Deriving force laws from a quaternionic reformulation of the Dirac equation

Luke A. D. Hutchison*
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
(Dated: October 24, 2026)

The energy of a stationary state is $\hbar$ times the frequency ω of its wave function, and for a particle of spin $\frac{1}{2}$ that frequency follows from the Dirac equation. This paper derives that frequency for a particle in electromagnetic and Lorentz-scalar potentials and a stationary curved metric. With each chiral component written as one quaternion, the Dirac equation becomes a second-order wave equation for one quaternion, whose operator separates over a table of quaternion units into the minimally coupled Klein–Gordon operator, the field strengths, and the spin couplings of the metric. Where the positive-frequency subspace is the graph of a bounded operator, a Riccati equation gives the exact frequency operator of the positive-frequency states. At zeroth order in $\hbar$ a particle at rest has the frequency of its potential energy plus its time-dilated rest energy. Expanded in powers of $1/c^2$ through order c^{-2}, the frequency is a sum of eight terms that contain the frequencies of the interactions, and a single product of the spatial operator with its adjoint gives the Darwin, spin-orbit, and geodetic terms. From these frequencies the force-frequency identity $\mathbf{F} = -\hbar\nabla\omega$ produces all of the force laws other than those of the strong and weak interactions, among them Newtonian and post-Newtonian gravity, the inertial forces, the Lorentz force, and the Stern–Gerlach and spin-orbit forces. Expanded in $\hbar$ instead, the frequency gives the spin precession at all speeds and, for fields that vary along one direction, Bohr–Sommerfeld levels proved accurate to order $\hbar^2$ under stated bounds on the fields.

I. INTRODUCTION

The energy of a stationary quantum state is $\hbar$ times the frequency of its wave function [1, 2], and the force on a particle is the negative gradient of its energy at fixed momentum. A companion paper [3] combines the two into the force-frequency identity $\mathbf{F} = -\hbar\nabla\omega$, where $\omega(\mathbf{x}, \mathbf{p})$ is the frequency as a function of position and momentum. The identity gives the force once the frequency is known, but does not say what the frequency is. For a particle of spin $\frac{1}{2}$ it follows from the Dirac equation [4]. This paper derives that frequency for a particle in electromagnetic and Lorentz-scalar potentials and a stationary curved metric, all acting together, and splits it into the frequencies of the known forces.

Parts of the answer are known. The Foldy–Wouthuysen transformation [5] separates the positive-energy states of the Dirac equation as a series in $1/c^2$ by successive unitary transformations, and Eriksen [6] gave the transformation in closed form through the sign operator $\operatorname{sgn} H = H(H^2)^{-1/2}$ of the Dirac Hamiltonian H. Hehl and Ni [7] derived the inertial and rotational couplings of a Dirac particle in an accelerated, rotating frame, and Silenko [8] and Obukhov *et al.* [9, 10, 11] extended the Foldy–Wouthuysen method to stationary gravitational fields. Heully *et al.* [12] and Kutzelnigg and Liu [13] decoupled the positive-energy states exactly by solving a Riccati equation. Quaternionic and Clifford-algebra forms of the Dirac equation go back to Lanczos [14] and were developed by Hestenes [15], De Leo and Rotelli [16], De Leo and Rodrigues [17], and, in curved spacetime, Lasenby *et al.* [18]. Feynman and Gell-Mann [19] wrote the Dirac equation in an electromagnetic field as a second-order equation for a two-component spinor. Where the $1/c^2$ series fails, at high speeds, Rubinow and Keller [20], Bolte and Keppeler [21], and Panati *et al.* [22] expanded the Dirac equation in $\hbar$ instead and obtained the relativistic spin precession, and Keppeler [23] obtained semiclassical quantization rules from that expansion.

This paper writes the Dirac equation with each chiral component as one quaternion and follows it to the frequency. The results, and what each adds to that prior work, are these.

(i) *A second-order quaternionic wave equation in curved spacetime.* The product of the two first-order chiral equations is one second-order equation for one quaternion, equivalent to the Dirac equation for a particle of nonzero mass in flat spacetime and in every stationary metric. In flat spacetime it is the second-order equation of Feynman and Gell-Mann [19], written in quaternions. Its operator expands over a four-by-four table of quaternion units, with the minimally coupled Klein–Gordon operator on the diagonal, the electric and gravitational field strengths as commutators on the border, and the magnetic and spatial-curvature spin couplings in the interior; the gravitomagnetic field enters all three kinds of entry. The Jacobi identity of the

* luke.hutch@alum.mit.edu

covariant derivatives gives the homogeneous Maxwell equations. In flat spacetime all four of Maxwell's equations are the four parts of one biquaternion equation [24]. Its operator is that of the massless Weyl equation with a second imaginary unit in place of i, the quaternion form of the comparison of Bialynicki-Birula [25].

(ii) *The exact frequency operator in all the fields at once.* A Riccati equation gives the exact frequency operator of the positive-frequency states, a Hermitian operator on two-component spinors, in electromagnetic potentials, a lapse, a shift, a general spatial metric, and a Lorentz-scalar potential together. For an electrostatic potential alone it is the result of Heully *et al.* [12]. Wherever a stated contraction bound holds, the Riccati equation has exactly one solution of norm at most 1.

(iii) *The time-dilated rest frequency from the Dirac equation.* The frequency is the potential plus the lapse times a relativistic root, with the magnetic spin coupling exact inside the root and a commutator term that vanishes where the momentum, the metric coefficients, and the potentials commute. At zeroth order in $\hbar$ a particle at rest has the frequency $(V + mc^2\alpha)/\hbar$, where V is its potential energy and α is the time-dilation factor. In the same metric the lapse times the root, together with the shift coupling, is the leading term of the Foldy–Wouthuysen Hamiltonian of Obukhov *et al.* [10], which is a series in $\hbar$; here the operator is exact, includes the electromagnetic and Lorentz-scalar potentials, and collects every correction in the commutator term. The companion paper takes the rest part, $mc^2\alpha/\hbar$, from local position invariance; here it follows from the Dirac equation.

(iv) *An expansion into the frequencies of the classical forces.* Expanded in powers of $1/c^2$ through order c^{-2}, the frequency operator is a sum of eight terms that contain the frequencies of the interactions, from Newtonian gravity and the inertial reactions to the Lorentz force, the spin couplings, and the fine structure. Where the momentum, lapse, shift and potential terms of the Riccati equation are bounded operators with norms small compared with mc^2, the series converges, with an error of order c^{-2N-2} after the terms of order c^{-2N}. With the force-frequency identity the frequency of each interaction gives its force law.

(v) *One product for the fine structure.* The Darwin term, the electric spin-orbit coupling with its Thomas factor, and the geodetic coupling with its coefficient $\gamma_{\mathrm{PPN}} + \frac{1}{2}$ all come from one dressed product of the spatial operator with its adjoint, in place of the nested commutators of the Foldy–Wouthuysen series.

(vi) *Spin precession and energy levels at all speeds.* An expansion in $\hbar$ instead of $1/c^2$ gives the spin precession at all speeds. It also gives Bohr–Sommerfeld rules for the levels in magnetic, electric, crossed, and static gravitational fields that vary along one direction. These rules are proved to approximate the Dirac levels with an error of order $\hbar^2$ when the magnetic field is bounded and nowhere zero. The approximation is one to one in the magnetic, electric, and gravitational cases, where the electrostatic potential is also bounded and the lapse and scale factor lie between two positive constants, and two way in crossed fields. Rules of this form, with half the spin rotation angle as the first-order term, are those of Keppeler [23]. For such 2×2 blocks Becker *et al.* [26] proved them as a Hausdorff distance between the spectrum and the solutions of the rule; that distance does not count levels. New here are the one-to-one count and constants whose dependencies are stated. In uniform crossed fields, and in crossed fields with a hyperbolic-tangent vector potential, the rules give the exact Dirac levels.

The companion paper uses the expansion of the frequency in powers of $1/c^2$ derived here to obtain the classical force laws term by term.

The paper first states the force-frequency identity and the three ways a potential can couple, and writes the spinor and its two covariant derivatives in quaternions. It then derives the quaternionic wave equation from the Dirac equation, the exact frequency operator from the wave equation, and the expansion of that operator in $1/c^2$ with its catalog of forces. The expansion in $\hbar$ follows, with the spin precession and the quantization rules. The section before the conclusion interprets the spinor as a rotating frame, compares the formulation with earlier ones, and extends it to general spatial metrics and to color, where the force law gains the Wong force on a precessing color charge. The appendices give the longer derivations, and state and prove the level theorems.

II. PHASE GRADIENTS, COUPLING CLASSES, AND THE QUATERNIONIC SPINOR

A particle has rest mass m and electric charge q; m_0 denotes its rest mass when a position-dependent Lorentz-scalar potential modifies the effective mass. The constants c and $\hbar$ are the speed of light in vacuum and the reduced Planck constant. An overdot denotes a total derivative d/dt along the particle's trajectory, so that $\mathbf{v} = \dot{\mathbf{x}}$ is the coordinate velocity, while ∂_t denotes the partial derivative with respect to coordinate time t at fixed spatial position $\mathbf{x}$. Unless stated otherwise, frequencies ω are angular frequencies measured per unit of coordinate time t, energies are $E = \hbar\omega$, the spacetime metric signature is $(-,+,+,+)$, and all equations are written in SI units.

A. The force-frequency identity in phase space and Hilbert space

The kinematic and operator relations of this section are those of the companion paper [3], restated to make this paper self-contained; they connect the frequency operator of a wave equation to mechanical forces and spin precession. A stationary quantum state of energy E oscillates with angular frequency ω satisfying the Planck–Einstein relation [1, 2]

$$E = \hbar\,\omega, \tag{1}$$

and a free particle at rest with rest energy mc^2 [27] has the *rest frequency*

$$\omega_0 = \frac{mc^2}{\hbar}. \tag{2}$$

For a matter wave with local phase $\theta(\mathbf{x}, t)$, the canonical momentum $\mathbf{p}$, wavevector $\mathbf{k} \equiv \nabla\theta$, and angular frequency ω are given by the de Broglie relations [28]

$$\begin{aligned} \mathbf{p} &= \hbar\,\nabla\theta, \\ \omega &= -\partial_t\theta. \end{aligned} \tag{3}$$

In a stationary spacetime where a static clock at position $\mathbf{x}$ records proper time $d\tau = \alpha(\mathbf{x})\,dt$, the dimensionless ratio $\alpha(\mathbf{x}) \equiv d\tau/dt$ is the *time-dilation factor* (or clock field). Rescaling the proper rest frequency ω_0 to coordinate time t gives the time-dilated rest frequency

$$\begin{aligned} \omega(\mathbf{x}) &= \omega_0\,\alpha(\mathbf{x}) \\ &= \frac{mc^2}{\hbar}\,\alpha(\mathbf{x}). \end{aligned} \tag{4}$$

The *gravitational acceleration field* is

$$\mathbf{g} \equiv -c^2\nabla\alpha. \tag{5}$$

In the weak-field Newtonian limit, $\alpha(\mathbf{x}) \approx 1 + \Phi_g(\mathbf{x})/c^2$ for the gravitational potential Φ_g, so $\mathbf{g} = -\nabla\Phi_g$. In a frame with uniform linear acceleration $\mathbf{a}$, $\alpha(\mathbf{x}) = 1 + \mathbf{a}\cdot\mathbf{x}/c^2$ [7, 29], so $\mathbf{g} = -\mathbf{a}$. The gravitational force on a particle of rest mass m at rest, $\mathbf{p} = \mathbf{0}$, is then, with no weak-field approximation,

$$\begin{aligned} \mathbf{F}_g &= -\hbar\nabla(\omega_0\,\alpha) \\ &= -mc^2\nabla\alpha \\ &= m\,\mathbf{g}. \end{aligned} \tag{6}$$

This reduces to Newtonian gravity $\mathbf{F}_g = -m\nabla\Phi_g = -(GMm/r^2)\hat{\mathbf{r}}$ around a spherical mass M, where G is Newton's constant [30], and to the d'Alembert inertial reaction force $\mathbf{F}_{\text{in}} = -m\mathbf{a}$ in an accelerated frame.

More generally, when the Hamiltonian on phase space $(\mathbf{x}, \mathbf{p})$ is written in frequency units as

$$\begin{aligned} H(\mathbf{x}, \mathbf{p}) &= E \\ &= \hbar\,\omega(\mathbf{x}, \mathbf{p}), \end{aligned} \tag{7}$$

Hamilton's canonical equations $\dot{\mathbf{x}} = \nabla_{\mathbf{p}}H$ and $\dot{\mathbf{p}} = -\nabla_{\mathbf{x}}H$ [31, 32], together with the definition of the canonical force $\mathbf{F} \equiv \dot{\mathbf{p}}$ and the notation $\nabla \equiv \nabla_{\mathbf{x}}|_{\mathbf{p}}$ for the spatial gradient at fixed canonical momentum $\mathbf{p}$, yield the *force-frequency identity* [3]:

$$\begin{aligned} \boxed{\mathbf{F} = -\hbar\nabla\omega}\,, \\ \dot{\mathbf{x}} = \hbar\nabla_{\mathbf{p}}\omega\,. \end{aligned} \tag{8}$$

(The force-frequency / SGM identity and group velocity)

At the operator level, for any state $|\psi\rangle$ evolving under a Hermitian Hamiltonian operator $\hat{H} = \hbar\hat{\omega}$, the rate of change of the expectation value $\langle\hat{A}\rangle \equiv \langle\psi|\hat{A}|\psi\rangle$ of an observable $\hat{A}$ is given by the Ehrenfest–Heisenberg relation [33]

$$\frac{d}{dt}\langle\hat{A}\rangle = \frac{i}{\hbar}\langle[\hat{H}, \hat{A}]\rangle + \Big\langle\frac{\partial\hat{A}}{\partial t}\Big\rangle, \tag{9}$$

where $[\hat{A},\hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$. The relation requires that the state be normalizable, that ψ and $\hat{A}\psi$ lie in the domain of $\hat{H}$, and that $\hat{H}\psi$ lie in the domain of $\hat{A}$, so that $\hat{H}$ can act on either factor of the integrand with no boundary term. Taking $\hat{A} = \hat{\mathbf{p}} = -i\hbar\nabla$ and using $[\hat{\omega},\hat{\mathbf{p}}] = i\hbar\nabla\hat{\omega}$ gives the exact expectation-value form of the force-frequency identity [3],

$$\frac{d\langle\hat{\mathbf{p}}\rangle}{dt} = -\hbar\,\langle\nabla\hat{\omega}\rangle, \tag{10}$$

while taking for $\hat{A}$ the spin operator $\mathbf{S} = \frac{1}{2}\hbar\boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = (\sigma_1,\sigma_2,\sigma_3)$ is the vector of Pauli matrices, gives the spin precession equation $d\langle\mathbf{S}\rangle/dt = (i/\hbar)\langle[\hat{H},\mathbf{S}]\rangle$.

Equation (8) concerns the function $\omega(\mathbf{x},\mathbf{p}) = H(\mathbf{x},\mathbf{p})/\hbar$ on phase space. The phase rate of Eq. (3) equals that function on the momentum field $\mathbf{p}(\mathbf{x}) = \hbar\nabla\theta$ only up to the *quantum potential* $V_Q \equiv -(\hbar^2/2m)\,\nabla^2|\psi|/|\psi|$ of the wave function ψ [34, 35]: for $H = mc^2 + \mathbf{p}^2/2m + V$, the Schrödinger equation written in amplitude and phase gives

$$-\partial_t\theta = \omega\big(\mathbf{x},\hbar\nabla\theta\big) + \frac{V_Q}{\hbar}. \tag{11}$$

V_Q is negligible where $|\psi|$ varies slowly on the scale of the de Broglie wavelength, and there Eq. (8) holds at each point of the wave. Elsewhere it holds only in the mean, Eq. (10): in the ground state of a harmonic oscillator the phase is uniform in space, so the force obtained from the phase vanishes at every point, while $-\nabla V$ vanishes only at the minimum of V. In the mean, for $H = mc^2 + \mathbf{p}^2/2m + V$, expand $\nabla V(\hat{\mathbf{x}})$ in a Taylor series about the mean position $\langle\hat{\mathbf{x}}\rangle$. Its first-order term is proportional to $\langle\hat{\mathbf{x}} - \langle\hat{\mathbf{x}}\rangle\rangle = \mathbf{0}$, so $\langle\nabla V\rangle = \nabla V(\langle\hat{\mathbf{x}}\rangle)$ up to terms of second order in the width of the state. The equality is exact when V is at most quadratic in position. From here on, ω is the function on phase space, or the Weyl symbol of the frequency operator.

B. External potentials, minimal coupling, and the three coupling classes

An external field shifts the frequency of the particle by an amount proportional to the charge that couples to the field, rather than to the mass:

$$\omega(\mathbf{x},\mathbf{v}) = \omega_0 + \frac{q\Phi(\mathbf{x}) - q\,\mathbf{v}\cdot\mathbf{A}(\mathbf{x})}{\hbar}. \tag{12}$$

Here Φ is the scalar potential, $\mathbf{A}$ is the vector potential, $\mathbf{v} = \dot{\mathbf{x}}$ is the velocity of the particle, and q is its electric charge; the coupling is written in the form of electromagnetic potentials, and any other potential that enters the phase of the wave function enters in the same way. In terms of the canonical momentum $\mathbf{p}$ and the *kinematic momentum*

$$\boldsymbol{\pi} \equiv \mathbf{p} - q\mathbf{A}(\mathbf{x},t), \tag{13}$$

the nonrelativistic Hamiltonian $H(\mathbf{x},\mathbf{p}) = |\mathbf{p} - q\mathbf{A}|^2/(2m) + q\Phi(\mathbf{x},t)$ has velocity $\mathbf{v} = \nabla_{\mathbf{p}}H = \boldsymbol{\pi}/m$, so that $\mathbf{p} = m\mathbf{v} + q\mathbf{A}$. Substituting $\mathbf{p}\cdot\mathbf{A} = m\,\mathbf{v}\cdot\mathbf{A} + q|\mathbf{A}|^2$ gives

$$H = \frac{\mathbf{p}^2}{2m} + q\Phi - q\,\mathbf{v}\cdot\mathbf{A} - \frac{q^2|\mathbf{A}|^2}{2m}. \tag{14}$$

Equation (12) is ω_0 plus the part of $H/\hbar$ that depends on position at fixed $\mathbf{p}$ and is of first order in the potentials: it omits $\mathbf{p}^2/2m$, which does not depend on position at fixed $\mathbf{p}$, and $-q^2|\mathbf{A}|^2/2m$, which is of second order. The omission does not change the force, because the gradient of $|\mathbf{p} - q\mathbf{A}|^2/2m$ at fixed $\mathbf{p}$ has the components $-q\,\mathbf{v}\cdot\partial_i\mathbf{A}$, which are those of the gradient of $-q\,\mathbf{v}\cdot\mathbf{A}$ at fixed $\mathbf{v}$. Applying Eq. (8) at fixed $\mathbf{p}$ gives $\dot{\mathbf{p}} = -\hbar\nabla\omega = -q\nabla\Phi + q\nabla(\mathbf{v}\cdot\mathbf{A})$, where the gradient acts on $\mathbf{A}(\mathbf{x},t)$ at fixed $\mathbf{v} = \boldsymbol{\pi}/m$. This rate of the canonical momentum depends on the gauge; the gauge-dependent parts cancel against $q\dot{\mathbf{A}}$ in the rate of the kinematic momentum. Since $m\dot{\mathbf{v}} = \dot{\mathbf{p}} - q\,\dot{\mathbf{A}} = \dot{\mathbf{p}} - q\,\partial_t\mathbf{A} - q(\mathbf{v}\cdot\nabla)\mathbf{A}$, combining this with the vector identity $\nabla(\mathbf{v}\cdot\mathbf{A}) - (\mathbf{v}\cdot\nabla)\mathbf{A} = \mathbf{v}\times(\nabla\times\mathbf{A})$ yields the *Lorentz force* [36, 37]

$$m\dot{\mathbf{v}} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}), \tag{15}$$

where $\mathbf{E} = -\nabla\Phi - \partial_t\mathbf{A}$ is the electric field and $\mathbf{B} = \nabla\times\mathbf{A}$ is the magnetic field. With the *electrostatic phase factor*

$$\alpha_e(\mathbf{x}) \equiv 1 + \frac{q\Phi(\mathbf{x})}{mc^2}, \tag{16}$$

a static particle ($\mathbf{v} = \mathbf{0}$) in an electrostatic field has frequency $\omega(\mathbf{x}) = \omega_0\,\alpha_e(\mathbf{x})$, so that Eq. (8) gives $\mathbf{F} = -\hbar\omega_0\nabla\alpha_e = -q\nabla\Phi = q\mathbf{E}$ [38], which has the form of the gravitational relation $\mathbf{F}_g = -\hbar\omega_0\nabla\alpha = -mc^2\nabla\alpha$ of Eq. (6). The two differ in that $\alpha(\mathbf{x})$ is a property of the spacetime metric alone, independent of m, whereas $\alpha_e(\mathbf{x})$ depends on the particle's charge-to-mass ratio q/m. Around a closed spatial loop, the vector potential $\mathbf{A}$ shifts the phase of the wave function by the Aharonov–Bohm phase [39, 40]

$$\begin{aligned}\Delta\theta_{AB} &= \frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{x} \\ &= \frac{q}{\hbar}\,\Phi_B,\end{aligned} \tag{17}$$

where $\Phi_B = \int \mathbf{B}\cdot d\mathbf{\Sigma}$ is the enclosed magnetic flux.

TABLE I. The three classes of coupling to the wave-function frequency.

Class	**How it enters ω**	**Proportionality**	**Interactions**
Multiplicative (geometric)	Rescales the rest frequency: $\omega = \omega_0\,\alpha(\mathbf{x})$, with α universal	$\Delta\omega \propto \omega_0 = mc^2/\hbar$; acceleration independent of m	Gravity; inertia
Additive (charge-coupled)	Adds a potential term: $\omega = \omega_0 + V(\mathbf{x})/\hbar$, with V proportional to the charge	$\Delta\omega = V/\hbar$, independent of m; acceleration $\propto 1/m$	Electromagnetism; weak and strong forces
Lorentz-scalar (mass-shifting)	Shifts the rest mass inside the root: $m(\mathbf{x}) = m_0 + V_{\rm sc}(\mathbf{x})/c^2$	$\Delta\omega_0 = V_{\rm sc}(\mathbf{x})/\hbar$ at rest; suppressed by $1/\gamma$ in motion	Higgs mass coupling; nuclear scalar mean field

Table I lists three ways in which an interaction can enter the frequency of a matter wave. In the *multiplicative (geometric) class*, the spacetime metric rescales the proper rate of time by $\alpha(\mathbf{x})$. The factor $\alpha(\mathbf{x})$ multiplies the entire rest frequency $\omega_0 = mc^2/\hbar$, so the force is proportional to m and the resulting acceleration is universal. In the *additive (charge-coupled) class*, a gauge potential $V(\mathbf{x})$ adds $V(\mathbf{x})/\hbar$ directly to the coordinate frequency ω independently of m. In the *Lorentz-scalar class* (analyzed in detail in Appendix A 1), a scalar potential $V_{\rm sc}(\mathbf{x})$ shifts the rest mass itself, $m(\mathbf{x}) = m_0 + V_{\rm sc}(\mathbf{x})/c^2$. The scalar meson field of nuclear matter shifts the nucleon mass in this way [41], and semiconductor heterostructures have a position-dependent effective mass [42].

Remark 1 (Local gauge invariance requires minimal coupling)**:** *By Eq. (3), the four-gradient of the phase $\theta(\mathbf{x},t)$ is $\partial_\mu\theta = (-\omega/c, \mathbf{k})$ in coordinates $x^\mu = (ct, \mathbf{x})$. Under a local phase rotation $\psi(\mathbf{x},t) \mapsto e^{i\mathcal{X}(\mathbf{x},t)}\psi(\mathbf{x},t)$ with a smooth real function $\mathcal{X}(\mathbf{x},t)$, the phase shifts by $\theta \mapsto \theta + \mathcal{X}$, so its four-gradient shifts additively by $\partial_\mu\theta \mapsto \partial_\mu\theta + \partial_\mu\mathcal{X}$: specifically, $\omega \mapsto \omega - \partial_t\mathcal{X}$ and $\mathbf{k} \mapsto \mathbf{k} + \nabla\mathcal{X}$. For the physical frequency and kinematic wavevector of a particle of charge q to remain invariant under this local rephasing, $\partial_t\theta$ and $\nabla\theta$ must be accompanied by a four-potential $A_\mu = (-\Phi/c, \mathbf{A})$ that transforms as $A_\mu \mapsto A_\mu + (\hbar/q)\partial_\mu\mathcal{X}$, that is, $\Phi \mapsto \Phi - (\hbar/q)\partial_t\mathcal{X}$ and $\mathbf{A} \mapsto \mathbf{A} + (\hbar/q)\nabla\mathcal{X}$ [43, 44]. The gauge-invariant combinations are then $\hbar(\partial_\mu\theta) - qA_\mu$, whose time and space components are $-(\hbar\omega - q\Phi)/c$ and $\hbar\mathbf{k} - q\mathbf{A} = \boldsymbol{\pi}$. Thus local phase invariance forces the scalar potential $V = q\Phi$ to enter $\hbar\omega$ additively and the vector potential $q\mathbf{A}$ to enter $\mathbf{p} = \hbar\mathbf{k}$ additively. This is the rule of* minimal coupling. *Terms built from the field strength, such as an anomalous magnetic moment, are allowed by the same invariance but contain no bare A_μ.*

A rephasing also excludes any other dependence on Φ. Suppose that $\omega = \omega_0\,f_\Phi(\Phi)$ for a differentiable function f_Φ. The rephasing $\mathcal{X} = -q\,t\,\delta\Phi/\hbar$, for a constant $\delta\Phi$, shifts Φ to $\Phi + \delta\Phi$ and adds $q\,\delta\Phi/\hbar$ to ω, so $\omega_0\,[f_\Phi(\Phi+\delta\Phi) - f_\Phi(\Phi)] = q\,\delta\Phi/\hbar$ for all Φ and $\delta\Phi$. Differentiating at $\delta\Phi = 0$ gives $df_\Phi/d\Phi = q/(mc^2)$, and with $f_\Phi(0) = 1$ this is Eq. (12) at $\mathbf{A} = \mathbf{0}$. The factor α couples multiplicatively because it adds no phase of its own: it rescales the proper time $d\tau = \alpha\,dt$ in which the existing phase advances.

C. The spin-$\frac{1}{2}$ wave function as a single quaternion

A particle of spin $\frac{1}{2}$ has a two-component complex wave function, a *spinor*,

$$\psi = \begin{pmatrix}\psi_1 \\ \psi_2\end{pmatrix}, \tag{18}$$

where $\psi_1(\mathbf{x},t)$ and $\psi_2(\mathbf{x},t)$ are complex functions of position and time, the amplitudes for the spin to be found along and against the third coordinate axis. Its four real parts, the real and imaginary parts of ψ_1 and ψ_2, can be written as a single *quaternion*, a number of the form $q_0 + q_1 e_1 + q_2 e_2 + q_3 e_3$ with four real components q_0, q_1, q_2, q_3 and three imaginary units e_1, e_2, e_3 that multiply by Hamilton's rules [45],

$$e_j\, e_k = -\delta_{jk} + \epsilon_{jkl}\, e_l, \tag{19}$$

for $j,k,l \in \{1,2,3\}$. Here δ_{jk} is 1 when $j = k$ and 0 otherwise. The symbol ϵ_{jkl} is $+1$ for $(j,k,l) = (1,2,3)$, $(2,3,1)$, $(3,1,2)$, -1 for $(3,2,1)$, $(1,3,2)$, $(2,1,3)$, and 0 whenever two indices are equal. A repeated index is summed over $1,2,3$. Each unit squares to -1, $e_1^2 = e_2^2 = e_3^2 = -1$, and distinct units anticommute: $e_1 e_2 = -e_2 e_1 = e_3$, $e_2 e_3 = -e_3 e_2 = e_1$, and $e_3 e_1 = -e_1 e_3 = e_2$. Multiplication is associative, $(ab)c = a(bc)$, so a factor on the left of ψ and a factor on the right of ψ may be applied in either order, $(a\psi)b = a(\psi b)$, even when a and b do not commute with each other. The *conjugate* of $q = q_0 + q_1 e_1 + q_2 e_2 + q_3 e_3$ reverses the signs of the three units, $\bar{q} = q_0 - q_1 e_1 - q_2 e_2 - q_3 e_3$. Conjugation reverses the order of a product, $\overline{ab} = \bar{b}\,\bar{a}$, and gives the squared norm $q\bar{q} = \bar{q}q = q_0^2 + q_1^2 + q_2^2 + q_3^2 = |q|^2$, a nonnegative real number. Every nonzero quaternion therefore has the inverse $q^{-1} = \bar{q}/|q|^2$. Because the three units anticommute, a product of a vector quaternion $\mathbf{u}\cdot\mathbf{e} = u_1 e_1 + u_2 e_2 + u_3 e_3$ with another vector quaternion $\mathbf{v}\cdot\mathbf{e}$ combines the dot product and the cross product into one operation,

$$(\mathbf{u}\cdot\mathbf{e})(\mathbf{v}\cdot\mathbf{e}) = -\mathbf{u}\cdot\mathbf{v} + (\mathbf{u}\times\mathbf{v})\cdot\mathbf{e}. \tag{20}$$

The spinor of Eq. (18) is written as a quaternion by identifying the complex imaginary unit i with right multiplication by $-e_3$ ($e_3^{-1} = -e_3$) and the second spinor component with left multiplication by e_2:

$$\psi = \mathrm{Re}\,\psi_1 - (\mathrm{Im}\,\psi_1)\, e_3 + e_2\,\big(\mathrm{Re}\,\psi_2 - (\mathrm{Im}\,\psi_2)\, e_3\big), \tag{21}$$

which expands via $e_2 e_3 = e_1$ to $\psi = \mathrm{Re}\,\psi_1 - (\mathrm{Im}\,\psi_2)\, e_1 + (\mathrm{Re}\,\psi_2)\, e_2 - (\mathrm{Im}\,\psi_1)\, e_3$. This mapping between two-component complex spinors and real quaternions is one-to-one and real-linear. Multiplying both ψ_1 and ψ_2 by i replaces $\mathrm{Re}\,\psi_a$ by $-\mathrm{Im}\,\psi_a$ and $\mathrm{Im}\,\psi_a$ by $\mathrm{Re}\,\psi_a$ ($a = 1,2$), which in Eq. (21) is right multiplication of the whole quaternion ψ by $-e_3 = e_3^{-1}$:

$$i\psi \;\longleftrightarrow\; -\psi\, e_3. \tag{22}$$

For any real angle ϕ, the series $\exp(e_3\phi) = \sum_{n=0}^{\infty} (e_3\phi)^n/n! = \cos\phi + e_3 \sin\phi$ has e_3 in place of $-i$, so multiplying the spinor of Eq. (18) by the complex phase factor $e^{-i\phi}$ is right multiplication of the quaternion ψ by $\exp(e_3\phi)$.

