcal{T}$ this implies

$$\mathcal{T}U(\Lambda, a)\mathcal{T}^{-1} = U(T\Lambda T^{-1}, Ta)\,, \tag{C.66}$$

and so for $\Lambda = I$ and infinitesimal a^μ this says $\mathcal{T}(\mathrm{i}P_\mu)\mathcal{T}^{-1} = T^\nu{}_\mu \mathrm{i}P_\nu$ and so $\mathcal{T}\mathrm{i}H\mathcal{T}^{-1} = -\mathrm{i}H$ and $\mathcal{T}\mathrm{i}\mathbf{P}\mathcal{T}^{-1} = \mathrm{i}\mathbf{P}$.

Now comes the main point. If $\mathcal{T}$ were unitary then it would satisfy $\mathcal{T}H\mathcal{T}^{-1} = -H$, which is inconsistent with H being bounded from below (as it typically is for stable systems). But if it is antiunitary then $\mathcal{T}\mathrm{i}H\mathcal{T}^{-1} = -\mathrm{i}\mathcal{T}H\mathcal{T}^{-1}$, allowing $\mathcal{T}$ to commute with H. For antiunitary $\mathcal{T}$ it then follows that $\mathcal{T}\mathbf{P}\mathcal{T}^{-1} = -\mathbf{P}$.

A third important discrete symmetry interchanges particles with antiparticles (with momenta and spins held fixed). This acts on creation and destruction operators by $C\,\mathfrak{a}_i\,C^{-1} = \eta_c\,\bar{\mathfrak{a}}_i$, an operation called charge-conjugation. Here η_c is a phase that can differ for different particle types. The action of charge conjugation on fields is found by applying the definition to the expansion of fields in terms of creation and annihilation operators. For example, for a scalar field this leads to

$$\begin{aligned} C\,\phi(x)\,C^{-1} &= \int \frac{\mathrm{d}^3p}{\sqrt{(2\pi)^3 2E_p}} \left[C\,\mathfrak{a}_{\mathbf{p}}\,C^{-1}e^{\mathrm{i}p\cdot x} + C\,\bar{\mathfrak{a}}^*_{\mathbf{p}}\,C^{-1}e^{-\mathrm{i}p\cdot x}\right] \\ &= \eta_c \int \frac{\mathrm{d}^3p}{\sqrt{(2\pi)^3 2E_p}} \left[\bar{\mathfrak{a}}_{\mathbf{p}}\,e^{\mathrm{i}p\cdot x} + \mathfrak{a}^*_{\mathbf{p}}\,e^{-\mathrm{i}p\cdot x}\right] = \eta_c\phi^*(x)\,, \end{aligned} \tag{C.67}$$

and so acts as complex conjugation.

The three discrete symmetries, C, P and T, are individually symmetries of electromagnetism, gravity and the strong interactions, but all three are separately broken by the weak interactions. The action of each of these three symmetries on various familiar physical quantities is summarized in Table C.1. (For A_μ the signs given in this table include the phases – like η_c in (C.67) – appearing in the parity, time-reveral and charge-conjugation transformations of the electromagnetic field.)

For relativisitic systems it turns out to be a theorem that any real and local action, $S = \int \mathrm{d}^4x\ \mathfrak{L}$, turns out to be always invariant under the combined combination of all three symmetries: CPT (a result called the CPT theorem). There is a simple reason for this. The action of C complex conjugates all the fields in $\mathfrak{L}$ and the anti-unitary nature of time-reversal complex conjugates all couplings in $\mathfrak{L}$, so their combined effect takes $\mathfrak{L} \to \mathfrak{L}^*$. But the lagrangian is hermitian so this has no effect. Time reversal and parity also together act to reverse the sign of all components of any 4-vector, $(PT)^\mu{}_\nu V^\nu = -V^\mu$. But this also has no effect because there are always an even number of such 4-vectors because this must be true if $\mathfrak{L}$ is to be a Lorentz scalar.

Table C.1 The transformation properties of common quantities under parity (P), time-reversal (T) and charge-conjugation (C).

quantity		P	T	C	quantity		P	T	C
position	$\mathbf{x}$	−	+	+	momentum	$\mathbf{p}$	−	−	+
spin	$\mathbf{s}$	+	−	+	helicity	$\mathbf{p}\cdot\mathbf{s}$	−	+	+
current	$\mathbf{j}$	−	−	−	charge density	j^0	+	+	−
vector potential	$\mathbf{A}$	−	−	−	scalar potential	A^0	+	+	−
electric field	$\mathbf{E}$	−	+	−	magnetic field	$\mathbf{B}$	+	−	−

C.5 Gauge interactions

Consider next a collection of quantum fields, ϕ^i, that transform under linearly realized infinitesimal internal symmetries of the form $\phi^i(x) \to \tilde{\phi}^i(x)$ with

$$\delta\phi^i(x) := \tilde{\phi}^i(x) - \phi^i(x) = \mathrm{i}\omega^a(\mathcal{T}_a)^i{}_j\,\phi^j(x)\,. \tag{C.68}$$

This is a global (or rigid) symmetry when the transformation parameter ω^a is independent of spacetime position.[9] But an important role is also played by local symmetries, for which $\omega^a = \omega^a(x)$ is a function of spacetime [475].

Since global symmetries are special cases of local ones, it is more difficult to make a theory invariant under a local symmetry than for a global one. To see this explicitly, consider an action for a collection of fields, ϕ^i, where $\mathfrak{L} = \mathfrak{L}(\phi, \partial\phi)$ is a function of both the fields and their first derivatives. The variation of $\mathfrak{L}$ under (C.68) is

$$\begin{aligned}\delta\mathfrak{L} &= \left(\frac{\partial\mathfrak{L}}{\partial\phi^i}\right)\mathrm{i}\omega^a(\mathcal{T}_a)^i{}_j\,\phi^j + \left(\frac{\partial\mathfrak{L}}{\partial(\partial_\mu\phi^i)}\right)\partial_\mu\left[\mathrm{i}\omega^a(\mathcal{T}_a)^i{}_j\,\phi^j\right] \\ &= \left(\frac{\partial\mathfrak{L}}{\partial\phi^i}\mathrm{i}\omega^a(\mathcal{T}_a)^i{}_j\,\phi^j + \frac{\partial\mathfrak{L}}{\partial(\partial_\mu\phi^i)}\mathrm{i}\omega^a(\mathcal{T}_a)^i{}_j\,\partial_\mu\phi^j\right) + \left(\frac{\partial\mathfrak{L}}{\partial(\partial_\mu\phi^i)}\right)\mathrm{i}\partial_\mu\omega^a(\mathcal{T}_a)^i{}_j\,\phi^j\,.\end{aligned} \tag{C.69}$$

The first two terms on the right-hand side vanish whenever $\mathfrak{L}$ is invariant under a global symmetry (with ω^a) like (C.68), but with ω^a spacetime independent.

Eq. (C.69) shows that even if a lagrangian is arranged to be invariant under global transformations, it is not invariant under local ones, but transforms in a universal way

$$\delta\mathfrak{L} = \left(\frac{\partial\mathfrak{L}}{\partial(\partial_\mu\phi^i)}\right)\mathrm{i}\partial_\mu\omega^a(\mathcal{T}_a)^i{}_j\,\phi^j = j_a^\mu\,\partial_\mu\omega^a\,, \tag{C.70}$$

where $j_a^\mu = [\partial\mathfrak{L}/\partial(\partial_\mu\phi^i)]\mathrm{i}(\mathcal{T}_a)^i{}_j\phi^j$ is the Noether current for the global symmetry, as defined by eq. (4.7) in §4.1.1. The universal form of this transformation suggests a way to build a locally invariant lagrangian. Juxtaposing the fact that (C.70) involves $\partial_\mu\omega^a$ and that massless spin-one particles can only be represented by a field A_μ if it transforms as $\delta A_\mu = \partial_\mu\omega$, as in eq. (C.40), suggests that a locally invariant lagrangian might be

[9] The attentive reader will notice the matrix $\mathcal{T}_a$ used in (C.68) is the transpose of the one used in (C.51). This is done so that the signs found in this section agree with those widely used in the literature.

constructed by adding a new term, $\mathfrak{L}_j = -j_a^\mu A_\mu^a$, in whose variation the transformation $\delta A_\mu^a = \partial_\mu \omega^a$ would cancel (C.70).

Adding $\mathfrak{L}_j$ need not be the whole story, because in general the current j_a^μ also transforms under the transformation (C.68). This transformation can also be inferred universally since the symmetry generators, T_a, themselves are obtained by integrating j_a^0 over all space. Since the generators satisfy $[T_a, T_b] = \mathrm{i} c^d{}_{ab}\, T_d$, the currents must transform (up to terms that vanish when integrated over space) as

$$\delta j_a^\mu = \mathrm{i}\omega^b \Big[T_b \,, j_a^\mu \Big] = c^d{}_{ab}\, \omega^b j_d^\mu \,. \tag{C.71}$$

This suggests modifying the transformation rule for A_μ^a to also transform in the adjoint representation:

$$\delta A_\mu^a = \partial_\mu \omega^a + c^a{}_{bc}\, \omega^b A_\mu^c \,. \tag{C.72}$$

so that $\delta(j_a^\mu A_\mu^a) = j_a^\mu \partial_\mu \omega^a$.

To see whether this works start with a more general lagrangian density $\mathfrak{L} = \mathfrak{L}(\phi, \partial\phi, A, \partial A)$ and ask whether it can be invariant under the transformations (C.68) and (C.72). The variation of $\mathfrak{L}$ then is

$$\begin{aligned} \delta\mathfrak{L} = &\left(\frac{\partial \mathfrak{L}}{\partial \phi^i}\right) \mathrm{i}\omega^a (\mathcal{T}_a)^i{}_j\, \phi^j + \left(\frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\right) \partial_\mu \left[\mathrm{i}\omega^a (\mathcal{T}_a)^i{}_j\, \phi^j \right] \\ &+ \left(\frac{\partial \mathfrak{L}}{\partial A_\mu^a}\right) \left(\partial_\mu \omega^a + c^a{}_{bc}\, \omega^b A_\mu^c \right) + \left(\frac{\partial \mathfrak{L}}{\partial(\partial_\nu A_\mu^a)}\right) \partial_\nu \left(\partial_\mu \omega^a + c^a{}_{bc}\, \omega^b A_\mu^c \right) \,, \end{aligned} \tag{C.73}$$

and for this to vanish for arbitrary local functions $\omega^a(x)$ the coefficients in it of ω^a, $\partial_\mu \omega^a$ and $\partial_\mu \partial_\nu \omega^a$ must separately vanish. The coefficient of $\partial_\mu \partial_\nu \omega^a$ vanishes if

$$\frac{\partial \mathfrak{L}}{\partial(\partial_\mu A_\nu^a)} = -\frac{\partial \mathfrak{L}}{\partial(\partial_\nu A_\mu^a)} \,, \tag{C.74}$$

which means that A_μ^a appears differentiated in $\mathfrak{L}$ only through the antisymmetric combination $f_{\mu\nu}^a := \partial_\mu A_\nu^a - \partial_\nu A_\mu^a$.

Changing independent variable from $\partial_\mu A_\nu^a$ to $f_{\mu\nu}^a$, the coefficient of $\partial_\mu \omega^a$ vanishes when

$$\left(\frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\right) \mathrm{i}(\mathcal{T}_a)^i{}_j\, \phi^j + \frac{\partial \mathfrak{L}}{\partial A_\mu^a} + 2\left(\frac{\partial \mathfrak{L}}{\partial f_{\mu\nu}^b}\right) c^b{}_{ac}\, A_\nu^c = 0 \,, \tag{C.75}$$

To extract the implications of this condition, consider first terms in $\mathfrak{L}$ that do not depend on ϕ^i or its derivative at all. In this case (C.75) states that $f_{\mu\nu}^a$ and A_μ^a can only appear together in $\mathfrak{L}$, though the one combination

$$F_{\mu\nu}^a = f_{\mu\nu}^a + c^a{}_{bc} A_\mu^b A_\nu^c = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + c^a{}_{bc} A_\mu^b A_\nu^c \,. \tag{C.76}$$

What is special about this quantity is that the dependence on $\partial_\mu \omega^a$ cancels when it is transformed using (C.72), leaving $\delta F_{\mu\nu}^a = c^a{}_{bc} \omega^b F_{\mu\nu}^c$.

Re-introducing a dependence on ϕ^i and trading $\partial\mathfrak{L}/\partial A_\mu^a$ for $(\partial\mathfrak{L}/\partial A_\mu^a)_F$ (with the subscript indicating the derivative is taken at fixed $F_{\mu\nu}^a$ instead of fixed $\partial_\mu A_\nu^a$ or $f_{\mu\nu}^a$), condition (C.75) becomes

$$\left(\frac{\partial \mathfrak{L}}{\partial(\partial_\mu \phi^i)}\right) \mathrm{i}(\mathcal{T}_a)^i{}_j\, \phi^j + \left(\frac{\partial \mathfrak{L}}{\partial A_\mu^a}\right)_F = 0 \,, \tag{C.77}$$

which implies $\partial_\mu \phi^i$ must always appear together with A^a_μ through the covariant-derivative combination

$$(D_\mu \phi)^i := \partial_\mu \phi^i - \mathrm{i}(\mathcal{T}_a)^i{}_j A^a_\mu \phi^j \,. \tag{C.78}$$

As is easily verified $\partial_\mu \omega^a$ also cancels in the transformation rule $\delta(D_\mu\phi)^i = \mathrm{i}\omega^a(\mathcal{T}_a)^i{}_j(D_\mu\phi)^j$. The lesson from the $\partial_\mu\omega^a$ term is that A^a_μ only appears in $\mathfrak{L}$ as part of $(D_\mu\phi)^i$ or $F^a_{\mu\nu}$, and never on its own. Notice that the covariant derivative and field strength are related by the following easily proven identity

$$[D_\mu, D_\nu]\phi = -\mathrm{i}(\mathcal{T}_a\phi)\, F^a_{\mu\nu} \,. \tag{C.79}$$

Finally, setting the coefficient of undifferentiated ω^a to zero gives

$$\left(\frac{\partial \mathfrak{L}}{\partial \phi^i}\right)\mathrm{i}(\mathcal{T}_a)^i{}_j\, \phi^j + \left(\frac{\partial \mathfrak{L}}{\partial (D_\mu \phi^i)}\right)\mathrm{i}(\mathcal{T}_a)^i{}_j\, (D_\mu\phi)^j + \left(\frac{\partial \mathfrak{L}}{\partial F^b_{\mu\nu}}\right) c^b{}_{ac}\, F^c_{\mu\nu} = 0 \,, \tag{C.80}$$

which simply states that $\mathfrak{L}$ should be built from ϕ^i, $(D_\mu\phi)^i$ and $F^a_{\mu\nu}$ in a way that is invariant under *global* transformations (for which ω^a is a constant).