On two-component spinors, the Pauli matrices [46]

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$
$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{23}$$

satisfy $\sigma_j \sigma_k = \delta_{jk} + i\epsilon_{jkl}\sigma_l$. Under the mapping of Eq. (21), the action of σ_k on ψ is left multiplication by e_k combined with right multiplication by $e_3^{-1} = -e_3$:

$$\begin{aligned} \sigma_k\, \psi &= e_k\, \psi\, e_3^{-1} \\ &= -e_k\, \psi\, e_3, \end{aligned} \tag{24}$$

for $k = 1,2,3$. Indeed, by associativity and $e_3^{-1} = -e_3$, applying Eq. (24) twice gives $\sigma_j(\sigma_k\psi) = e_j(e_k\psi e_3^{-1})e_3^{-1} = (e_j e_k)\psi(e_3^{-2}) = -e_j e_k \psi = \delta_{jk}\psi - \epsilon_{jkl} e_l \psi = \delta_{jk}\psi + \epsilon_{jkl}(e_l \psi e_3^{-1})e_3^{-1} = \delta_{jk}\psi + i\epsilon_{jkl}\sigma_l\psi$, which is the Pauli algebra. Since i is right multiplication by e_3^{-1}, multiplying Eq. (24) by $-i$ (right multiplication by e_3) gives the simpler rule

$$-i\sigma_k\, \psi = e_k\, \psi. \tag{25}$$

The probability density of the spinor is $\rho = |\psi_1|^2 + |\psi_2|^2 = \psi^\dagger\psi$, which equals the quaternion norm squared $\rho = \psi\bar{\psi} = \bar{\psi}\psi = |\psi|^2$. The inner product $\psi^\dagger\chi$ of two spinors has real part $\mathrm{Re}(\psi^\dagger\chi) = \mathrm{Sc}(\bar{\psi}\chi)$, the scalar part of $\bar{\psi}\chi$. The local spin density vector $\mathbf{s}(\mathbf{x},t) \equiv \psi^\dagger \boldsymbol{\sigma} \psi$ (in units of $\frac{1}{2}\hbar$) is represented as a pure vector quaternion by

$$\mathbf{s}\cdot\mathbf{e} = \psi\, e_3\, \bar{\psi}. \tag{26}$$

Indeed, $\overline{\psi e_3\bar{\psi}} = \psi\bar{e}_3\bar{\psi} = -\psi e_3\bar{\psi}$, so $\psi e_3\bar{\psi}$ has zero scalar part, and its kth component is $-\mathrm{Sc}(e_k\psi e_3\bar{\psi}) = \mathrm{Sc}(\bar{\psi}(-e_k\psi e_3)) = \mathrm{Re}(\psi^\dagger\sigma_k\psi) = \psi^\dagger\sigma_k\psi = s_k$.

A spatial rotation by angle ϑ about the unit axis $\hat{\mathbf{m}}$ acts on a two-component spinor by the $SU(2)$ matrix $\exp(-i\frac{1}{2}\vartheta\,\hat{\mathbf{m}}\cdot\boldsymbol{\sigma})$, which by Eq. (25) is left multiplication by the unit quaternion (rotor)

$$\begin{aligned}\mathcal{R} &= \exp\big(\tfrac{1}{2}\vartheta\,\hat{\mathbf{m}}\cdot\mathbf{e}\big)\\ &= \cos\frac{\vartheta}{2} + (\hat{\mathbf{m}}\cdot\mathbf{e})\,\sin\frac{\vartheta}{2},\end{aligned} \tag{27}$$

which maps ψ to $\mathcal{R}\,\psi$. Under this left action, the spin vector of Eq. (26) transforms as $\mathbf{s}\cdot\mathbf{e} \mapsto (\mathcal{R}\psi)e_3(\overline{\mathcal{R}\psi}) = \mathcal{R}(\psi e_3\bar{\psi})\bar{\mathcal{R}} = \mathcal{R}(\mathbf{s}\cdot\mathbf{e})\bar{\mathcal{R}}$. This rotates $\mathbf{s}$ about $\hat{\mathbf{m}}$ through the full angle ϑ, while $\mathcal{R}$ itself turns through half the angle. At $\vartheta = 2\pi$, $\mathcal{R} = -1$ and $\psi \mapsto -\psi$, the spinor double cover of $SO(3)$. A $U(1)$ phase rotation, by contrast, acts on ψ from the right,

$$\psi \mapsto \psi\,\exp(e_3\phi), \tag{28}$$

which leaves the spin density $\psi\exp(e_3\phi)e_3\exp(-e_3\phi)\bar{\psi} = \psi e_3\bar{\psi}$ invariant because e_3 commutes with $\exp(e_3\phi)$. By associativity of quaternion multiplication, every spatial rotation on the left commutes with every phase transformation on the right: $(\mathcal{R}\psi)\exp(e_3\phi) = \mathcal{R}(\psi\exp(e_3\phi))$. Placing the two symmetry groups on opposite sides of the single quaternion ψ separates spatial rotations (left multiplication) from internal gauge operations (right multiplication).

D. The two covariant derivatives: gauge connection on the right, spacetime geometry on the left

A gauge transformation, the rephasing $\psi \mapsto e^{i\mathcal{X}}\psi$ with a real function $\mathcal{X}(\mathbf{x},t)$, multiplies ψ on the right by $\exp(-\mathcal{X}(\mathbf{x},t)\,e_3)$, while a local spatial frame rotation multiplies ψ on the left by $\mathcal{R}(\mathbf{x},t)$. Therefore the two covariant derivatives of a spin-$\frac{1}{2}$ particle in curved spacetime and external gauge potentials act on opposite sides of ψ.

On the right, the *gauge-covariant derivative* couples to the additive potential $V(\mathbf{x},t)$ (with $V = q\Phi$ for an electrostatic potential Φ) in time and to the vector potential $\mathbf{A}(\mathbf{x},t)$ in space:

$$D_t\psi \equiv \partial_t\psi - \frac{V}{\hbar}\,\psi\,e_3, \tag{29}$$

(Time component of the gauge-covariant derivative)

$$D_k\psi \equiv \partial_k\psi + \frac{q}{\hbar}\,A_k\,\psi\,e_3 \qquad (k=1,2,3). \tag{30}$$

(Spatial components of the gauge-covariant derivative)

Since i is right multiplication by $-e_3$, the kinematic momentum operator $\hat{\pi}_k = -i\hbar\partial_k - qA_k$ acts on ψ by

$$\begin{aligned}\hat{\pi}_k\,\psi &= \hbar\,(\partial_k\psi)\,e_3 - qA_k\,\psi\\ &= \hbar\,(D_k\psi)\,e_3,\end{aligned} \tag{31}$$

so that $\hbar\,D_k\psi = (\hat{\pi}_k\psi)\,e_3^{-1} = i\,\hat{\pi}_k\psi$. Contracting the spatial gauge-covariant derivative $\mathbf{D} = (D_1, D_2, D_3)$ on the left with the quaternion units $\mathbf{e} = (e_1, e_2, e_3)$,

$$\mathbf{e}\cdot\mathbf{D}\,\psi \equiv \sum_{k=1}^{3} e_k\,D_k\psi, \tag{32}$$

and using $\sigma_k\psi = e_k\psi e_3^{-1}$ (Eq. (24)) together with $e_3e_3^{-1} = 1$ yields the quaternionic form of the Dirac kinetic momentum operator $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$:

$$\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\,\psi = \hbar\,\mathbf{e}\cdot\mathbf{D}\,\psi. \tag{33}$$

In the same way, the gauge-covariant Laplacian is the sum of the three repeated spatial derivatives,

$$\begin{aligned}\mathbf{D}^2\psi &\equiv (D_1^2 + D_2^2 + D_3^2)\,\psi\\ &= -\frac{1}{\hbar^2}\,\hat{\boldsymbol{\pi}}^2\psi.\end{aligned} \tag{34}$$

Because the potentials V and A_k in Eqs. (29) and (30) are real scalars and multiply ψ on the right by e_3, both D_t and D_k commute with left multiplication by any constant quaternion.

On the left, the spacetime geometry enters through the Arnowitt–Deser–Misner (ADM) decomposition of a stationary spacetime metric [47–49]. For spatial slices $t = \text{const}$ that are conformally flat, the line element is

$$\begin{aligned} ds^2 &= -\alpha_n^2 c^2\, dt^2 + W^2 \left(d\mathbf{x} + \mathbf{A}_g\, dt\right)^2 \\ &= -\left(\alpha_n^2 c^2 - W^2 |\mathbf{A}_g|^2\right) dt^2 + 2\, W^2\, \mathbf{A}_g \cdot d\mathbf{x}\, dt + W^2\, d\mathbf{x} \cdot d\mathbf{x}, \end{aligned} \tag{35}$$

(The stationary lapse-and-shift metric)

where $\alpha_n(\mathbf{x}) > 0$ is the *lapse*, $\mathbf{A}_g(\mathbf{x})$ is the *shift* vector, which has dimensions of velocity, and $W(\mathbf{x}) > 0$ is the *spatial scale factor*. The slices are flat when $W = 1$, and the proper volume element of a slice is $W^3\, d^3x$. The Eulerian *normal observer*, whose worldline is orthogonal to the slice $t = \text{const}$, moves at velocity $-\mathbf{A}_g$ relative to the spatial coordinates, $d\mathbf{x} = -\mathbf{A}_g\, dt$. Its proper time is $d\tau_n = \alpha_n\, dt$, and the directional derivative along the unit future normal is $c^{-1}\partial_{\tau_n} = \partial_n \equiv (c\alpha_n)^{-1}(\partial_t - \mathbf{A}_g \cdot \nabla)$. A clock held at fixed spatial coordinates ($d\mathbf{x} = \mathbf{0}$) has proper time $d\tau = \alpha\, dt$, where the time-dilation factor $\alpha(\mathbf{x})$ is related to the lapse, shift and scale factor by

$$\alpha = \sqrt{\alpha_n^2 - \frac{W^2 |\mathbf{A}_g|^2}{c^2}}. \tag{36}$$

In isotropic coordinates, the Schwarzschild metric of a mass M has $\mathbf{A}_g = \mathbf{0}$, $\alpha_n = (1 - GM/2rc^2)/(1 + GM/2rc^2)$, and $W = (1 + GM/2rc^2)^2$ [49]. In the weak-field parametrized post-Newtonian (PPN) metric, $\alpha_n = 1 + \Phi_g/c^2 + (\beta_{\text{PPN}} - \frac{1}{2})\Phi_g^2/c^4$ and $W^{-1} = 1 + \gamma_{\text{PPN}}\Phi_g/c^2$ [50], where Φ_g is the Newtonian potential, γ_{PPN} measures the spatial curvature per unit mass, and β_{PPN} measures the nonlinearity in superposition ($\gamma_{\text{PPN}} = \beta_{\text{PPN}} = 1$ in general relativity). Around a slowly rotating mass M of angular momentum $\mathbf{J}_{\text{src}}$, the same metric has $\Phi_g = -GM/r$ and the shift $\mathbf{A}_g = -(2G/c^2r^3)(\mathbf{J}_{\text{src}} \times \mathbf{x})$ [48, 51]. In the Schwarzschild metric in Painlevé–Gullstrand coordinates [52, 53], the spatial slices are flat, $W = 1$, $\alpha_n = 1$, and gravity enters only through the radial free-fall shift $\mathbf{A}_g = \sqrt{2GM/r}\,\hat{\mathbf{r}}$; for this shift Eq. (36) gives the exact Schwarzschild time-dilation factor $\alpha = \sqrt{1 - 2GM/(rc^2)}$. In a frame rotating at angular velocity $\boldsymbol{\Omega}$ in flat Minkowski spacetime, $W = 1$, $\alpha_n = 1$, and $\mathbf{A}_g = \boldsymbol{\Omega} \times \mathbf{x}$, so $\alpha = \sqrt{1 - |\boldsymbol{\Omega} \times \mathbf{x}|^2/c^2}$ generates the centrifugal potential while $\mathbf{A}_g$ generates the Coriolis and spin-rotation couplings [7]. Flat spacetime in an inertial frame is $W = \alpha_n = 1$ and $\mathbf{A}_g = \mathbf{0}$.

When a flow $\boldsymbol{\xi}(\mathbf{x})$ drags the spatial coordinates by $\mathbf{x} \mapsto \mathbf{x} + \boldsymbol{\xi}(\mathbf{x})\,\delta t$ during a short time δt, the Jacobian matrix of the displacement is $\delta_{jk} + \partial_k \xi_j\, \delta t$. Its antisymmetric part $\frac{1}{2}(\partial_k \xi_j - \partial_j \xi_k)\,\delta t = -\frac{1}{2}\epsilon_{jkl}(\nabla \times \boldsymbol{\xi})_l\,\delta t$ rotates the local spatial axes by the angle $\delta\boldsymbol{\vartheta} = \frac{1}{2}(\nabla \times \boldsymbol{\xi})\,\delta t$, half the curl (vorticity) of the flow. Because the spinor is referred to the local orthonormal axes, dragging the spinor along $\boldsymbol{\xi}$ both translates its argument, which changes ψ by $(\boldsymbol{\xi} \cdot \mathbf{D}\psi)\,\delta t$, and rotates its frame by $-\delta\boldsymbol{\vartheta}$ via the left rotor $\exp(-\frac{1}{2}\delta\boldsymbol{\vartheta} \cdot \mathbf{e}) = 1 - \frac{1}{4}\,\mathbf{e} \cdot (\nabla \times \boldsymbol{\xi})\,\delta t$ of Eq. (27). The rate of change per unit time is the gauge-covariant *spinor Lie derivative* [10, 54]:

$$\mathcal{L}_{\boldsymbol{\xi}}\, \psi = \boldsymbol{\xi} \cdot \mathbf{D}\, \psi - \tfrac{1}{4}\, \mathbf{e} \cdot (\nabla \times \boldsymbol{\xi})\, \psi. \tag{37}$$

(The spinor Lie derivative along a spatial vector field $\boldsymbol{\xi}$)

The factor $\frac{1}{4}$ is the product of two halves: $\frac{1}{2}$ from the vorticity $\frac{1}{2}\nabla \times \boldsymbol{\xi}$ of a velocity field, and $\frac{1}{2}$ from the half-angle spinor rotation rotor of Eq. (27). For a rigid rotation $\boldsymbol{\xi} = \boldsymbol{\Omega} \times \mathbf{x}$ with $\mathbf{A} = \mathbf{0}$, $\nabla \times \boldsymbol{\xi} = 2\boldsymbol{\Omega}$, $(\boldsymbol{\Omega} \times \mathbf{x}) \cdot \nabla\psi = (i/\hbar)\boldsymbol{\Omega} \cdot (\mathbf{x} \times \hat{\mathbf{p}})\psi = -(1/\hbar)(\boldsymbol{\Omega} \cdot \mathbf{L}\psi)e_3$, and $-\frac{1}{4}\mathbf{e} \cdot (2\boldsymbol{\Omega})\psi = -\frac{1}{2}(\boldsymbol{\Omega} \cdot \boldsymbol{\sigma}\psi)e_3 = -(1/\hbar)(\boldsymbol{\Omega} \cdot \mathbf{S}\psi)e_3$, so that

$$\mathcal{L}_{\boldsymbol{\Omega} \times \mathbf{x}}\, \psi = -\frac{1}{\hbar}\left(\boldsymbol{\Omega} \cdot (\mathbf{L} + \mathbf{S})\, \psi\right) e_3. \tag{38}$$

Thus the spinor Lie derivative couples rotation and gravitomagnetic frame dragging $\mathbf{B}_g \equiv \nabla \times \mathbf{A}_g$ equally to the orbital angular momentum $\mathbf{L} = \mathbf{x} \times \hat{\boldsymbol{\pi}}$ and the spin $\mathbf{S} = \frac{1}{2}\hbar\boldsymbol{\sigma}$, as for a classical body [7]. A Dirac particle therefore has no anomalous gravitomagnetic moment [55].

The spinor is referred to the local orthonormal axes $\alpha_n c\, dt$ and $W(dx^k + A_g^k\, dt)$ of the metric (35). Along the normal, the spatial rotation part of the spin connection of these axes is proportional to $\nabla \times \mathbf{A}_g$ alone, with no term in ∇W, so the rotation term of Eq. (37) holds unchanged when $W \neq 1$. Let ψ_o be the spinor in these axes, with density $|\psi_o|^2$ per unit proper volume. Its covariant derivative along the unit normal is $(c\alpha_n)^{-1}$ times the sum of three terms: the gauge-covariant time derivative $D_t\psi_o$ of Eq. (29), the spinor Lie derivative $\mathcal{L}_{\mathbf{A}_g}\psi_o$ along the shift, and the volume-element divergence term $-\frac{1}{2}W^{-3}\nabla \cdot (W^3\mathbf{A}_g)\,\psi_o$, which makes the shift flow preserve the norm $\int |\psi_o|^2 W^3 d^3x$. The spinor used from here on is the rescaled spinor $\psi \equiv \alpha_n^{1/2} W^{3/2}\psi_o$, whose norm is $\int |\psi|^2 \alpha_n^{-1} d^3x$. With the factor $W^{3/2}$ the transport and divergence terms take their flat-slice form, since

$$W^{3/2}\left[\mathbf{A}_g \cdot \nabla + \tfrac{1}{2}W^{-3}\nabla \cdot (W^3\mathbf{A}_g)\right]W^{-3/2} = \mathbf{A}_g \cdot \nabla + \tfrac{1}{2}\,\nabla \cdot \mathbf{A}_g,$$

and the factor $\alpha_n^{1/2}$ divides the factor α_n^{-1} equally between the two sides of the operator. On ψ the *covariant derivative along the unit normal* is therefore

$$\begin{aligned} D_n\,\psi &= \frac{1}{c\,\alpha_n^{1/2}}\left[D_t - \mathcal{L}_{\mathbf{A}_g} - \tfrac{1}{2}\,(\nabla\cdot\mathbf{A}_g)\right]\left(\alpha_n^{-1/2}\psi\right) \\ &= \frac{1}{c\alpha_n}\left[D_t\psi - \mathbf{A}_g\cdot\mathbf{D}\,\psi + \tfrac{1}{4}\,(\mathbf{e}\cdot\mathbf{B}_g)\,\psi - \tfrac{1}{2}\,(\nabla\cdot\mathbf{A}_g)\,\psi + \tfrac{1}{2}\,(\mathbf{A}_g\cdot\nabla\ln\alpha_n)\,\psi\right], \\ &\qquad \textit{(The covariant derivative along the unit normal)} \end{aligned} \tag{39}$$

where the second line uses Eq. (37) and $\mathbf{A}_g\cdot\nabla(\alpha_n^{-1/2}\psi) = \alpha_n^{-1/2}\left[\mathbf{A}_g\cdot\nabla\psi - \frac{1}{2}(\mathbf{A}_g\cdot\nabla\ln\alpha_n)\psi\right]$ for a static lapse. The only term of D_n in which a quaternion unit multiplies ψ from the left is $\frac{1}{4}(\mathbf{e}\cdot\mathbf{B}_g)\psi/(c\alpha_n)$.

III. THE QUATERNIONIC WAVE EQUATION IN CURVED SPACETIME

A. From the Weyl–Dirac equation to the chiral quaternionic pair

The phase θ counts wave crests, so it is a Lorentz scalar, and its four-gradient with respect to $x^\mu = (ct, \mathbf{x})$ is a Lorentz covector. By Eq. (3), $\omega = -\partial_t\theta$ and $\mathbf{k} = \nabla\theta$, so

$$\partial_\mu\theta = \left(-\frac{\omega}{c},\ \mathbf{k}\right). \tag{40}$$

Its squared Minkowski norm $\mathbf{k}^2 - \omega^2/c^2$ is the same in every frame. In the rest frame $\mathbf{k} = \mathbf{0}$ and $\omega = \omega_0$ (Eq. (2)), so the norm is $-\omega_0^2/c^2$ in every frame, and solving for ω^2 gives the dispersion relation of a free particle,

$$\omega^2 = \omega_0^2 + c^2\,\mathbf{k}^2. \tag{41}$$

Since $\exp(e_3\phi) = \cos\phi + e_3\sin\phi$ (Section II C) has derivative $\exp(e_3\phi)\,e_3$ with respect to ϕ, a quaternion plane wave

$$\psi = \psi_{\mathrm{pw}}\exp\left(e_3\,(\omega t - \mathbf{k}\cdot\mathbf{x})\right), \tag{42}$$

where ψ_{pw} is a constant quaternion, has $\partial_t^2\psi = \psi\,e_3^2\,\omega^2 = -\omega^2\psi$ and $\nabla^2\psi = -\mathbf{k}^2\psi$. Multiplying Eq. (41) by $-\psi/c^2$ and replacing $-\omega^2\psi$ by $\partial_t^2\psi$ and $-\mathbf{k}^2\psi$ by $\nabla^2\psi$ gives the free Klein–Gordon equation [56, 57]

$$\frac{1}{c^2}\,\partial_t^2\psi - \nabla^2\psi + \kappa^2\,\psi = 0, \tag{43}$$

where

$$\kappa \equiv \frac{mc}{\hbar} \tag{44}$$

is the inverse reduced Compton wavelength, equal to ω_0/c by Eq. (2). Conversely, substituting Eq. (42) into the left side of Eq. (43) gives $(-\omega^2 + \omega_0^2 + c^2\mathbf{k}^2)\,\psi/c^2$, so a plane wave solves Eq. (43) only at the frequency of Eq. (41). Applying $\mathbf{e}\cdot\nabla$ twice gives $(\mathbf{e}\cdot\nabla)^2\psi = \sum_{j,k} e_j e_k\,\partial_j\partial_k\psi$, with the products $e_j e_k$ given by the 3×3 multiplication table of the quaternion units:

$$(e_j\,e_k) = \begin{pmatrix} -1 & e_3 & -e_2 \\ -e_3 & -1 & e_1 \\ e_2 & -e_1 & -1 \end{pmatrix}. \tag{45}$$

The three diagonal entries -1 yield $-\nabla^2\psi$, while the six off-diagonal entries are antisymmetric ($e_j e_k = -e_k e_j$ for $j \neq k$), so their terms cancel in pairs because $\partial_j\partial_k = \partial_k\partial_j$, leaving

$$(\mathbf{e}\cdot\nabla)^2\,\psi = -\nabla^2\psi, \tag{46}$$

and the free spinor Klein–Gordon equation

$$\frac{1}{c^2}\,\partial_t^2\psi + (\mathbf{e}\cdot\nabla)^2\,\psi + \kappa^2\,\psi = 0. \tag{47}$$

Interactions are switched on by minimal coupling. The time derivative is replaced by the normal derivative, $c^{-1}\partial_t \to D_n$ (Eq. (39)). The gradient is replaced by the gauge-covariant gradient per unit proper length, $W^{-1}\mathbf{D}$ (Eq. (30)). Its factor W^{-1} is divided equally between the two sides, as the factor α_n^{-1} is in D_n: $\nabla \to W^{-1/2}\mathbf{D}\,W^{-1/2}$. This replacement is an assumption: field-strength terms such as an anomalous-moment term are allowed by local gauge invariance (Remark 1) and are not included, so the equations that follow describe a particle with no anomalous magnetic moment. Define the normal and spatial first-order quaternion operators (all with dimensions of inverse length, like κ):

$$\begin{aligned} \mathsf{N}\,\psi &\equiv (D_n\psi)\,e_3, \\ \mathsf{X}\,\psi &\equiv \mathbf{e}\cdot\mathbf{D}\,\psi, \\ \mathsf{X}_W\,\psi &\equiv W^{-1/2}\,\mathsf{X}\big(W^{-1/2}\psi\big). \end{aligned} \tag{48}$$

(The operators of the quaternionic wave equation)

By the product rule,

$$\mathsf{X}_W\,\psi = W^{-1}\,\mathbf{e}\cdot\big(\mathbf{D} - \tfrac{1}{2}\nabla\ln W\big)\,\psi, \tag{49}$$

which reduces to X when $W = 1$. In flat spacetime ($W = \alpha_n = 1$, $\mathbf{A}_g = \mathbf{0}$), $\mathsf{X}_W = \mathsf{X}$ and $D_n = c^{-1}D_t$, and with $\hat{\mathcal{E}} \equiv i\hbar\partial_t - V$, Eqs. (22) and (33) give $\mathsf{N} = -\hat{\mathcal{E}}/(\hbar c)$ and $\mathsf{X} = \boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}/\hbar$. Since right multiplication by e_3 commutes with D_n and $e_3^2 = -1$, $\mathsf{N}^2\psi = (D_n^2\psi)\,e_3^2 = -D_n^2\psi$. For a free particle in flat spacetime N and X commute, since $\mathsf{N}(\mathsf{X}\psi) = c^{-1}\sum_k e_k\,(\partial_t\partial_k\psi)\,e_3 = \mathsf{X}(\mathsf{N}\psi)$ by associativity and $\partial_t\partial_k = \partial_k\partial_t$. When N and X_W do not commute, as in an electric field or a nonuniform lapse, the three second-order operators $\mathsf{N}^2 - \mathsf{X}_W^2$, $(\mathsf{N}-\mathsf{X}_W)(\mathsf{N}+\mathsf{X}_W) = \mathsf{N}^2 - \mathsf{X}_W^2 + [\mathsf{N},\mathsf{X}_W]$, and $(\mathsf{N}+\mathsf{X}_W)(\mathsf{N}-\mathsf{X}_W) = \mathsf{N}^2 - \mathsf{X}_W^2 - [\mathsf{N},\mathsf{X}_W]$ differ by multiples of the commutator $[\mathsf{N},\mathsf{X}_W]$. Minimal coupling in Eq. (47) gives $\mathsf{N}^2 - \mathsf{X}_W^2$ and does not select either factored product.

The Dirac equation [4] selects a factored product for each chiral component, with the two factors in opposite orders for the two components. In the Weyl (chiral) representation [58], a four-component bispinor $\Psi = (\psi_L, \psi_R)^T$ in electromagnetic potentials $(\Phi, \mathbf{A})$ obeys

$$i\hbar\,\partial_t \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} = \begin{pmatrix} q\Phi - c\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}} & mc^2 \\ mc^2 & q\Phi + c\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}} \end{pmatrix} \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}, \tag{50}$$

(The Dirac equation, Weyl representation)

which splits into the two coupled two-component spinor equations

$$i\hbar\,\partial_t\psi_L = \big(q\Phi - c\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\big)\,\psi_L + mc^2\,\psi_R, \tag{51}$$

$$i\hbar\,\partial_t\psi_R = mc^2\,\psi_L + \big(q\Phi + c\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\big)\,\psi_R. \tag{52}$$

In the quaternion representation of Eq. (21), $(i\hbar\partial_t - q\Phi)\psi = -\hbar c\,\mathsf{N}\psi$ and $c\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\psi = \hbar c\,\mathsf{X}\psi$, and $mc^2/(\hbar c) = \kappa$. With these identities, Eqs. (52) and (51) become $(\mathsf{N}+\mathsf{X})\psi_R = -\kappa\psi_L$ and $(\mathsf{N}-\mathsf{X})\psi_L = -\kappa\psi_R$. Defining the chiral quaternion pair with the sign of the left-handed component reversed,

$$\begin{aligned} \varphi_R &\equiv \psi_R, \\ \varphi_L &\equiv -\psi_L, \end{aligned} \tag{53}$$

puts the Dirac equation into a symmetric first-order quaternionic pair. Written with the minimally coupled operators of Eq. (48), which are N and X in flat spacetime, this is the *quaternionic Weyl pair*:

$$(\mathsf{N} + \mathsf{X}_W)\,\varphi_R = \kappa\,\varphi_L, \tag{54}$$

$$(\mathsf{N} - \mathsf{X}_W)\,\varphi_L = \kappa\,\varphi_R. \tag{55}$$

Here each chiral component is a single real quaternion, as in the two-quaternion formulations of Hestenes [15] and De Leo and Rodrigues [17], and the inverse Compton wavelength $\kappa = mc/\hbar$ enters both lines with the same positive sign. De Leo and Rodrigues [17] write the Dirac equation with two quaternions in the standard basis; in the chiral basis used here each quaternion is one chiral component. Appendix A 2 proves that every first-order pair that factors the Klein–Gordon operator on $\mathbb{H}$ and has translation invariance, spatial rotation covariance under $\psi \mapsto \mathcal{R}\psi$, and $U(1)$ phase invariance under $\psi \mapsto \psi\exp(e_3\phi)$ has, up to a phase-preserving change of variables, the form of Eqs. (54) and (55) once each derivative is replaced with its covariant derivative. It also proves that these symmetries forbid a nonzero mass term on a single phase-invariant quaternion [59].

Right multiplication by e_3 generates $U(1)$ phase rotations. Right multiplication by e_2, which anticommutes with e_3 ($e_2e_3 = -e_3e_2$), implements charge conjugation, $\psi \mapsto \psi e_2$, the quaternionic form of $-i\sigma_2\psi^*$. It reverses the signs of q and V in Eqs. (29) and (30) and exchanges the chiral components via $(\varphi_R, \varphi_L) \mapsto (\varphi_L e_2, -\varphi_R e_2)$ (Appendix A 3). For a neutral particle ($q = 0$, $V = 0$), imposing the Majorana reality condition $\varphi_L = \varphi_R e_1$ (with $e_1 = e_2e_3$) makes Eqs. (54) and (55) coincide and reduces the pair to a single first-order equation for one quaternion φ,

$$(D_n\,\varphi)\,e_3 + W^{-1/2}\,\mathbf{e}\cdot\nabla\big(W^{-1/2}\varphi\big) = \kappa\,\varphi\,e_1, \tag{56}$$

the curved-spacetime Majorana equation [60, 61], which describes a particle that is its own antiparticle. Its right factor e_1 anticommutes with e_3 and breaks the $U(1)$ phase symmetry.

B. Comparison with the Obukhov–Silenko–Teryaev Hamiltonian

Equations (54) and (55) reproduce the exact curved-spacetime Dirac Hamiltonian of Obukhov *et al.* [10, 11] for the metric (35). That Hamiltonian is Hermitian in the flat inner product $\int \Psi^\dagger\Psi\, d^3x$, so it acts on $W^{3/2}$ times the bispinor in the local orthonormal axes. By the definition of the rescaled spinor in Sec. II D, its chiral components are therefore $\alpha_n^{-1/2}\varphi_R$ and $\alpha_n^{-1/2}\varphi_L$, and the conserved inner product for each of φ_R and φ_L is $\langle\phi,\chi\rangle = \int \phi^\dagger\chi\,\alpha_n^{-1}\,d^3x$. The spatial kinetic term of that Hamiltonian is $\frac{1}{2}c\{\alpha_n/W, \boldsymbol{\alpha}\cdot\hat{\boldsymbol{\pi}}\}$, where $\boldsymbol{\alpha}$ is the vector of Dirac matrices α_i and $\{\hat{A},\hat{B}\} \equiv \hat{A}\hat{B} + \hat{B}\hat{A}$. The shift term of that Hamiltonian is $-\frac{1}{2}(\mathbf{A}_g\cdot\hat{\mathbf{p}} + \hat{\mathbf{p}}\cdot\mathbf{A}_g)$, since that work writes the metric with the shift vector $c\mathbf{K} = -\mathbf{A}_g$. Conjugating the kinetic term and the negative of the shift term by $\alpha_n^{1/2}$ gives, for any two-component spinor χ,

$$\begin{aligned}
\alpha_n^{1/2}\,\tfrac{1}{2}\big\{\alpha_n/W, \boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\big\}\big(\alpha_n^{-1/2}\chi\big) &= \alpha_n\,W^{-1/2}\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\big(W^{-1/2}\chi\big),\\
\alpha_n^{1/2}\,\tfrac{1}{2}\big(\mathbf{A}_g\cdot\hat{\mathbf{p}} + \hat{\mathbf{p}}\cdot\mathbf{A}_g\big)\big(\alpha_n^{-1/2}\chi\big) &= \mathbf{A}_g\cdot\hat{\mathbf{p}}\,\chi - \frac{i\hbar}{2}\,(\nabla\cdot\mathbf{A}_g)\,\chi + \frac{i\hbar}{2}\,(\mathbf{A}_g\cdot\nabla\ln\alpha_n)\,\chi.
\end{aligned}$$

In the first line, the two orderings produce the lapse-derivative terms $\pm\frac{i\hbar}{4}\,\alpha_n W^{-1}(\boldsymbol{\sigma}\cdot\nabla\ln\alpha_n)\,\chi$, which cancel. For any other power of α_n they do not cancel. Divided by $\hbar\alpha_n$, the right side of the first line is X_W of Eq. (49) in Pauli form, through Eq. (33). Divided by $i\hbar c\alpha_n$, the second line gives the transport, divergence and lapse terms of D_n in Eq. (39). The mass term of the Hamiltonian, $mc^2\alpha_n$ times the Dirac matrix β, gives κ after division by $\hbar c\alpha_n$. The spin-rotation term of the Hamiltonian is $-\frac{1}{4}\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g$ on each chiral component and gives the rotation term of D_n. In the inner product $\langle\phi,\chi\rangle$ the operator $\alpha_n\mathsf{X}_W$ is Hermitian, since $\alpha_n^{-1}\alpha_n = 1$ and $W^{-1/2}\,\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\,W^{-1/2}$ is Hermitian in $\int\phi^\dagger\chi\,d^3x$; without the weight α_n^{-1} it is not.

C. The second-order quaternionic wave equation and the 4×4 conjugate table

Applying $\mathsf{N} - \mathsf{X}_W$ to both sides of Eq. (54) and substituting Eq. (55) on the right eliminates φ_L and yields, with $\psi \equiv \varphi_R$,

$$\boxed{(\mathsf{N} - \mathsf{X}_W)(\mathsf{N} + \mathsf{X}_W)\,\psi = \kappa^2\,\psi}\,. \tag{57}$$

(The quaternionic wave equation)

Here ψ is one quaternion, the four real components of a two-component spinor (Eq. (21)), and $\kappa = mc/\hbar$ (Eq. (44)). The operator $\mathsf{N}\psi = (D_n\psi)\,e_3$ (Eq. (48)) contains the potential V in D_t (Eq. (29)) and the lapse and shift in D_n (Eq. (39)). The operator X_W contains the vector potential $\mathbf{A}$ in $\mathbf{D}$ (Eq. (30)) and the scale factor W (Eq. (49)). In flat spacetime $\mathsf{X}_W = \mathsf{X}$ and Eq. (57) reads $(\mathsf{N} - \mathsf{X})(\mathsf{N} + \mathsf{X})\psi = \kappa^2\psi$. In the composition $(\mathsf{N} - \mathsf{X}_W)(\mathsf{N} + \mathsf{X}_W)$, the operator $\mathsf{N} + \mathsf{X}_W$ acts first. Inside these operators, $e_k\,D_k\psi$ and $(D_n\psi)\,e_3$ are quaternion products whose factor order matters (Eq. (19)). Conversely, let ψ be any solution of Eq. (57) for a particle of nonzero mass ($\kappa > 0$), and define $\varphi_R \equiv \psi$ and $\varphi_L \equiv \kappa^{-1}(\mathsf{N} + \mathsf{X}_W)\psi$. This pair satisfies Eq. (54) by definition, and it satisfies $(\mathsf{N} - \mathsf{X}_W)\varphi_L = \kappa^{-1}(\mathsf{N}-\mathsf{X}_W)(\mathsf{N}+\mathsf{X}_W)\psi = \kappa\psi$, which is Eq. (55). Thus, for a constant mass $m > 0$, the single-quaternion second-order wave equation (57) and the four-component Dirac equation are in exact one-to-one correspondence. When κ depends on position, as it does with a Lorentz-scalar potential, the same elimination gives

$$(\mathsf{N} - \mathsf{X}_W)\big[\kappa^{-1}(\mathsf{N} + \mathsf{X}_W)\psi\big] = \kappa\,\psi,$$

and $\varphi_L = \kappa^{-1}(\mathsf{N}+\mathsf{X}_W)\psi$ again. Equation (57) is second order in time, so its initial data are ψ and $\partial_t\psi$; at one instant they give the pair (φ_R, φ_L) and are given by it. The conserved density of the Dirac equation is $(|\varphi_R|^2 + |\varphi_L|^2)/\alpha_n$,

with the weight α_n^{-1} of Section II D. For a solution of Eq. (57) it is therefore $(|\psi|^2 + \kappa^{-2}|(\mathsf{N}+\mathsf{X}_W)\psi|^2)/\alpha_n$, and the part contributed by φ_R is $|\psi|^2/\alpha_n$, the density $\rho = |\psi|^2$ of Section II C divided by α_n.

Multiplying out Eq. (57) with $\mathsf{N}^2 = -D_n^2$ separates the wave equation into its Klein–Gordon part and its first-order commutator part:

$$D_n^2\,\psi + \mathsf{X}_W^2\,\psi + \kappa^2\,\psi = [\mathsf{N}, \mathsf{X}_W]\,\psi. \tag{58}$$

The left side is Eq. (47) with each derivative replaced by its minimally coupled form. Applying $\mathsf{N}+\mathsf{X}_W$ to Eq. (55) in the same way gives $(\mathsf{N}+\mathsf{X}_W)(\mathsf{N}-\mathsf{X}_W)\,\varphi_L = \kappa^2\varphi_L$, so φ_L obeys Eq. (58) with $-[\mathsf{N}, \mathsf{X}_W]$ on the right: the two chiral components share the left side, and the commutator enters them with opposite signs.

In the spatial square $(\mathbf{e}\cdot\mathbf{D})^2\psi = \sum_{j,k} e_j e_k\, D_j D_k \psi$, the 3×3 table of Eq. (45) puts $-\mathbf{D}^2\psi$ on the diagonal. The off-diagonal pairs no longer cancel, because the components of $\mathbf{D}$ need not commute: the pair $(1,2)$ gives $e_1e_2\,D_1D_2\psi + e_2e_1\,D_2D_1\psi = e_3\,[D_1, D_2]\,\psi$, and likewise for $(2,3)$ and $(3,1)$:

$$(\mathbf{e}\cdot\mathbf{D})^2\,\psi = -\mathbf{D}^2\psi + e_1\,[D_2, D_3]\,\psi + e_2\,[D_3, D_1]\,\psi + e_3\,[D_1, D_2]\,\psi. \tag{59}$$

Expanding $D_1(D_2\psi)$ by Eq. (30) and the product rule, with $e_3^2 = -1$, gives

$$D_1D_2\,\psi = \partial_1\partial_2\psi + \frac{q}{\hbar}\,(\partial_1 A_2)\,\psi\,e_3 + \frac{q}{\hbar}\,A_2\,(\partial_1\psi)\,e_3 + \frac{q}{\hbar}\,A_1\,(\partial_2\psi)\,e_3 - \frac{q^2}{\hbar^2}\,A_1A_2\,\psi. \tag{60}$$

Exchanging the indices 1 and 2 gives $D_2D_1\psi$: the first and last terms are unchanged, the third and fourth exchange with each other, and only the second changes, into $(q/\hbar)(\partial_2 A_1)\,\psi\,e_3$. Subtracting $D_2D_1\psi$ therefore leaves the magnetic curvature

$$[D_1, D_2]\,\psi = \frac{q}{\hbar}\,B_3\,\psi\,e_3, \tag{61}$$

where $B_3 = \partial_1 A_2 - \partial_2 A_1 = (\nabla\times\mathbf{A})_3$. Cyclic permutation gives $[D_2, D_3]\psi = (q/\hbar)B_1\psi e_3$ and $[D_3, D_1]\psi = (q/\hbar)B_2\psi e_3$. Substituting these into Eq. (59) gives the magnetic Fueter-square identity:

$$(\mathbf{e}\cdot\mathbf{D})^2\,\psi = -\mathbf{D}^2\psi + \frac{q}{\hbar}\,(\mathbf{e}\cdot\mathbf{B})\,\psi\,e_3. \tag{62}$$

The magnetic field $\mathbf{e}\cdot\mathbf{B} = B_1e_1 + B_2e_2 + B_3e_3$ multiplies ψ from the left and rotates the spatial spin frame, while the phase generator e_3 multiplies ψ from the right, as in the internal covariant derivative commutator of Lasenby *et al.* [18].

By Eq. (48), the spatial square is $\mathsf{X}_W^2\psi = W^{-1/2}\mathsf{X}\big(W^{-1}\mathsf{X}(W^{-1/2}\psi)\big)$, with a position-dependent factor between the two X. For any positive function $g(\mathbf{x})$ and any quaternion ϕ, the product rule gives $\mathsf{X}(g\,\mathsf{X}\phi) = \sum_{j,k} e_j e_k\,(\partial_j g)\,D_k\phi + g\,\mathsf{X}^2\phi$, since the D_j commute with left multiplication by e_k. By Eq. (45), the first term is $-\nabla g\cdot\mathbf{D}\phi + \mathbf{e}\cdot(\nabla g\times\mathbf{D})\phi$, where $(\nabla g\times\mathbf{D})_l \equiv \epsilon_{ljk}(\partial_j g)\,D_k$. With Eq. (62) for X^2,

$$\mathsf{X}\big(g\,\mathsf{X}\phi\big) = g\Big[-\mathbf{D}^2\phi + \frac{q}{\hbar}\,(\mathbf{e}\cdot\mathbf{B})\,\phi\,e_3\Big] - \nabla g\cdot\mathbf{D}\,\phi + \mathbf{e}\cdot(\nabla g\times\mathbf{D})\,\phi. \tag{63}$$

With $g = W^{-1}$ and $\phi = W^{-1/2}\psi$, Eq. (63) gives $W^{1/2}\mathsf{X}_W^2\psi$. Its last term is the *spatial-curvature spin coupling*: like the magnetic term, it acts on ψ through the left units e_l, here with the vector operator $\nabla W^{-1}\times\mathbf{D}$.

The time-space commutator $[D_t, D_k]\psi$ follows from Eq. (60) by replacing $\partial_1 \to \partial_t$ and $qA_1 \to -V$:

$$[D_t, D_k]\,\psi = \frac{1}{\hbar}\,\big(\partial_k V + q\,\partial_t A_k\big)\,\psi\,e_3, \tag{64}$$

which for $V = q\Phi$ is $-(q/\hbar)E_k\,\psi\,e_3$ with $E_k = -\partial_k\Phi - \partial_t A_k$. In flat spacetime, where $\mathsf{N}\psi = c^{-1}(D_t\psi)e_3$, the left factors e_k of X pass through D_t and the right factor e_3 of N passes through D_k, so $[\mathsf{N}, \mathsf{X}]\psi = c^{-1}\sum_k e_k([D_t, D_k]\psi)e_3$, and $e_3^2 = -1$ yields

$$\begin{aligned}[\mathsf{N}, \mathsf{X}]\,\psi &= \frac{q}{\hbar c}\,(\mathbf{e}\cdot\mathbf{E})\,\psi \\ &= -i\,\frac{q}{\hbar c}\,\boldsymbol{\sigma}\cdot\mathbf{E}\,\psi,\end{aligned} \tag{65}$$

the second line by Eq. (25). Written in Pauli matrices via Eq. (24), Eq. (57) in flat spacetime is the two-component second-order Dirac equation of Feynman and Gell-Mann [19]. The field term of Eq. (62) is its magnetic term $\boldsymbol{\sigma}\cdot\mathbf{B}$,

and the commutator of Eq. (65) is its electric term $\boldsymbol{\sigma}\cdot\mathbf{E}$, the one term that minimal coupling in Eq. (47) does not produce.

In Eq. (58) the magnetic commutators $[D_j, D_k]$ are part of X_W^2, while $[\mathsf{N}, \mathsf{X}_W]$ appears separately on the right, because the three spatial derivatives have the three anticommuting units e_1, e_2, e_3 as coefficients, whereas N has the unit 1, which commutes with every e_k. No fourth unit could take its place: any element of an associative algebra that anticommutes with e_2 and e_3 commutes with their product $e_1 = e_2 e_3$. Instead, as in the squared norm $\bar{Y}Y = (Y_0 - \mathbf{Y}\cdot\mathbf{e})(Y_0 + \mathbf{Y}\cdot\mathbf{e})$ of a quaternion $Y = Y_0 + \mathbf{Y}\cdot\mathbf{e}$, the left factor of Eq. (57) reverses the sign of the three vector units. To collect the products by unit, each operator must commute with left multiplication by a constant unit. In N the only term that does not is the rotation term of D_n (Eq. (39)), which with $\mathbf{b} \equiv \mathbf{B}_g/(4c\alpha_n)$ is $(\mathbf{e}\cdot\mathbf{b})\psi$. Let $\mathring{D}_n$ be D_n without this term, $\mathring{\mathsf{N}}\psi \equiv (\mathring{D}_n\psi)\, e_3$, and $\tilde{D}_k \equiv W^{-1}(D_k - \frac{1}{2}\partial_k \ln W)$, so that $\mathsf{N}\psi = \mathring{\mathsf{N}}\psi + \sum_k e_k\, b_k\, \psi\, e_3$ and, by Eq. (49), $\mathsf{X}_W\psi = \sum_k e_k\, \tilde{D}_k\psi$. With the quaternion basis $e_\mu = (1, e_1, e_2, e_3)$, its conjugate $\bar{e}_\mu = (1, -e_1, -e_2, -e_3)$, and the operators

$$\begin{aligned}
\mathsf{N}_0^\pm\, \psi &\equiv \mathring{\mathsf{N}}\psi, \\
\mathsf{N}_k^\pm\, \psi &\equiv \tilde{D}_k\psi \pm b_k\, \psi\, e_3 \qquad (k = 1, 2, 3),
\end{aligned}$$

the two factors of Eq. (57) are $\mathsf{N} + \mathsf{X}_W = \sum_{\mu=0}^3 e_\mu \mathsf{N}_\mu^+$ and $\mathsf{N} - \mathsf{X}_W = \sum_{\mu=0}^3 \bar{e}_\mu \mathsf{N}_\mu^-$. Each $\mathsf{N}_\mu^\pm$ commutes with left multiplication by a constant unit, since D_k does (Section II D), the remaining terms of $\mathring{D}_n$ and $\tilde{D}_k$ are scalar, and b_k multiplies $\psi\, e_3$ from the left by a real number. The products of units in the composition of the two factors are therefore the entries of the 4×4 conjugate quaternion product table:

$$(\bar{e}_\mu\, e_\nu) = \begin{pmatrix} 1 & e_1 & e_2 & e_3 \\ -e_1 & 1 & -e_3 & e_2 \\ -e_2 & e_3 & 1 & -e_1 \\ -e_3 & -e_2 & e_1 & 1 \end{pmatrix}. \tag{66}$$

The four diagonal entries are $\bar{e}_\mu e_\mu = 1$, while the twelve off-diagonal entries are antisymmetric ($\bar{e}_\nu e_\mu = -\bar{e}_\mu e_\nu$), with $\bar{e}_0 e_k = e_k$ on the $0k$ border and $\bar{e}_j e_k = -\epsilon_{jkl} e_l$ in the lower-right 3×3 block. Pairing each upper-triangular entry (μ, ν) with its transpose (ν, μ) gives

$$(\mathsf{N} - \mathsf{X}_W)(\mathsf{N} + \mathsf{X}_W)\, \psi = \sum_{\mu=0}^3 \mathsf{N}_\mu^- \mathsf{N}_\mu^+\, \psi + \sum_{0 \le \mu < \nu \le 3} \bar{e}_\mu\, e_\nu \left(\mathsf{N}_\mu^- \mathsf{N}_\nu^+ - \mathsf{N}_\nu^- \mathsf{N}_\mu^+\right) \psi. \tag{67}$$

The three kinds of entry follow from the product rule, $e_3^2 = -1$ and $\partial_t \mathbf{b} = \mathbf{0}$. On the diagonal, $\mathsf{N}_0^- \mathsf{N}_0^+ \psi = -\mathring{D}_n^2 \psi$ and $\mathsf{N}_k^- \mathsf{N}_k^+ \psi = \tilde{D}_k^2 \psi + W^{-1}(\partial_k b_k)\psi e_3 + b_k^2 \psi$, so

$$\sum_{\mu=0}^3 \mathsf{N}_\mu^- \mathsf{N}_\mu^+\, \psi = -\mathring{D}_n^2 \psi + \sum_{k=1}^3 \tilde{D}_k^2 \psi + |\mathbf{b}|^2\, \psi + W^{-1}(\nabla\cdot\mathbf{b})\, \psi\, e_3.$$

The border entry $0k$ is

$$\mathsf{N}_0^- \mathsf{N}_k^+ \psi - \mathsf{N}_k^- \mathsf{N}_0^+ \psi = [\mathring{\mathsf{N}}, \tilde{D}_k]\, \psi - 2\, b_k\, \mathring{D}_n \psi + (c\alpha_n)^{-1}(\mathbf{A}_g \cdot \nabla b_k)\, \psi.$$

The interior entry jk is

$$\mathsf{N}_j^- \mathsf{N}_k^+ \psi - \mathsf{N}_k^- \mathsf{N}_j^+ \psi = [\tilde{D}_j, \tilde{D}_k]\, \psi + \left[W^{-1}(\partial_j b_k - \partial_k b_j)\, \psi + 2\, b_k\, \tilde{D}_j \psi - 2\, b_j\, \tilde{D}_k \psi\right] e_3,$$

whose first term, by the product rule and the symmetry of $\partial_j \partial_k \ln W$, is

$$[\tilde{D}_j, \tilde{D}_k]\, \psi = W^{-2}\, [D_j, D_k]\, \psi + W^{-1}\big(\partial_k \ln W\, \tilde{D}_j \psi - \partial_j \ln W\, \tilde{D}_k \psi\big).$$

Contracted with $\bar{e}_j e_k = -\epsilon_{jkl} e_l$, and with $\tilde{\mathbf{D}} \equiv (\tilde{D}_1, \tilde{D}_2, \tilde{D}_3)$, the interior entries give three spin couplings: the magnetic term $-W^{-2}(q/\hbar)(\mathbf{e}\cdot\mathbf{B})\psi e_3$ of Eq. (61); the spatial-curvature term $W^{-1}\mathbf{e}\cdot(\nabla \ln W \times \tilde{\mathbf{D}})\psi$, which matches the last term of Eq. (63) in $-\mathsf{X}_W^2$; and the gravitomagnetic terms $-W^{-1}\big(\mathbf{e}\cdot(\nabla\times\mathbf{b})\big)\psi e_3 + 2\big(\mathbf{e}\cdot(\mathbf{b}\times\tilde{\mathbf{D}})\psi\big)e_3$. The border entries, contracted with e_k, give the commutator $[\mathring{\mathsf{N}}, \mathsf{X}_W]$ and the gravitomagnetic terms $-2(\mathbf{e}\cdot\mathbf{b})\mathring{D}_n\psi + (c\alpha_n)^{-1}\sum_k e_k(\mathbf{A}_g\cdot\nabla b_k)\psi$.

When $\mathbf{B}_g = \mathbf{0}$ and $W = 1$, $\mathbf{b} = \mathbf{0}$, $\mathring{\mathsf{N}} = \mathsf{N}$ and $\tilde{D}_k = D_k$, so $\mathsf{N}^+_\mu = \mathsf{N}^-_\mu$ and every off-diagonal pair is a commutator:

$$(\mathsf{N} - \mathsf{X})(\mathsf{N} + \mathsf{X})\,\psi = \big(-D_n^2 + \mathbf{D}^2\big)\,\psi + \sum_{k=1}^{3} e_k\,[\mathsf{N}, D_k]\,\psi - \sum_{1\le j<k\le 3} \epsilon_{jkl}\,e_l\,[D_j, D_k]\,\psi. \tag{68}$$

The four diagonal 1's produce the minimally coupled Klein–Gordon operator $-D_n^2 + \mathbf{D}^2$, the three $0k$ border pairs produce the electric and gravitational commutator $[\mathsf{N}, \mathsf{X}]\psi$, and the three jk interior pairs produce the magnetic spin coupling $-(q/\hbar)(\mathbf{e}\cdot\mathbf{B})\psi e_3$ of Eq. (62). In flat spacetime the border pairs give the electric term of Eq. (65). The commutator $[\mathsf{N}, \mathsf{X}_W]$ is evaluated directly for a shift in Appendix A 4 and for a static lapse in Appendix A 5.

D. Maxwell's equations as one biquaternion equation

The border and interior entries of the table contain the electric and magnetic fields as the commutators of Eqs. (61) and (64). Two of Maxwell's equations follow from these commutators alone. Let the index a be t or $1, 2, 3$. Since D_a commutes with right multiplication by e_3, for a real function f the commutator of D_a with the operator $\psi \mapsto f\,\psi\,e_3$ is $\psi \mapsto (\partial_a f)\,\psi\,e_3$: the terms in the derivatives of ψ and in the potential cancel. Each of the commutators of Eqs. (61) and (64) is an operator of this form, with $f = qB_3/\hbar$ in the first and, for $V = q\Phi$, $f = -qE_k/\hbar$ in the second. Any three operators obey the Jacobi identity

$$[D_a, [D_b, D_c]] + [D_b, [D_c, D_a]] + [D_c, [D_a, D_b]] = 0, \tag{69}$$

since the twelve products of its expansion cancel in pairs. For $a, b, c = 1, 2, 3$ its left side applied to ψ is $(q/\hbar)(\nabla\cdot\mathbf{B})\,\psi\,e_3$. For $a, b, c = t, 1, 2$ it is $(q/\hbar)\big[\partial_t B_3 + (\nabla \times \mathbf{E})_3\big]\,\psi\,e_3$, and $t, 2, 3$ and $t, 3, 1$ give the other two components. Since $\psi\,e_3$ is nonzero wherever ψ is, for $q \neq 0$

$$\nabla \cdot \mathbf{B} = 0, \tag{70}$$

$$\nabla \times \mathbf{E} + \partial_t \mathbf{B} = \mathbf{0}, \tag{71}$$

the magnetic Gauss law and Faraday's law. They hold for any potentials Φ and $\mathbf{A}$: together they are the *Bianchi identity* of the connection of Eqs. (29) and (30).