Here is the point: a global symmetry can be promoted to a local symmetry by introducing a new massless spin-one particle for each symmetry generator, and then building the lagrangian out of undifferentiated fields like ϕ^i, covariant derivatives like $(D_\mu\phi)^i$, and covariant field strengths like $F^a_{\mu\nu}$.

C.5.1 Higgs Mechanism

Historically, when promoting global to local symmetries the need for massless spin-one particles was seen as a handicap. Although it worked splendidly for the massless photon in quantum electrodynamics, the prospects for applications elsewhere seemed limited (the phenomenon of confinement kept prevented the discovery of massless gluons until later).

The modern understanding wherein all fundamental spin-one particles, massless or not, are gauge bosons had to await the discovery (by Brout and Englert [55], Guralnik, Hagan and Kibble [56, 57], Higgs [58, 59] and others, building on earlier work by Anderson [60] for non-relativistic systems) of the *Higgs* mechanism. This mechanism shows why the spin-one particles can be massive, provided they are associated with local symmetries that are spontaneously broken.

It is fundamental that systems with spontaneously broken symmetries do not have unique ground states, because by assumption the action of $U(g)$ on one ground state gives a different state, $|\tilde{0}\rangle := U|0\rangle \neq |0\rangle$. But because a symmetry satisfies $UH = HU$ the state $|\tilde{0}\rangle$ has precisely the same energy as does $|0\rangle$, making it a second 'ground' state.

Related to this, the operators $U(g) = \exp[\mathrm{i}Q]$, and their generators Q, are less useful when dealing with spontaneously broken symmetries in field theories, particularly in situations where the spatial directions are infinitely large and the ground state is translation invariant [43]. This is because in a field theory Noether's theorem ensures that

$$Q = \int \mathrm{d}^3x\; j^0(x) \,, \tag{C.81}$$

arises as the integral over a local current density. It follows then that the state $Q|0\rangle$ is not normalizable (and so does not lie within the Fock space built apon $|0\rangle$), because

$$||Q|0\rangle||^2 = \langle 0|QQ|0\rangle = \int \mathrm{d}^3x\, \langle 0|Qj^0(x)|0\rangle = \int \mathrm{d}^3x\, \langle 0\,|Qj^0(0)|0\rangle\,, \tag{C.82}$$

where the first equality uses $Q^* = Q$, the second equality uses (C.81) and the last equality uses the representation of translation symmetries to write $j^0(x) = e^{\mathrm{i}P\cdot x} j^0(0) e^{-\mathrm{i}P\cdot x}$, together with the translation invariance of Q and the ground state: $[Q, P_\mu] = 0$ and $P_\mu|0\rangle = 0$. The final result diverges like the volume of space because the integrand does not depend at all on x^μ.

Because of this it is preferable to have a more useful proxy for spontaneous symmetry breaking than the evaluation of $U(g)|0\rangle$ or $Q|0\rangle$. The existence of a nonzero *order parameter* fills this role, by providing a simpler-to-use criterion for the non-invariance of the vacuum. For example, imagine two fields that are related by a symmetry, such as if[10]

$$\psi(x) = \mathrm{i}[Q, \phi(x)]\,. \tag{C.83}$$

Then $\psi(x)$ is an order parameter for the symmetry generated by Q if its vacuum expectation value (*vev*) is nonzero: $v := \langle 0|\psi(x)|0\rangle \neq 0$. A nonzero *vev* is a proxy for spontaneous symmetry breaking because an unbroken symmetry implies $Q|0\rangle = 0$ — and its conjugate $\langle 0|Q = 0$ — and both of these arise in the right-hand side of (C.83) once its vacuum expectation-value is taken. Since unbroken symmetry implies $v = 0$ it follows that nonzero v implies the symmetry must be broken.

In order not to break any spacetime symmetries the order-parameter field must be a Lorentz-scalar and independent of x^μ. To see how this works at weak couplings, where semiclassical reasoning is valid, consider then a collection of scalar fields, ϕ^i, $i = 1, \cdots, N$, which without loss of generality can be chosen to be real. Suppose the particles represented by these fields couple to a collection of spin-one particles represented by A^a_μ, with local symmetry group $\delta\phi^i = \mathrm{i}\omega^a (\mathcal{T}_a)^i{}_j\, \phi^j$ and $\delta A^a_\mu = \partial_\mu \omega^a + c^a{}_{bc}\omega^b A^c_\mu$. A lagrangian density for these particles involving only up to two derivatives is

$$\mathfrak{L} = -V(\phi) - \frac{1}{2} Z_{ij} (D_\mu\phi)^i (D^\mu\phi)^j - \frac{1}{4g^2}\gamma_{ab} F^a_{\mu\nu} F^{b\mu\nu}\,, \tag{C.84}$$

where Z_{ij} are a collection of numerical coefficients (that can be set to δ_{ij} by appropriately redefining the fields, the covariant derivative is $D_\mu\phi = \partial_\mu\phi - \mathrm{i}\mathcal{T}_a A^a_\mu \phi$ and $F^a_{\mu\nu}$ is as defined in (C.76) with $c^a{}_{bc}$ the structure constants associated with the generators $\mathcal{T}_a$.

For the present purposes the important feature is to have a potential energy, $V(\phi)$, whose minimum occurs for $\phi^i \neq 0$. This is easily arranged following the example of the toy model of §1.1. For instance for symmetries that preserve the quantity $\phi^T\phi = \sum_i (\phi^i)^2$, the potential

$$V(\phi) = \frac{\lambda}{4}\left(\phi^T\phi - v^2\right)^2\,, \tag{C.85}$$

does the job. For positive real parameters λ and v^2 this potential is strictly non-negative and

[10] Notice that commutators like $\delta\phi(x) = \mathrm{i}[Q, \phi(x)]$ are usually well-defined even if the action of Q on $|0\rangle$ is not. This is because the equal-time commutators of fields are usually local, such as the canonical commutation relations $\Pi(\mathbf{x}, t), \phi(\mathbf{y}, t)] = -\mathrm{i}\delta^3(\mathbf{x} - \mathbf{y})$, for which the delta-function removes the otherwise diverging spatial integration.

vanishes for the minimizing surface $\phi^T\phi = v^2$. This does not pick a unique solution for ϕ^i because it contains all configurations related by the symmetries that preserve $\phi^T\phi$. Each of these provides an equally good vacuum, and all are equivalent to the extent that they are related by symmetries.

For concreteness' sake choose the vacuum to be the one with $\langle\phi^1\rangle = v$ and all others zero, and expand all quantum fields about this semi-classical vacuum expectation value: $\phi^1 = v + \hat{\phi}^1$ and $\phi^i = \hat{\phi}^i$ for $i \neq 1$. The leading correction to the classical limit keeps only terms quadratic in the $\hat{\phi}^i$ and A^a_μ.

The revealing terms in this expansion are those arising within the scalar kinetic term,

$$\mathfrak{L}_{\rm sk} = -\frac{1}{2}Z_{ij}D^\mu\phi^i D_\mu\phi^j = -\frac{1}{2}Z_{ij}v^2(\mathcal{T}_a)^i{}_1(\mathcal{T}_b)^j{}_1A^a_\mu A^{b\mu} + \mathrm{i}Z_{ij}v(\mathcal{T}_a)^i{}_1A^a_\mu\partial^\mu\hat{\phi}^j + \cdots . \quad \text{(C.86)}$$

The second term on the right-hand side is unusual inasmuch as it mixes scalar and vector degrees of freedom. The good news is that it is always possible to perform a gauge transformation to completely remove this term (a choice called 'unitary gauge'). The gauge transformation required to reach this choice absorbs one scalar degree of freedom into A^a_μ for each independent symmetry generator that is broken by the vacuum. It is the addition of these new states that provides the missing longitudinal spin states required to promotes the two spin-states of a massless spin-one particle to the three spin states of a massive one.

Once this is removed, the first term on the right-hand side is revealed as a spin-one mass term – compare with (C.35). Canonically normalizing fields (which sets $Z_{ij} = \delta_{ij}$) and computing the particle energies at zero momentum gives the spin-one mass matrix

$$\mu^2_{ab} = V^T\mathcal{T}_a\mathcal{T}_bV\,, \quad \text{(C.87)}$$

where $V^i = \langle\phi^i\rangle$ denotes the field-vector containing the field vacuum expectation values. The spin-one particles indeed acquire a mass when their associated gauge symmetry becomes spontaneously broken.

C.5.2 General relativity

A short summary of the basics of General Relativity (GR) is also appropriate here, since gravitational interactions arise at several points within the main text. Although a proper discussion goes well beyond the scope of this book, this section suffices to collect some of the main formulae.

There is a strong analogy between GR and gauge theories of massless nonabelian spin-one particles, like QCD. Both involve massless states (though the gluons of QCD, unlike the graviton of GR, are prevented from escaping to infinity as massless states due to the growth of the strong force with distance). Both involve nonabelian local symmetries: for QCD these are the local $SU_c(3)$ colour transformations of the Standard Model, while for GR these are a combination of local diffeomorphisms (and local Lorentz transformations, when coupled to fields with spin). Both also involve nonlinear self-interactions wherein the force carriers themselves carry charges (that is, gluons carry colour and gravitons carry energy and momentum). This makes them unlike abelian massless spin-one particles like photons, which do not carry electric charge.

The basic field for GR is the spacetime metric itself: $g_{\mu\nu}(x)$. The local symmetries in this case correspond to local diffeomorphisms of the type $x^\mu \to x^\mu + \xi^\mu(x)$, under which $g_{\mu\nu}$ transforms linearly, like a covariant rank-two tensor,

$$\delta g_{\mu\nu} = \mathcal{L}_\xi g_{\mu\nu} := \xi^\lambda \partial_\lambda g_{\mu\nu} + \partial_\mu \xi^\lambda g_{\lambda\nu} + \partial_\nu \xi^\lambda g_{\mu\lambda}\,, \tag{C.88}$$

where the right-hand side defines the *Lie derivative* $\mathcal{L}_\xi$ of the metric. Other fields, such as scalar or vector fields, similarly transform under diffeomorphisms as their index content suggests

$$\delta\phi = \mathcal{L}_\xi \phi := \xi^\lambda \partial_\lambda \phi \quad \text{and} \quad \delta V_\mu = \mathcal{L}_\xi V_\mu = \xi^\lambda \partial_\lambda V_\mu + \partial_\mu \xi^\lambda V_\lambda\,, \tag{C.89}$$

and so on.

Covariant derivatives and curvatures

Just like for local gauge invariance, local lagrangian densities that are invariant under these transformations can be built by starting with a lagrangian that is invariant under a global symmetry (in this case the spacetime symmetry of Poincaré invariance) with two provisos: all ordinary derivatives get promoted to covariant derivatives, $\partial_\mu \to D_\mu$, and the gauge field itself (in this case $g_{\mu\nu}$) appears through a covariant field strength – in this case the Riemann tensor (see below) – and its (covariant) derivatives.

The covariant derivatives appropriate for scalars and vectors transforming as in (C.89) are

$$D_\mu\phi := \partial_\mu\phi\,, \quad D_\mu V_\nu := \partial_\mu V_\nu - \Gamma^\lambda_{\mu\nu} V_\lambda \quad \text{and} \quad D_\mu V^\nu := \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda\,, \tag{C.90}$$

where the Christoffel symbol is defined by (A.7), reproduced for convenience here:

$$\Gamma^\mu_{\nu\lambda} = \frac{1}{2} g^{\mu\alpha}\left[\partial_\nu g_{\alpha\lambda} + \partial_\lambda g_{\alpha\nu} - \partial_\alpha g_{\nu\lambda}\right]. \tag{C.91}$$

Here $g^{\mu\nu}$ denotes the inverse metric, defined by the condition $g^{\mu\nu} g_{\nu\lambda} = \delta^\mu_\lambda$. With the above definitions the metric is covariantly constant:

$$D_\mu g_{\nu\lambda} = 0 = D_\mu g^{\nu\lambda}\,. \tag{C.92}$$

Notice that the definitions (C.90) ensure that covariant derivatives satisfy the usual product rule for derivatives: *e.g.*

$$\partial_\mu(V_\lambda W^\lambda) = D_\mu(V_\lambda W^\lambda) = (D_\mu V_\lambda) W^\lambda + V_\lambda (D_\mu W^\lambda)\,. \tag{C.93}$$

Notice also that antisymmetrized ordinary derivatives are already covariant, inasmuch as

$$D_\mu V_\nu - D_\nu V_\mu = \partial_\mu V_\nu - \partial_\nu V_\mu\,, \tag{C.94}$$

so (for example) the relation between electromagnetic field strength and vector potential does not change in the presence of a gravitational field. It is this observation about how Christoffel symbols cancel in antisymmetric tensors that underlies the study of differential forms and exterior derivatives: covariant quantities that can be defined without making reference to a metric.