The other two equations, taken here in flat spacetime ($W = \alpha_n = 1$, $\mathbf{A}_g = \mathbf{0}$), contain the charge density ρ_q and the current density $\mathbf{J}_q$. The vacuum permittivity is ε_0, and the vacuum permeability is $\mu_0 = 1/(\varepsilon_0 c^2)$. Let i be a second imaginary unit, with $\mathsf{i}^2 = -1$, that commutes with e_1, e_2, and e_3, and form the biquaternion

$$\mathfrak{F} = \mathbf{E}\cdot\mathbf{e} + \mathsf{i}\,c\,\mathbf{B}\cdot\mathbf{e}. \tag{72}$$

It is $\sqrt{2/\varepsilon_0}$ times the vacuum *Riemann–Silberstein vector* $\sqrt{\varepsilon_0/2}\,\mathbf{E} + \mathsf{i}\,\mathbf{B}/\sqrt{2\mu_0}$ written as a quaternion, since $\sqrt{2/\varepsilon_0}/\sqrt{2\mu_0} = c$ [25]. For a real vector field $\mathbf{w}$, Eq. (19) gives $(\mathbf{e}\cdot\nabla)(\mathbf{w}\cdot\mathbf{e}) = -\nabla\cdot\mathbf{w} + (\nabla\times\mathbf{w})\cdot\mathbf{e}$: the scalar part is minus the divergence and the vector part is the curl. Applied to $\mathfrak{F}$, with $\mathsf{i}^2 = -1$, the operator $\mathbf{e}\cdot\nabla - (\mathsf{i}/c)\,\partial_t$ therefore gives the scalar part $-\nabla\cdot\mathbf{E} - \mathsf{i}\,c\,\nabla\cdot\mathbf{B}$ and the vector part $(\nabla\times\mathbf{E} + \partial_t\mathbf{B})\cdot\mathbf{e} + \mathsf{i}\,c\,(\nabla\times\mathbf{B} - c^{-2}\,\partial_t\mathbf{E})\cdot\mathbf{e}$. Gauss's law $\nabla\cdot\mathbf{E} = \rho_q/\varepsilon_0$, the Ampère–Maxwell law $\nabla\times\mathbf{B} - c^{-2}\,\partial_t\mathbf{E} = \mu_0\mathbf{J}_q$, and Eqs. (70) and (71) reduce the two parts to $-\rho_q/\varepsilon_0$ and $\mathsf{i}\,\mu_0 c\,\mathbf{J}_q\cdot\mathbf{e}$:

$$\Big(\mathbf{e}\cdot\nabla - \frac{\mathsf{i}}{c}\,\partial_t\Big)\,\mathfrak{F} = -\frac{\rho_q}{\varepsilon_0} + \mathsf{i}\,\mu_0 c\,\mathbf{J}_q\cdot\mathbf{e}. \tag{73}$$

Conversely, since $\mathbf{E}$, $\mathbf{B}$, ρ_q, and $\mathbf{J}_q$ are real, the real and i components of the scalar part are Gauss's law and Eq. (70). Those of the vector part are Eq. (71) and the Ampère–Maxwell law. All four of Maxwell's equations are thus the four parts of one biquaternion equation. Silberstein wrote the two curl equations of the vacuum as one equation for the complex vector $\mathbf{E} + \mathsf{i}\mathbf{B}$ in units with $c = 1$, with the divergence of that vector zero [62], and the four equations with sources as one quaternion equation [24].

For a free particle of zero mass in flat spacetime, $\mathsf{N}\psi = c^{-1}(\partial_t\psi)\,e_3$ and $\mathsf{X} = \mathbf{e}\cdot\nabla$ (Eq. (48)). Since i is right multiplication by $-e_3$, $\mathsf{N}\psi = -(i/c)\,\partial_t\psi$, and the operator of Eq. (54) is $\mathsf{N} + \mathsf{X} = \mathbf{e}\cdot\nabla - (i/c)\,\partial_t$. This is the operator of Eq. (73) with i in place of i. Both units commute with left multiplication by e_k, so the Weyl equation $(\mathsf{N} + \mathsf{X})\,\varphi_R = 0$ and Maxwell's equations in vacuum differ only in what the operator acts on: a spinor, whose phase is a right multiplication, or the field $\mathfrak{F}$. In matrix form the Weyl equation has the Pauli matrices where Maxwell's equations have the spin-1 matrices [25].

The conjugate operator $\mathbf{e}\cdot\nabla + (\mathsf{i}/c)\,\partial_t$ commutes with that of Eq. (73), and by Eq. (19) the product of the two is $-\nabla^2 + c^{-2}\partial_t^2$. Applying the conjugate operator to Eq. (73) gives $(-\nabla^2 + c^{-2}\partial_t^2)\,\mathfrak{F}$ on the left, which has no scalar part, and the scalar part $-\mathsf{i}\,\mu_0 c\,(\partial_t\rho_q + \nabla\cdot\mathbf{J}_q)$ on the right. The equation is therefore consistent only if

$$\partial_t\rho_q + \nabla\cdot\mathbf{J}_q = 0. \tag{74}$$

In vacuum each component of $\mathfrak{F}$ obeys Eq. (43) with $\kappa = 0$, the wave equation with speed c. In the Lorenz gauge, $\nabla\cdot\mathbf{A} + c^{-2}\partial_t\Phi = 0$, the conjugate operator applied to the potential biquaternion $-\Phi + \mathsf{i}\,c\,\mathbf{A}\cdot\mathbf{e}$ gives $\mathfrak{F}$, so Eq. (73) becomes $(-\nabla^2 + c^{-2}\partial_t^2)(-\Phi + \mathsf{i}\,c\,\mathbf{A}\cdot\mathbf{e}) = -\rho_q/\varepsilon_0 + \mathsf{i}\,\mu_0 c\,\mathbf{J}_q\cdot\mathbf{e}$. Its operator, like that of Eq. (57), is a product of two first-order operators, here acting on the potentials of Eqs. (29) and (30).

Equations (70) and (71) are identities of the covariant derivatives. Gauss's law and the Ampère–Maxwell law are the field's own equations, and the Dirac equation does not produce them. The charge density and current of a Dirac particle are q times its conserved density (Section III C) and q times its current $c\,\Psi^\dagger\boldsymbol{\alpha}\Psi$, and they satisfy Eq. (74), which Eq. (73) requires of its sources ρ_q and $\mathbf{J}_q$.

IV. THE EXACT MASTER FREQUENCY EQUATION

A. Stationary states and the shift and potential operators

Consider a stationary state of angular frequency ω, $\psi(\mathbf{x},t) = \psi_0(\mathbf{x})\exp(e_3\,\omega t)$, which corresponds to the standard energy eigenstate $\psi_0(\mathbf{x})e^{-i\omega t}$. Differentiating with respect to t gives

$$\partial_t\psi = \psi\, e_3\,\omega, \tag{75}$$

while by Eq. (31) every spatial covariant derivative acts as

$$\begin{aligned} D_k\,\psi &= \frac{i}{\hbar}\,\hat\pi_k\,\psi \\ &= -\frac{1}{\hbar}\,(\hat\pi_k\,\psi)\,e_3. \end{aligned} \tag{76}$$

In the spatial square of Eq. (62), $D_k^2\psi = -\hat\pi_k^2\psi/\hbar^2$ and $(q/\hbar)B_k e_k\psi e_3 = -(q/\hbar)B_k\sigma_k\psi$ by Eq. (24). Substituting these gives the Dirac–Pauli momentum square [4]

$$\begin{aligned} (\mathbf{e}\cdot\mathbf{D})^2\,\psi &= \frac{1}{\hbar^2}\,(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2\,\psi \\ &= \frac{1}{\hbar^2}\,\big(\hat{\boldsymbol{\pi}}^2 - q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}\big)\,\psi, \end{aligned} \tag{77}$$

whose spin term divided by $2m$ is $-q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}/(2m) = -g_s(q/2m)\mathbf{S}\cdot\mathbf{B}$ with the Dirac spin g-factor $g_s = 2$.

In the normal derivative $D_n\psi$ of Eq. (39) on a stationary state, $D_t\psi = ((\hbar\omega - V)\psi)e_3/\hbar$, and the Lie transport term is $-\mathbf{A}_g\cdot\mathbf{D}\psi = ((\mathbf{A}_g\cdot\hat{\boldsymbol{\pi}})\psi)e_3/\hbar$. The divergence term is $-\frac{1}{2}(\nabla\cdot\mathbf{A}_g)\psi = -(\frac{i\hbar}{2}(\nabla\cdot\mathbf{A}_g)\psi)e_3/\hbar$. The rotation term is $+\frac{1}{4}(\mathbf{e}\cdot\mathbf{B}_g)\psi = (\frac{1}{4}\boldsymbol{\sigma}\cdot\mathbf{B}_g\psi)e_3$. The lapse term is $+\frac{1}{2}(\mathbf{A}_g\cdot\nabla\ln\alpha_n)\psi = (\frac{i\hbar}{2}(\mathbf{A}_g\cdot\nabla\ln\alpha_n)\psi)e_3/\hbar$. Since $\mathbf{A}_g\cdot\hat{\boldsymbol{\pi}} - \frac{i\hbar}{2}(\nabla\cdot\mathbf{A}_g) = \frac{1}{2}\{\mathbf{A}_g,\hat{\boldsymbol{\pi}}\}$ and $\frac{1}{2}\{\mathbf{A}_g,\hat{\boldsymbol{\pi}}\}(\alpha_n^{-1/2}\phi) = \alpha_n^{-1/2}[\frac{1}{2}\{\mathbf{A}_g,\hat{\boldsymbol{\pi}}\} + \frac{i\hbar}{2}(\mathbf{A}_g\cdot\nabla\ln\alpha_n)]\phi$, the shift enters through the *shift coupling*

$$\hat\Pi_g \equiv \alpha_n^{1/2}\,\tfrac{1}{2}\{\mathbf{A}_g,\hat{\boldsymbol{\pi}}\}\,\alpha_n^{-1/2}, \tag{78}$$

and defining the *potential-energy operator*

$$\hat V \equiv V - \hat\Pi_g - \tfrac{1}{4}\,\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g, \tag{79}$$

puts the normal derivative on any stationary state into the exact form

$$\begin{aligned} D_n\,\psi &= \frac{1}{\hbar c\alpha_n}\,\big((\hbar\,\omega - \hat V)\,\psi\big)\,e_3, \\ \mathsf{N}\,\psi &= -\frac{1}{\hbar c\alpha_n}\,(\hbar\,\omega - \hat V)\,\psi. \end{aligned} \tag{80}$$

Define the scaled spatial momentum operator $\mathsf{P} \equiv \alpha_n \hbar \mathsf{X}_W = \alpha_n\, W^{-1/2}(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})W^{-1/2}$. Multiplying the chiral Weyl pair of Eqs. (54) and (55) by $-\hbar c\alpha_n$ and forming the Dirac-basis *large component* $u \equiv (\varphi_R - \varphi_L)/\sqrt{2}$ and *small component* $w \equiv (\varphi_R + \varphi_L)/\sqrt{2}$ yields the exact stationary block equation

$$\hbar\,\omega \begin{pmatrix} u \\ w \end{pmatrix} = \mathsf{H} \begin{pmatrix} u \\ w \end{pmatrix}, \qquad \mathsf{H} \equiv \begin{pmatrix} \hat{V} + \alpha_n mc^2 & c\,\mathsf{P} \\ c\,\mathsf{P} & \hat{V} - \alpha_n mc^2 \end{pmatrix}, \tag{81}$$

where H is Hermitian in the conserved weighted inner product $\langle \phi, \chi\rangle = \int \phi^\dagger \chi\, \alpha_n^{-1}\, d^3x$.

B. Exact decoupling via the operator Riccati equation and contraction mapping proof

Suppose that on the positive-frequency subspace of H the small component is related to the large component by a state-independent operator R, $w = \mathsf{R}u$. Applying R to the first row of Eq. (81) gives $\hbar\omega\,\mathsf{R}u = \mathsf{R}(\hat{V}+\alpha_n mc^2)u + c\,\mathsf{RPR}u$, and the second row is $\hbar\omega\,\mathsf{R}u = c\mathsf{P}u + (\hat{V} - \alpha_n mc^2)\mathsf{R}u$. Write $\alpha_n mc^2 = mc^2 + \mathsf{G}$, where $\mathsf{G} \equiv mc^2(\alpha_n - 1)$ is the *lapse energy*. Equating the two expressions for $\hbar\omega\,\mathsf{R}u$ shows that R satisfies the exact operator Riccati equation [12, 13]:

$$2mc^2\,\mathsf{R} = c\,\mathsf{P} - \{\mathsf{G}, \mathsf{R}\} - [\mathsf{R}, \hat{V}] - c\,\mathsf{RPR}. \tag{82}$$

When a Lorentz-scalar potential $V_{\rm sc}(\mathbf{x}) = [m(\mathbf{x}) - m_0]c^2$ is present, $\alpha_n m(\mathbf{x})c^2 = m_0c^2 + \mathsf{G} + \alpha_n V_{\rm sc}$, so m in Eq. (82) is m_0 and $\mathsf{G} + \alpha_n V_{\rm sc}$ takes the place of G; the same replacement holds in everything that follows from Eq. (82). Because the norm of a state $(u, \mathsf{R}u)^T$ is $\langle u, u\rangle + \langle \mathsf{R}u, \mathsf{R}u\rangle = \langle u, \mathsf{W}u\rangle$ with $\mathsf{W} \equiv 1 + \mathsf{R}^\dagger\mathsf{R} \geq 1$, the norm-preserving two-component spinor is $\chi \equiv \mathsf{W}^{1/2}u$. Multiplying the first row $\hbar\omega\, u = (\hat{V} + \alpha_n mc^2 + c\mathsf{PR})u$ on the left by $\mathsf{W} = 1 + \mathsf{R}^\dagger\mathsf{R}$, using the Riccati relation $\mathsf{R}(\hat{V} + \alpha_n mc^2 + c\mathsf{PR}) = c\mathsf{P} + (\hat{V} - \alpha_n mc^2)\mathsf{R}$, and conjugating by $\mathsf{W}^{-1/2}$ yields the manifestly Hermitian *exact frequency equation*:

$$\boxed{\hbar\,\omega\,\chi = \hat{h}\,\chi}, \qquad \textit{(The exact frequency equation)} \tag{83}$$

where the *frequency operator* $\hat{h}$ is

$$\hat{h} \equiv \mathsf{W}^{-1/2}\left[\hat{V} + \alpha_n mc^2 + c\,(\mathsf{PR} + \mathsf{R}^\dagger\mathsf{P}) + \mathsf{R}^\dagger\,(\hat{V} - \alpha_n mc^2)\,\mathsf{R}\right]\mathsf{W}^{-1/2} \tag{84}$$

$$= \mathsf{W}^{1/2}\left(\hat{V} + \alpha_n mc^2 + c\,\mathsf{PR}\right)\mathsf{W}^{-1/2}. \tag{85}$$

Equation (84) applies the block-operator construction of Heully *et al.* [12] and Kutzelnigg and Liu [13], used there for an electrostatic potential, to arbitrary static vector potentials, lapse, shift, spatial metric, and Lorentz-scalar potentials, and writes $\hat{h}$ in manifestly Hermitian form.

Suppose first that $c\mathsf{P}$, G, and $\hat{V}$ are bounded operators, as they are on any finite set of basis states, and let $\epsilon \equiv (c\|\mathsf{P}\| + \|\mathsf{G}\| + \|\hat{V}\|)/mc^2 < 1$, where $\|\cdot\|$ is the operator norm. Let $\mathcal{F}(\mathsf{R})$ be the right side of Eq. (82) divided by $2mc^2$, an operator-valued function of the operator R. For $\|\mathsf{R}\| \leq 1$, the triangle inequality and $\|\mathsf{AB}\| \leq \|\mathsf{A}\|\,\|\mathsf{B}\|$ give

$$\|\mathcal{F}(\mathsf{R})\| \leq \frac{c\|\mathsf{P}\| + 2\|\mathsf{G}\|\,\|\mathsf{R}\| + 2\|\hat{V}\|\,\|\mathsf{R}\| + c\|\mathsf{P}\|\,\|\mathsf{R}\|^2}{2mc^2} \leq \epsilon. \tag{86}$$

For two such operators R and $\tilde{\mathsf{R}}$ with difference $\Delta \equiv \mathsf{R} - \tilde{\mathsf{R}}$, the identity $\mathsf{RPR} - \tilde{\mathsf{R}}\mathsf{P}\tilde{\mathsf{R}} = \Delta\,\mathsf{PR} + \tilde{\mathsf{R}}\mathsf{P}\,\Delta$ gives

$$\mathcal{F}(\mathsf{R}) - \mathcal{F}(\tilde{\mathsf{R}}) = -\frac{\{\mathsf{G}, \Delta\} + [\Delta, \hat{V}] + c\,\Delta\,\mathsf{PR} + c\,\tilde{\mathsf{R}}\mathsf{P}\,\Delta}{2mc^2},$$

$$\|\mathcal{F}(\mathsf{R}) - \mathcal{F}(\tilde{\mathsf{R}})\| \leq \frac{2\|\mathsf{G}\| + 2\|\hat{V}\| + 2c\|\mathsf{P}\|}{2mc^2}\,\|\Delta\| = \epsilon\,\|\Delta\|. \tag{87}$$

So $\mathcal{F}$ maps the closed unit ball of operators into itself and contracts distances by the factor $\epsilon < 1$. By the contraction mapping theorem Eq. (82) has exactly one solution with $\|\mathsf{R}\| \le 1$, and its norm is at most ϵ. By Eq. (85) the operator $\hat{h}$ has the eigenvalues of $\hat{V} + \alpha_n mc^2 + c\,\mathsf{PR}$, and these lie within ϵmc^2 of mc^2, so they are positive. In the same way the negative-frequency states have $u = \mathsf{R}_- w$ for an operator R_- with $\|\mathsf{R}_-\| \le \epsilon$, and energies within ϵmc^2 of $-mc^2$. Every pair (u, w) is the sum $(u_+, \mathsf{R}u_+) + (\mathsf{R}_- w_-, w_-)$ of one state of each kind, since eliminating $w_- = w - \mathsf{R}u_+$ leaves $(1 - \mathsf{R}_-\mathsf{R})\,u_+ = u - \mathsf{R}_- w$, which has exactly one solution because $\|\mathsf{R}_-\mathsf{R}\| \le \epsilon^2 < 1$. So the states with $w = \mathsf{R}u$ are all of the positive-frequency subspace. A number V_0 added to $\hat{V}$ drops out of Eq. (82), so $\|\hat{V}\|$ may be replaced in ϵ by $\|\hat{V} - V_0\|$ for any number V_0, and the energies of the two kinds then lie within ϵmc^2 of $V_0 + mc^2$ and of $V_0 - mc^2$. For the unbounded momentum operator of a free particle the same holds with $\mathsf{R} = c\,\boldsymbol{\sigma}\cdot\hat{\mathbf{p}}\left(mc^2 + \sqrt{(mc^2)^2 + c^2\hat{\mathbf{p}}^2}\right)^{-1}$, which at each momentum is a matrix of norm less than 1, although its operator norm, the supremum over momenta, is 1. For a particle in weak fields, in a state built from momenta small compared with mc, the existence of such an operator R is assumed.

Write $\beta \equiv \mathrm{diag}(1, -1)$ and J for the 2×2 exchange matrix, with 1 in both off-diagonal entries, so that $\beta^2 = \mathsf{J}^2 = 1$, $\beta\mathsf{J} = -\mathsf{J}\beta$, and $\mathsf{H} = \hat{V} + \beta\,\alpha_n mc^2 + \mathsf{J}\,c\mathsf{P}$. A 2×2 matrix X of operators is *even* if it commutes with β, that is, if it has no off-diagonal entries, and *odd* if it anticommutes with β, that is, if it has no diagonal entries; its odd part is $\frac{1}{4}[\beta, [\beta, X]]$. Let the *frequency sign* β_ω be 1 on the positive-frequency subspace and -1 on the negative-frequency subspace, its orthogonal complement, so that β_ω is Hermitian, $\beta_\omega^2 = 1$, and β_ω commutes with H. Put $\mathsf{S} \equiv \mathsf{W}^{-1/2}$ and $\mathsf{T} \equiv (1 + \mathsf{R}\mathsf{R}^\dagger)^{-1/2}$. From $(\mathsf{R}\mathsf{R}^\dagger)\mathsf{R} = \mathsf{R}(\mathsf{R}^\dagger\mathsf{R})$ follows $\mathsf{T}\mathsf{R} = \mathsf{R}\mathsf{S}$ and its adjoint $\mathsf{R}^\dagger\mathsf{T} = \mathsf{S}\mathsf{R}^\dagger$, and with these the matrix

$$\mathsf{U} \equiv \begin{pmatrix} \mathsf{S} & \mathsf{R}^\dagger\mathsf{T} \\ -\mathsf{R}\mathsf{S} & \mathsf{T} \end{pmatrix} \tag{88}$$

satisfies $\mathsf{U}^\dagger\mathsf{U} = 1$. It takes a positive-frequency state $(u, \mathsf{R}u)$ to $(\mathsf{S}\mathsf{W}u, 0) = (\chi, 0)$, and a negative-frequency state, which by the Hermiticity of H has the form $(-\mathsf{R}^\dagger w, w)$, to $(0, \mathsf{T}^{-1}w)$. So $\mathsf{U}\beta_\omega\mathsf{U}^\dagger = \beta$, and $\mathsf{U}\mathsf{H}\mathsf{U}^\dagger$ commutes with β, so it is even, with upper diagonal entry $\hat{h}$. Also $\beta\mathsf{U}\beta = \mathsf{U}^\dagger$, by inspection, so $\beta_\omega = \mathsf{U}^\dagger\beta\,\mathsf{U} = \mathsf{U}^{-2}\beta$, that is, $\mathsf{U}^2 = \beta\beta_\omega$. The Hermitian part $\frac{1}{2}(\mathsf{U} + \mathsf{U}^\dagger) = \mathrm{diag}(\mathsf{S}, \mathsf{T})$ is positive, so U is the square root of $\beta\beta_\omega$ whose eigenvalues have positive real part, and $\mathsf{U} = e^{\mathsf{Z}}$, where $\mathsf{Z} \equiv \frac{1}{2}\ln(\beta\beta_\omega)$ is defined with the principal logarithm. Then

$$e^{\mathsf{Z}}\,\mathsf{H}\,e^{-\mathsf{Z}} = \begin{pmatrix} \hat{h} & 0 \\ 0 & \hat{h}_- \end{pmatrix}, \tag{89}$$

where $\hat{h}_-$ is the operator of the negative frequencies. Since $\beta(\beta\beta_\omega)\beta = (\beta\beta_\omega)^{-1}$, Z is odd, and it is anti-Hermitian, since U is unitary. For $\beta_\omega = \mathrm{sgn}\,\mathsf{H}$, Eriksen [6] gives this rotation as $\frac{1}{2}(1 + \beta\beta_\omega)\,[1 + \frac{1}{4}(\beta\beta_\omega + \beta_\omega\beta - 2)]^{-1/2}$, which equals $(\beta\beta_\omega)^{1/2}$, since $\beta\beta_\omega$ is unitary and the principal square root of a unit complex number $\mathfrak{z} \ne -1$ is $(1 + \mathfrak{z})/|1 + \mathfrak{z}|$.

Expand H/c^2 in powers of c^{-1}, with first term βm, and write $\mathsf{Z} = \sum_{n\ge1} c^{-n}\mathsf{Z}_n$ with every Z_n odd. Suppose $\mathsf{Z}_1, \ldots, \mathsf{Z}_{n-1}$ make $e^{\mathsf{Z}}\mathsf{H}e^{-\mathsf{Z}}/c^2$ even through order $c^{-(n-1)}$, and let D_n be the odd part of its term of order c^{-n}. At that order the term $c^{-n}\mathsf{Z}_n$ adds only $[\mathsf{Z}_n, \beta m]$, and for $\mathsf{Z}_n = [\beta, \mathsf{D}_n]/4m$ this is $-\frac{1}{4}[\beta, [\beta, \mathsf{D}_n]] = -\mathsf{D}_n$. This choice removes the odd part at that order, and it is the only odd one that does, since $[\mathsf{Z}_n, \beta] = -2\beta\mathsf{Z}_n$ for odd Z_n, which is zero only if Z_n is. Each Z_n is therefore a sum of nested commutators of β with the terms of H, and by Eq. (89) so is each term of $\hat{h}$ other than $\hat{V} + \alpha_n mc^2$, the upper entry of the even part of H. These commutators are not zero because H couples the particle to the antiparticle.

C. The master frequency equation

Whenever P, G, and $\hat{V}$ commute with one another, $[\mathsf{R}, \hat{V}] = 0$ and $\{\mathsf{G}, \mathsf{R}\} = 2\mathsf{G}\mathsf{R}$, so Eq. (82) reduces to the algebraic quadratic $c\mathsf{P}\mathsf{R}^2 + 2\alpha_n mc^2\mathsf{R} - c\mathsf{P} = 0$. With $\mathsf{E} \equiv \sqrt{(\alpha_n mc^2)^2 + c^2\mathsf{P}^2}$, its contractive root is $\mathsf{R} = c\mathsf{P}/(\alpha_n mc^2 + \mathsf{E})$, which gives $\hat{h} = \hat{V} + \mathsf{E}$ and

$$(\hbar\,\omega - \hat{V})^2\,\chi = \alpha_n^2\Big[\big(mc^2\big)^2 + c^2\,W^{-2}\big(\hat{\boldsymbol{\pi}}^2 - q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}\big)\Big]\,\chi. \tag{90}$$

This exact commuting reduction occurs, among other cases, in two settings with constant α_n, W, and V: (a) a static magnetic field $\mathbf{B}(\mathbf{x})$ of arbitrary strength and spatial profile with $\mathbf{A}_g = \mathbf{0}$, and (b) a rigid-flow shift $\mathbf{A}_g = \mathbf{u} + \boldsymbol{\Omega}\times\mathbf{x}$ with $\mathbf{A} = \mathbf{0}$, for which $\mathbf{B}_g = 2\boldsymbol{\Omega}$ and

$$\hat{V} = V - \mathbf{u}\cdot\hat{\mathbf{p}} - \boldsymbol{\Omega}\cdot(\mathbf{L} + \mathbf{S}), \tag{91}$$

which commutes with $\boldsymbol{\sigma}\cdot\hat{\mathbf{p}}$ because $\hat{\mathbf{p}}$ generates translations and $\mathbf{L}+\mathbf{S}$ generates simultaneous orbital and spin rotations.

Equation (90) is also obeyed by the negative-frequency states, for which $\hbar\omega-\hat{V}$ is $-\mathsf{E}$. Of its two roots the positive one is the particle's: at rest, free of every field, and in flat spacetime it gives $\hbar\omega=mc^2$, so that ω is the rest frequency ω_0 of Eq. (2). Under the same conditions the negative root, which describes the antiparticle, gives $\hbar\omega=-mc^2$. The two roots differ by $2\mathsf{E}$, whose values are at least $2\alpha_n mc^2$, so the positive root never meets the negative one as the fields are switched on, and a state on it stays on it when the fields change slowly on the time scale $\hbar$ divided by that difference.

In a general static background, the operator $W^{-1}(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2W^{-1}=\hbar^2(\mathsf{X}W^{-1})^\dagger(\mathsf{X}W^{-1})$ is nonnegative and Hermitian in the unweighted inner product, so the positive square root of $(mc^2)^2+c^2W^{-1}(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2W^{-1}$ exists, and α_n times it is Hermitian in the weighted product. Define the *commutator term* C as $\hat{h}$ minus $\hat{V}$ and minus α_n times this root. Writing out $\hat{V}$ by Eq. (79) and $(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2$ by Eq. (77) in Eq. (83) gives the exact *master frequency equation*:

$$\boxed{\hbar\,\omega\,\chi=\left[V+\alpha_n\sqrt{\left(mc^2\right)^2+c^2\,W^{-1}\left(\hat{\boldsymbol{\pi}}^2-q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}\right)W^{-1}}-\hat{\Pi}_g-\tfrac{1}{4}\,\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g+\mathsf{C}\right]\chi}\,. \tag{92}$$

(The master frequency equation)

Here $\hbar\omega$ is the energy of the stationary state of Eq. (75). The additive potential energy V ($q\Phi$ for a charge in an electrostatic potential) is taken in coordinate time, in which it is α times the value $V_{\rm loc}$ that a static observer measures in proper time. So $V=\alpha V_{\rm loc}$ already contains the factor α and appears outside the factor α_n; a second factor α would add the spurious energy $V(\alpha-1)$. The lapse α_n, shift $\mathbf{A}_g$, and spatial scale factor W are those of Eq. (35). The rest mass m equals $m(\mathbf{x})=m_0+V_{\rm sc}(\mathbf{x})/c^2$ when a Lorentz-scalar potential is present (Section II B). The kinematic momentum is $\hat{\boldsymbol{\pi}}=-i\hbar\nabla-q\mathbf{A}$, the magnetic and gravitomagnetic fields are $\mathbf{B}$ and $\mathbf{B}_g$, and $\hat{\Pi}_g$ is the shift coupling of Eq. (78).

Equation (92) holds for every positive-frequency state in every static field on conformally flat slices for which the positive-frequency subspace is the set of pairs $(u,\mathsf{R}u)$ with R bounded. Section IV B proves this when $c\mathsf{P}$, G, and $\hat{V}$ are bounded with $\epsilon<1$, and for a free particle, and assumes it for a particle in weak fields in a state built from momenta small compared with mc. The commutator term C vanishes whenever $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$, W, G, and $\hat{V}$ commute. Equation (81) is the Weyl pair itself, so keeping C keeps every term of Eq. (58), the commutator $[\mathsf{N},\mathsf{X}_W]$ included. A change of spinor keeps the eigenvalues, and outside the commuting cases the eigenvalues of $\hat{V}$ plus α_n times the root are in general not those of $\hat{h}$, so no change of spinor removes C.

When the potentials depend on time t, the instantaneous rotation $e^{\mathsf{Z}(t)}$ adds the term $-i\hbar\,e^{\mathsf{Z}(t)}(\partial_t e^{-\mathsf{Z}(t)})$, whose leading part $i\hbar\,\partial_t\mathsf{Z}$ is odd and couples the two blocks, so Eq. (92) does not by itself give an exact two-component evolution. Removing the odd part order by order, as Foldy and Wouthuysen [5] do for the operator $H-i\hbar\partial_t$, adds $-q\,\partial_t\mathbf{A}$ to $-\nabla V$ inside $q\mathbf{E}$ in the Darwin and spin-orbit terms at $O(c^{-2})$. Then $\nabla\times\mathbf{E}=-\partial_t\mathbf{B}$ need not vanish, so each product $\mathbf{E}\times\hat{\boldsymbol{\pi}}$ in those terms stands for its Hermitian form $\frac{1}{2}(\mathbf{E}\times\hat{\boldsymbol{\pi}}-\hat{\boldsymbol{\pi}}\times\mathbf{E})=\mathbf{E}\times\hat{\boldsymbol{\pi}}+\frac{1}{2}i\hbar\nabla\times\mathbf{E}$.