The covariant field strength containing derivatives of $g_{\mu\nu}$ appropriate for diffeomorphisms is the Riemann tensor, $R^{\mu}{}_{\nu\lambda\rho}$, as defined by (A.6), again reproduced here:

$$R^{\mu}{}_{\nu\lambda\rho} = \partial_\rho \Gamma^{\mu}_{\nu\lambda} + \Gamma^{\mu}_{\rho\sigma}\Gamma^{\sigma}_{\nu\lambda} - (\rho \leftrightarrow \lambda)\,. \tag{C.95}$$

This definition implies the covariant version of this tensor, $R_{\mu\nu\lambda\rho} = g_{\mu\sigma}R^{\sigma}{}_{\nu\lambda\rho}$, has the important symmetry properties $R_{\mu\nu\lambda\rho} = R_{\lambda\rho\mu\nu} = -R_{\nu\mu\lambda\rho} = -R_{\mu\nu\rho\lambda}$ as well as the 'Bianchi' identities

$$R_{\mu\nu\lambda\rho} + R_{\mu\lambda\rho\nu} + R_{\mu\rho\nu\lambda} = 0\,, \tag{C.96}$$

and

$$D_\sigma R_{\mu\nu\lambda\rho} + D_\lambda R_{\mu\nu\rho\sigma} + D_\rho R_{\mu\nu\sigma\lambda} = 0\,. \tag{C.97}$$

Finally, the Riemann tensor is related to the commutator of two covariant derivatives; a straightforward use of the definitions implies the gravitational analog of (C.79),

$$[D_\mu, D_\nu]V^\lambda = R^{\lambda}{}_{\nu\rho\mu}V^\rho\,. \tag{C.98}$$

Generally covariant actions

A local action arises as an integral over a lagrangian density, $S = \int \mathrm{d}^4x\, \mathfrak{L}$. The lagrangian density cannot be a scalar under diffeomorphisms, however, because $\mathfrak{L}$ must transform in such a way as to cancel the transformation of the measure d^4x. This is accomplished if $\mathfrak{L} = \sqrt{-g}\, L$ where $g = \det(g_{\mu\nu}) < 0$ is the determinant of the metric and L is a scalar under diffeomorphisms (*i.e.* transforms as a scalar field).

The appearance of $\sqrt{-g}$ in the lagrangian density makes the following identity very useful:

$$\partial_\mu(\sqrt{-g}\, V^\mu) = D_\mu(\sqrt{-g}\, V^\mu) = \sqrt{-g}\, D_\mu V^\mu \tag{C.99}$$

for any 4-vector V^μ. This shows that integrals of the form $\int \mathrm{d}^4x\, \sqrt{-g}\, D_\mu V^\mu$ are total divergences and so depend only on boundary information.

Because the Riemann tensor already involves two derivatives of the metric, it should appear linearly in the kinetic term for the metric. Because of the symmetries there are two types of tensors that can be built by taking traces of the Riemann tensor. The first is the Ricci tensor, $R_{\mu\nu} := R^{\lambda}{}_{\mu\lambda\nu} = R_{\nu\mu}$, and the second is the Ricci scalar $R = g^{\mu\nu}R_{\mu\nu}$. The Einstein-Hilbert lagrangian for gravity coupled to matter is then given by

$$\mathfrak{L} = \sqrt{-g}\left[-\frac{1}{2\kappa^2}R + L_m(\phi, A_\mu)\right], \tag{C.100}$$

where $\kappa^2 = 8\pi G_N$ and L_m denotes the generally covariant action for matter fields, given (for example) for a charged scalar field and electromagnetism by

$$L_m = -V(\phi^*\phi) - g^{\mu\nu}D_\mu\phi^*\, D_\nu\phi - \frac{1}{4}g^{\mu\nu}g^{\lambda\rho}F_{\mu\lambda}F_{\nu\rho}\,, \tag{C.101}$$

where $D_\mu\phi = (\partial_\mu - \mathrm{i}qA_\mu)\phi$ and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

The Einstein equation obtained from varying this action is

$$R^{\mu\nu} - \frac{1}{2} R\, g^{\mu\nu} + \kappa^2 T^{\mu\nu} = 0\,, \tag{C.102}$$

where

$$T^{\mu\nu} := \frac{2}{\sqrt{-g}} \left(\frac{\delta S_m}{\delta g_{\mu\nu}} \right), \tag{C.103}$$

and so on.

C.5.3 Spacetime symmetries reloaded

Once the metric is recognized as being a dynamical field it is worth revisiting the idea of a spacetime symmetry. Recall that in the bulk of this book spacetime symmetries are regarded as those transformations

$$\delta x^\mu = \xi^\mu(x) \tag{C.104}$$

that leave the Minkowski metric invariant: $\delta(\eta_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu) = 0$, or

$$\delta\eta_{\mu\nu} = \mathcal{L}_\xi \eta_{\mu\nu} := \xi^\lambda \partial_\lambda \eta_{\mu\nu} + \partial_\mu \xi^\lambda \eta_{\lambda\nu} + \partial_\nu \xi^\lambda \eta_{\mu\lambda} = 0\,, \tag{C.105}$$

(compare with eq. (C.88) which defines the transformation for a general metric). The general solution to this condition led to the Poincaré group: $\xi^\mu = a^\mu + \omega^\mu{}_\nu x^\nu$ where a^μ and $\omega_{\mu\nu} = -\omega_{\nu\mu}$ (with $\omega_{\mu\nu} = \eta_{\mu\lambda}\omega^\lambda{}_\nu$) are constant parameters representing translations and Lorentz transformations.

Generally covariant theories provide a new context for these transformations, because for these the action is invariant under a much broader set of transformations: general diffeomorphisms corresponding to (C.104) and (C.88) for general $g_{\mu\nu}(x)$ and $\xi^\mu(x)$. Within this new context $g_{\mu\nu} = \eta_{\mu\nu}$ is a specific solution to the field equations and so can be regarded as being the analog of a field expectation-value: $\langle g_{\mu\nu}(x)\rangle = \eta_{\mu\nu}$ in much the same way that the field ϕ acquires a nonzero expectation value $\langle\phi(x)\rangle = v$ in the ground state of the toy model of §1.1.

From this point of view eq. (C.105) simply identifies that subset of symmetry transformations that leaves the metric's expectation value unchanged — that is, are not spontaneously broken by $\langle g_{\mu\nu}\rangle = \eta_{\mu\nu}$. More generally, the diffeomorphisms that leave a generic metric unchanged are called *isometries* and must satisfy $\delta g_{\mu\nu} = \mathcal{L}_\xi g_{\mu\nu} = 0$ and so

$$\xi^\lambda \partial_\lambda g_{\mu\nu} + \partial_\mu \xi^\lambda g_{\lambda\nu} + \partial_\nu \xi^\lambda g_{\mu\lambda} = D_\mu \xi_\nu + D_\nu \xi_\mu = 0\,, \tag{C.106}$$

where the first equality follows from the definition of the covariant derivative and uses the definition $\xi_\mu := g_{\mu\lambda}\xi^\lambda$. Any solution ξ^μ to (C.106) is called a Killing vector field, and such fields need not exist for arbitrary metrics. From this point of view (C.105) states that Poincaré transformations are the isometries of Minkowski spacetime.

Conserved currents

These observations provide another way to identify (and count) conserved currents, at least for gauge symmetries whose transformation parameters are spacetime-independent.

Consider first (for simplicity) an abelian internal local symmetry that acts only on some matter fields through a transformation rule, $\delta\phi^i = \omega(x) f^i(\phi)$, and on the gauge potential $\delta A_\mu = \partial_\mu \omega$. The matter action, $S_m[\phi, A_\mu]$ for the fields ϕ^i must be invariant under the gauge symmetry, and this makes it depend on A_μ through the covariant derivative $D_\mu\phi^i$. Invariance means S_m satisfies

$$\delta S_m = \int \mathrm{d}^4x \left[\frac{\delta S_m}{\delta\phi^i(x)}\, \omega(x) f^i + \frac{\delta S_m}{\delta A_\mu(x)} \partial_\mu \omega(x) \right] = 0 \tag{C.107}$$

for any field configurations $\phi^i(x)$ and $A_\mu(x)$ and for any symmetry parameter $\omega(x)$. If this is specialized to a solution to the ϕ^i field equation, $\delta S_m/\delta\phi^i = 0$, then the first term vanishes leaving the result[11]

$$0 = \int \mathrm{d}^4x\, J^\mu \partial_\mu \omega = -\int \mathrm{d}^4x\, \omega\, (\partial_\mu J^\mu)\,, \tag{C.108}$$

where the second equality performs an integration by parts (and discards the surface term), and defines the current

$$J^\mu := \frac{\delta S_m}{\delta A_\mu(x)}\,. \tag{C.109}$$

Since (C.108) must vanish for *any* $\omega(x)$ it must be true that J^μ as defined in (C.109) is conserved, in the sense that the ϕ^i equations of motion imply $\partial_\mu J^\mu = 0$. It is easy to verify in simple examples that this definition of the current agrees with the Noether-current derivation for internal symmetries given in §4.1.1.

The same logic also goes through for spacetime symmetries in generally covariant systems, and provides a more systematic way to count currents. In this case it is the metric, $g_{\mu\nu}$ that plays the role of the gauge potential, but otherwise the argument goes through identically. Consider then a matter action $S_m[\phi^i, g_{\mu\nu}]$ that is generally covariant in the sense that it is unchanged by some transformation $\delta\phi^i = \pounds_\xi \phi^i$ and $\delta g_{\mu\nu} = \pounds_\xi g_{\mu\nu}$:

$$\delta S_m = \int \mathrm{d}^4x \left[\frac{\delta S_m}{\delta\phi^i(x)}\, \pounds_\xi \phi^i + \frac{\delta S_m}{\delta g_{\mu\nu}(x)}\, \pounds_\xi g_{\mu\nu} \right] = 0\,. \tag{C.110}$$

Specializing to configurations satisfying the ϕ^i equations of motion and using the definition of $\pounds_\xi g_{\mu\nu}$ given in the first equality of (C.106) then allows (after integration by parts) eq. (C.110) to be rewritten as

$$0 = \int \mathrm{d}^4x\ \sqrt{-g}\, T^{\mu\nu} D_\mu \xi_\nu = -\int \mathrm{d}^4x\ \sqrt{-g}\, \xi_\nu\, D_\mu T^{\mu\nu}\,, \tag{C.111}$$

and so (because ξ_ν is arbitrary) the stress-energy tensor, $T^{\mu\nu}$, defined by (C.103), must be covariantly conserved, in the sense that

$$D_\mu T^{\mu\nu} = 0\,. \tag{C.112}$$

This definition of the conserved stress-energy has the enormous advantage that it is what appears in the Einstein equations, (C.102), and so is precisely what gravity couples to.

For any specific metric that has an isometry, in the sense that the second equality of

[11] The gauge field does not in general satisfy $\delta S_m/\delta A_\mu = 0$ even for classical fields because S_m only consists of the matter action and does not include, for example, the Maxwell action $-\frac{1}{4} F_{\mu\nu}F^{\mu\nu}$.

(C.106) is satisfied for some ξ_μ, a standard conserved current can be defined for each isometry,

$$j^\mu = T^{\mu\nu}\xi_\nu\,, \tag{C.113}$$

that satisfies $D_\mu j^\mu = 0$ — as can be seen using eqs. (C.112) and $D_\mu\xi_\nu + D_\nu\xi_\mu = 0$ (*i.e.* eq. (C.106)). In the special case of flat Minkowski space these are the conserved currents for Poincaré invariance, but (C.113) shows that they are all really built from one basic quantity: the stress-energy tensor, $T^{\mu\nu}$.

C.6 Nonlinear realizations

The nonlinear realization used to implement spontaneously broken symmetries in an effective theory can be less intuitive than is the linear realization used for unbroken symmetries. But it is worth understanding given the widespread appearance of Goldstone bosons throughout physics.

This appendix derives the 'standard' nonlinear realization for the general case of an internal symmetry group G spontaneously broken down to a subgroup $H \subset G$. Following steps initially taken by [12, 13] and using the notation of [107] this is done by generalizing the arguments used for the abelian broken symmetry presented in the toy model of the main text. Since half the art of constructing nonlinear realizations involves choosing variables that transform conveniently, the first steps in this construction motivate the choices to be made by describing a simple nonabelian version of the toy model.

C.6.1 A nonabelian toy model

To set up the standard transformation law, consider N real scalar fields, $\phi^i, i = 1, \ldots, N$, arranged for convenience into an N-component column vector, Φ. There is no loss of generality in using real fields, since any complex fields can be decomposed into real and imaginary parts.

The nonabelian toy model is defined by the lagrangian density

$$\mathfrak{L} = -\frac{1}{2}\,\partial_\mu\Phi^T\partial^\mu\Phi - V(\Phi), \tag{C.114}$$

where the superscript 'T' denotes the transpose, and where $V(\Phi)$ is a potential whose detailed form is not important in what follows. The lagrangian's kinetic term is manifestly invariant under the $O(N)$ group ($N \times N$ orthogonal matrices, $O^T O = 1$) of global rotations: $\Phi \to O\Phi$, where the O's are independent of spacetime position, $\partial_\mu O = 0$. Because the fields are chosen to be real all generators of these symmetries are simultaneously hermitian, imaginary and antisymmetric: $T_a^\dagger = T_a = -T_a^T = -T_a^*$.