An anomalous magnetic moment $\boldsymbol{\mu}_{\rm anom}=\mu_a\boldsymbol{\sigma}$ requires a Pauli tensor term in the Dirac equation, which Eq. (57) does not contain. For a charge, $\mu_a=(g_s-2)q\hbar/(4m)$; for a neutral particle, μ_a is the whole moment. In flat spacetime the Pauli term adds the coupling $-\boldsymbol{\mu}_{\rm anom}\cdot\mathbf{B}$ to $\hat{h}$, and at order c^{-2} it adds three more terms. The first is the contact term $-\mu_a\hbar\,\nabla\cdot\mathbf{E}/(2mc^2)$, which is $-(g_s-2)\,q\hbar^2\,\nabla\cdot\mathbf{E}/(8m^2c^2)$ for a charge [63, 64]. The second is the spin-orbit term $-(\boldsymbol{\mu}_{\rm anom}\times\mathbf{E})\cdot\hat{\boldsymbol{\pi}}/(mc^2)$, with $\mathbf{E}\times\hat{\boldsymbol{\pi}}$ in its Hermitian form; it has this coefficient because the spin-orbit coefficient of a moment $g_s(q/2m)\mathbf{S}$ is proportional to g_s-1 [37]. The third comes from the double commutator of the Pauli term with $c\boldsymbol{\alpha}\cdot\hat{\boldsymbol{\pi}}$, and in a uniform $\mathbf{B}$ it is $\mu_a(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})(\hat{\boldsymbol{\pi}}\cdot\mathbf{B})/(2m^2c^2)$. For a neutral particle the spin-orbit term gives the Aharonov–Casher phase in an electric field [65, 66].

To extract the classical limit at all speeds and field strengths, each operator is replaced by its Weyl symbol on phase space $(\mathbf{x},\mathbf{p})$, where operator products map to the Moyal star product [67]:

$$f_1\star f_2=f_1\,\exp\!\left[\frac{i\hbar}{2}\left(\overleftarrow{\nabla}_{\mathbf{x}}\cdot\overrightarrow{\nabla}_{\mathbf{p}}-\overleftarrow{\nabla}_{\mathbf{p}}\cdot\overrightarrow{\nabla}_{\mathbf{x}}\right)\right]f_2. \tag{93}$$

Each arrow marks the factor its gradient acts on, and matrix factors keep their order, so $f_1\star f_2$ is f_1f_2 plus terms with one or more powers of $\hbar$. Assume that the symbols of R, $\mathsf{W}^{-1/2}$, and $\hat{h}$ are the sums of their series in $\hbar$, and that the symbol η of R at order $\hbar^0$ has norm at most 1 at each point, as R has. At order $\hbar^0$ the symbols of G and $\hat{V}$ are scalars, and that of P is $\alpha_nW^{-1}\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$. Moving $\{\mathsf{G},\mathsf{R}\}$ to the left side of Eq. (82) and dividing both sides by α_nc then gives $2mc\,\eta=W^{-1}\boldsymbol{\sigma}\cdot\boldsymbol{\pi}-\eta(W^{-1}\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\eta$. At $\boldsymbol{\pi}=\mathbf{0}$ this gives $\eta=0$. For $\boldsymbol{\pi}\neq\mathbf{0}$, work in the eigenbasis of $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$. A solution with a nonzero off-diagonal entry has determinant 1, and its diagonal entries differ by the nonzero real number $-2mcW/|\boldsymbol{\pi}|$. A unitary 2×2 matrix of determinant 1 has complex-conjugate diagonal entries, so this solution is not unitary. Its two singular values have product 1 and are not both 1, so its norm exceeds 1. Each diagonal

entry of a diagonal solution is a root of a quadratic whose two real roots have product -1 and a sum of modulus $2mcW/|\boldsymbol{\pi}| \neq 0$, so exactly one of them has modulus less than 1. The only solution of norm at most 1 is therefore $\eta = W^{-1}\boldsymbol{\sigma}\cdot\boldsymbol{\pi}/(mc + \sqrt{(mc)^2 + W^{-2}|\boldsymbol{\pi}|^2})$. Since the symbols of $\mathsf{W}^{1/2}$ and $\mathsf{W}^{-1/2}$ at order $\hbar^0$ are inverse matrices, Eq. (85) gives the symbol of $\hat{h}$ at that order as the scalar symbol of $\hat{V} + \alpha_n mc^2 + c\,\mathsf{PR}$, and with $(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2 = |\boldsymbol{\pi}|^2$ this is the relativistic geodesic Hamiltonian in the lapse-shift form of the metric [48]. The magnetic spin term under the root is of order $\hbar$, so the symbol of C vanishes at order $\hbar^0$, as shown in flat spacetime for bounded smooth potentials [22]:

$$\hbar\,\omega\big|_{\hbar^0} = V + \alpha_n\sqrt{\left(mc^2\right)^2 + c^2\,W^{-2}|\boldsymbol{\pi}|^2} - \mathbf{A}_g\cdot\boldsymbol{\pi}. \tag{94}$$

With $\mathcal{H} \equiv \alpha_n\sqrt{(mc^2)^2 + c^2W^{-2}|\boldsymbol{\pi}|^2}$, a particle held at rest has zero velocity,

$$\begin{aligned}\dot{\mathbf{x}} &= \nabla_{\mathbf{p}}(\hbar\omega)\\ &= \alpha_n^2 c^2 W^{-2}\boldsymbol{\pi}/\mathcal{H} - \mathbf{A}_g\\ &= \mathbf{0},\end{aligned} \tag{95}$$

and solving for $\boldsymbol{\pi} = \mathcal{H}W^2\mathbf{A}_g/(\alpha_n^2c^2)$ gives $\mathcal{H} = mc^2\alpha_n^2/\alpha$, where α is the time-dilation factor of a stationary clock, Eq. (36). Thus $\mathcal{H} - \mathbf{A}_g\cdot\boldsymbol{\pi} = mc^2\alpha$, and at rest in every stationary metric of the form (35) Eq. (94) is, with no expansion in $1/c^2$,

$$\hbar\,\omega = V + mc^2\,\alpha, \tag{96}$$

whose negative spatial gradient via Eq. (8) gives the electrostatic and gravitational force $\mathbf{F} = -\nabla V - mc^2\nabla\alpha$ on a particle at rest. Equation (96) is Eq. (94) evaluated at the rest momentum at each point. Its gradient equals the gradient of Eq. (94) at fixed $\mathbf{p}$ because $\nabla_{\mathbf{p}}(\hbar\omega)$ vanishes at the rest momentum, by Eq. (95). The spin couplings $\boldsymbol{\sigma}\cdot\mathbf{B}$ and $\boldsymbol{\sigma}\cdot\mathbf{B}_g$ and the commutator term C add to Eq. (96) at order $\hbar$.

V. THE c^{-2} EXPANSION AND THE CATALOG OF FORCES

A. Convergence of the Riccati series and the dressed spatial operator

For a slow particle in weak fields, the rest energy mc^2 is large compared to $c\mathsf{P}$, the lapse energy $\mathsf{G} = m_0c^2(\alpha_n - 1)$, and $\hat{V}$. Weak gravity means that $\alpha_n - 1$ and $W - 1$ are of order c^{-2}, as in the post-Newtonian metric of Section II D. Expanding $\mathsf{R} = \sum_{k\geq 0}\mathsf{R}_k\,c^{-2k-1}$ in powers of c^{-2} at fixed P, G, and $\hat{V}$ and substituting into Eq. (82) yields the recursion

$$\begin{aligned}\mathsf{R}_0 &= \frac{\mathsf{P}}{2m},\\ 2m\,\mathsf{R}_{k+1} &= -\{\mathsf{G},\mathsf{R}_k\} - [\mathsf{R}_k,\hat{V}] - \sum_{l=0}^{k}\mathsf{R}_l\,\mathsf{P}\,\mathsf{R}_{k-l}\quad (k\geq 0).\end{aligned} \tag{97}$$

In the bounded-operator setting of Section IV, let λ be a complex parameter whose physical value is $1/(2mc^2)$, and let $\mathsf{Q} \equiv c\mathsf{R}$ and $Q_{\max} \equiv \|\mathsf{P}\|/m$. For every λ in the disk $|\lambda| \leq \lambda_0$ with $\lambda_0(2\|\mathsf{G}\| + 2\|\hat{V}\| + 2\|\mathsf{P}\|^2/m) \leq \frac{1}{2}$, the Riccati map $\mathsf{Q} \mapsto \mathsf{P}/(2m) - \lambda(\{\mathsf{G},\mathsf{Q}\} + [\mathsf{Q},\hat{V}] + \mathsf{QPQ})$ maps the ball $\|\mathsf{Q}\| \leq Q_{\max}$ into itself with contraction factor at most $\frac{1}{2}$. Its unique fixed point $\mathsf{Q}(\lambda)$ is the uniform limit of polynomial iterates and is therefore analytic in $|\lambda| < \lambda_0$. The adjoint $\mathsf{Q}(\lambda)^\dagger$ is not analytic in λ. Define $\mathsf{Q}^{\#}(\lambda) \equiv \mathsf{Q}(\bar{\lambda})^\dagger$, which is analytic, and $\mathsf{W}(\lambda) \equiv 1 + 2m\lambda\,\mathsf{Q}^{\#}\mathsf{Q}$. For real λ, $\mathsf{Q}^{\#} = \mathsf{Q}^\dagger$. At the physical value $\lambda = 1/(2mc^2)$, $\mathsf{W} = 1 + \mathsf{R}^\dagger\mathsf{R}$. Since $\|\mathsf{W} - 1\| \leq 2m\lambda_0 Q_{\max}^2 \leq \frac{1}{2}$ in the disk, $\mathsf{W}^{-1/2}$ is a convergent binomial series. The operator $\hat{h}(\lambda) - mc^2$, written with Q and $\mathsf{Q}^{\#}$, is analytic in $|\lambda| < \lambda_0$ and bounded there by some $\mathcal{K}$. By Cauchy's estimate on the circle $|\lambda| = \lambda_1 < \lambda_0$, truncating the c^{-2} expansion of $\hat{h}$ after λ^N incurs, for $|\lambda| \leq \lambda_1/2$, an operator-norm error of at most $2\mathcal{K}(|\lambda|/\lambda_1)^{N+1} = O(c^{-2N-2})$. The physical value $\lambda = 1/(2mc^2)$ lies in this range when $\|\mathsf{G}\| + \|\hat{V}\| + \|\mathsf{P}\|^2/m < mc^2/4$, since λ_0 and λ_1 can then be chosen with $1/(mc^2) < \lambda_1 < \lambda_0$.

To evaluate $\hat{h}$ through order c^{-2}, combine mc^2 from $\alpha_n mc^2$ with $-mc^2\mathsf{R}^\dagger\mathsf{R}$ inside $\mathsf{W}^{-1/2}[\cdots]\mathsf{W}^{-1/2}$ into $mc^2(2\mathsf{W}^{-1} - 1) = mc^2 - \mathsf{P}^2/(2m) - (\mathsf{PR}_1 + \mathsf{R}_1^\dagger\mathsf{P})/c^2 + \mathsf{P}^4/(8m^3c^2) + O(c^{-4})$. Adding $c(\mathsf{PR} + \mathsf{R}^\dagger\mathsf{P}) = \mathsf{P}^2/m + (\mathsf{PR}_1 + \mathsf{R}_1^\dagger\mathsf{P})/c^2$ cancels the R_1 terms identically. With $\mathsf{R}_0 = \mathsf{P}/(2m)$, $\mathsf{R}^\dagger(\hat{V} - \mathsf{G})\mathsf{R} = \mathsf{P}(\hat{V} - \mathsf{G})\mathsf{P}/(4m^2c^2)$, and $\mathsf{W}^{-1/2} = 1 - \mathsf{P}^2/(8m^2c^2) + O(c^{-4})$, splitting $\hat{V} - \mathsf{G} = (\hat{V} + \mathsf{G}) - 2\mathsf{G}$ combines $-\mathsf{PGP}/(2m^2c^2)$ with $\mathsf{P}^2/(2m)$ via $\alpha_n(2 - \alpha_n) = 1 - (\alpha_n - 1)^2$

into $\alpha_n\hbar^2\mathsf{X}_W^2/(2m) + O(c^{-4})$. When a Lorentz-scalar potential $V_{\rm sc}(\mathbf{x}) = [m(\mathbf{x}) - m_0]c^2$ is present, $\alpha_n m(\mathbf{x})c^2 = m_0c^2 + \mathsf{G} + \alpha_n V_{\rm sc}$. In the split $-\{\hbar^2\mathsf{X}^2, V_{\rm sc}\}/(8m_0^2c^2) - \hbar\mathsf{X}\,V_{\rm sc}\,\hbar\mathsf{X}/(4m_0^2c^2) = -\{\hbar^2\mathsf{X}^2, V_{\rm sc}\}/(4m_0^2c^2) + (\{\hbar^2\mathsf{X}^2, V_{\rm sc}\} - 2\hbar\mathsf{X}\,V_{\rm sc}\,\hbar\mathsf{X})/(8m_0^2c^2)$, the anticommutator combines with $\alpha_n\hbar^2\mathsf{X}_W^2/(2m_0)$ into $\frac{1}{2}\{m(\mathbf{x})^{-1}, \frac{1}{2}\alpha_n\hbar^2\mathsf{X}_W^2\}$. The remaining difference subtracts $V_{\rm sc}$ from $\hat{V} + \mathsf{G}$ in the commutator dressing terms. Define $f \equiv \hat{V} + \mathsf{G} - V_{\rm sc}$ and the *dressed spatial operators*

$$\begin{aligned} \tilde{\mathsf{X}} &\equiv e^{-f/4mc^2}\,\mathsf{X}\,e^{f/4mc^2}, \\ \tilde{\mathsf{X}}_W &\equiv W^{-1/2}\,\tilde{\mathsf{X}}\,W^{-1/2}. \end{aligned} \tag{98}$$

With $\mathsf{Y} = f/(4mc^2)$, the identity $e^{-\mathsf{Y}}\mathsf{X}e^{2\mathsf{Y}}\mathsf{X}e^{-\mathsf{Y}} = \mathsf{X}^2 - (\{\mathsf{X}^2, \mathsf{Y}\} - 2\mathsf{X}\mathsf{Y}\mathsf{X}) + O(\mathsf{Y}^2)$ expands the product of $\tilde{\mathsf{X}}_W$ and its adjoint as

$$\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger = \mathsf{X}_W^2 - \frac{\{\mathsf{X}^2, f\} - 2\,\mathsf{X}\,f\,\mathsf{X}}{4mc^2} + O(c^{-4}). \tag{99}$$

The dressed operators collect the $O(c^{-2})$ dressing terms from $\mathsf{W}^{-1/2}$ and $\mathsf{R}^\dagger(\hat{V} - \mathsf{G} - V_{\rm sc})\mathsf{R}$ into one operator product:

$$\hat{h} = \alpha_n mc^2 + \hat{V} + \frac{\alpha_n\hbar^2}{4}\left\{m^{-1}, \tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger\right\} - \frac{\hbar^4\mathsf{X}^4}{8m^3c^2} + O(c^{-4}) \tag{100}$$

$$= \hat{V} + \alpha_n\sqrt{m^2c^4 + c^2\hbar^2\,\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger} + O(c^{-4}). \tag{101}$$

Here m is $m(\mathbf{x})$ when $V_{\rm sc} \neq 0$, and $m = m_0$ in the terms of order c^{-2}, since $m - m_0 = V_{\rm sc}/c^2$ changes those terms only at $O(c^{-4})$. The terms of order c^{-2} also have X in place of X_W and no factor α_n, since $\alpha_n - 1$ and $W - 1$ are of order c^{-2} and so change those terms only at $O(c^{-4})$. Writing $W^{-1/2} = W^{-1}W^{1/2}$ on the left of $\tilde{\mathsf{X}}_W$, $W^{-1/2} = W^{1/2}W^{-1}$ on the right of $\tilde{\mathsf{X}}_W^\dagger$, and $W^{1/2} = e^{\frac{1}{2}\ln W}$ absorbs the spatial scale factor W into the *shifted dressing potential*

$$\begin{aligned} f_W &\equiv f - 2mc^2\ln W \\ &= \hat{V} + \mathsf{G} - V_{\rm sc} - 2mc^2\ln W, \end{aligned} \tag{102}$$

so that $\tilde{\mathsf{X}}_W = W^{-1}e^{-f_W/4mc^2}\mathsf{X}e^{f_W/4mc^2}$ up to $O(c^{-4})$. Splitting $\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger$ into the undressed curved-slice square $W^{-1}\mathsf{X}^2W^{-1}$ plus the dressing difference $\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger - W^{-1}\mathsf{X}^2W^{-1}$ and using $\hbar^2\mathsf{X}^2 = \hat{\boldsymbol{\pi}}^2 - q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}$ gives the *Taylor expansion of the frequency equation*, with each term written as one operator:

$$\begin{aligned} \hat{h} = {} & \underbrace{mc^2\,\alpha_n}_{\substack{=\hbar\,\omega_{\rm geom}\\(\text{geometric})}} + \underbrace{V}_{\substack{=\hbar\,\omega_{\rm ext}\\(\text{extrinsic})}} + \underbrace{\frac{\alpha_n}{4}\left\{m^{-1}, W^{-1}\hat{\boldsymbol{\pi}}^2W^{-1}\right\}}_{\substack{=\hbar\,\omega_{\rm kin}\\(\text{kinetic})}} + \underbrace{\left(-\hat{\Pi}_g\right)}_{\substack{=\hbar\,\omega_{\rm gm}\\(\text{gravitomagnetic})}} + \underbrace{\left(-\frac{\alpha_n q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}}{2mW^2}\right)}_{\substack{=\hbar\,\omega_{\rm Pauli}\\(\text{Pauli})}} \\ & + \underbrace{\left(-\frac{\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g}{4}\right)}_{\substack{=\hbar\,\omega_{\rm rot}\\(\text{spin-rotation})}} + \underbrace{\left(-\frac{\hbar^4\mathsf{X}^4}{8m^3c^2}\right)}_{\substack{=\hbar\,\omega_{\rm rel}\\(\text{relativistic})}} + \underbrace{\frac{\hbar^2}{2m}\left(\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger - W^{-1}\mathsf{X}^2W^{-1}\right)}_{\substack{=\hbar\,\omega_{\rm so}\\(\text{Darwin \& spin-orbit})}} + O(c^{-4}) \end{aligned} \tag{103}$$

(The Taylor expansion of the frequency equation)

Let $\hbar\hat{\omega} \equiv \alpha_n^{-1/2}\hat{h}\alpha_n^{1/2}$, which is Hermitian in the unweighted inner product. In its Weyl symbol each braced term in Eq. (103) is $\hbar$ times one additive part of the frequency $\omega(\mathbf{x}, \mathbf{p})$. Because Eq. (103) is linear in the additive parts and the force-frequency identity $\mathbf{F} = -\hbar\nabla\omega$ of Eq. (8) is linear in ω, the individual forces superpose. Equation (103) is complete through $O(c^{-2})$; the $+O(c^{-4})$ remainder contains the second post-Newtonian kinetic term $+\hbar^6\mathsf{X}^6/(16m^5c^4)$ and $O(v^2/c^2)$ kinetic dressings of the Darwin and spin-orbit terms. Table II lists the force or spin effect of each part.

B. Quaternion-table expansion of the dressed product and the shift-strain Moyal calculus

The eighth term of Eq. (103) is the leading part of the commutator operator C in Eq. (92):

$$\mathsf{C} = \frac{\hbar^2}{2m}\left(\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger - W^{-1}\mathsf{X}^2W^{-1}\right) + O(c^{-4}). \tag{104}$$

TABLE II. The force or spin effect of each part of the frequency operator of Eq. (103), from $\mathbf{F} = -\hbar\nabla\omega$ (Eq. (8)) and, for the spin, $d\mathbf{S}/dt = (i/\hbar)[\hat{h}, \mathbf{S}]$ (Eq. (9)). The second column names the fields that the part contains.

Part	Fields	Force or spin effect
ω_{geom}	α_n, V_{sc}	Newtonian gravity, first post-Newtonian nonlinear gravity, inertial reaction $-m_0\mathbf{a}$, Lorentz-scalar force
ω_{ext}	V	Coulomb force, forces of external traps
ω_{kin}	$\mathbf{A}$, α_n, W, V_{sc}	magnetic Lorentz force, diamagnetism, ponderomotive force, first post-Newtonian velocity-dependent gravity, inertial reaction of the kinetic energy, velocity-dependent effective-mass force
ω_{gm}	$\mathbf{A}_g$, $\mathbf{A}$	Lense–Thirring force; Coriolis, Euler, and centrifugal forces
ω_{Pauli}	$\mathbf{B}$, α_n, W, V_{sc}	Stern–Gerlach force, Larmor precession
ω_{rot}	$\mathbf{B}_g$	spin-rotation coupling, Barnett effect, spin-vorticity force, Lense–Thirring spin precession
ω_{rel}	$\mathbf{A}$, $\mathbf{B}$	relativistic mass-velocity correction, $O(v^2/c^2)$ correction to the Stern–Gerlach force, quadratic magnetic shift (no force: it cancels in the Weyl symbol)
ω_{so}	all	Darwin terms; electric spin-orbit coupling with the Thomas factor $\frac{1}{2}$, inertial Thomas coupling, geodetic coupling with coefficient $\gamma_{\mathrm{PPN}} + \frac{1}{2}$, Lorentz-scalar spin-orbit coupling, shift-strain Darwin and geodetic terms

Every term in the all-orders expansion of C is a polynomial in $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$, W, G, and $\hat{V}$ that vanishes when those operators commute, and is therefore a sum of products each containing a commutator. Yet C is not determined by the wave-equation commutator $[\mathsf{N}, \mathsf{X}_W]$ alone: shifting $V \mapsto V + \delta\,\alpha_n$ for a constant δ adds the constant $\delta/(\hbar c)$ to N and leaves $[\mathsf{N}, \mathsf{X}_W]$ unchanged at every frequency, whereas the $O(c^{-2})$ part $\delta\,\mathsf{G}/(mc^2)$ of $\delta\,\alpha_n$ changes the leading spin-orbit part of C by $(\delta\hbar/4m^3c^4)\boldsymbol{\sigma}\cdot(\nabla\mathsf{G}\times\hat{\boldsymbol{\pi}}) \neq 0$ wherever $\nabla\alpha_n \neq \mathbf{0}$ (Appendix A 6).

When the shift vanishes ($\mathbf{A}_g = \mathbf{0}$), $f_W = V + \mathsf{G} - V_{\mathrm{sc}} - 2mc^2\ln W$ is a scalar function of position alone. Let m also be constant, as it is when $V_{\mathrm{sc}} = 0$. By the product rule $e^{-f_W/4mc^2}\mathbf{D}(e^{f_W/4mc^2}\psi) = (\mathbf{D} + \nabla f_W/4mc^2)\psi$ exactly, and since $e_k^\dagger = -e_k$ and $D_k^\dagger = -D_k$,

$$\tilde{\mathsf{X}}_W = W^{-1}\,\mathbf{e}\cdot\left(\mathbf{D} + \frac{\nabla f_W}{4mc^2}\right),$$
$$\tilde{\mathsf{X}}_W^\dagger = \mathbf{e}\cdot\left(\mathbf{D} - \frac{\nabla f_W}{4mc^2}\right)W^{-1}. \tag{105}$$

The product of the two inner factors follows from the 3×3 quaternion multiplication table $e_j e_k = -\delta_{jk} + \epsilon_{jkl}e_l$ of Eq. (45). In the three diagonal entries $j = k$, where $e_j^2 = -1$, the cross terms $\pm(\partial_j f_W/4mc^2)D_j\psi$ cancel each other, and the remaining terms give $-\mathbf{D}^2\psi + (\nabla^2 f_W/4mc^2)\psi + (|\nabla f_W|^2/16m^2c^4)\psi$. In the six off-diagonal entries $j \neq k$, the symmetric terms $\partial_j\partial_k f_W$ and $(\partial_j f_W)(\partial_k f_W)$ give zero when contracted with the antisymmetric ϵ_{jkl}. The $D_jD_k\psi$ terms give the magnetic spin term of Eq. (62). The cross terms combine into $2\epsilon_{jkl}e_l(\partial_j f_W/4mc^2)D_k\psi = \mathbf{e}\cdot(\nabla f_W\times\mathbf{D})\psi/(2mc^2) = \boldsymbol{\sigma}\cdot(\nabla f_W\times\hat{\boldsymbol{\pi}})\psi/(2\hbar mc^2)$. Thus the dressed product expands exactly as

$$\frac{\hbar^2}{2m}\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger = W^{-1}\left[\frac{(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2}{2m} + \frac{\hbar^2}{8m^2c^2}\nabla^2 f_W + \frac{1}{2m^2c^2}\,\mathbf{S}\cdot(\nabla f_W\times\hat{\boldsymbol{\pi}}) + \frac{\hbar^2}{32m^3c^4}\,|\nabla f_W|^2\right]W^{-1}. \tag{106}$$

When $W = 1$, $f_W = f$ and $\tilde{\mathsf{X}}_W = \tilde{\mathsf{X}}$, and Eq. (106) becomes

$$\frac{\hbar^2}{2m}\tilde{\mathsf{X}}\tilde{\mathsf{X}}^\dagger = \frac{(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2}{2m} + \frac{\hbar^2}{8m^2c^2}\nabla^2 f + \frac{1}{2m^2c^2}\,\mathbf{S}\cdot(\nabla f\times\hat{\boldsymbol{\pi}}) + \frac{\hbar^2}{32m^3c^4}\,|\nabla f|^2. \tag{107}$$

The last term of Eqs. (106) and (107) is of order c^{-4} and lies in the remainder of Eq. (103). For a scalar potential with $W = 1$ that remainder also contains $-\hbar^2|\nabla f|^2/(32m^3c^4)$, which cancels it, so $\hat{h}$ has no term quadratic in the potential gradient at order c^{-4}. This cancellation depends on the frame of Eq. (84), which is that of Eriksen [6]: the

unitary transformation $e^{i\hat{S}}$ with $\hat{S} = \mathcal{Z}\hbar(\nabla V\cdot\hat{\mathbf{p}} + \hat{\mathbf{p}}\cdot\nabla V)/(m^3c^4)$ and a real number $\mathcal{Z}$ leaves every term through order c^{-2} unchanged. Through $i[\hat{S}, V] = 2\mathcal{Z}\hbar^2|\nabla V|^2/(m^3c^4)$ it changes the coefficient of $\hbar^2|\nabla V|^2/(m^3c^4)$ by $2\mathcal{Z}$.

The second-order wave equation (57) and the dressed product (105) both use the identity $(A-B)(A+B) = A^2 - B^2 + [A,B]$. Equation (67) takes $A = \mathsf{N}$ and $B = \mathsf{X}_W$ on the 4×4 table $(\bar{e}_\mu e_\nu)$, which gives the first-order commutator $[\mathsf{N}, \mathsf{X}_W]$. Equation (105) takes $A = \mathsf{X}$ and $B = -[\mathsf{X}, f_W/4mc^2]$ on the 3×3 block $(e_j e_k)$, which gives the nested Foldy–Wouthuysen commutator $-[\mathsf{X}, [\mathsf{X}, f_W/4mc^2]]$.