In general the potential $V(\Phi)$ need not also be $O(N)$-invariant, but it is assumed to preserve a subgroup $G \subset O(N)$, in the sense that

$$V(g\,\Phi) = V(\Phi) \qquad \text{for all } g \in G \text{ and all } \Phi. \tag{C.115}$$

The potential V is assumed to be satisfy two properties. First, its parameters are assumed to be chosen to allow a weak-coupling semiclassical treatment of the model's predictions. Second, it is assumed to be minimized at field values $\langle\Phi\rangle \neq 0$, for which the symmetry group G is generically spontaneously broken to a subgroup $H \subset G$ defined by: $h\langle\Phi\rangle = \langle\Phi\rangle$, for all $h \in H$. It is convenient to choose generators, t_i, of H as part of the basis of the Lie algebra of G, writing $\{T_a\} = \{t_i, X_\alpha\}$ where T_a are a basis of generators of the algebra of G while X_α are the broken generators of the coset G/H (for more about the nomenclature see §C.4.1).

A choice of variables

The idea is to identify the Goldstone and non-Goldstone degrees of freedom in this model and to identify how these each realize the model's symmetries. Within a semiclassical framework this involves sorting the fields $\Phi = \{\phi^i\}$ into a set of Goldstone modes, $\Xi = \{\xi^\alpha\}$, plus an orthogonal set of remaining physical fields, $\mathcal{X} = \{\chi^n\}$.

As usual, Goldstone modes are obtained by performing symmetry transformations on the ground state, and for infinitesimal transformations this corresponds to the directions $X_\alpha\langle\Phi\rangle$ in field space. That is, the components of Φ in this direction, $\langle\Phi\rangle^T X_\alpha \Phi$, are the ones that create and destroy Goldstone particles. It is straightforward to verify that the G-invariance of the potential ensures the masslessness of these modes in the semiclassical approximation. This gives precisely one Goldstone mode for each generator of G/H.

Experience with the abelian symmetries of the toy model of §1.1 suggests that the variables $\langle\Phi\rangle^T X_\alpha \Phi$ need not be the most efficient for making Goldstone properties manifest, however. In particular, the low-energy decoupling of Goldstone modes are most manifest if the freedom to redefine fields is used to arrange that they do not appear at all in the scalar potential. This is most easily arranged by writing

$$\Phi = U(\xi)\,\mathcal{X}\,, \tag{C.116}$$

where

$$U(\xi) = \exp[\mathrm{i}\xi^\alpha(x)X_\alpha]\,, \tag{C.117}$$

is a spacetime-dependent symmetry transformation in the direction of the broken generators, X_α. Since $U(\xi)$ is an element of G, this definition ensures the ξ^α drop out of the scalar potential because G-invariance requires the potential must satisfy $V(U\mathcal{X}) = V(\mathcal{X})$. Consequently all terms in $\mathfrak{L}$ involving the Goldstone bosons, ξ^α, vanish when $\partial_\mu\xi^\alpha = 0$, and eq. (C.116) is the change of variables that makes low-energy properties of Goldstone bosons most manifest.

In order for eq. (C.116) not to over-count the N original fields in Φ the variables $\mathcal{X}$ must satisfy a constraint that keeps them orthogonal (in field space) to the Goldstone directions, such as:

$$\langle\Phi\rangle^T X_\alpha \mathcal{X} = 0, \qquad \text{for all } X_\alpha \tag{C.118}$$

everywhere in spacetime. As a reality check, notice that this constraint ensures the vanishing of the cross terms, proportional to $\partial_\mu\xi^\alpha\partial^\mu\hat{\chi}^n$, in the quadratic part of the expansion of

the kinetic terms about the ground state configuration: $\mathcal{X} = \langle\Phi\rangle + \hat{\mathcal{X}}$. (Proving this uses the identity $\langle\Phi\rangle^T X_\alpha \langle\Phi\rangle = 0$, that is a consequence of the antisymmetry of the X_α's.) It can be shown [107] that it is always possible to change to variables satisfying (C.118) from any original smooth configuration for Φ.

C.6.2 The Nonlinear Realization

The next step asks how the variables ξ^α and χ^n transform under the group G given the simple linear representation of G carried by Φ,

$$\Phi \to \widetilde{\Phi} := g\,\Phi \qquad \text{where} \qquad g = \exp[\mathrm{i}\omega^a T_a] \in G\,. \tag{C.119}$$

This leads to the standard transformation rules widely used when studying Goldstone boson properties.

The transformation rule implied for the new variables, $\xi^\alpha \to \tilde{\xi}^\alpha$ and $\chi^n \to \tilde{\chi}^n$, is found by writing $\Phi = U(\xi)\mathcal{X}$ and $\widetilde{\Phi} = U(\tilde{\xi})\widetilde{\mathcal{X}}$ in (C.119), and so

$$gU(\xi)\mathcal{X} = U(\tilde{\xi})\widetilde{\mathcal{X}}\,, \tag{C.120}$$

for any $g \in G$.

The standard nonlinear transformation law therefore becomes:

$$\xi^\alpha \to \tilde{\xi}^\alpha(\xi, g) \qquad \text{and} \qquad \chi^n \to \tilde{\chi}^n(\xi, g, \chi), \tag{C.121}$$

where

$$g\; e^{\mathrm{i}\xi^\alpha X_\alpha} = e^{\mathrm{i}\tilde{\xi}^\alpha X_\alpha}\, e^{\mathrm{i}u^i t_i} \quad \text{and} \quad \widetilde{\mathcal{X}} = e^{\mathrm{i}u^i t_i}\,\mathcal{X}\,. \tag{C.122}$$

The first of eqs. (C.122) should be read as defining the nonlinear functions $\tilde{\xi}^\alpha(\xi, g)$ and $u^i(\xi, g)$. One first finds the element, $g\, e^{\mathrm{i}\xi\cdot X} \in G$, and then defines the functions $\tilde{\xi}^\alpha$ and u^i by decomposing this matrix into the product of a factor, $e^{\mathrm{i}\tilde{\xi}\cdot X}$, lying in G/H times an element, $e^{\mathrm{i}u\cdot t}$, in H. The second of eqs. (C.122) then defines the transformation rule for the non-Goldstone fields, χ^n.

These transformation laws are generically nonlinear in the Goldstone fields, ξ^α. They nonetheless realize the symmetry group G in that $\tilde{\xi}(\theta, g_1 g_2) = \tilde{\xi}(\tilde{\xi}(\xi, g_2), g_1)$, as can be verified using the definitions of eqs. (C.122) or by noticing that this property is inherited from the original linear representation of G on Φ.

The transformations (C.121) and (C.122) remain linear in the special case where $g = h$ lies in the unbroken sector H. In this case, the solution for u^i and $\tilde{\xi}^\alpha$ are easily seen to be: $e^{\mathrm{i}u\cdot t} = h$ and $\widetilde{U} = hUh^{-1}$ since in this case $hU = \widetilde{U}e^{\mathrm{i}u\cdot t}$, as required. Both χ^n and ξ^α therefore transform *linearly* under the unbroken symmetry transformations of H, with:

$$\begin{aligned} \xi^\alpha X_\alpha \to \tilde{\xi}^\alpha X_\alpha = h(\xi^\alpha X_\alpha)h^{-1} = \xi^\alpha L^\beta{}_\alpha X_\beta\,, \\ \mathcal{X} \to \widetilde{\mathcal{X}} = h\mathcal{X}\,, \end{aligned} \tag{C.123}$$

where the last equality in the first line uses (C.61).

It is harder to be equally explicit for general $g \in G/H$, but closed forms are possible for infinitesimal transformations, $g = 1 + \mathrm{i}\omega^\alpha X_\alpha + \cdots$, if one works with a basis of generators that satisfy (C.55). In this case writing $\gamma = 1 + \mathrm{i}u^i(\xi, \omega)t_i + \cdots$, and $U(\tilde{\xi}) = U(\xi)[1 +$

$i\Delta^\alpha(\xi,\omega)X_\alpha + \cdots]$ and using (C.122) implies $u^i(\xi,\omega)$ and $\Delta^\alpha(\xi,\omega)$ are given (at linear order in ω^a) explicitly by:

$$\Delta_\alpha = \mathrm{Tr}\left[X_\alpha e^{-i\xi\cdot X}(\omega\cdot X)e^{i\xi\cdot X}\right] \simeq \omega_\alpha - c_{\alpha\beta\gamma}\omega^\beta\xi^\gamma + O(\omega\,\xi^2)\,, \tag{C.124}$$

and

$$u_i = \mathrm{Tr}\left[t_i e^{-i\xi\cdot X}(\omega\cdot X)e^{i\xi\cdot X}\right] \approx -c_{i\alpha\beta}\omega^\alpha\xi^\beta + O(\omega\,\xi^2)\,. \tag{C.125}$$

These expressions liberally use the conventional choices $\mathrm{Tr}\,(X_\alpha X_\beta) = \delta_{\alpha\beta}$, $\mathrm{Tr}\,(t_i t_j) = \delta_{ij}$ and $\mathrm{Tr}\,(t_i X_\alpha) = 0$ for the basis of generators of the Lie algbra of G.

In particular, the transformation rules for the ξ^α under broken symmetries implied by (C.124) are

$$\delta\xi^\alpha = \omega^\alpha - c^\alpha{}_{\beta\gamma}\,\omega^\beta\xi^\gamma + O(\omega\,\xi^2)\,. \tag{C.126}$$

This transformation rule is both inhomogeneous (*i.e.* includes a shift) and acts nonlinearly on the fields ξ^α. Inhomogeneous transformations are characteristic of Goldstone bosons because shifts show that a symmetry necessarily changes the vacuum (it changes because the *vev* of the Goldstone boson field – *i.e.* the relevant order parameter – changes). It is the shift component of the symmetry that precludes ξ^α from appearing undifferentiated in the lagrangian and so enforces the low-energy decoupling of Goldstone states. The nonlinearity allows low-energy interactions to arise involving two derivatives; fewer than are possible in the abelian case studied in the toy model of §1.1.

C.6.3 Invariant Lagrangians

The transformation rules allow the construction of G-invariant Lagrangians built directly using the ξ^α and χ^n fields. The main complication arises from the construction of the kinetic terms, since the nonlinearity of the transformation rules for the fields makes them more like local than global transformations due to the spacetime-dependence of the fields.

Connections and vielbeins

The toy model provides insight into how to construct G-invariant lagrangians. The kinetic term of the toy model is proportional to $\partial_\mu\Phi^T\partial^\mu\Phi$ and so is manifestly G invariant. This must remain so after performing the change of variables to ξ^α and $\mathcal{X}^n$, and it is instructive to see how this comes about.

To this end notice that the replacement $\Phi = U(\xi)\mathcal{X}$ implies $\partial_\mu\Phi = U(\partial_\mu\mathcal{X} + U^{-1}\partial_\mu U\mathcal{X})$. This suggests defining the combination

$$\mathcal{D}_\mu\mathcal{X} = \partial_\mu\mathcal{X} + U^{-1}\partial_\mu U\mathcal{X}\,, \tag{C.127}$$

as a covariant derivative for $\mathcal{X}$. Applying the transformations (C.121) and (C.122) to this shows that it transforms covariantly: $\mathcal{D}_\mu\mathcal{X} \to \mathtt{h}\,\mathcal{D}_\mu\mathcal{X}$, where $\mathtt{h} := e^{i\omega\cdot t}$. It does so because $U^{-1}\partial_\mu U$ transforms like a gauge potential:

$$U^{-1}\partial_\mu U \to \tilde{U}^{-1}\partial_\mu\tilde{U} = \mathtt{h}\,(U^{-1}\partial_\mu U)\,\mathtt{h}^{-1} - (\partial_\mu\mathtt{h})\,\mathtt{h}^{-1}\,. \tag{C.128}$$

More information emerges if $U^{-1}\partial_\mu U$ is separated into a piece proportional to X_α plus one proportional to t_i since the inhomogeneous term, $(\partial_\mu \mathrm{h})\mathrm{h}^{-1}$, is purely proportional to t_i. Defining $\mathcal{A}^i_\mu$ and e^α_μ by

$$U^{-1}\partial_\mu U = -\mathrm{i}\mathcal{A}^i_\mu t_i + \mathrm{i}e^\alpha_\mu X_\alpha, \tag{C.129}$$

eq. (C.128) implies each of $\mathcal{A}^i_\mu(\xi)$ and $e^\alpha_\mu(\xi)$ have separate transformation rules,

$$\begin{aligned} &\mathcal{A}^i_\mu(\xi)t_i \to \mathcal{A}^i_\mu(\tilde{\xi})t_i = \mathrm{h}\,[\mathcal{A}^i_\mu(\xi)t_i]\,\mathrm{h}^{-1} - \mathrm{i}\partial_\mu \mathrm{h}\,\mathrm{h}^{-1}, \\ \text{and}\quad &e^\alpha_\mu(\xi)X_\alpha \to e^\alpha_\mu(\tilde{\xi})X_\alpha = \mathrm{h}\,[e^\alpha_\mu(\xi)X_\alpha]\,\mathrm{h}^{-1}\,. \end{aligned} \tag{C.130}$$

The quantity $\mathcal{A}^i_\mu$ therefore transforms as if it were a gauge potential for local H transformations. To see this more explicitly, for infinitesimal $g \simeq 1 + \mathrm{i}\omega^\alpha X_\alpha$ and $\mathrm{h}(\xi, g) \simeq 1 + \mathrm{i}u^i(\xi,\omega)\,t_i$ the above definitions give (compare with[12] eq. (C.72))

$$\delta\mathcal{A}^i_\mu(\xi) = \partial_\mu u^i(\xi,\omega) - c^i{}_{jk}u^j(\xi,\omega)\mathcal{A}^k_\mu(\xi)\,, \tag{C.131}$$

for structure constants $c^i{}_{jk}$ purely within the Lie algebra of H.

Similarly, $e^\alpha_\mu(\xi)$ transforms covariantly under the transformations, with

$$\delta e^\alpha_\mu(\xi) = -c^\alpha{}_{i\beta}u^i(\xi,\omega)\; e^\beta_\mu(\xi). \tag{C.132}$$

In this last expression, the structure constants define representation matrices, $(\mathcal{T}_i)^\alpha{}_\beta = \mathrm{i}c^\alpha{}_{i\beta}$.

More explicit formulae for $\mathcal{A}^i_\mu$ and e^α_μ can be found by first extracting the overall factor of $\partial_\mu\xi^\alpha$ – so that $\mathcal{A}^i_\mu = \mathcal{A}^i_\alpha(\xi)\,\partial_\mu\xi^\alpha$ and $e^\alpha_\mu = e^\alpha{}_\beta(\xi)\,\partial_\mu\xi^\beta$. Then the useful identity[13]

$$e^{-\mathrm{i}A}e^{\mathrm{i}(A+B)} = 1 + \mathrm{i}\int_0^1 \mathrm{d}s\; e^{-\mathrm{i}sA}B\,e^{\mathrm{i}s(A+B)} = 1 + \mathrm{i}\int_0^1 \mathrm{d}s\; e^{-\mathrm{i}sA}B\,e^{\mathrm{i}sA} + O(B^2) \tag{C.133}$$

for square matrices A and B leads to the following expressions

$$\mathcal{A}^i_\alpha(\xi) = -\int_0^1 ds\;\; \mathrm{Tr}\left[t^i e^{-\mathrm{i}s\,\xi\cdot X}X_\alpha e^{\mathrm{i}s\,\xi\cdot X}\right] \simeq \frac{1}{2}\,c^i{}_{\alpha\beta}\xi^\beta + O(\xi^2)\,, \tag{C.134}$$

and

$$e^\alpha{}_\beta(\xi) = \int_0^1 ds\;\; \mathrm{Tr}\left[X^\alpha e^{-\mathrm{i}s\,\xi\cdot X}X_\beta e^{\mathrm{i}s\,\xi\cdot X}\right] \simeq \delta^\alpha{}_\beta - \frac{1}{2}\,c^\alpha{}_{\beta\gamma}\xi^\gamma + O(\xi^2)\,, \tag{C.135}$$

where the approximate equalities expand in powers of ξ^α.

In the same way that $\mathcal{A}^i_\alpha$ is used to build G-covariant derivatives like $\mathcal{D}_\mu X$, the n-bein $e^\alpha{}_\beta$ can also be used to build G-invariant self-interactions for the ξ^α. To see how, notice that the covariant quantity, $e^\alpha_\mu = e^\alpha{}_\beta\,\partial_\mu\xi^\beta$, transforms very simply under G: $e_\mu\cdot X \to \mathrm{h}(e_\mu\cdot X)\mathrm{h}^{-1}$. Its covariant derivative is constructed from $\mathcal{A}^i_\mu t_i$:

$$(\mathcal{D}_\mu e_\nu)^\alpha = \partial_\mu e^\alpha_\nu + c^\alpha{}_{i\beta}\mathcal{A}^i_\mu\, e^\beta_\nu, \tag{C.136}$$

which transforms in the same way as does e^α_μ: $\delta(\mathcal{D}_\mu e_\nu)^\alpha = -c^\alpha{}_{i\beta}u^i(\mathcal{D}_\mu e_\nu)^\beta$.

[12] The sign mismatch between these equations is to do with representing the group using generators that are the transpose of those used for matter in (C.72) (regarding which, see also footnote [9] after eq. (C.68)).

[13] This identity is derived by setting up and solving a first-order differential equation for $U(s) := e^{-\mathrm{i}sA}e^{\mathrm{i}s(A+B)}$.

The most general G-invariant lagrangian then is $\mathfrak{L}(e_\mu, \mathcal{D}_\mu e_\nu, \dots)$, where the ellipses denote terms involving higher covariant derivatives and the lagrangian is constrained to be globally H invariant:

$$\mathfrak{L}(he_\mu h^{-1}, h\mathcal{D}_\mu e_\nu h^{-1}, \dots) \equiv \mathfrak{L}(e_\mu, \mathcal{D}_\mu e_\nu, \dots)\,. \tag{C.137}$$

Whenever $\mathfrak{L}$ satisfies (C.137) for constant h, the definitions of $e^\alpha{}_\beta$ and $\mathcal{A}^i_\alpha$ ensure it is also *automatically* invariant under global G transformations of the form of eqs. (C.121) and (C.122).

For a Poincaré invariant system, the term involving the fewest derivatives found in this way is

$$\mathfrak{L}_{GB} = -\frac{1}{2}\, f_{\alpha\beta}\, \eta^{\mu\nu}\, e^\alpha_\mu\, e^\beta_\nu = -\frac{1}{2}\, g_{\alpha\beta}(\xi)\, \partial^\mu \xi^\alpha \partial_\mu \xi^\beta\,, \tag{C.138}$$

where the second equality defines the target-space metric $g_{\alpha\beta} := f_{\gamma\delta}\, e^\gamma{}_\alpha e^\delta{}_\beta$. Here global H-invariance requires the constant positive-definite matrix $f_{\alpha\beta}$ must satisfy

$$f_{\lambda\beta} c^\lambda{}_{i\alpha} + f_{\alpha\lambda} c^\lambda{}_{i\beta} = 0\,. \tag{C.139}$$

Eq. (C.139) can be solved fairly generally. To see how recall the discussion around (C.61), where it is pointed out that the matrices X_α fill out a linear representation of the unbroken subgroup H with representation matrices given by $(\mathcal{T}_i)^\alpha{}_\beta = c^\alpha{}_{i\beta}$. In terms of these matrices (C.139) states that the commutators, $[\mathcal{T}_i, f]$ vanish in this representation, for all of the generators, $\mathcal{T}_i$, in the Lie algebra of H. If this representation of H is irreducible then, by Schur's lemma, this implies $f_{\alpha\beta}$ must be proportional to the unit matrix, with positive coefficient: $f_{\alpha\beta} = F^2 \delta_{\alpha\beta}$. Otherwise, if this representation can be reduced into n irreducible blocks, then $f_{\alpha\beta}$ need only be block-diagonal, with each diagonal element proportional to a unit matrix:

$$f_{\alpha\beta} = \begin{pmatrix} F_1^2 \delta_{\alpha_1\beta_1} & & \\ & \ddots & \\ & & F_n^2 \delta_{\alpha_n\beta_n} \end{pmatrix}, \tag{C.140}$$

for n independent positive constants, F_n^2.

A similar construction gives the action for the $\mathcal{X}$ fields (and for any other fields that happen to be present at low energies). Because the symmetry H is unbroken, these fields all transform linearly under H: $\mathcal{X} \to h\mathcal{X}$, where the constant matrices $\{h\}$ form a (possibly reducible) representation of H.

In this case the general coupling of these fields to the Goldstone bosons again starts with an arbitrary, globally H-invariant lagrangian: $\mathfrak{L}(\mathcal{X}, \partial_\mu \mathcal{X}, \dots) = \mathfrak{L}(h\mathcal{X}, h\partial_\mu \mathcal{X}, \dots)$, for constant $h \in H$. This lagrangian is automatically promoted to become G-invariant by appropriately coupling the Goldstone bosons.

The promotion to G invariance proceeds by assigning to χ the nonlinear G-transformation rule: $\mathcal{X} \to \mathtt{h}\mathcal{X}$, where $\mathtt{h} = \mathtt{h}(\xi, g) = e^{i\omega\cdot t} \in H$ is the field-dependent H matrix which is defined by the nonlinear realization, eq. (C.122). An arbitrary globally H-invariant $\mathcal{X}$-lagrangian then becomes G invariant if all derivatives are replaced by the ξ-dependent covariant derivative: $\partial_\mu \mathcal{X} \to \mathcal{D}_\mu \mathcal{X} = \partial_\mu \mathcal{X} - i\mathcal{A}^i t_i \mathcal{X}$ since this ensures $\mathcal{D}_\mu \mathcal{X} \to \mathtt{h}\mathcal{D}_\mu \mathcal{X}$ and so transforms covariantly under G transformations.

Combining all of the above constructions, a general G-invariant lagrangian has the form $\mathfrak{L}(e_\mu, \mathcal{X}, \mathcal{D}_\mu e_\nu, \mathcal{D}_\mu \mathcal{X}, \dots)$, provided only that $\mathfrak{L}$ is constrained to be invariant under global H transformations:

$$\mathfrak{L}(he_\mu h^{-1}, h\,\mathcal{X}, h\mathcal{D}_\mu e_\nu h^{-1}, h\,\mathcal{D}_\mu \mathcal{X}, \dots) \equiv \mathfrak{L}(e_\mu, \mathcal{X}, \mathcal{D}_\mu e_\nu, \mathcal{D}_\mu \mathcal{X}, \dots)\,. \tag{C.141}$$

In summary, the general statement for nonlinear realizations is this: when a global internal symmetry group G is broken to a subgroup H then the low-energy action is found by constructing the most general H-invariant local lagrangian built from the low-energy field content. This lagrangian is then 'for free' promoted to be G-invariant by coupling the Goldstone bosons in the way dictated by replacing ordinary derivative by covariant derivatives, $\partial_\mu \mathcal{X} \to \mathcal{D}_\mu \mathcal{X}$ and $\partial_\mu e^\alpha_\nu \to \mathcal{D}_\mu e^\alpha_\nu$.

Uniqueness

Although the above construction defines a G-invariant local lagrangian for the fields ξ^α and $\mathcal{X}$, is this the most general way such an action can be built? This section closes with a proof of uniqueness for the construction.

To prove uniqueness assume the existence of a general lagrangian density of the form, $\mathfrak{L}(\xi, \partial_\mu \xi, \mathcal{X}, \partial_\mu \mathcal{X})$, involving the fields ξ^α, χ^n and their derivatives. (The extension to lagrangians depending on second and higher derivatives is straightforward.) It is actually more convenient to trade the dependence of $\mathfrak{L}$ on $\partial_\mu \xi$ for a dependence on the combinations $e^\alpha_\mu = e^\alpha{}_\beta(\xi)\ \partial_\mu \xi^\beta$ and $\mathcal{A}^i_\mu = \mathcal{A}^i_\alpha(\xi)\ \partial_\mu \xi^\alpha$. There is no loss of generality in doing so, since any function of ξ and $\partial_\mu \xi$ can always be written as a function of ξ, e^α_μ and $\mathcal{A}^i_\mu$. This equivalence is most easily seen in terms of the matrix variable $U(\xi) = e^{i\xi\cdot X}$ since any function of ξ and $\partial_\mu \xi$ can equally well be written as a function of U and $\partial_\mu U$, or equivalently as a function of U and $U^{-1}\partial_\mu U$. But expression (C.129) shows that an arbitrary function of $U^{-1}\partial_\mu U$ is equivalent to a general function of e^α_μ and $\mathcal{A}^i_\mu$.

The condition that a general function, $\mathfrak{L}(\xi^\alpha, e^\alpha_\mu, \mathcal{A}^i_\mu, \chi, \partial_\mu \chi)$, be invariant with respect to G transformations then is

$$\delta\mathfrak{L} = \frac{\partial \mathfrak{L}}{\partial \xi^\alpha}\ \delta\xi^\alpha + \frac{\partial \mathfrak{L}}{\partial e^\alpha_\mu}\ \delta e^\alpha_\mu + \frac{\partial \mathfrak{L}}{\partial \mathcal{A}^i_\mu}\ \delta\mathcal{A}^i_\alpha + \frac{\partial \mathfrak{L}}{\partial \chi^n}\ \delta\chi^n + \frac{\partial \mathfrak{L}}{\partial(\partial_\mu \chi^n)}\ \delta\partial_\mu \chi^n = 0\,. \tag{C.142}$$

To see what this means, first specialize to the special case where the symmetry transformation lies in H: $g = e^{i\eta\cdot t} \in H$ by using in eq. (C.142), the transformations:

$$\begin{aligned} &\delta\xi^\alpha = -c^\alpha{}_{i\beta}\eta^i\xi^\beta, \qquad \delta e^\alpha_\mu = -c^\alpha{}_{i\beta}\eta^i e^\beta_\mu, \qquad \delta\mathcal{A}^i_\mu = -c^i{}_{jk}\eta^i\mathcal{A}^k_\mu, \\ &\quad \text{and} \qquad \delta\chi^n = i\eta^i(t_i\chi)^n, \qquad \delta\partial_\mu\chi^n = i\eta^i(t_i\partial_\mu\chi)^n. \end{aligned} \tag{C.143}$$

Requiring $\delta\mathfrak{L} = 0$ for all possible transformation parameters, η^i, then implies

$$\frac{\partial \mathfrak{L}}{\partial \xi^\alpha}\ c^\alpha{}_{i\beta}\xi^\beta + \frac{\partial \mathfrak{L}}{\partial e^\alpha_\mu}\ c^\alpha{}_{i\beta}e^\beta_\mu + \frac{\partial \mathfrak{L}}{\partial \mathcal{A}^j_\mu}\ c^j{}_{ik}\mathcal{A}^k_\mu - \frac{\partial \mathfrak{L}}{\partial \chi^n}\ i(t_i\chi)^n - \frac{\partial \mathfrak{L}}{\partial(\partial_\mu\chi^n)}\ i(t_i\partial_\mu\chi)^n = 0\,, \tag{C.144}$$

which simply states that $\mathfrak{L}$ must be an H-invariant function of its arguments for global linear H transformations.