When a nonvanishing shift $\mathbf{A}_g(\mathbf{x})$ is present (with $\alpha_n = W = 1$, $V = V_{\rm sc} = 0$, $\mathbf{A} = \mathbf{0}$), f_W has the momentum-dependent Weyl symbol $f_W(\mathbf{x},\mathbf{p}) = u_g - \frac{1}{4}\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g$ where $u_g \equiv -\mathbf{A}_g\cdot\mathbf{p}$. Decompose the shift velocity gradient $\partial_i A_{g,j} = \Sigma_{ij} + \frac{1}{2}\epsilon_{ijk}B_{g,k}$ into its symmetric strain-rate tensor $\Sigma_{ij} \equiv \frac{1}{2}(\partial_i A_{g,j} + \partial_j A_{g,i})$ and antisymmetric vorticity $\frac{1}{2}\epsilon_{ijk}B_{g,k}$. The Moyal commutator $(\boldsymbol{\sigma}\cdot\mathbf{p})\star f_W - f_W\star(\boldsymbol{\sigma}\cdot\mathbf{p})$ in Eq. (98) has a matrix commutator part $[\boldsymbol{\sigma}\cdot\mathbf{p}, -\frac{1}{4}\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g] = -\frac{1}{2}i\hbar\,\boldsymbol{\sigma}\cdot(\mathbf{p}\times\mathbf{B}_g)$. This part cancels the antisymmetric vorticity part $+\frac{1}{2}i\hbar\,\boldsymbol{\sigma}\cdot(\mathbf{p}\times\mathbf{B}_g)$ of $-i\hbar\sigma_i\partial_i u_g$ identically. Define the *shift strain force* $\mathbf{F}_\Sigma$, with components $F_{\Sigma,i} \equiv \Sigma_{ij}p_j$. It vanishes for a rigid rotation $\mathbf{A}_g = \boldsymbol{\Omega}\times\mathbf{x}$, for which $\tilde{\mathsf{X}} = \mathsf{X}$ of Section IV, and equals $(\mathbf{p}\cdot\nabla)A_{g,i}$ whenever $\mathbf{B}_g = \mathbf{0}$. The Weyl symbols of $\hbar\tilde{\mathsf{X}}$ and $\hbar\tilde{\mathsf{X}}^\dagger$ are $\boldsymbol{\sigma}\cdot(\mathbf{p}\pm i\hbar\mathbf{F}_\Sigma/4mc^2)$, and their Moyal product gives the shift-strain Darwin and spin-orbit contributions $-(\hbar^2/8m^2c^2)\nabla\cdot\mathbf{F}_\Sigma + (2m^2c^2)^{-1}\mathbf{S}\cdot(\mathbf{p}\times\mathbf{F}_\Sigma)$. In Painlevé–Gullstrand coordinates for the Schwarzschild metric ($\alpha_n = W = 1$, $\mathbf{A}_g = v_{\rm ff}\hat{\mathbf{r}}$ with $v_{\rm ff} = \sqrt{2GM/r}$),

$$\mathbf{F}_\Sigma = \frac{v_{\rm ff}}{r}\,\mathbf{p} - \frac{3\,v_{\rm ff}}{2r^2}\,(\mathbf{p}\cdot\hat{\mathbf{r}})\,\mathbf{x}, \tag{108}$$

so that $\mathbf{p}\times\mathbf{F}_\Sigma = (3v_{\rm ff}/2r^2)(\mathbf{p}\cdot\hat{\mathbf{r}})(\mathbf{x}\times\mathbf{p})$.

C. Full expansion into individual forces and spin couplings

The entries of Table II follow from the eight terms of Eq. (103), taken in order and each written in *bare form*: every function of position stands to the left of every momentum operator $\hat{\mathbf{p}} = -i\hbar\nabla$ or $\hat{\boldsymbol{\pi}} = \hat{\mathbf{p}} - q\mathbf{A}$, and every derivative of a function of position is written out. Each $\hbar\omega$ includes the sign with which it enters $\hat{h}$. In the terms of order c^{-2}, m is replaced by m_0, as stated after Eq. (101). The forces follow from the force-frequency identity $\mathbf{F} = -\hbar\nabla\omega$ (Eq. (8)) applied to the Weyl symbol of $\hbar\hat{\omega}$, and the spin precessions from the Heisenberg commutator $d\mathbf{S}/dt = (i/\hbar)[\hat{h}, \mathbf{S}]$ (Eq. (9)).

1. *Geometric term* $\hbar\omega_{geom} = mc^2\alpha_n$

This term multiplies the wave function by a function of position, so it is already in bare form and has no operator to expand. Only its two factors need their weak-field forms: the PPN lapse $\alpha_n = 1 + \Phi_g/c^2 + (\beta_{\rm PPN} - \frac{1}{2})\Phi_g^2/c^4 + \mathbf{a}\cdot\mathbf{x}/c^2$ [7, 50] and the position-dependent mass $m(\mathbf{x}) = m_0 + V_{\rm sc}(\mathbf{x})/c^2$ (Table I). They give

$$\hbar\,\omega_{\rm geom} = \underbrace{m_0c^2}_{\substack{\text{rest}\\ \text{energy}}} + \underbrace{m_0\,\Phi_g(\mathbf{x})}_{\substack{\text{Newtonian}\\ \text{gravity}}} + \underbrace{\left(\beta_{\rm PPN} - \tfrac{1}{2}\right)\frac{m_0\,\Phi_g(\mathbf{x})^2}{c^2}}_{\substack{\text{1PN nonlinear}\\ \text{gravity}}} + \underbrace{m_0\,\mathbf{a}\cdot\mathbf{x}}_{\substack{\text{frame-acceleration}\\ \text{potential}}} + \underbrace{\alpha_n\,V_{\rm sc}(\mathbf{x})}_{\substack{\text{Lorentz-scalar /}\\ \text{effective-mass potential}}} + O(c^{-4}). \tag{109}$$

The constant rest energy m_0c^2 exerts no force. The second term gives Newtonian gravity $\mathbf{F}_g = -m_0\nabla\Phi_g$ [30], and the third the 1PN nonlinear gravitational force $-(2\beta_{\rm PPN} - 1)(m_0\Phi_g/c^2)\nabla\Phi_g$ [50, 68]. The fourth gives the d'Alembert inertial reaction force $\mathbf{F}_{\rm in} = -m_0\mathbf{a}$ in a linearly accelerated frame [7]. The fifth gives the Lorentz-scalar gradient force $-\nabla(\alpha_n V_{\rm sc}) = -\alpha_n c^2\nabla m - V_{\rm sc}\nabla\alpha_n$, whose first term is the force of a position-dependent mass, as on a nucleon in the scalar mean field of a nucleus [69].

2. *Extrinsic term* $\hbar\omega_{ext} = V$

The additive potential is a function of position, so this term too is already in bare form, and it has no factor of α_n, W or m to expand. Its parts are the electrostatic and external potentials:

$$\hbar\,\omega_{\rm ext} = \underbrace{q\,\Phi(\mathbf{x})}_{\substack{\text{electrostatic}\\ \text{potential}}} + \underbrace{V_{\rm ext}(\mathbf{x})}_{\substack{\text{external scalar}\\ \text{potential}}}. \tag{110}$$

The first gives the Coulomb force $-q\nabla\Phi$ [38], which is $q\mathbf{E}$ when $\mathbf{A}$ does not depend on time. The second gives the conservative force $-\nabla V_{\rm ext}$ of an optical dipole trap, a Casimir–Polder potential, or a coherent forward-scattering matter potential [70, 71].

3. *Kinetic term* $\hbar\,\omega_{kin} = \frac{1}{4}\alpha_n\{m^{-1}, W^{-1}\hat{\boldsymbol{\pi}}^2 W^{-1}\}$

Let $\eta_g \equiv \alpha_n^{1/2} W^{-1}$. Since α_n commutes with m^{-1}, the conjugation by $\alpha_n^{-1/2}$ to $\hbar\hat{\omega}$ takes $\hbar\,\omega_{\rm kin}$ to $\frac{1}{4}\{m^{-1}, \eta_g\hat{\boldsymbol{\pi}}^2\eta_g\}$. In the weak field, $\eta_g^2 = 1 + [(1+2\gamma_{\rm PPN})\Phi_g + \mathbf{a}\cdot\mathbf{x}]/c^2 + O(c^{-4})$ [50]. The coefficient 1 of Φ_g/c^2 comes from the lapse $\alpha_n - 1$ and $2\gamma_{\rm PPN}$ from the spatial curvature $W^{-2} - 1$. Moving $\hat{\mathbf{p}}$ to the right of $\mathbf{A}$ by $[\hat{p}_j, f] = -i\hbar\,\partial_j f$ gives

$$\hat{\boldsymbol{\pi}}^2 = \hat{\mathbf{p}}^2 - 2q\,\mathbf{A}\cdot\hat{\mathbf{p}} + iq\hbar\,(\nabla\cdot\mathbf{A}) + q^2|\mathbf{A}|^2. \tag{111}$$

Moving $\hat{\boldsymbol{\pi}}$ to the right of a function $g(\mathbf{x})$ in the same way gives, exactly,

$$\eta_g\,\hat{\boldsymbol{\pi}}^2\,\eta_g = \eta_g^2\,\hat{\boldsymbol{\pi}}^2 - i\hbar\,\nabla(\eta_g^2)\cdot\hat{\boldsymbol{\pi}} - \hbar^2\,\eta_g\nabla^2\eta_g \tag{112}$$

and

$$\{g, \hat{\boldsymbol{\pi}}^2\} = 2g\,\hat{\boldsymbol{\pi}}^2 - 2i\hbar\,\nabla g\cdot\hat{\boldsymbol{\pi}} - \hbar^2\,\nabla^2 g. \tag{113}$$

With the inverse mass $m(\mathbf{x})^{-1} = m_0^{-1}(1 - V_{\rm sc}/m_0c^2) + O(c^{-4})$, Eq. (113) for $g = V_{\rm sc}$, and $\eta_g\nabla^2\eta_g = \nabla^2\eta_g + O(c^{-4})$, the kinetic term in bare form is

$$\begin{aligned}\hbar\,\omega_{\rm kin} = {} & \underbrace{\frac{\hat{\mathbf{p}}^2}{2m_0}}_{\substack{\text{free kinetic}\\ \text{energy}}} \underbrace{-\frac{q}{m_0}\,\mathbf{A}\cdot\hat{\mathbf{p}} + \frac{iq\hbar}{2m_0}\,\nabla\cdot\mathbf{A}}_{\substack{\text{magnetic Lorentz \&}\\ \text{orbital Zeeman coupling}}} + \underbrace{\frac{q^2|\mathbf{A}|^2}{2m_0}}_{\substack{\text{diamagnetic /}\\ \text{ponderomotive energy}}} \\ & + \underbrace{\frac{\eta_g^2 - 1}{2m_0}\,\hat{\boldsymbol{\pi}}^2 - \frac{i\hbar}{2m_0}\,\nabla(\eta_g^2)\cdot\hat{\boldsymbol{\pi}} - \frac{\hbar^2}{2m_0}\,\nabla^2\eta_g}_{\substack{\text{1PN spatial-curvature \&}\\ \text{redshift kinetic coupling}}} \\ & \underbrace{-\frac{V_{\rm sc}}{2m_0^2c^2}\,\hat{\boldsymbol{\pi}}^2 + \frac{i\hbar}{2m_0^2c^2}\,\nabla V_{\rm sc}\cdot\hat{\boldsymbol{\pi}} + \frac{\hbar^2}{4m_0^2c^2}\,\nabla^2 V_{\rm sc}}_{\substack{\text{effective-mass}\\ \text{kinetic coupling}}} + O(c^{-4}),\end{aligned} \tag{114}$$

where $\hat{\boldsymbol{\pi}}^2$ in the terms of order c^{-2} stands for the bare operator of Eq. (111). The terms with a factor $i\hbar$ make each group Hermitian. In the Weyl symbol, with $\boldsymbol{\pi} = \mathbf{p} - q\mathbf{A}$, the metric group becomes $(\eta_g^2 - 1)\boldsymbol{\pi}^2/(2m_0) - \hbar^2\nabla^2\eta_g/(4m_0)$ and the effective-mass group becomes $-V_{\rm sc}\boldsymbol{\pi}^2/(2m_0^2c^2) + \hbar^2\nabla^2 V_{\rm sc}/(8m_0^2c^2)$, both up to $O(c^{-4})$.

At fixed canonical momentum $\mathbf{p}$, the free term $\mathbf{p}^2/(2m_0)$ has zero spatial gradient, and the free, magnetic and diamagnetic terms give the group velocity $\dot{\mathbf{x}} = \nabla_{\mathbf{p}}(\hbar\omega) = (\mathbf{p} - q\mathbf{A})/m_0$. Combining $\dot{\mathbf{p}} = -\hbar\nabla\omega$ on the magnetic and diamagnetic terms with $m_0\ddot{\mathbf{x}} = \dot{\mathbf{p}} - q\,\partial_t\mathbf{A} - q(\dot{\mathbf{x}}\cdot\nabla)\mathbf{A}$ yields the magnetic Lorentz force $q\,\dot{\mathbf{x}}\times\mathbf{B}$ and the Faraday induction force $-q\,\partial_t\mathbf{A}$ [36, 37]. The Faraday induction force adds to the Coulomb force of Eq. (110) to give $q\mathbf{E}$. In a uniform field $\mathbf{B}$, the vector potential $\mathbf{A} = \frac{1}{2}\mathbf{B}\times\mathbf{x}$ has $\nabla\cdot\mathbf{A} = 0$. The magnetic and diamagnetic terms then give the orbital Zeeman coupling $-(q/2m_0)\mathbf{B}\cdot(\mathbf{x}\times\hat{\mathbf{p}})$ and the diamagnetic restoring force [72].

In a high-frequency field $\mathbf{A}(\mathbf{x}, t)$ in the radiation gauge, with a slowly varying envelope and a nonrelativistic quiver motion, the diamagnetic term averaged over a period gives the ponderomotive force $-\nabla\langle q^2|\mathbf{A}|^2/2m_0\rangle$ [73].

For $\mathbf{A} = \mathbf{0}$, the negative gradient of the symbol $(\eta_g^2 - 1)\boldsymbol{\pi}^2/(2m_0)$ at fixed $\mathbf{p}$ gives two forces. Its Φ_g part gives the 1PN velocity-dependent gravitational force $-(\frac{1}{2} + \gamma_{\rm PPN})m_0(v^2/c^2)\nabla\Phi_g$, which enters the 1PN perihelion advance of bound orbits [50, 68]. Its $\mathbf{a}\cdot\mathbf{x}$ part gives $-(m_0v^2/2c^2)\,\mathbf{a}$, the inertial reaction of the kinetic energy in an accelerated frame.

The effective-mass group is the change of the kinetic energy with a position-dependent mass. The Hermitian form $\frac{1}{2}\hat{\mathbf{p}}\cdot m^{-1}\hat{\mathbf{p}}$ used in semiconductor heterostructures [74] differs from the ordering here by a term of order $\hbar^2$. This group gives the velocity-dependent effective-mass force $+(v^2/2c^2)\nabla V_{\rm sc}$.

4. *Gravitomagnetic term* $\hbar\,\omega_{gm} = -\hat{\Pi}_g$

The conjugation by $\alpha_n^{-1/2}$ takes $-\hat{\Pi}_g$ of Eq. (78) to $-\frac{1}{2}\{\mathbf{A}_g, \hat{\boldsymbol{\pi}}\}$. Moving $\hat{\boldsymbol{\pi}}$ to the right of $\mathbf{A}_g$ gives the bare form

$$\hbar\,\omega_{\rm gm} = \underbrace{-\mathbf{A}_g\cdot\hat{\boldsymbol{\pi}} + \frac{i\hbar}{2}\,\nabla\cdot\mathbf{A}_g}_{\substack{\text{gravitomagnetic / frame-}\\ \text{dragging / Coriolis coupling}}}, \tag{115}$$

whose second term makes it Hermitian. The Weyl symbol is $-\mathbf{A}_g\cdot\boldsymbol{\pi}$, which adds $-\mathbf{A}_g$ to the group velocity. For a uniform gravitomagnetic field $\mathbf{B}_g=\nabla\times\mathbf{A}_g$, the shift $\mathbf{A}_g=\frac{1}{2}\mathbf{B}_g\times\mathbf{x}$ has $\nabla\cdot\mathbf{A}_g=0$, and the gravitomagnetic term is

$$\hbar\,\omega_{\rm gm}=-\tfrac{1}{2}\,\mathbf{B}_g\cdot\mathbf{L},\tag{116}$$

with $\mathbf{L}=\mathbf{x}\times\hat{\boldsymbol{\pi}}$. For a general shift $\mathbf{A}_g(\mathbf{x},t)$ and $\mathbf{A}=\mathbf{0}$, the group velocity gives $\mathbf{p}=m_0(\dot{\mathbf{x}}+\mathbf{A}_g)$. Combining $\dot{\mathbf{p}}=\nabla(\mathbf{p}\cdot\mathbf{A}_g)$ with $m_0\ddot{\mathbf{x}}=\dot{\mathbf{p}}-m_0\partial_t\mathbf{A}_g-m_0(\dot{\mathbf{x}}\cdot\nabla)\mathbf{A}_g$ yields $m_0\ddot{\mathbf{x}}=m_0\,\dot{\mathbf{x}}\times\mathbf{B}_g-m_0\,\partial_t\mathbf{A}_g+\frac{1}{2}m_0\nabla(|\mathbf{A}_g|^2)$. The first term is the gravitomagnetic Lense–Thirring orbital force $m_0\,\dot{\mathbf{x}}\times\mathbf{B}_g$ [51]. In a frame rotating at $\boldsymbol{\Omega}(t)$ ($\mathbf{A}_g=\boldsymbol{\Omega}\times\mathbf{x}$, $\mathbf{B}_g=2\boldsymbol{\Omega}$), the same equation gives the Coriolis force $-2m_0\,\boldsymbol{\Omega}\times\dot{\mathbf{x}}$, the Euler force $-m_0\,\dot{\boldsymbol{\Omega}}\times\mathbf{x}$, and the centrifugal force $-m_0\,\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{x})$ [7, 32].

5. *Pauli term* $\hbar\,\omega_{Pauli}=-\alpha_n q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}/(2mW^2)$

This term is a sum of Pauli matrices times functions of position and contains no momentum operator, so it is already in bare form, and the conjugation by $\alpha_n^{-1/2}$ leaves it unchanged. With $\mathbf{S}=\frac{1}{2}\hbar\boldsymbol{\sigma}$, $\alpha_nW^{-2}=1+[(1+2\gamma_{\rm PPN})\Phi_g+\mathbf{a}\cdot\mathbf{x}]/c^2+O(c^{-4})$ and $m(\mathbf{x})^{-1}=m_0^{-1}(1-V_{\rm sc}/m_0c^2)+O(c^{-4})$, it expands into

$$\hbar\,\omega_{\rm Pauli}=\underbrace{-\frac{q}{m_0}\,\mathbf{S}\cdot\mathbf{B}}_{\substack{\text{Dirac Pauli}\\ \text{coupling }(g_s=2)}}\underbrace{-\frac{(1+2\gamma_{\rm PPN})\Phi_g+\mathbf{a}\cdot\mathbf{x}}{c^2}\,\frac{q}{m_0}\,\mathbf{S}\cdot\mathbf{B}}_{\substack{\text{gravitational redshift /}\\ \text{curvature of magnetic moment}}}+\underbrace{\frac{V_{\rm sc}}{m_0c^2}\,\frac{q}{m_0}\,\mathbf{S}\cdot\mathbf{B}}_{\substack{\text{effective-mass change}\\ \text{of magnetic moment}}}+O(c^{-4}).\tag{117}$$

Through Eq. (8), Eq. (117) gives the Stern–Gerlach force $\nabla(\boldsymbol{\mu}\cdot\mathbf{B})$ on the magnetic moment $\boldsymbol{\mu}=(\alpha_nW^{-2}q/m)\mathbf{S}$ [4, 75], and $d\mathbf{S}/dt=(i/\hbar)[\hat{h},\mathbf{S}]=-\alpha_nW^{-2}(q/m)\mathbf{B}\times\mathbf{S}$, the Larmor precession. Per unit of angular momentum a uniform magnetic field couples to the spin twice as strongly as to the orbital motion, since $\hat{\boldsymbol{\pi}}^2/2m_0$ contains $-(q/2m_0)\mathbf{B}\cdot\mathbf{L}$ in the gauge $\mathbf{A}=\frac{1}{2}\mathbf{B}\times\mathbf{x}$ (Eq. (114)), compared with $-(q/m_0)\mathbf{B}\cdot\mathbf{S}$ here.

6. *Spin-rotation term* $\hbar\,\omega_{rot}=-\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g/4$

This term too is a sum of Pauli matrices times functions of position, so it is already in bare form, and it has no factor of α_n, W or m to expand. With $\mathbf{S}=\frac{1}{2}\hbar\boldsymbol{\sigma}$ it is

$$\hbar\,\omega_{\rm rot}=\underbrace{-\frac{1}{2}\,\mathbf{S}\cdot\mathbf{B}_g}_{\substack{\text{Mashhoon spin-rotation /}\\ \text{Lense–Thirring spin coupling}}}.\tag{118}$$

Through Eq. (8) it gives the gravitomagnetic Stern–Gerlach spin-curvature force $\nabla(\frac{1}{2}\mathbf{S}\cdot\mathbf{B}_g)$ [11, 76], and it precesses the spin as $d\mathbf{S}/dt=-\frac{1}{2}\mathbf{B}_g\times\mathbf{S}$, the Mashhoon–Lense–Thirring precession [76, 77]. Rotation and frame dragging couple to the spin and to the orbital motion with the same coefficient $\frac{1}{2}$, through $-\frac{1}{2}\,\mathbf{B}_g\cdot(\mathbf{L}+\mathbf{S})$ (Eqs. (116) and (118)), unlike the magnetic field. In a frame rotating at $\boldsymbol{\Omega}=\frac{1}{2}\mathbf{B}_g$, $-\boldsymbol{\Omega}\cdot\mathbf{S}$ splits the spin levels by $\hbar\Omega$ [78, 79]. This splitting gives the Barnett magnetization at equivalent field $\mathbf{B}_{\rm Barnett}=(m_0/q)\boldsymbol{\Omega}$ [80]. In a spatially varying acoustic vorticity field $\boldsymbol{\Omega}=\frac{1}{2}\nabla\times\dot{\mathbf{u}}_{\rm lat}$ [81] the term $-\boldsymbol{\Omega}\cdot\mathbf{S}$ exerts the spin-vorticity force

$$\mathbf{F}=\pm\tfrac{1}{2}\,\hbar\nabla\Omega,\tag{119}$$

where $\Omega=|\boldsymbol{\Omega}|$ and the upper sign is for the spin parallel to $\boldsymbol{\Omega}$. Equation (119) is Eq. (8) applied to one eigenvalue $\mp\frac{1}{2}\hbar\Omega$ of $-\boldsymbol{\Omega}\cdot\mathbf{S}$. It holds while the splitting $\hbar\Omega$ is large enough that the spin stays in that level and the direction of $\boldsymbol{\Omega}$ is fixed. Where that direction turns with position, the spin state in that level can acquire a Berry phase [82] and the force can acquire Berry-curvature corrections [83].

7. *Relativistic term* $\hbar\,\omega_{rel} = -\hbar^4\mathsf{X}^4/(8m^3c^2)$

Squaring $\hbar^2\mathsf{X}^2 = \hat{\boldsymbol{\pi}}^2 - q\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}$ and using $(\boldsymbol{\sigma}\cdot\mathbf{B})^2 = |\mathbf{B}|^2$ expands the seventh term of Eq. (103) into three parts:

$$\hbar\,\omega_{\rm rel} = \underbrace{-\frac{\hat{\boldsymbol{\pi}}^4}{8m_0^3c^2}}_{\substack{\text{relativistic mass-}\\ \text{velocity energy}}} + \underbrace{\frac{q}{4m_0^3c^2}\left\{\hat{\boldsymbol{\pi}}^2,\,\mathbf{S}\cdot\mathbf{B}\right\}}_{\substack{O(v^2/c^2)\text{ kinetic correction}\\ \text{to Pauli / Stern–Gerlach}}} \underbrace{-\frac{q^2\hbar^2|\mathbf{B}|^2}{8m_0^3c^2}}_{\substack{\text{quadratic-in-}\mathbf{B}\\ \text{relativistic shift}}}. \tag{120}$$

The first and third parts are already bare: the first because $\hat{\boldsymbol{\pi}}$ is one of the two momentum operators of the definition, the third because it contains no momentum operator. Equation (113) holds for $g = \mathbf{S}\cdot\mathbf{B}$, since $\mathbf{S}$ commutes with $\hat{\boldsymbol{\pi}}$, and puts the second part in the bare form

$$\frac{q}{4m_0^3c^2}\left\{\hat{\boldsymbol{\pi}}^2,\,\mathbf{S}\cdot\mathbf{B}\right\} = \frac{q}{2m_0^3c^2}\left[(\mathbf{S}\cdot\mathbf{B})\,\hat{\boldsymbol{\pi}}^2 - i\hbar\,\nabla(\mathbf{S}\cdot\mathbf{B})\cdot\hat{\boldsymbol{\pi}} - \frac{\hbar^2}{2}\,\nabla^2(\mathbf{S}\cdot\mathbf{B})\right], \tag{121}$$

where $\nabla(\mathbf{S}\cdot\mathbf{B})$ has components $S_a\,\partial_j B_a$. For $\mathbf{A} = \mathbf{0}$ the first part is $-\hbar^4\nabla^4/(8m_0^3c^2)$, the special-relativistic mass-velocity correction $-p^4/(8m_0^3c^2)$ [5, 84]. It has no gradient at fixed $\mathbf{p}$ and exerts no force; it lowers the group velocity to $(\mathbf{p}/m_0)(1 - p^2/2m_0^2c^2)$.

The Weyl symbol of the second part is $(q/2m_0^3c^2)\big[(\mathbf{S}\cdot\mathbf{B})\boldsymbol{\pi}^2 - \frac{1}{4}\hbar^2\nabla^2(\mathbf{S}\cdot\mathbf{B})\big]$. At leading order in $\hbar$, the part of its negative gradient at fixed $\mathbf{p}$ that contains the gradient of $\mathbf{B}$ is $-(v^2/2c^2)\nabla(\boldsymbol{\mu}\cdot\mathbf{B})$, with $\boldsymbol{\mu} = (q/m_0)\mathbf{S}$ and $\mathbf{v} = \boldsymbol{\pi}/m_0$. This part reduces the Stern–Gerlach force of Eq. (117) by the factor $1 - v^2/2c^2$.

The third part is the quadratic relativistic magnetic shift $-(q\hbar/2m_0)^2|\mathbf{B}|^2/(2m_0c^2)$. The Weyl symbol of $\hat{\boldsymbol{\pi}}^4$ is $(\boldsymbol{\pi}^2)^2 - \hbar^2q^2|\mathbf{B}|^2 + \hbar^2q\,\boldsymbol{\pi}\cdot\nabla^2\mathbf{A}$, so in the Weyl symbol of $\hbar\hat{\omega}$ the third part cancels the term $+\hbar^2q^2|\mathbf{B}|^2/(8m_0^3c^2)$ in the symbol of the first, and exerts no force.

8. *Darwin and spin-orbit term* $\hbar\,\omega_{so}$

In $f_W = V + \mathsf{G} - V_{\rm sc} - 2m_0c^2\ln W - \hat{\Pi}_g - \frac{1}{4}\hbar\,\boldsymbol{\sigma}\cdot\mathbf{B}_g$, Eq. (106) expands the position-dependent part (with $-\nabla V \mapsto q\mathbf{E} = -\nabla V - q\,\partial_t\mathbf{A}$ and $\mathbf{E}\times\hat{\boldsymbol{\pi}}$ in the Hermitian form of Section IV when $\mathbf{A}$ depends on time), and Eq. (108) gives the shift-strain Moyal part. Together, when the shift and the magnetic field are not both present, they give the expansion of the eighth term of Eq. (103):

$$\begin{aligned}
\hbar\,\omega_{\rm so} &= \frac{\hbar^2}{2m_0}\left(\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger - W^{-1}\mathsf{X}^2W^{-1}\right)\\
&= \underbrace{-\frac{q\hbar^2}{8m_0^2c^2}\,\nabla\cdot\mathbf{E}}_{\substack{\text{electric Darwin}\\ \text{(Zitterbewegung) contact term}}} - \underbrace{\frac{q}{2m_0^2c^2}\,\mathbf{S}\cdot(\mathbf{E}\times\hat{\boldsymbol{\pi}})}_{\substack{\text{electric spin-orbit /}\\ \text{Thomas precession } (g_s=2)}}\\
&\quad + \underbrace{\frac{\hbar^2}{8m_0^2c^2}\,\nabla^2\big(\mathsf{G} - V_{\rm sc} - 2m_0c^2\ln W\big)}_{\substack{\text{gravitational, scalar \&}\\ \text{curvature Darwin contact term}}} + \underbrace{\big(\gamma_{\rm PPN} + \tfrac{1}{2}\big)\,\frac{1}{m_0c^2}\,\mathbf{S}\cdot(\nabla\Phi_g\times\hat{\boldsymbol{\pi}})}_{\substack{\text{gravitational spin-orbit /}\\ \text{Fokker–de Sitter geodetic coupling}}}\\
&\quad + \underbrace{\frac{1}{2m_0c^2}\,\mathbf{S}\cdot(\mathbf{a}\times\hat{\boldsymbol{\pi}})}_{\substack{\text{inertial frame-acceleration}\\ \text{Thomas spin-orbit coupling}}} - \underbrace{\frac{1}{2m_0^2c^2}\,\mathbf{S}\cdot(\nabla V_{\rm sc}\times\hat{\boldsymbol{\pi}})}_{\substack{\text{Lorentz-scalar nuclear /}\\ \text{effective-mass spin-orbit}}}\\
&\quad - \underbrace{\frac{\hbar^2}{8m_0^2c^2}\,\nabla\cdot\mathbf{F}_\Sigma}_{\substack{\text{shift-strain}\\ \text{Darwin term}}} + \underbrace{\frac{1}{2m_0^2c^2}\,\mathbf{S}\cdot(\hat{\mathbf{p}}\times\mathbf{F}_\Sigma)}_{\substack{\text{shift-strain geodetic}\\ \text{spin-orbit coupling}}} + O(c^{-4}).
\end{aligned} \tag{122}$$

The first six sub-terms of Eq. (122) are already in bare form: each is a function of position, or a function of position times $\mathbf{S}$ standing to the left of $\hat{\boldsymbol{\pi}}$. The product $\mathbf{E}\times\hat{\boldsymbol{\pi}}$ is Hermitian when $\nabla\times\mathbf{E} = \mathbf{0}$, since $\hat{\boldsymbol{\pi}}\times\mathbf{E} = -\mathbf{E}\times\hat{\boldsymbol{\pi}} - i\hbar\,\nabla\times\mathbf{E}$. When $\mathbf{B}$ depends on time, so that $\nabla\times\mathbf{E} = -\partial_t\mathbf{B} \neq \mathbf{0}$, the Hermitian form of $\mathbf{E}\times\hat{\boldsymbol{\pi}}$ adds the bare term $-(iq\hbar/4m_0^2c^2)\,\mathbf{S}\cdot(\nabla\times\mathbf{E})$ to the electric spin-orbit term. The two shift-strain sub-terms are Weyl-ordered operators. The sub-terms act as follows, through $\mathbf{F} = -\hbar\nabla\omega$ and $d\mathbf{S}/dt = (i/\hbar)[\hat{h},\mathbf{S}]$:

(a) *Electric Darwin contact term* $-(q\hbar^2/8m_0^2c^2)\nabla\cdot\mathbf{E} = +(\hbar^2/8m_0^2c^2)\nabla^2 V$ [5, 84]. It arises from the diagonal entries of the 3×3 quaternion table (45), smears the electrostatic potential over the Compton wavelength $\hbar/(m_0c)$, and shifts the atomic s levels, whose wave functions do not vanish at the nucleus.