Next consider transformations that are not in H, $g = e^{i\omega\cdot X} \in G/H$, instead evaluate eq. (C.142) using the transformations

$$\delta\xi^\alpha = \xi^\alpha{}_\beta\omega^\beta, \qquad \delta e^\alpha_\mu = -c^\alpha{}_{i\beta}u^i e^\beta_\mu, \qquad \delta\mathcal{A}^i_\mu = \partial_\mu u^i - c^i{}_{jk}u^j\mathcal{A}^k_\mu,$$
$$\text{and} \qquad \delta\chi^n = iu^i(t_i\chi)^n \qquad \delta\partial_\mu\chi^n = iu^i(t_i\partial_\mu\chi)^n\,, \tag{C.145}$$

where $\xi^\alpha = \xi^\alpha{}_\beta(\xi)\,\omega^\beta$ and $u^i = u^i_\alpha(\xi)\,\omega^\alpha$ are the nonlinear functions of ξ defined by eq. (C.122), or (C.125) and (C.126). Using these in eq. (C.142), and simplifying the resulting expression using eq. (C.144), leads to the remaining condition for G invariance:

$$\frac{\partial\Omega}{\partial\xi^\alpha}\left(\xi^\alpha{}_\beta + c^\alpha{}_{i\gamma}u^i_\beta\xi^\gamma\right) + \frac{\partial\Omega}{\partial\mathcal{A}^j_\mu}\,\partial_\mu u^i_\beta + \frac{\partial\Omega}{\partial(\partial_\mu\chi^n)}\,i\partial_\mu u^i_\beta(t_i\chi)^n = 0. \tag{C.146}$$

To see what this means, first specialize to $\xi^\alpha = 0$, in which case $\partial_\mu u^i_\beta = \partial_\alpha u^i_\beta\,\partial_\mu\xi^\alpha$ vanishes. Then since eq. (C.126) implies $\xi^\alpha{}_\beta(\xi = 0) = \delta^\alpha_\beta$, it follows that

$$\left.\frac{\partial\Omega}{\partial\xi^\alpha}\right|_{\xi=0} = 0\,. \tag{C.147}$$

But since the group transformation law for ξ^α is inhomogeneous, it is always possible to perform a symmetry transformation to ensure that $\xi^\alpha = 0$ for *any* point within G/H, and so eq. (C.147) also implies the more general result

$$\frac{\partial\Omega}{\partial\xi^\alpha} \equiv 0 \quad \text{for all } \xi^\alpha \in G/H\,. \tag{C.148}$$

The rest of the information contained in eq. (C.146) is extracted by simplifying using $\partial\Omega/\partial\xi^\alpha = 0$. This leads to

$$\left(\frac{\partial\Omega}{\partial\mathcal{A}^j_\mu} + \frac{\partial\Omega}{\partial(\partial_\mu\chi^n)}\,i(t_i\chi)^n\right)\partial_\mu u^i_\beta = 0\,, \tag{C.149}$$

which states that the two variables, $\mathcal{A}^i_\mu$ and $\partial_\mu\chi^n$, can only appear in Ω through the one combination: $(\mathcal{D}_\mu\chi)^n \equiv \partial_\mu\chi^n - i\mathcal{A}^i_\mu(t_i\chi)^n$. That is, $\mathcal{X}$ can appear differentiated in Ω only through the covariant derivative, $\mathcal{D}_\mu\mathcal{X}$.

We see from these arguments that the G-invariance of Ω is equivalent to the statement that Ω must be an H-invariant function constructed from the covariantly-transforming variables e^α_μ, $\mathcal{X}$ and $\mathcal{D}_\mu\mathcal{X}$. If higher derivatives of ξ had been considered, then the vanishing of the terms in $\delta\Omega$ that are proportional to more than one derivative of u^i would similarly imply that derivatives of e^α_μ must also only appear through its covariant derivative, $(\mathcal{D}_\mu e_\nu)^\alpha$, defined by eq. (C.136).

Since these are the constructions for invariant lagrangians used in earlier sections, this earlier construction must be unique.

C.7 LSZ reduction and bound-state energies

In §12.2.4 of the main text conclusions are drawn about the size of particular contributions to bound-state energies for positronium. These conclusions are drawn using Feynman

rules for a correlation function $\langle T\,[\Psi^*_{i_1}(x_1)\,\Phi^*_{i_2}(x_2)\,\Psi_{i_3}(x_3)\,\Phi_{i_4}(x_4)]\rangle$, and this section of the appendix aims to fill in some of the missing steps that relate the correlation function to bound-state energies.

The main observation is this: the Fourier transform of a time-ordered vacuum correlation function of a field operator has poles at the positions of the energies of states that can be created from the vacuum by the operator in question. To see how this works, consider the vacuum time-ordered correlation function for any local field operator $\mathcal{O}(x)$:

$$\mathrm{i}G(\omega,\mathbf{q}) := \frac{1}{(2\pi)^3}\int \mathrm{d}^4x\,\langle\Omega|T[\mathcal{O}(0)\,\mathcal{O}^*(x)]|\Omega\rangle\,e^{-\mathrm{i}\omega x^0+\mathrm{i}\mathbf{q}\cdot\mathbf{x}}\,. \tag{C.150}$$

Imagine evaluating this by inserting a complete set of momentum eigenstates:

$$\begin{aligned}\mathrm{i}G(\omega,\mathbf{q}) &= \int \mathrm{d}^4x\,e^{\mathrm{i}q\cdot x}\sum_N\int\frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\Big[\Theta(x^0)\,\langle\Omega|\mathcal{O}(0)|N(\mathbf{k})\rangle\,\langle N(\mathbf{k})|\mathcal{O}^*(x)|\Omega\rangle\\ &\qquad\qquad +\Theta(-x^0)\,\langle\Omega|\mathcal{O}^*(x)|N(\mathbf{k})\rangle\,\langle N(\mathbf{k})|\mathcal{O}(0)|\Omega\rangle\Big]\\ &= \int \mathrm{d}^4x\,e^{\mathrm{i}q\cdot x}\sum_N\int\frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,\Theta(x^0)\,\langle\Omega|\mathcal{O}(0)|N(\mathbf{k})\rangle\,\langle N(\mathbf{k})|\mathcal{O}^*(x)|\Omega\rangle\,,\end{aligned} \tag{C.151}$$

where N contains all other labels besides momentum and $\Theta(u) = \{0 \text{ if } u<0 \text{ and } 1 \text{ if } u>0\}$ is the usual Heaviside step function. The last equality assumes $\mathcal{O}(x)$ carries a conserved charge and the quantum numbers are such that it is $\langle N(\mathbf{k})|\mathcal{O}^*(x)|\Omega\rangle$ that is nonzero (and so $\langle\Omega|\mathcal{O}^*(x)|N(\mathbf{k})\rangle$ vanishes). Spacetime translation invariance implies $\langle N(\mathbf{k})|\mathcal{O}^*(x)|\Omega\rangle = \langle N(\mathbf{k})|\mathcal{O}^*(0)|\Omega\rangle\,e^{-\mathrm{i}k_N\cdot x}$ and so

$$\mathrm{i}G(\omega,\mathbf{q}) = \sum_N\int\frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,\big|\langle\Omega|\mathcal{O}(0)|N(\mathbf{k})\rangle\big|^2\int\mathrm{d}^4x\,\Theta(x^0)\,e^{\mathrm{i}(q-k_N)\cdot x} = \mathrm{i}\sum_N\frac{\big|\langle\Omega|\mathcal{O}(0)|N(\mathbf{q})\rangle\big|^2}{E_N(\mathbf{q})-\omega+\mathrm{i}\epsilon}\,, \tag{C.152}$$

where $q^\mu = (\omega,\mathbf{q})$ and $k^\mu_N = [E_N(\mathbf{k}),\mathbf{k}]$, and the Fourier representation of $\Theta(u)$ is used:

$$\Theta(u) = \int\frac{\mathrm{d}w}{2\pi}\left(\frac{\mathrm{i}}{w+\mathrm{i}\epsilon}\right)e^{-\mathrm{i}wu}\,. \tag{C.153}$$

For the present purposes what is important about (C.152) is the pole it reveals at $\omega = E_N(\mathbf{q}) + \mathrm{i}\epsilon$. This argument as applied to many-field correlation functions is related to the 'Lehmann-Symanzik-Zimmermann' (LSZ) reduction formula [324], which further argues that the residue at these poles gives S-matrix elements for transitions amongst states corresponding to the fields involved in the correlation function.

Notice that nothing in the above derivation assumes $\mathcal{O}(x)$ is a particular 'elementary' field for particle type N; *any* operator for which $\langle\Omega|\mathcal{O}(0)|N(\mathbf{k})\rangle$ is nonzero – usually called an 'interpolating field' for N – will do. In particular, for applications to two-body bound states it is usually convenient to focus on interpolating fields that are bilinears of the 'fundamental' fields:

$$\mathcal{O}(x_1,x_2) = \Psi(x_1)\,\Phi(x_2)\,, \tag{C.154}$$

and choose equal times, $x^0_1 = x^0_2 =: x^0$, so that $\mathcal{O}(x_1,x_2) = \mathcal{O}(x^0,\mathbf{x}\ \mathbf{X})$ where $\mathbf{X}$ is the

centre-of-mass coordinate for $\mathbf{x}_1$ and $\mathbf{x}_2$ while $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ is the relative separation. In this case the starting point for the above argument would be the correlation function

$$\mathrm{i}G(\omega, \mathbf{q}) := \frac{1}{(2\pi)^6} \int \mathrm{d}x^0 \mathrm{d}^3\mathbf{x}\, \mathrm{d}^3\mathbf{X}\, \langle\Omega|T[O(0,0;0)\, O^*(x^0, \mathbf{x}; \mathbf{X})]|\Omega\rangle\, e^{-\mathrm{i}\omega x^0 + \mathrm{i}\mathbf{q}\cdot\mathbf{x}}\,. \tag{C.155}$$

rather than (C.150), where the integration over $\mathbf{X}$ has the effect of projecting onto zero centre-of-mass momentum.

In practice the above correlation function is computed by perturbing about an approximate solution for the bound state, in which case it is useful to write the near-pole behaviour as

$$G(\omega, \mathbf{q}) \simeq \frac{|R(\omega, \mathbf{q})|^2}{\omega - E_N} \tag{C.156}$$

where $R(\omega, \mathbf{q}) = \langle\Omega|O(0)|N(\mathbf{q})\rangle$ has no pole at $\omega = \varepsilon_N$, and expand $E_N = \varepsilon_N + \delta E_N$ and $R(\omega, \mathbf{q}) = \mathcal{R}(\mathbf{q}) + \delta R(\omega, \mathbf{q})$. Then

$$\begin{aligned} G(\omega, \mathbf{q}) &\simeq \frac{|\mathcal{R}(\mathbf{q}) + \delta R(\omega, \mathbf{q})|^2}{\omega - \varepsilon_N - \delta E_N} \\ &\simeq \frac{|\mathcal{R}(\mathbf{q})|^2}{\omega - \varepsilon_N} + \frac{\mathcal{R}^*(\mathbf{q})\, \delta R(\omega, \mathbf{q}) + \mathcal{R}(\mathbf{q})\, \delta R^*(\omega, \mathbf{q})}{\omega - \varepsilon_N} + \frac{\mathcal{R}^*(\mathbf{q})}{\omega - \varepsilon_N}\, \delta E_N\, \frac{\mathcal{R}(\mathbf{q})}{\omega - \varepsilon_N} + \cdots\,, \end{aligned} \tag{C.157}$$

which shows that the leading corrections to δE_N can be read off by amputating the two external bound-state propagators — that is to say, by multiplying by a factor of $(\omega - \varepsilon_N)/\mathcal{R}$ and its complex conjugate — and evaluating the result at $\omega = \varepsilon_N$. In the non-relativistic applications of Chapter 12 this amounts to evaluating the amputated graph and taking the expectation value of the result using the zeroeth-order (Schrödinger-Coulomb) wavefunction.

D Appendix D Further reading

This book touches only briefly on each of the applications of effective field theories (EFTs) throughout physics in order to emphasize the great generality of EFT techniques. But this also means that many readers are likely to be dissatisfied with the level of detail used to describe each application. This section aims to help with this by providing some further reading for those interested in quenching a more fundamental thirst for knowledge in each of the areas touched.

The bibliography given here is not meant to be an exhaustive survey of the literature, about parts of which I am sure I am relatively poorly informed (and I apologize in advance for any gems I may have missed). Instead I list references that I have found useful myself, and include review articles to which the reader should go for more detailed referencing in each area.

Many of these papers (at least those published since the development of the World-Wide Web in the early 1990s) are available for free online. In particular, references like

[arXiv:hep-ph/9708416] or [arXiv : 1704.02751]

are shorthands (respectively) for the links:

https://arxiv.org/abs/hep-ph/9708416

or https://arxiv.org/abs/1704.02751 .

D.1 Quantum field theory

The main prerequisite for reading this book is an understanding of quantum field theory (QFT). At face value QFT is only a convenient formalism for handing many-particle quantum mechanics, including in particular processes like emission and absorption, that change the number of particles. But it is also the ubiquitous language of physics, since it makes it simple to bake in basic properties like unitarity and cluster decomposition from the get-go when trying to guess a system's dynamics. QFT is particularly useful for relativistic applications because it is a basic fact of relativistic quantum mechanics that *all* interactions involve components that change the number of particles (due to the inevitable presence of antiparticles).

For this reason the appendices are largely devoted to providing a very brief summary of the basic facts of quantum field theory. Inevitably an interested reader will want more, and here are a few suggestions for further reading (organized roughly by topics).

There are very many good textbooks on quantum field theory, not all of which can be named here. Some useful textbooks on general-purpose relativistic quantum field theory that I have used are:

- "An Introduction to Quantum Field Theory," by M. Peskin and D. Schroeder, Westview, 1995. An excellent and readable book on field theory (though with an unfortunate choice of metric conventions). The first 4 chapters of the book give all required background preparation for this book, and the remaining chapters give the tools needed to take the material presented in this book to the next level (renormalization, higher order effects, and a more solid theoretical foundation).
- "The Quantum Theory of Fields, I–III", by S. Weinberg, Cambridge Press, 2000. An original and encyclopedic presentation of quantum field theory from one of the masters who helped systematize much of it. In particular Volume I addresses many of the foundational arguments that underpin quantum field theory, while volumes II and III are more dedicated to applications. Most of the field theoretical arguments alluded to in this book are laid out in detail here. You will learn something new every time you read it, probably for the rest of your life. But it is likely not as good for novices as is an introductory text like Peskin and Schroeder's book.
- "Quantum Field Theory in a Nutshell," by A. Zee, Princeton Press 2010. This book fits into a special niche in that its emphasis is more on concepts and less on calculational tools. Not a bad place for a learner to start, but probably also not enough in itself for someone seeking a practical hands-on calculational ability.
- "Quantum Field Theory," by L. Ryder, Cambridge Press, 1996 (2'nd ed). An older and somewhat more introductory text on field theory, providing more than enough background material to understand this book.
- "Quantum Field Theory," by L. Brown, Cambridge Press, 1994. A clear introduction to quantum field theory with an interestingly novel choice of topics, that gives an extremely solid underpinning (though does not cover nonabelian gauge theories).
- "Quantum Field Theory", by G. Sterman, Cambridge Press, 1993. A clear and systematic exposition of modern field theoretic techniques which includes a number of topics (like infrared divergences and factorization) not covered in other texts.
- For detailed (but advanced and somewhat more mathematical) discussions of C,P, and T symmetries, the spin-statistics theorem, and related topics, try "PCT, Spin and Statistics, and All That," by A. Wightman and R. Streater, Princeton University Press, 2000.

Other books that are more aimed at particle physics are also listed below in the section devoted to the Standard Model. Many of these books (and those mentioned later), particularly Weinberg's, advocate an effective field theory point of view though this is usually not their main focus.

D.2 EFT framework

In this book Part I develops the main EFT formalism used throughout the rest of the text. The main logic explored throughout the entire EFT program is largely laid out in the paper entitled:

- "Phenomenological Lagrangians" [*Physica A* **96** (1979) 327] by Steven Weinberg [2]. This paper has held up remarkably well to the passage of time and remains worth reading as a statement of purpose for those taking up the subject anew.

A book with similar goals to the one you are reading, whose scope directly aims at EFT methods (but with a complementary choice of topics) is:

- "Effective Field Theories" by Alexei Petrov and Andrew Blechman, World Scientific, 2015. This book aims more directly at high-energy physics and non-relativistic applications like NRQED and effective theories of gravity than the one you are now reading.

Quantum actions

The framework of generating functionals goes back into the mists of time in the mid-twentieth century when quantum field theory was relatively young. The specific use of the 1PI quantum action (often in the old days also called an 'effective action' though this term is now normally reserved for the Wilson action) came in the mid-1960s, where it was introduced within perturbation theory as the formal sum over 1PI graphs [5]. The non-perturbative definition used here came a bit later in [15].

In retrospect, much of the formalism of field theory used today was systematized in the 1960s. A comprehensive one-stop-shopping source for much of these developments is:

- "The Quantum Theory of Fields, vol I", by Steven Weinberg, Cambridge Press, 2000. This book (already mentioned above) authoritatively lays out the foundations of quantum field theory, straight from the proverbial horse's mouth. Unlike most books on quantum field theory, this book ('Vol-I', for short) does not start off assuming quantum field theory is the right subject to study. The goal instead is to study what it means for quantum mechanics to be consistent with special relativity (Poincaré-invariance), and Vol-I systematically makes the case that this *is* quantum field theory.
- "Aspects of Symmetry", by Sidney Coleman, Cambridge Press 2010. This is a collection of lectures given by Sidney Coleman over the years at the summer school in Erice in Sicily. All of these are well-known as masterful expositions of different topics in field theory, and include a very clear explanation of the generators $W[J]$ and $\Gamma[\varphi]$ of connected and 1PI correlation functions.
- "What is quantum field theory, and what did we think it is?", also by Steven Weinberg [a contribution to the proceedings of the conference *Conceptual foundations of quantum field theory*, Boston 1996, p 241-251 [408]), also available online at hep-th/9702027]. This is less of a 'shut up and calculate' description of quantum field theory, and more

of a retrospective view of what quantum field theory is and how this has changed over the years. In particular it provides a chatty and easy to read summary of the modern picture wherein quantum field theory is what emerges when you combine special relativity, quantum mechanics and 'cluster decomposition' (the principle that probabilities for independent events widely separated in space must factorize).

This last article explicitly enunciates the basic modern point of view: quantum field theory in itself has very little content, in that the most general field theory consistent with the analyticity properties of scattering amplitudes is the same as the most general physics that is consistent with these properties (subject to a few motherhood principles like conservation of probability (*i.e.* unitarity) and cluster decomposition). Although largely taken for granted now, this was controversial in the 1960s when it was felt that quantum field theory could not describe the strong interactions. This led to a program that based itself only on the analytic properties of the S-matrix, a summary of which can be found in the review

- "Regge Poles and S-matrix Theory," by Steven Frautschi, New York: W. A. Benjamin, Inc., 1963,

and which has echoes in more recent lines of research [476].

The formalism of coarse-graining short distances and the related renormalization group also has a long history. It starts off with the study of renormalization and scaling in particle physics (and QED in particular) in the early 1950s [477, 478]. A big improvement in generality came with more explicit formulations of how to split low- and high-energy degrees of freedom (coarse-graining) in the 1960s and early 70s, starting within condensed-matter physics [479, 480, 481] and moving from there back to particle physics [139, 140]. Extensive reviews of these developments can be found in

- "The Renormalization group and the epsilon expansion" by Ken Wilson [*Physics Reports* **12** (1974) 75] provides an excellent contemporary survey of these techniques by an inventor.
- "Field Theory, the Renormalization Group and Critical Phenomena" by Daniel Amit, World Scientific 1984. This book aims more at condensed matter applications of renormalization methods, and later editions (with Victor Martin-Mayor) also include discussions of strong-coupling, lattice models and numerical methods.
- "Why The Renormalization Group Is A Good Thing", published in *Asymptotic Realms Of Physics,* 1-19 Cambridge 1981 [40]. This is a contribution to the proceedings of the festschrift for Francis Low by Steven Weinberg, clearly summarizing some of the history and ideas. (A bonus is the statement found in this article of the Three Laws of Theoretical Physics.)

Later progress in formulating and using coarse-grained techniques starts with Polchinski's formulation of the exact renormalization group [25], and continuing with later refinements [26, 27, 482]. In the meantime parallel developments separately begin to apply EFT ideas to more and more areas of physics (aspects of which are largely the subject of this book and so are described below).

Much of the discussion in this chapter is based on the presentation given in some review lectures on effective theories

- "Introduction to Effective Field Theory," by C.P. Burgess [*Annual Reviews of Nuclear & Particle Science* **57** (2007) 329, arXiv:hep-th/0701053]. These are my own lectures and so not surprisingly they overlap in their layout with what is found in this book, including the development of the 1LPI action and the use of the toy model as a useful vehicle for illustrating more general features.
- "Five lectures on effective field theory" by David B Kaplan [arXiv:nucl-th/0510023], which is a very clear survey that sets up the framework quite broadly and then narrows in to applications more focussed on nuclear and non-relativistic physics. In particular the discussion of scaling given here is largely as presented in these lectures.

Power counting

Chapter 3 deals with power-counting with effective lagrangians, using dimensional analysis to estimate the dependence of generic Feynman graphs in terms of the scales appearing in the couplings of the effective theory. The arguments made parallel the power-counting arguments used when deciding the superficial degree of divergence of Feynman graphs, such as when proving the renormalizability of a field theory (like QED) [127, 483]. The dimensional analysis likely comes across as cavalier inasmuch as the relevant graphs really give multidimensional integrals and one might worry whether their behaviour is well-captured by naive one-dimensional estimates. As usual, Weinberg's textbook (Vol-I *ibid*) is an invaluable – though fairly compact – resource for these arguments.

The justification for these arguments ultimately rely on Weinberg's theorem [484], which underpins the proofs of renormalizability, and clarify why naive arguments properly capture the multidimensional complications. An authoritative summary of the issues, with historical commentary, can be found in Vol I of Weinberg's 'Quantum Field Theory' trilogy, cited above. A more recent (though also not that recent) and exhaustive treatment can also be found in

- "Renormalization" by John Collins, Cambridge Press, 1984. This book provides a very detailed treatment of renormalization in the post-dimensional regularization age.

The spirit of power-counting from an EFT framework is already in the "Phenomenological Lagrangians" paper cited above [2], though I follow in this book the notation and presentation outlined in my own review [24].

The 'method of regions' is a very useful technique for identifying how different scales can enter a calculation when using dimensional regularization. This is described in some detail in the book

- "Introduction to Soft-Collinear Effective Theory" by Thomas Becher, Alessandro Broggio and Andrea Ferroglia, (Springer, 2015). More generally this book is a useful handbook for techniques that arise when using dimensional regularization within an EFT analysis.

Symmetries

For some reason physicists tend to pick up much of their group theory on the streets. Two very useful introductions to the theory of Lie groups for physicists, and a very useful reference with extensive tables, are

- "Lie Algebras in Particle Physics," by H. Georgi, Perseus, 1999 (2nd ed). This is a very thorough treatment of group theory for particle physics, essential for those who find Appendix C.4.1 either too telegraphic or incomplete.
- "Semi-Simple Lie Algebras and Their Representations", by R. N. Cahn, Benjamin-Cummings, 1984. This book picks up where the previous suggestion leaves off, presenting more of the properties of groups and their representations.
- "Group Theory For Unified Model Building", R. Slansky, *Physics Reports* **79** (1981) 1-128. This provides a very useful summary of the properties of the representations of Lie groups, including detailed tables showing how representations decompose in terms of representations of subgroups.

The development of the theory of nonlinear realizations starts with the nonlinear sigma model for pion physics [11], which was then generalized to general groups in the standard form used today in [12, 13].

An in-depth discussion of nonlinear realizations and their historical development, as well as a systematic derivation of anomalies both from the point of view of triangle diagrams and of path-integral measure, including also the Wess-Zumino consistency relations [485, 486] and their solution using descent equations [487], is given in

- "The Quantum Theory of Fields, vol II", by Steven Weinberg, Cambridge Press, 2000. This, the second volume of the Quantum Theory of Fields trilogy ('Vol-II' for short) picks up where Vol-I leaves off, touching on most of the higher topics of quantum field theory. Besides detailed derivations this book has many historical commentaries from one of the central participants of the time.

A summary of consistency conditions and the descent equations that emphasizes more geometrical methods is given in the review by Bruno Zumino in "Relativity, Groups and Topology II", edited by B.S. de Witt and R. Stora (Elsevier, Amsterdam, 1984).

Time-dependent backgrounds

The power of making low-energy arguments and the relative simplicity of the low-energy limit in quantum mechanics has been known for a long time, going back to the Born-Oppenheimer approximation [313]. Effective theories were first systematically developed for field theories with applications to particle scattering in mind [48] (though a parallel line of development was also underway in condensed matter physics [404]). Because time-dependent backgrounds often do not arise in these applications work to develop a formalism for describing classical time evolution within a Wilsonian effective theory was historically not a priority.

The beginnings of precision cosmology with the measurement of primordial fluctuations

[488] provided a big incentive for having controlled low-energy approximations in time-dependent environments, and most developments trace back to this. (See the discussion below about Chapter 10 for more references to the cosmology literature.) More recently the prospect of measuring gravitational waves also stimulated a Wilsonian reformulation [360] for calculations of time-dependent classical motion, such as those describing the radiation of inspiraling, nonrelativistic, gravitating objects like black holes or neutron stars.

The discussion in Chapter 6 follows the logic of my own reviews [24] and [489]. This differs somewhat from much of the cosmology literature for which EFTs often zero in more specifically to the study of fluctuations about a cosmological background along the lines developed in [104]. In particular the toy-model discussion follows [89], which was itself stimulated by related work on cosmological fluctuations [490].

The discussion of well-posedness of the initial-value problem is only now starting to sink in to the EFT community, largely driven by the desire to describe and test modifications to general relativity in the strong-field regime revealed by gravitational wave observations. References [98] and [99] quoted in the main text provide good summaries of these issues both for gravitational and fluid physics.

D.3 Relativistic applications

Part II begins a discussion of relativistic applications. The Fermi theory of the weak interactions is the poster child for how effective theories arise in nature, and so is discussed in a variety of EFT reviews such as:

- "Weak Interactions and Modern Particle Theory", by H. Georgi, Benjamin Cummings, 1984. A very physical discussion of much of the standard model and some of the techniques used to compute with it, with EFT methods squarely in mind. The treatment of the weak interactions includes various loop corrections to the Fermi lagrangian and survives well despite its age.

The notation and description of the weak interactions and QED used in this section is partly taken from my own book

- "The Standard Model: A Modern Primer" by Guy Moore and me, Cambridge Press, 2007, post-Higgs-discovery revision 2013 uses the Standard Model as a vehicle for learning quantum field theory, mostly at the level of tree graphs, but includes sections on QED, infrared effects, hadrons and chiral perturbation theory. Modern EFT methods are also included, and used to organize the treatment of Beyond the Standard Model (BSM) physics.

There are many classic texts on Quantum Electrodynamics, though usually of a vintage that predates the widespread adoption of EFT reasoning. More modern discussions can be found in some of the books listed above on quantum field theory.

A discussion of the $U_A(1)$ problem and the role of anomalies in resolving it is given by

[491], as well as in the review "Uses of Instantons" by Coleman in Aspects of Symmetry [17].

Chiral Perturbation Theory

Chiral perturbation theory was the place where EFT methods, and the low-energy treatment of Goldstone bosons, were first systematized in the particle-physics literature. A discussion appears in Weinberg (Quantum Theory of Fields Vol II *ibid*) with historical notes. A very comprehensive and instructive book on the subject is

- "Dynamics of the Standard Model," by J. Donoghue, E. Golowich, and B. Holstein, Cambridge Press, 1992. This contains an advanced discussion of the standard model, with particular emphasis on bound states in QCD, chiral symmetry, and radiative corrections.