(b) *Electric Thomas spin-orbit coupling* $-(q/2m_0^2c^2)\mathbf{S}\cdot(\mathbf{E}\times\hat{\boldsymbol{\pi}}) = (2m_0^2c^2)^{-1}\mathbf{S}\cdot(\nabla V\times\hat{\boldsymbol{\pi}})$ [4, 5, 85]. It arises, with the exact Thomas factor $\frac{1}{2}$, from the off-diagonal entries of the quaternion table. It combines with $\hbar\omega_{\rm rel}$ and the Darwin term to give the hydrogen fine structure (Appendix A 7) [72, 84]. With a coupling renormalized by nearby bands, it has the form $\boldsymbol{\sigma}\cdot(\mathbf{k}\times\nabla V)$ that underlies the extrinsic spin Hall effect in solids [86], where $\hbar\mathbf{k}$ is the crystal momentum.

(c) *Gravitational, scalar, and curvature Darwin term* $+(\hbar^2/8m_0^2c^2)\nabla^2(\mathsf{G}-V_{\rm sc}-2m_0c^2\ln W)$ [9, 10]. It is the electric Darwin term with V replaced by $\mathsf{G}-V_{\rm sc}-2m_0c^2\ln W$, from the same diagonal entries. With $\mathsf{G}=m_0\Phi_g+m_0\mathbf{a}\cdot\mathbf{x}$ and $-2m_0c^2\ln W = 2\gamma_{\rm PPN}m_0\Phi_g$ at this order, it is $(\hbar^2/8m_0^2c^2)\big[(1+2\gamma_{\rm PPN})m_0\nabla^2\Phi_g-\nabla^2V_{\rm sc}\big]$. The $\hbar^2$ parts of the Weyl symbols of the metric and effective-mass groups of Eq. (114) add up to the negative of this, $-(\hbar^2/8m_0^2c^2)\big[(1+2\gamma_{\rm PPN})m_0\nabla^2\Phi_g-\nabla^2V_{\rm sc}\big]$, since $\nabla^2(\mathbf{a}\cdot\mathbf{x})=0$. In the Weyl symbol of $\hbar\hat{\omega}$ the two cancel, so through order c^{-2} the symbol has no gravitational, curvature or scalar contact term, and only the electric Darwin term of item (a) remains. The term belongs to the operator in the ordering of Eq. (103), as it does to the Foldy–Wouthuysen Hamiltonian of Obukhov *et al.* [10], but it exerts no contact force of its own.

(d) *Geodetic spin-orbit coupling* $+(\gamma_{\rm PPN}+\frac{1}{2})(m_0c^2)^{-1}\mathbf{S}\cdot(\nabla\Phi_g\times\hat{\boldsymbol{\pi}})$ [9, 77, 87, 88]: on static isotropic slices ($\mathbf{A}_g=\mathbf{0}$), $\mathsf{G}=m_0\Phi_g$ contributes $\frac{1}{2}$ (gravitational Thomas precession) and $-2m_0c^2\ln W = 2\gamma_{\rm PPN}m_0\Phi_g$ contributes $\gamma_{\rm PPN}$ from spatial curvature. The term precesses a gyroscope at the Fokker–de Sitter geodetic rate $\boldsymbol{\Omega}_{\rm spin}=(\gamma_{\rm PPN}+\frac{1}{2})(GM/c^2r^3)(\mathbf{x}\times\mathbf{v})$ measured by Gravity Probe B [77], and through $-\hbar\nabla\omega$ it exerts the Mathisson–Papapetrou spin-curvature force.

(e) *Inertial frame-acceleration Thomas spin-orbit coupling* $+(2m_0c^2)^{-1}\mathbf{S}\cdot(\mathbf{a}\times\hat{\boldsymbol{\pi}})$ [7, 85]: from $\nabla\mathsf{G}=m_0\mathbf{a}$ in a uniformly accelerated frame.

(f) *Lorentz-scalar nuclear spin-orbit coupling* $-(2m_0^2c^2)^{-1}\mathbf{S}\cdot(\nabla V_{\rm sc}\times\hat{\boldsymbol{\pi}})$ [69, 89]. Because $V_{\rm sc}$ enters f_W with a minus sign relative to V, an attractive scalar potential ($V_{\rm sc}<0$) and a repulsive vector potential ($V>0$) largely cancel in the central potential $V+V_{\rm sc}$ of Eqs. (109) and (110) and add in $V-V_{\rm sc}$ inside Eq. (122), which produces the strong nuclear shell-model spin-orbit splitting [69].

(g) *Shift-strain Darwin and geodetic spin-orbit terms* $-(\hbar^2/8m_0^2c^2)\nabla\cdot\mathbf{F}_\Sigma+(2m_0^2c^2)^{-1}\mathbf{S}\cdot(\hat{\mathbf{p}}\times\mathbf{F}_\Sigma)$ [10]: on spatially flat Painlevé–Gullstrand slices ($\alpha_n=W=1$, $\mathbf{A}_g=v_{\rm ff}\hat{\mathbf{r}}$), where $\nabla\Phi_g$ and $\nabla\ln W$ vanish, Eq. (108) with the canonical momentum $\mathbf{p}=m_0(\mathbf{v}+\mathbf{A}_g)$ of the group velocity $\mathbf{v}=\dot{\mathbf{x}}$ gives $\boldsymbol{\Omega}_{\rm spin}=(3GM/2c^2r^3)(\mathbf{x}\times\mathbf{v})$ on a circular orbit, and on general orbits after averaging over the orbit (Appendix A 7). This is the leading general-relativistic geodetic precession ($\gamma_{\rm PPN}=1$), obtained from the shift strain alone. Here $\nabla\cdot\mathbf{F}_\Sigma$ and $\hat{\mathbf{p}}\times\mathbf{F}_\Sigma$ denote the Weyl-ordered operators of the symbols $(\partial_i\Sigma_{ij})p_j$ and $\mathbf{p}\times\mathbf{F}_\Sigma$. Let $\mathbf{d}_\Sigma$ be the vector with components $(d_\Sigma)_j\equiv\partial_i\Sigma_{ij}$, so that $\mathbf{d}_\Sigma=\nabla(\nabla\cdot\mathbf{A}_g)-\frac{1}{2}\nabla\times\mathbf{B}_g$, and let $\hat{\mathbf{F}}_\Sigma$ be the operator with components $\Sigma_{ij}\hat{p}_j$. Writing each Weyl-ordered product with the functions of position on the left gives the bare forms

$$\nabla\cdot\mathbf{F}_\Sigma=\mathbf{d}_\Sigma\cdot\hat{\mathbf{p}}-\frac{i\hbar}{2}\nabla\cdot\mathbf{d}_\Sigma \tag{123}$$

and

$$\mathbf{S}\cdot(\hat{\mathbf{p}}\times\mathbf{F}_\Sigma)=-\mathbf{S}\cdot(\hat{\mathbf{F}}_\Sigma\times\hat{\mathbf{p}})-\frac{i\hbar}{4}\nabla(\mathbf{S}\cdot\mathbf{B}_g)\cdot\hat{\mathbf{p}}+\frac{i\hbar}{2}\mathbf{S}\cdot(\mathbf{d}_\Sigma\times\hat{\mathbf{p}})-\frac{\hbar^2}{8}\mathbf{S}\cdot\nabla^2\mathbf{B}_g, \tag{124}$$

where the derivation uses $\epsilon_{abc}\,\partial_b\Sigma_{cd}=\frac{1}{2}\partial_d B_{g,a}$ and $\nabla\times\mathbf{d}_\Sigma=\frac{1}{2}\nabla^2\mathbf{B}_g$. On the Painlevé–Gullstrand slices $\mathbf{B}_g=\mathbf{0}$ and $\nabla\cdot\mathbf{A}_g=3v_{\rm ff}/(2r)$, so $\mathbf{d}_\Sigma=\nabla(3v_{\rm ff}/2r)$ and only the first and third terms of Eq. (124) remain.

When the shift and the magnetic field are both present, the eighth term is $-[\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}},[\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}},f_W]]/(8m_0^2c^2)+O(c^{-4})$, and the part $-\hat{\Pi}_g$ of f_W adds mixed terms to Eq. (122). For a uniform translation $\mathbf{A}_g=u_{\rm tr}\,\hat{\mathbf{x}}$ at a constant speed $u_{\rm tr}$ in a uniform field $\mathbf{B}=B\,\hat{\mathbf{z}}$, $[\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}},[\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}},-u_{\rm tr}\hat{\pi}_x]]=2q\hbar u_{\rm tr}B(\sigma_x\hat{\pi}_z-\sigma_z\hat{\pi}_x)$, so the mixed term is $-q\hbar u_{\rm tr}B(\sigma_x\hat{\pi}_z-\sigma_z\hat{\pi}_x)/(4m_0^2c^2)$, although Σ_{ij} and $\mathbf{B}_g$ both vanish.

D. Orbital and spin parts

Where $\mathbf{E}$, $\mathbf{B}$, $\mathbf{B}_g$, the gradients of G, $V_{\rm sc}$ and W, and the symmetric part of the gradient of $\mathbf{A}_g$ vanish, every term of Eq. (103) that contains $\boldsymbol{\sigma}$ vanishes, and the spin is conserved. In general the terms of Eq. (103) are of two kinds, orbital and spin. The orbital terms contain no $\boldsymbol{\sigma}$: $\omega_{\rm geom}$, $\omega_{\rm ext}$, $\omega_{\rm kin}$, $\omega_{\rm gm}$, the scalar $-\hat{\boldsymbol{\pi}}^4/8m_0^3c^2$ and $-q^2\hbar^2|\mathbf{B}|^2/8m_0^3c^2$ parts of $\omega_{\rm rel}$, whose $|\mathbf{B}|^2$ terms cancel each other in the Weyl symbol, and the Darwin parts of $\omega_{\rm so}$. Their Weyl symbols are scalar functions of $\mathbf{x}$ and $\boldsymbol{\pi}$, and these terms act on the wave function through its phase. Their gradients $-\hbar\nabla\omega$ can change the wavevector $\mathbf{k}$, the momentum $\mathbf{p}$, and the orbital angular momentum $\mathbf{L}$ while leaving the spin unchanged. The spin terms ($\omega_{\rm Pauli}$, $\omega_{\rm rot}$, the $\{\hat{\boldsymbol{\pi}}^2, \mathbf{S}\cdot\mathbf{B}\}$ part of $\omega_{\rm rel}$, and the spin-orbit parts of $\omega_{\rm so}$) are linear in $\boldsymbol{\sigma}$ and multiply the wave function from the left. They precess $\mathbf{S}$, since $[\sigma_j, \sigma_k] = 2i\epsilon_{jkl}\sigma_l$, and produce Stern–Gerlach and spin-orbit forces wherever their fields vary in space.

The spin-orbit coupling $\zeta_{\rm so}\,\mathbf{S}\cdot\mathbf{L}$, with $\zeta_{\rm so} = (2m^2c^2r)^{-1}(dV/dr)$, changes the operators $\mathbf{S}$ at the rate $\zeta_{\rm so}\,\mathbf{L}\times\mathbf{S}$ and $\mathbf{L}$ at the rate $\zeta_{\rm so}\,\mathbf{S}\times\mathbf{L}$. These rates sum to zero, so the coupling exchanges angular momentum between the two and conserves $\mathbf{L}+\mathbf{S}$. Where $\mathbf{A}$ vanishes, $\mathbf{L} = \mathbf{x}\times\hbar\nabla\theta$ by Eq. (3), the moment of $\hbar$ times the phase gradient, and a force contributes the torque $\mathbf{x}\times\mathbf{F}$ to the rate of change of $\mathbf{L}$. A measurement of the spin along any axis gives $\pm\frac{1}{2}\hbar$, whatever forces act, and $\mathbf{S}$ changes only under the terms that multiply ψ from the left. Each of these terms turns $\mathbf{S}$, and the spin-orbit term also exchanges it with $\mathbf{L}$.

Every field in Eq. (92) and in its Taylor expansion Eq. (103) enters through one of the two covariant derivatives of the Dirac equation or through the mass. The potentials V and $\mathbf{A}$ are the coefficients of the time and space components of the gauge-covariant derivative (Eqs. (29) and (30)), which acts from the right and adds to $\hbar\omega$ (Remark 1). The lapse α_n, the shift $\mathbf{A}_g$, and the spatial scale factor W are the coefficients of the metric in the normal derivative D_n (Eq. (39)) and the curved-slice spatial operator X_W (Eq. (49)). By Eq. (90), the lapse multiplies the whole root of Eq. (92), the rest energy included; this is the multiplicative class of Table I. The mass $m(\mathbf{x}) = m_0 + V_{\rm sc}(\mathbf{x})/c^2$ belongs to neither covariant derivative: it couples φ_R to φ_L and gives the local rest frequency $\omega_0 = \kappa c = mc^2/\hbar$ that the lapse multiplies.

VI. ALL-SPEEDS WKB DYNAMICS AND BOHR–SOMMERFELD QUANTIZATION RULES

A. The exact $\hbar$ expansion and the all-speeds spin precession

Expanding the static equation in powers of $\hbar$, in place of c^{-2}, keeps all orders in v/c and in the field strengths, and requires that the fields vary slowly on the scale of the wavelength. On the curved conformally flat spatial slices of Eq. (35) with zero shift,

$$ds^2 = -\alpha_n^2c^2\,dt^2 + W^2\,d\mathbf{x}\cdot d\mathbf{x}, \tag{125}$$

the Hermitian Dirac Hamiltonian of Obukhov *et al.* [10] with minimal coupling to static potentials $V = q\Phi$ and $\mathbf{A}$ is

$$H = V + \beta\,mc^2\,\alpha_n + \tfrac{1}{2}\,c\,\{\alpha_n/W, \boldsymbol{\alpha}\cdot\hat{\boldsymbol{\pi}}\}, \tag{126}$$

where β is the Dirac matrix of the mass term, which is off-diagonal in the Weyl representation. Let $\tilde{\Psi} = \alpha_n^{1/2}\Psi = (\tilde{\psi}_L, \tilde{\psi}_R)^T$ be the rescaled bispinor. Define the local proper energy $\mathcal{E} \equiv (\hbar\omega - V)/\alpha_n$ and the symmetric operator $\mathfrak{B} \equiv W^{-1/2}(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})W^{-1/2} = \hbar\mathsf{X}_W$. The two rows of $H\Psi = \hbar\omega\,\Psi$ in the Weyl representation are equivalent to the pair

$$(\mathcal{E} - c\,\mathfrak{B})\,\tilde{\psi}_R = mc^2\,\tilde{\psi}_L, \tag{127}$$

$$(\mathcal{E} + c\,\mathfrak{B})\,\tilde{\psi}_L = mc^2\,\tilde{\psi}_R, \tag{128}$$

which, by Section III B, are Eqs. (54) and (55) on a stationary state with $\varphi_R = \tilde{\psi}_R$ and $\varphi_L = -\tilde{\psi}_L$. Applying $\mathcal{E} + c\mathfrak{B}$ to Eq. (127) and substituting Eq. (128) gives the exact second-order equation

$$\mathcal{E}^2\,\tilde{\psi}_R = (mc^2)^2\,\tilde{\psi}_R + c^2\,\mathfrak{B}^2\,\tilde{\psi}_R + c\,[\mathcal{E}, \mathfrak{B}]\,\tilde{\psi}_R, \tag{129}$$

whose commutator is

$$[\mathcal{E}, \mathfrak{B}] = i\hbar\,W^{-1}\,\boldsymbol{\sigma}\cdot\nabla\mathcal{E}, \tag{130}$$

where $\nabla\mathcal{E} = -(\nabla V + \mathcal{E}\nabla\alpha_n)/\alpha_n$. For any positive function g, $\hbar^2$ times Eq. (63), written in Pauli form by Eqs. (25), (22), and (33), is

$$\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}\;g\;\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}} = g\,\hat{\boldsymbol{\pi}}^2 - i\hbar\,(\nabla g)\cdot\hat{\boldsymbol{\pi}} + \hbar\,\boldsymbol{\sigma}\cdot(\nabla g\times\hat{\boldsymbol{\pi}}) - q\hbar\,g\,\boldsymbol{\sigma}\cdot\mathbf{B}, \tag{131}$$

whose third term is the spatial-curvature spin coupling of $\mathfrak{B}^2 = W^{-1/2}(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})W^{-1}(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})W^{-1/2}$. In the weak-field PPN metric ($\alpha_n = 1+\Phi_g/c^2$, $W^{-1} = 1+\gamma_{\rm PPN}\Phi_g/c^2$, $V=0$, $\Phi_g = -GM/r$), the spatial-curvature term of $c^2\mathfrak{B}^2$ and the lapse term of $c[\mathcal{E},\mathfrak{B}]$ combine to give the PPN geodetic precession of item (d) of Eq. (122) [50]:

$$\boldsymbol{\Omega}_{\rm spin} = \left(\gamma_{\rm PPN}+\tfrac{1}{2}\right)\frac{GM}{c^2 r^3}\,(\mathbf{x}\times\mathbf{v}). \tag{132}$$

Under the WKB ansatz $\tilde{\psi}_R = a(\mathbf{x})\,e^{iS(\mathbf{x})/\hbar}$, with real phase $S(\mathbf{x})$, two-component spinor amplitude $a(\mathbf{x})$, and $\boldsymbol{\pi}\equiv\nabla S - q\mathbf{A}$, the expansion of Eq. (129) in powers of $\hbar$ terminates exactly at order $\hbar^2$, because $\mathfrak{B}$ is a first-order differential operator. With $\mathfrak{B}_0\equiv W^{-1}\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$ and $\mathfrak{D}\chi\equiv W^{-1/2}\boldsymbol{\sigma}\cdot\nabla(W^{-1/2}\chi)$, dividing Eq. (129) by $e^{iS/\hbar}$ gives, with no approximation,

$$\left(\mathcal{E}^2-(mc^2)^2-\frac{c^2\boldsymbol{\pi}^2}{W^2}\right)a + i\hbar\left[c^2\left(\mathfrak{B}_0\mathfrak{D}+\mathfrak{D}\mathfrak{B}_0\right)-c\,W^{-1}\,\boldsymbol{\sigma}\cdot\nabla\mathcal{E}\right]a + \hbar^2c^2\,\mathfrak{D}^2 a = 0. \tag{133}$$

At order $\hbar^0$, the first term yields the relativistic Hamilton–Jacobi equation for $H_0(\mathbf{x},\mathbf{p}) = V+\alpha_n R$ with $R\equiv\sqrt{(mc^2)^2+c^2\boldsymbol{\pi}^2/W^2}$,

$$\hbar\omega = V+\alpha_n R, \tag{134}$$

whose characteristic rays satisfy $\dot{\mathbf{x}} = \alpha_n c^2\boldsymbol{\pi}/(W^2\mathcal{E})$ and $\dot{\boldsymbol{\pi}} = \alpha_n\nabla\mathcal{E} + (\alpha_n c^2\boldsymbol{\pi}^2/W^3\mathcal{E})\nabla W + q\,\dot{\mathbf{x}}\times\mathbf{B}$ with $R=\mathcal{E}$. At order $\hbar^1$, dividing the second term of Eq. (133) by $-2i\hbar\mathcal{E}/\alpha_n$ yields the amplitude transport equation along the characteristics:

$$\dot{\mathbf{x}}\cdot\nabla a = -\gamma_{\rm amp}\,a + \boldsymbol{\sigma}\cdot\mathbf{h}\,a - i\,\boldsymbol{\sigma}\cdot\mathbf{f}\,a - \tfrac{i}{2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\Omega}_B\,a, \tag{135}$$

where $\gamma_{\rm amp}\equiv(\alpha_n/2\mathcal{E})\nabla\cdot(\mathcal{E}\dot{\mathbf{x}}/\alpha_n)$ is the rate, of either sign, at which the amplitude decreases along the ray, $\mathbf{h}\equiv\alpha_n c\,\nabla\mathcal{E}/(2W\mathcal{E})$, $\mathbf{f}\equiv\frac{1}{2}\dot{\mathbf{x}}\times\nabla\ln W$, and $\boldsymbol{\Omega}_B\equiv -q\alpha_n c^2\mathbf{B}/(W^2\mathcal{E})$. Factor out the Lorentz boost by writing $a = e^K a_0$, where a_0 is the amplitude in the rest frame and $K\equiv\frac{1}{2}\zeta_{\rm r}\,\boldsymbol{\sigma}\cdot\hat{\mathbf{k}}$. The boost is along $\hat{\mathbf{k}}\equiv\boldsymbol{\pi}/|\boldsymbol{\pi}|$, with rapidity $\zeta_{\rm r}$ defined by $\cosh\zeta_{\rm r}\equiv\mathcal{E}/(mc^2)$ and $\sinh\zeta_{\rm r} = c|\boldsymbol{\pi}|/(Wmc^2)$. This change cancels the non-unitary boost term $\boldsymbol{\sigma}\cdot\mathbf{h}$ at every speed and reduces Eq. (135) to $da_0/dt = -\gamma_{\rm amp}a_0 - \frac{i}{2}\boldsymbol{\sigma}\cdot\boldsymbol{\Omega}_{\rm spin}a_0$. The magnitude of a_0 obeys $d|a_0|/dt = -\gamma_{\rm amp}|a_0|$, and the second term rotates the spin unitarily at the all-speeds precession rate

$$\boldsymbol{\Omega}_{\rm spin} = \dot{\mathbf{x}}\times\nabla\ln\left[W\left(\mathcal{E}+mc^2\right)\right] - \frac{q\,\alpha_n c^2}{W^2\mathcal{E}}\,\mathbf{B}. \tag{136}$$

Thus the unit spin vector $\hat{\mathbf{s}}\equiv a_0^\dagger\boldsymbol{\sigma}a_0/(a_0^\dagger a_0)$ obeys $d\hat{\mathbf{s}}/dt = \boldsymbol{\Omega}_{\rm spin}\times\hat{\mathbf{s}}$ and the probability current obeys $\nabla\cdot(a_0^\dagger a_0\,\mathcal{E}\dot{\mathbf{x}}/\alpha_n) = 0$. In flat spacetime ($\alpha_n = W = 1$), Eq. (136) is the relativistic Thomas–Bargmann–Michel–Telegdi (BMT) precession $\nabla V\times\mathbf{v}/(\mathcal{E}+mc^2) - qc^2\mathbf{B}/\mathcal{E}$ [21, 23]. For $V=0$ and $\mathbf{B}=\mathbf{0}$, Eq. (136) is the all-speeds gravitational spin precession in a static isotropic metric [10, 90, 91], which reduces to Eq. (132) at low speed. The effective semiclassical Hamiltonian is $H_0+\boldsymbol{\Omega}_{\rm spin}\cdot\mathbf{S}$; in flat spacetime it matches the first-order symbol of the electron block of Panati *et al.* [22].

Let (S,a) be a global solution of Eqs. (134) and (135) with a not identically zero, for which $\alpha_n^{-1/2}a$ and $\alpha_n^{1/2}\mathfrak{D}^2 a$ are square integrable. Let $\tilde{\psi}_R = a\,e^{iS/\hbar}$ and $\tilde{\psi}_L = (\mathcal{E}-c\mathfrak{B})\tilde{\psi}_R/(mc^2)$, and suppose that $\tilde{\Psi} = (\tilde{\psi}_L,\tilde{\psi}_R)^T$ lies in the domain of $\tilde{H}\equiv\alpha_n^{1/2}H\alpha_n^{-1/2}$. The map $\Psi\mapsto\alpha_n^{1/2}\Psi$ is unitary from the flat inner product to the inner product with weight α_n^{-1}, since $\int(\alpha_n^{1/2}\Psi)^\dagger(\alpha_n^{1/2}\Psi)\,\alpha_n^{-1}\,d^3x = \int\Psi^\dagger\Psi\,d^3x$. Thus $\tilde{H}$ is self-adjoint in the weighted inner product and has the spectrum of H. The equation $\tilde{H}\tilde{\Psi} = \hbar\omega\tilde{\Psi}$ is the pair of Eqs. (127) and (128). Because Eq. (133) terminates at order $\hbar^2$, $\tilde{\Psi}$ satisfies the exact quasimode spectral bound

$$\operatorname{dist}\left(\hbar\omega,\operatorname{spec}H\right)\le\frac{\hbar^2}{m}\,\frac{\left\|\alpha_n^{1/2}\,\mathfrak{D}^2 a\right\|}{\left\|\alpha_n^{-1/2}\,a\right\|}. \tag{137}$$

B. Relativistic Bohr–Sommerfeld quantization rules

For fields that vary along one Cartesian coordinate x, with turning points present, chiral elimination reduces the stationary Dirac problem on the whole line to an exact scalar second-order operator equation, whose Bohr–Sommerfeld rule approximates the Dirac eigenvalues to order $\hbar^2$.

Definition 1 (Two-way $O(\hbar^2)$ spectral approximation)**:** *Let J be a compact interval, $S_\hbar$ the eigenvalues of an operator that depends on $\hbar$, and $E_n^0(\hbar)$, for integers n, the solutions of a quantization rule. The set $S_\hbar$ is a* two-way $O(\hbar^2)$ spectral approximation *of the E_n^0 on J if for every compact J' inside the interior of J there are C and $\hbar_0 > 0$, independent of n, such that for $\hbar < \hbar_0$ every eigenvalue in J' lies within $C\hbar^2$ of some E_n^0 in J, and every E_n^0 in J' lies within $C\hbar^2$ of some eigenvalue in J. It is* one-to-one *if in addition exactly one eigenvalue, counted with multiplicity, lies within $C\hbar^2$ of each E_n^0 in J'.*

For a magnetic field $\mathbf{B} = B(x)\hat{\mathbf{z}}$ in flat spacetime with $\mathbf{A} = A_y(x)\hat{\mathbf{y}}$ ($A_y' = B$, where a prime denotes d/dx) and constant V, $[\mathcal{E}, \mathfrak{B}] = 0$. Let $\psi = e^{i(p_y y + p_z z)/\hbar} F(x)\check{\chi}_s$, where $\sigma_z \check{\chi}_s = s\check{\chi}_s$ ($s = \pm 1$) and the phases are chosen so that $\check{\chi}_{-s} = \sigma_x \check{\chi}_s$, and let $\varrho(x) \equiv qA_y(x) - p_y$ and $\hat{p} \equiv -i\hbar\, d/dx$. This separation reduces $(\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}})^2$ to $p_z^2 + \mathcal{P}_s$, where $\mathcal{P}_s \equiv \hat{p}^2 + \varrho^2 - s\,q\hbar B$ factors as $\mathcal{P}_{+1} = Q^\dagger Q$ and $\mathcal{P}_{-1} = QQ^\dagger$ with $Q \equiv \hat{p} - i\varrho$. Each eigenvalue Λ of $\mathcal{P}_s$ gives the exact Dirac energy

$$\hbar\omega = V + \sqrt{(mc^2)^2 + c^2\,(p_z^2 + \Lambda)}, \tag{138}$$

at every field strength. For a uniform field with $qB > 0$ and $A_y = Bx$, $\varrho = qB(x - p_y/qB)$, so $\hat{p}^2 + \varrho^2$ is a harmonic oscillator with eigenvalues $(2n_{\mathrm{L}} + 1)\,q\hbar B$ for integers $n_{\mathrm{L}} \geq 0$, and

$$\Lambda = (2n_{\mathrm{L}} + 1 - s)\,q\hbar B. \tag{139}$$

The level $n_{\mathrm{L}} = 0$, $s = +1$ has $\Lambda = 0$ and the frequency $(V + mc^2)/\hbar$ at every field, a consequence of the spin g-factor 2 of the Dirac equation. With $N \equiv n_{\mathrm{L}} + \frac{1}{2}(1 - s)$ and $p_z = 0$, Eq. (138) is $\hbar\omega = V + mc^2(1 + 2Nq\hbar B/m^2c^2)^{1/2}$. Its series in powers of $2Nq\hbar B/m^2c^2$ is the c^{-2} series of Section V for this level; it converges for $2Nq\hbar B \leq m^2c^2$ and diverges beyond. So for $B > m^2c^2/(2q\hbar)$ the series diverges for every level except the lowest, while Eq. (138) still holds. For an electron this field is 2.2×10^9 T. The carriers at the L point of bismuth obey the two-band Hamiltonian of Wolff [92], which has the form of a Dirac Hamiltonian with the half-gap Δ_{gap} in place of mc^2 and an anisotropic velocity in place of c. For a magnetic field in a fixed direction, its Landau levels at zero momentum along the field are Eq. (138) with Δ_{gap} in place of mc^2 and $\Delta_{\mathrm{gap}}/m_c$ in place of c^2, where m_c is the cyclotron mass of that direction [93]. The field $m^2c^2/(2q\hbar)$ then becomes $m_c\Delta_{\mathrm{gap}}/(2q\hbar)$, which is 0.30 T for the values $\Delta_{\mathrm{gap}} = 7$ meV and m_c one hundredth of the electron mass given by Fuseya *et al.* [93]. So in bismuth the series diverges for every level except the lowest once B exceeds 0.30 T.