Many good review articles also exist on chiral perturbation theory, a selection of which is listed in reference [492].

Standard Model

There are many books on the Standard Model, which is a well-developed subject. In addition to the ones listed above two other noteworthy examples are

- "Quarks and Leptons: An Introductory Course in Particle Physics," by F. Halzen and A. Martin, Wiley, 1984. This book is an elementary introduction aimed at developing the computational tools and getting people calculating, with a minimum of formal baggage. An excellent introduction to utilitarian field theory.
- "Quantum Field Theory and the Standard Model," by Matthew Schwartz, Cambridge Press 2014, is a more recent and modern treatment of quantum field theory as applied to the Standard Model, that also draws heavily on EFT methods to organize calculations.

Supersymmetry is a well-developed topic in its own right and there are a number of books that review its various aspects. Among those aimed at possible implications for particle phenomenology are

- "The Quantum Theory of Fields, vol III", by S. Weinberg, Cambridge Press, 2000. This presentation has the nice feature that it builds directly from the tools built in vols. I and II of this sequence. This is particularly nice for supersymmetry, since many of the other presentations of supersymmetry use completely different notation for spinors used in supersymmetry compared to spinors used elsewhere in physics. Weinberg's vols I–III are uniform in their treatment of spinors in all aspects of their use.
- "Supersymmetry in Particle Physics," by Ian Aitchison, Cambridge Press 2007, is a more recent and modern treatment of supersymmetry as aimed at particle physics applications.

- "Theory and Phenomenology of Sparticles: An Account of Four-Dimensional N=1 Supersymmetry in High Energy Physics," by Manuel Drees, Rohini Godbole and Probir Roy, World Scientific Press, 2005 provides a phenomenologist's eye view of N=1 supersymmetry in 4 dimensions.

The above are complemented by discussions of supersymmetry that emphasize the more formal strong-coupling, gravity and string theory connections.

- "Modern Supersymmetry, Dynamics and Duality," by John Terning, Oxford Press 2006, is a treatment of supersymmetry as aimed at applications to dualities and many of the modern issues associated with supersymmetric systems.
- "Supersymmetry and String Theory, Beyond the Standard Model," by Michael Dine, Cambridge Press 2007, is a treatment of supersymmetry as aimed at more fundamental applications, such as to string theory.
- "Supergravity," by A. van Proeyen and D.Z. Freedman, Cambridge Press 2012, is a modern treatment of supergravity by those that invented much of it, covering many topics not usually treated.
- "Introduction to Supersymmetry and Supergravity," by Peter West, World Scientific 1990. This is a treatment aimed at the framework for supergravity, a topic often skirted over quickly in textbooks (but not here).
- "Supersymmetry and Supergravity," by Julius Wess and Jon Bagger, Princeton Press 1992. This is one of the standard textbooks by some of the authors who helped define the subject.

General Relativity and Cosmology

More and more, physicists in all fields are expected to be knowledgeable about gravitational physics, and the geometrical techniques used in its study. Some useful texts for these purposes are:

- "Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity", S. Weinberg, Wiley, 1972. An oldie but a goodie: a very physical introduction to general relativity and its applications in astrophysics, the solar system and cosmology.
- "Gravitation", C.W. Misner, K.S. Thorne and J.A. Wheeler, Freeman, 1973. The classic book with the quirky style, which sets the standard for its comprehensive and modern treatment of geometrical techniques.
- "General Relativity", R.M. Wald, University of Chicago Press, 1984. A modern update of the two previous classics, containing more of the modern mathematical techniques.
- "Spacetime and Geometry: An Introduction to General Relativity", S. Carroll, Cambridge Press, 2019. A re-release of a modern and very readable book that is a good place to start.

The last few decades have seen cosmology turn from a very speculative to a data-rich subject. Much of the evidence that the Standard Model is incomplete comes from the unified picture of cosmology that this data has spurred, making a good knowledge of this area

also mandatory for many areas of physics. Some of the books I have learned from myself are

- "Principles of Physical Cosmology", P.J.E. Peebles, Princeton Press 1993. This is a classic written by one of the inventors of the modern picture of physical cosmology.
- "Physical Foundations of Cosmology", V. Mukhanov, Cambridge Press 2005. This is a comprehensive and very clear description of the theory of fluctuations and structure formation, by one of its inventors.
- "Modern Cosmology," S. Dodelson, Elsevier 2003. This is a modern textbook on cosmology including the discussion of fluctuations.
- "Cosmology", S. Weinberg, Oxford 2008. This is a modern treatment that definitively updates Weinberg's earlier book on Gravitation and Cosmology.
- "Introduction to Cosmology," B. Ryden, Pearson 2002. This introductory book is aimed at undergraduates, and so does not presuppose as much background. Yet it is also thorough and detailed so a good place to start for beginners.

The presentation of the effective theory of gravity and cosmology used in this book follows some of my own review articles on EFTs in cosmology, mentioned above [489]. String theory provides a concrete example of what might take place at the highest energies, and this is partly what makes it interesting. Some references are

- "Superstring Theory, vols I and II", by M. Green, J. Schwarz and E. Witten, Cambridge Press, 1987. Strings and superstrings described by the masters. This book has provided the first exposure to the field for many novice learners.
- "String Theory, vols I and II", by Joe Polchinski, Cambridge Press, 1998. Strings and branes (including the origins of the word 'brane' for membranes [87]) most often arise in supergravity and string theory, and this book is the classic textbook by the discoverer of D-branes [88].
- "String Theory and M-Theory", by K. Becker, M. Becker and J. Schwarz, Cambridge Press, 2007. A modern update on the subject.
- "String Theory and Particle Physics," by L.E. Ibáñez and A.M. Uranga, Cambridge Press 2012. This is a recent and high-level introduction to the phenomenological aspects of string theory and branes.
- "A First Course in String Theory," by B. Zweibach, Cambridge Press 2009. This is an undergraduate level introduction to string theory, aimed at those who do not already have an exposure to quantum field theory.

D.4 Nonrelativistic applications

The last two parts of the book are aimed at non-relativistic applications of EFT methods, both for collections of just a few slowly moving particles and for full-blown many body systems. Given here are a selection of references that I have found useful when learning

these areas, though the further the area is from my own area of expertise the more likely I am to have missed other gems (and my apologies if this is so).

HQET, NRQED and all that

There are a variety of review articles for heavy-particle EFTs, some of which are included among the reviews of EFT methods quoted earlier. Some others are

- "Review of selected topics in HQET," A.I. Vainshtein (hep-ph/9512419).
- "Heavy-quark effective theory," M. Neubert, in 20th Johns Hopkins Workshop on Current Problems in Particle Theory (hep-ph/9610385).
- "Heavy-quark and Soft-collinear Effective Field Theory," by C.W. Bauer and M. Neubert, in PDG review K.A. Olive et al.,*Chin. Phys.* **C38** (2014) 090001 (http://pdg.lbl.gov).
- "An Introduction to the heavy quark effective theory," F. Hussain, and G. Thompson, (hep-ph/9502241).
- "Precision study of positronium: Testing bound state QED theory," S.G. Karshenboim, *International Journal of Modern Physics* **A19** (2004) 3879 (hep-ph/0310099).
- "An Introduction to NRQED," G. Paz, *Mod. Phys. Lett.* **A30** (2015) 1550128 (arXiv:1503.07216 (hep-ph)).

First-quantized methods

There are fewer surveys of of first-quantized methods within an effective field theory context. Discussions of collective coordinates can be found in books on solitons, such as in Coleman's "Aspects of Symmetry" mentioned earlier, or (for example)

- "Magnetic Monopoles," by Ja. Schnir, Springer-Verlag 2005.

For reviews on the quantum mechanics of the inverse-square potential and the phenomenon of 'fall to the centre', see for example

- B. Holstein, "Anomalies for Pedestrians," *Am. J. Phys.* **61** (1993) 142;
- A.M. Essin and D.J. Griffiths, "Quantum Mechanics of the $1/x^2$ Potential," *Am. J. Phys.* **74** (2006) 109.
- M. W. Frank, D. J. Land and R. M. Spector, "Singular Potentials", *Rev. Mod. Phys.* **43** (1971) 36.

Atomic physics

A number of books on atomic methods have proven useful over the years, including

- "Quantum Mechanics of One- and Two-Electron Atoms," by H.A. Bethe and E.E. Salpeter, Springer-Verlag 1957; Plenum Publishing 1977;
- "Rydberg Atoms," by T.F. Gallagher, Cambridge Press, 1994;
- "Many-body atomic physics," ed. by J.J. Boyle and M.S. Pindzola, Cambridge Press, 1998.

- "Atom-Photon Interactions", C. Cohen-Tannoudji, J. Dupont-Roc and G. Grynberg, Wiley Press 2004.
- "Bose-Einstein Condensation in Dilute Gases," C.J. Pethick and H. Smith, (Cambridge Press, 2001).

Goldstone bosons in nonrelativistic systems

The different types of Goldstone counting for spacetime symmetries in non-relativistic systems is another one of those things that has long been in the air (since at least the early 1980s, when I was a graduate student), and because these issues arise more commonly for condensed matter systems were appreciated there much earlier. See for example

- "Concepts in Solids: Lectures on the Theory of Solids" by Philip W. Anderson, World Scientific, Singapore 1997

For particle physicists perhaps the most familiar examples where these issues arise are spin waves in ferromagnets and antiferromagnets [107] (see also Chapter 14.1), whose unusual properties eventually became systematized in [105] and [108].

Condensed Matter Surveys

Condensed matter physics is a vast area of research, for which there are a number of good textbook treatments, including:

- "Principles of Condensed Matter Physics", by Paul Chaikin and Tom Lubensky, Cambridge Press, 1995. This both gives a thorough treatment of condensed matter physics, with an emphasis on its 'soft' side, and it is also very accessible to those of the unwashed who are not professional condensed matter physicists (myself included). This book includes many instances of topological defects and domain walls of various types.
- "Solid State Physics," N.W. Ashcroft, N.D. Mermin, Harcourt, 1976. This is a classic undergraduate textbook on condensed matter physics.
- "Introduction to Solid State Physics, 8th Edition," C. Kittel, Wiley, 2004. This is the other classic undergraduate textbook.
- "Introduction to Superconductivity," M. Tinkham, McGraw-Hill Press, 1975. This is a book aimed more explicitly at the phenomenon of superconductivity.
- "Quantum Field Theory and Condensed Matter," R. Shankar, Cambridge Press, 2017. This is a more modern treatment of many of the ideas handled in the older texts.
- "Quantum Field Theory of Many Body Systems," X.G. Wen, Oxford Press 2004. Another modern treatment of quantum field theory for a modern condensed matter audience.

Degenerate systems

The treatment of degenerate systems described here follows the wonderful reviews,

- "Effective field theory and the Fermi surface." J. Polchinski, In the proceedings of the TASI school *Recent directions in particle theory* (hep-th/9210046) [405].
- "Renormalization group approach to interacting fermions", R. Shankar, *Reviews of Modern Physics* **66** (1994) 129 (cond-mat/9307009) [406].

A classic treatment of Fermi liquids that predates EFT methods is

- "Quantum Theory of Many-Particle Systems," A.L. Fetter and J.D. Walecka, McGraw-Hill 1971 (Dover, 2002).

There are a number of good reviews about Quantum Hall systems, both from the point of view of EFT methods as well as from a more fundamental point of view. Some textbook descriptions are in

- "Field Theories of Condensed Matter Physics.", E. Fradkin, Cambridge Press, 2013.
- "The Quantum Hall Effect," R.E. Prange and S.M. Girvin, Springer-Verlag, 1987.
- "Perspectives in Quantum Hall Effects: Novel Quantum Liquids in Low-Dimensional Semiconductor Structures," S. Das Sarma and A. Pinczuk, John Wiley & Sons, 2004.

Some lecture notes that I have found very useful are

- "The Quantum Hall Effect: Novel Excitations and Broken Symmetries," S.M. Girvin, Lectures delivered at *École d'Éte des Houches*, July 1998 (arXiv:cond-mat/9907002 (cond-mat.mes-hall)).
- "Quantum Hall Fluids," A. Zee, (cond-mat/9501022).
- "Topological Orders and Edge Excitations in FQH States," X.G. Wen (cond-mat/9506066).
- "Introduction to the Physics of the Quantum Hall Regime," A.H. MacDonald (cond-mat/9410047).
- "Lectures on the Quantum Hall Effect", D. Tong, (arXiv:1606.06687 (hep-th)).
- "Three Lectures On Topological Phases Of Matter," E. Witten. lectures given at the PITP school 2015. Published in *Riv. Nuovo Cim.* **39** (2016) 313.

Some other useful references are cited in the main text.

Open systems

The treatment of fluids goes back to the 19th century, and is the birthplace of many EFT methods. Very useful textbooks are

- "An Introduction to Fluid Dynamics," G.K. Batchelor, Cambridge Press, 1967.
- "Fluid mechanics," L.D. Landau and E.M. Lifshitz, *A Course of Theoretical Physics* (2nd revised ed.) Vol 6, (Pergamon Press, 1987).

A textbook study of how electromagnetic fields interact with media is [458]:

- "Electrodynamics of Continuous Media," L.D. Landau, and E.M. Lifshitz, in *A Course of Theoretical Physics* Vol 8, (Pergamon Press, 1960).

The theory of open systems is also a well-studied field, and some textbook treatments are given by

- "The Theory of Open Quantum Systems", H.P. Breuer and F. Petruccione, Oxford Press, 2002. This is a book from which I have learned some of the open-system techniques described in this book. It is very user-friendly to those not in the area.

Two other good textbooks for this area are

- "An Open Systems Approach to Quantum Optics", H. Carmichael, Springer-Verlag, 1991.
- Quantum Dynamical Semigroups and Applications", R. Alicki and K. Lendi, Springer, 2007.