For a real elliptic Weyl symbol $\mathfrak{p} = \mathfrak{p}_0 + \hbar\mathfrak{p}_1 + O(\hbar^2)$ whose level sets $\mathfrak{p}_0 = E$ are closed curves without critical points, the Bohr–Sommerfeld rule of Ifa *et al.* [94, 95] is

$$\oint_{\mathfrak{p}_0 = E} \xi\, dx = 2\pi\hbar\left(n + \tfrac{1}{2}\right) + \hbar \oint \mathfrak{p}_1\, dt_{\mathfrak{p}} + O(\hbar^2), \tag{140}$$

where ξ is the momentum conjugate to x and $t_{\mathfrak{p}}$ is the time of the Hamiltonian flow of $\mathfrak{p}_0$. For $\mathcal{P}_s$, $\mathfrak{p}_0 = \xi^2 + \varrho^2$ and $\mathfrak{p}_1 = -s\,qB$. Along the flow of $\mathfrak{p}_0$, $dx/dt_{\mathfrak{p}} = 2\xi$ and $d\xi/dt_{\mathfrak{p}} = -2\varrho\varrho'$ with $\varrho' = qA_y' = qB$, so $d\varrho/dt_{\mathfrak{p}} = 2qB\,\xi$ and the angle of (ξ, ϱ) about the origin increases at the rate $2qB$. One period is one turn, so $\oint qB\, dt_{\mathfrak{p}} = \pi$ for every profile with $qB(x) > 0$, and Eq. (140) becomes

$$\oint_{\mathfrak{p}_0 = \Lambda} \xi\, dx = 2\pi\hbar\left(n + \tfrac{1}{2} - \tfrac{1}{2}s\right) + O(\hbar^2). \tag{141}$$

The squared one-dimensional Dirac operator with a mass profile has blocks of the form of $\mathcal{P}_s$, with the mass in place of ϱ, and Sutton and Watson [96] derive the same shifted rule for them by a formal WKB construction. For massless electrons in graphene in a field $B(x)$, Kormányos *et al.* [97] find the corresponding phase as an orbit integral of the field, equal to π for piecewise constant profiles.

The supersymmetric pairing of $\mathcal{P}_{+1} = Q^\dagger Q$ and $\mathcal{P}_{-1} = QQ^\dagger$ [98] makes the $n = 0$, $s = +1$ level $\hbar\omega = V + \sqrt{(mc^2)^2 + c^2p_z^2}$ exact and independent of the field profile $B(x)$, and maps the exact Foldy–Wouthuysen wave function $F\check{\chi}_s$ to the Dirac state with the relativistic centroid offset

$$\langle x\rangle_{\mathrm{Dirac}} = \langle x\rangle - \frac{s\,\hbar c^2\,\langle\varrho\rangle}{2\mathcal{E}(\mathcal{E} + mc^2)}. \tag{142}$$

When an electric potential $V(x)$ is added at $p_z = 0$, set $\tilde{\psi}_+ \equiv \tilde{\psi}_R + \tilde{\psi}_L = F\check{\chi}_s$ and $\tilde{\psi}_- \equiv \tilde{\psi}_R - \tilde{\psi}_L = c\,g\,(Q_sF)\check{\chi}_{-s}$, with $g \equiv 1/[W(\mathcal{E} + mc^2)]$, $Q_{+1} \equiv Q$, and $Q_{-1} \equiv Q^\dagger$. Then $\hbar\omega$ is a Dirac eigenvalue if and only if 0 is an eigenvalue of the energy-dependent scalar operator $\mathsf{K}(\hbar\omega) \equiv c^2Q_s^\dagger g Q_s - W(\mathcal{E} - mc^2)$. Here $\alpha_n = W = 1$, so $\mathcal{E} = \hbar\omega - V$ and

$g = 1/(\hbar\omega - V + mc^2)$. The Weyl symbol $\mathsf{K}_0 + \hbar\mathsf{K}_1 + \hbar^2\mathsf{K}_2$ of K has $\mathsf{K}_0 = c^2 g(\xi^2 + \varrho^2) - (\hbar\omega - V - mc^2)$, $\mathsf{K}_1 = -sc^2(g\varrho)'$, and $\mathsf{K}_2 = \frac{1}{4}c^2 g''$. Let t_0 be the time of the flow of K_0. Along the orbit $H_0 = \hbar\omega$ ($\mathsf{K}_0 = 0$), $\mathsf{K}_1\, dt_0 = \frac{1}{2} s\, \Omega_z\, dt$, where Ω_z is the z component of $\mathbf{\Omega}_{\text{spin}}$ in Eq. (136) and t is the time of the flow of H_0. This relation gives the quantization rule

$$\oint_{H_0 = \hbar\omega} \xi\, dx = 2\pi\hbar \left(n + \tfrac{1}{2}\right) + \tfrac{1}{2}\, s\, \hbar\, \Theta_{\text{spin}} + O(\hbar^2), \tag{143}$$

where $\Theta_{\text{spin}} \equiv \oint \Omega_z\, dt$ is the *spin angle* [23, 26, 99, 100]. In terms of the polar angles $(\theta_{\mathbf{n}}, \phi_{\mathbf{n}})$ of the unit vector $\mathbf{n} \equiv (c\xi, c\varrho, mc^2)/\mathcal{E}$ on the orbit ($\cos\theta_{\mathbf{n}} = mc^2/\mathcal{E}$, $\phi_{\mathbf{n}} = \arg(\xi + i\varrho)$), Θ_{spin} splits into the solid angle enclosed by $\mathbf{n}$ on the upper hemisphere (the Berry phase) and an orbital-moment integral proportional to mc^2 (the Rammal–Wilkinson phase):

$$\Theta_{\text{spin}} = -\oint (1 - \cos\theta_{\mathbf{n}})\, d\phi_{\mathbf{n}} - \oint \frac{mc^4\, qB}{\mathcal{E}^2}\, dt. \tag{144}$$

In crossed fields $\mathbf{B} = B(x)\hat{\mathbf{z}}$ and $V = \beta_{\text{d}} c\, qA_y$ ($0 < \beta_{\text{d}} < 1$), consider a Lorentz boost along $\hat{\mathbf{y}}$ at the drift speed $\beta_{\text{d}} c$, where $\gamma_{\text{d}} \equiv (1 - \beta_{\text{d}}^2)^{-1/2}$ and $E_{\text{d}} \equiv \hbar\omega - \beta_{\text{d}} c\, p_y$. This boost, used by Lukose *et al.* [101] for massless electrons in graphene in uniform crossed fields, eliminates the electric field in the drift frame and maps each orbit to $\xi^2 + \varrho_{\text{d}}^2 = \Lambda$ for the field B/γ_{d}. Here $\varrho_{\text{d}} \equiv qA_y/\gamma_{\text{d}} - b_{\text{d}}$, $b_{\text{d}} \equiv p_y/\gamma_{\text{d}} - \gamma_{\text{d}}\beta_{\text{d}} E_{\text{d}}/c$, and $c^2\Lambda = (\gamma_{\text{d}} E_{\text{d}})^2 - (mc^2)^2$. For a uniform magnetic field B, Eq. (143) without its remainder gives the exact Dirac levels

$$E_{\text{d}}^2 = (1 - \beta_{\text{d}}^2) \left[(mc^2)^2 + (2n + 1 - s)\, \hbar q B\, c^2\, (1 - \beta_{\text{d}}^2)^{1/2}\right], \tag{145}$$

and for the inhomogeneous profile $qA_y = \mathfrak{a} \tanh(x/L)$ ($\mathfrak{a}, L > 0$), shape invariance of the drift-frame operator [98] makes Eq. (143) exact with zero remainder as well. Set $\mathfrak{a}_{\text{d}} = \mathfrak{a}/\gamma_{\text{d}}$, $\mathcal{N}_k = \mathfrak{a}_{\text{d}} - k\hbar/L$, and $N = n + \frac{1}{2}(1 - s)$. Here N is the integer defined for the uniform field, with n in place of n_{L}. The levels of the branch $E_{\text{d}} > 0$ are

$$(\gamma_{\text{d}} E_{\text{d}})^2 = (mc^2)^2 + c^2 \left(\mathfrak{a}_{\text{d}}^2 + b_{\text{d}}^2 - \mathcal{N}_N^2 - \frac{\mathfrak{a}_{\text{d}}^2 b_{\text{d}}^2}{\mathcal{N}_N^2}\right) \tag{146}$$

for the integers $n \geq 0$ with $\mathcal{N}_N > \sqrt{\mathfrak{a}_{\text{d}} |b_{\text{d}}|}$. For the massless tilted-cone Hamiltonian of a two-dimensional Dirac material, O-Campa *et al.* [102] solve exactly the case of electric and magnetic fields both proportional to $\operatorname{sech}^2(x/L)$, in terms of Heun functions.

Finally, in the static curved metric of Eq. (125) with smooth positive $\alpha_n(x)$ and $W(x)$ bounded above and below with bounded derivatives, setting $\tilde{\psi}_+ = W^{1/2} F \check{\chi}_s$ reduces the stationary Dirac equation to $\mathsf{K}F = 0$, with K and g defined as in the paragraph before Eq. (143), now with general α_n and W. With $\Upsilon \equiv \ln[W(\mathcal{E} + mc^2)]$ from Eq. (136), the spin angle is $\Theta_{\text{spin}} = \oint \Omega_z\, dt = \oint (\varrho\Upsilon' - qB)\, \xi^{-1} dx$, and the unit vector on the orbit is $\mathbf{n} \equiv (c\xi/W, c\varrho/W, mc^2)/\mathcal{E}$.

Appendix B states and proves four theorems on how closely these rules approximate the Dirac levels, each under hypotheses on the smoothness and boundedness of the fields stated with it. The proofs use Olver's uniform Airy approximation and Sturm comparison. For a magnetic field along one direction, the eigenvalues of $\mathcal{P}_s$ in each compact interval above 0 and below the threshold of its continuum, and hence the Dirac levels through Eq. (138), are simple and are a one-to-one $O(\hbar^2)$ spectral approximation (Definition 1) of the solutions of Eq. (141) without its remainder (Theorem 1). The same holds for Eq. (143) with an electric potential $V(x)$ (Theorem 2), and in the static metric with $H_0 = V + \alpha_n R$ (Theorem 4). In crossed fields with a general profile the approximation is two way, and $\Theta_{\text{spin}} = -2\pi$ on every closed orbit with $\mathcal{E} \geq mc^2$ (Theorem 3).

Theorem 4 applies equally to a weak static gravitational field and to a gapped two-dimensional Dirac material such as strained graphene whose Fermi velocity $v_{\text{F}}(x)$ varies with position [103]. Setting $\alpha_n = 1$, $W = c/v_{\text{F}}(x)$, and $\Delta_{\text{gap}} = mc^2 > 0$ in Eq. (126) gives for each $s = \pm 1$ the operator $V\mathbf{1} + \frac{1}{2}\{v_{\text{F}}, \hat{p}\}\sigma_1 - s v_{\text{F}} \varrho \sigma_2 + \Delta_{\text{gap}}\sigma_3$ on a pair of functions of x. Theorem 4 gives its Landau levels one-to-one to $O(\hbar^2)$ for any smooth velocity profile $v_{\text{F}}(x)$ that lies between two positive constants and has bounded derivatives, and for any potential $V(x)$, magnetic profile $B(x)$, and energy window that meet the hypotheses of the theorem, with $\Theta_{\text{spin}} = -2\pi$ whenever V is constant.

The *root rule* is the Bohr–Sommerfeld quantization of the root of Eq. (92) without C. At $p_z = 0$, on the block of spin s, the root acts as a scalar operator in x. The operator $\alpha_n^{1/2}$ times the root times $\alpha_n^{1/2}$ has the spectrum of α_n times the root and is Hermitian in the unweighted inner product. Its Weyl symbol is $\alpha_n R - s\hbar\, qB\, \alpha_n c^2/(2W^2 R) + O(\hbar^2)$. Along the orbit $R = \mathcal{E}$ and $dt = W^2 \mathcal{E}\, dx/(\alpha_n c^2 \xi)$, so Eq. (140) applied to V plus this symbol gives Eq. (143) with Θ_{spin} replaced by its magnetic part $-\oint qB\, \xi^{-1} dx$. The root rule therefore omits the orbit integral of the electric and gravitational spin-orbit terms,

$$\Theta_W \equiv \oint \varrho\, \Upsilon'\, \xi^{-1} dx.$$

Both rules quantize the same action $\oint_{H_0=E} \xi\, dx$, whose derivative with respect to E is the period T of the orbit. So the solutions E_{root} of the root rule and E_{full} of Eq. (143), both without their remainders and with the same n, differ by

$$E_{\text{root}} - E_{\text{full}} = -\frac{s\hbar\,\Theta_W}{2T} + O(\hbar^2). \tag{147}$$

Taking T and Θ_W on the orbit of either level changes the right side only at $O(\hbar^2)$. The shift vanishes at first order exactly when $\Theta_W = 0$, as for a magnetic field with constant V, α_n and W.

Consider an orbit small compared with the lengths over which qB, V, α_n and W vary, with $qB > 0$. Let x_c be the midpoint of its turning points, and normalize α_n and W to 1 at x_c. Let $\mathcal{E}_c$ and $p_\perp$ be the values of $\mathcal{E}$ and ξ at x_c, and let ϱ_c be the mean of the values of ϱ at the two turning points. To first order in $x - x_c$, $\varrho - \varrho_c = qB\,(x - x_c)$, and the orbit $\xi^2 + \varrho^2 = W^2(\mathcal{E}^2 - m^2c^4)/c^2$ is the circle $\xi^2 + (\varrho - \varrho_c)^2 = p_\perp^2$ in the (ξ, ϱ) plane, with

$$\varrho_c = \frac{W'(\mathcal{E}_c^2 - m^2c^4) + \mathcal{E}_c\mathcal{E}'}{c^2 qB}$$

at x_c. The center is moved off $\varrho = 0$ by the drift of the orbit in the electric and gravitational fields. The second-order terms make the circle an ellipse and change its period. They are negligible when $\left|\frac{1}{2}[W^2(\mathcal{E}^2 - m^2c^4)/c^2]'' - \varrho_c(qB)'\right| \ll (qB)^2$ at x_c. For a slow particle in a uniform magnetic field in flat spacetime this requires $m|V''| \ll (qB)^2$, an electric force whose change across the orbit is small compared with the magnetic force, and $|V'| \ll c\,qB$, a slow drift. On the circle $dx = d\varrho/(qB)$, $\oint d\varrho/\xi = 2\pi$ and $\oint(\varrho - \varrho_c)^2\, d\varrho/\xi = \pi p_\perp^2$. Expanding $\Upsilon'/(qB)$ to first order in $x - x_c$ then gives

$$\Theta_W = \frac{\pi}{qB}\left[p_\perp^2\left(\frac{\Upsilon'}{qB}\right)' + 2\varrho_c\Upsilon'\right],$$

and the period of the circle is $T = 2\pi\mathcal{E}_c/(c^2qB)$. Equation (147) gives

$$|E_{\text{root}} - E_{\text{full}}| \simeq \frac{\hbar\,c^2}{4\mathcal{E}_c}\left|p_\perp^2\left(\frac{\Upsilon'}{qB}\right)' + 2\varrho_c\Upsilon'\right|.$$

For a slow particle in a uniform magnetic field in flat spacetime, $\Upsilon' = -V'/(\mathcal{E} + mc^2)$ and $\varrho_c = -mV'/(qB)$, and the shift is $\hbar\,|p_\perp^2 V'' - 2mV'^2|/(8m^2c^2qB)$. It is of the order of the spin-orbit energy and is nonzero for a generic potential. In the weak static field of Eq. (132) with constant V, a slow particle has $\Upsilon' = -(\gamma_{\text{PPN}} + \frac{1}{2})\,\Phi_g'/c^2$ and $\varrho_c = -m^2\Phi_g'/(qB)$, and the shift in a uniform magnetic field is $(\gamma_{\text{PPN}} + \frac{1}{2})\,\hbar\,|p_\perp^2\Phi_g'' - 2m^2\Phi_g'^2|/(4mc^2qB)$. In a uniform gravitational field only the drift term remains. At leading order the coefficient $\gamma_{\text{PPN}} + \frac{1}{2}$ is also the one in the geodetic precession rate of Eq. (132).

VII. ROTOR INTERPRETATION, GENERAL METRICS, AND NON-ABELIAN GAUGE FIELDS

A. Polar rotor decomposition, phase frame, and causal wavefronts

Every nonzero quaternion factors into its positive scalar norm $\rho^{1/2}$ ($\rho = \psi\bar{\psi} = |\psi|^2$) and a unit quaternion (*rotor*) $\mathcal{R}_\psi = \psi/\rho^{1/2}$ ($\mathcal{R}_\psi\bar{\mathcal{R}}_\psi = 1$):

$$\psi = \rho^{1/2}\,\mathcal{R}_\psi. \tag{148}$$
(The polar form of the wave function)

Substituting Eq. (148) into the spin density vector of Eq. (26) gives

$$\mathbf{s}\cdot\mathbf{e} = \rho\,\mathcal{R}_\psi\, e_3\,\bar{\mathcal{R}}_\psi. \tag{149}$$

The rotor $\mathcal{R}_\psi$ maps each unit e_k of the fixed reference triad to $\mathcal{R}_\psi e_k\bar{\mathcal{R}}_\psi$ ($k = 1, 2, 3$), a local orthonormal triad at each point in space. The third axis is the local spin direction $\mathbf{s}/\rho$. A right phase factor $\psi \mapsto \psi\exp(e_3\phi)$ rotates the two transverse axes $\mathcal{R}_\psi e_1\bar{\mathcal{R}}_\psi$ and $\mathcal{R}_\psi e_2\bar{\mathcal{R}}_\psi$ in the plane perpendicular to $\mathbf{s}$ through the angle 2ϕ [15, 104]. On a stationary state $\psi = \psi_0\exp(e_3\omega t)$, the transverse axes rotate about the spin axis at the angular rate 2ω, the same at every point ($2mc^2/\hbar$ at rest in flat spacetime). Here ω is the phase rate of the chosen state. The force-frequency identity of Eq. (8) instead differentiates the phase-space function $\omega(\mathbf{x}, \mathbf{p})$ at fixed canonical momentum. For $H = mc^2 + \mathbf{p}^2/2m + V$, Eq. (11) shows that the phase rate differs from $\omega(\mathbf{x}, \hbar\nabla\theta)$ by $V_Q/\hbar$.

In the 4×4 table $(\bar{e}_\mu e_\nu)$ of Eq. (66), the four diagonal scalars 1 propagate all four components of ψ without turning the frame $\mathcal{R}_\psi$, whereas the twelve off-diagonal entries $\pm e_l$ correspond to the six coordinate planes of spacetime (three time-space planes $0k$ on the border and three space-space planes jk in the interior). Each pair of these entries multiplies the term $\mathsf{N}_\mu^-\mathsf{N}_\nu^+\psi-\mathsf{N}_\nu^-\mathsf{N}_\mu^+\psi$ of Eq. (67) on the left by $\pm e_l$, which turns the frame $\mathcal{R}_\psi$.

Because $[\mathsf{N},\mathsf{X}_W]$ in Eq. (58) contains at most first derivatives, the principal symbol of the second-order quaternionic wave equation (57) is that of $\mathsf{N}^2-\mathsf{X}_W^2$, whose second-derivative terms are those of $-D_n^2+W^{-2}\mathbf{D}^2$. For a wavefront $\psi=\psi_{\text{front}}(\hat{\mathbf{n}}\cdot\mathbf{x}-v_{\text{front}}t)$ with unit normal $\hat{\mathbf{n}}$, the leading second-derivative part vanishes when $(v_{\text{front}}+\mathbf{A}_g\cdot\hat{\mathbf{n}})^2=(c\alpha_n/W)^2$. Relative to the local Eulerian normal observer, with proper time $d\tau_n=\alpha_n\,dt$ and proper distance $d\ell=W\,|d\mathbf{x}|$, every wavefront therefore propagates at the speed c [105].

B. Comparison with earlier formulations

Table III compares the present quaternionic formulation with five standard formulations of relativistic spin-$\frac{1}{2}$ dynamics [4–6, 8–11, 14–19, 104, 106].

TABLE III. Comparison of the present quaternionic formulation with five earlier formulations of relativistic spin-$\frac{1}{2}$ dynamics.

Formulation	**Algebra and state**	**Field and spin-orbit mechanism**	**Curved metric and $\hat{h}$**
Dirac / Foldy–Wouthuysen [4, 5]	4×4 complex matrices; 4-component bispinor (8 real)	Unitary series; nested commutator $-[\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}},[\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}},\hat{V}]]/8m^2c^2$	Flat spacetime; closed form [6], c^{-2} series in practice
Feynman–Gell-Mann [19]	2×2 complex Pauli matrices; 2-component spinor (4 real)	Second-order equation with complex $\boldsymbol{\sigma}\cdot(\mathbf{B}-i\mathbf{E}/c)$	Flat spacetime; no reduction to $\hat{h}$
Hestenes / Lasenby et al. [15, 18]	16-real-dim spacetime algebra; 8-real even multivector	First-order multivector equation; left rotor, right phase	Tetrad gauge gravity; no folded $\tilde{\mathsf{X}}\tilde{\mathsf{X}}^\dagger$ or $\hat{h}$
Lanczos / De Leo–Rodrigues [14, 17]	8-real biquaternions or 2 real quaternions (8 real)	First-order pair in Dirac basis	Flat spacetime; no exact Riccati or dressed root
Obukhov–Silenko–Teryaev [10]	4×4 complex matrices; 4-component bispinor (8 real)	4×4 Foldy–Wouthuysen transformation	Exact curved Dirac H; perturbative FW reduction
Present work (Eqs. (57), (67), (101))	4-real division algebra of quaternions; 1 quaternion $\psi=\rho^{1/2}\mathcal{R}_\psi$ (4 real)	$(\mathsf{N}-\mathsf{X}_W)(\mathsf{N}+\mathsf{X}_W)$ on 4×4 table $(\bar{e}_\mu e_\nu)$; dressed $\tilde{\mathsf{X}}_W\tilde{\mathsf{X}}_W^\dagger$ on 3×3 block	Lapse, shift, and spatial metric in D_n and the spatial operator; exact Riccati $\hat{h}$

C. General anisotropic spatial metrics

Consider an arbitrary anisotropic spatial metric $h_{jk}(\mathbf{x})$ in place of the conformally flat spatial metric $W^2\delta_{jk}$ of Eq. (35), with inverse $h^{jk}=\sum_{\hat{a}=1}^3\mathfrak{e}_{\hat{a}}^j(\mathbf{x})\mathfrak{e}_{\hat{a}}^k(\mathbf{x})$ built from an orthonormal spatial triad $\mathfrak{e}_{\hat{a}}^k(\mathbf{x})$ $(\hat{a}=1,2,3)$ [11]. Rotating sources need this case: the Kerr spacetime has no axisymmetric conformally flat slices that reduce smoothly, in the Schwarzschild limit, to the constant-time slices of the Schwarzschild spacetime [107]. Absorbing $(\det h)^{1/4}$ alongside $\alpha_n^{1/2}$ into ψ keeps the conserved inner product in the form $\int\phi^\dagger\chi\,\alpha_n^{-1}\,d^3x$, with no factor $(\det h)^{1/2}$. The constant quaternion units $e_{\hat{a}}$ belong to the local orthonormal triad frame, and $\mathsf{X}_W\psi$ of Eq. (49) generalizes to

$\mathsf{X}_h\psi \equiv (\det h)^{1/4}\sum_{\hat a,k} e_{\hat a}\,\mathfrak{e}^k_{\hat a}\,(D_k+\Gamma_k)\big[(\det h)^{-1/4}\psi\big]$, where Γ_k is the spatial spin connection, which acts on the left and rotates $\mathcal{R}_\psi$ as the triad turns in space. For $\mathfrak{e}^k_{\hat a}=W^{-1}\delta^k_{\hat a}$, $(\det h)^{1/4}=W^{3/2}$ and $\sum_k e_k W^{-1}\Gamma_k = W^{-2}\,\mathbf{e}\cdot\nabla W$, and the operator reduces to $\frac12\{W^{-1},\mathsf{X}\}$, which is Eq. (49). Let D_n in $\mathsf{N}\psi=(D_n\psi)\,e_3$ be the normal derivative built from the unit normal and this triad. With X_h in place of X_W, the chiral pair (54)–(55) and its product, Eq. (57), hold on these slices.

After the right gauge terms of D_k are set aside, the terms of X_h without a derivative multiply ψ on the left by a quaternion-valued function. Let $\mathcal{S}(\mathbf{x})$ be the real part of this function, the coefficient of the unit 1. Its vector part joins the derivative terms to give $\mathcal{S}\,\psi+\sum_{\hat a} e_{\hat a}\,\tilde D_{\hat a}\psi$; for conformally flat slices $\tilde D_{\hat a}$ is the $\tilde D_k$ of Section III C. Each $\tilde D_{\hat a}$ is a first-order operator with real coefficients plus the right gauge terms of D_k, so it commutes with left multiplication by a constant unit. The normal derivative D_n built from this triad is split in the same way: $D_n\psi=\mathring D_n\psi+\sum_{\hat a} e_{\hat a}\,b_{\hat a}\,\psi$, where the last term collects the terms that multiply ψ on the left by a vector. As in Eq. (39), $\mathring D_n$ has real coefficients apart from the right gauge terms, so it also commutes with left multiplication by a constant unit. Let $\mathring{\mathsf{N}}\psi\equiv(\mathring D_n\psi)\,e_3$ with this $\mathring D_n$. The function $\mathcal{S}$ therefore joins $\mathring{\mathsf{N}}$ as the coefficient of the unit 1 in the table (66), and Eq. (67) holds with $\mathsf{N}^\pm_0\psi=\mathring{\mathsf{N}}\psi\pm\mathcal{S}\,\psi$ and $\mathsf{N}^\pm_{\hat a}\psi=\tilde D_{\hat a}\psi\pm b_{\hat a}\,\psi\,e_3$. For the triad $\mathfrak{e}^k_{\hat a}=W^{-1}\delta^k_{\hat a}$ of conformally flat slices, $\sum_{\hat a} e_{\hat a}\,b_{\hat a}=\mathbf{e}\cdot\mathbf{b}$, and $\mathcal{S}=0$, since the quaternion-valued function is then $-\frac12 W^{-2}\,\mathbf{e}\cdot\nabla W$, a pure vector.

Turning the triad by a stationary rotor field $U(\mathbf{x})$, with ψ replaced by $U\psi$, changes $\sum_{\hat a} e_{\hat a}\,b_{\hat a}$ to $U^{-1}\big(\sum_{\hat a} e_{\hat a}\,b_{\hat a}\big)U - U^{-1}(\mathbf{A}_g\cdot\nabla U)/(c\alpha_n)$, so $b_{\hat a}$ depends on the triad as well as on the metric.

In flat space, consider the helical triad $\mathfrak{e}^k_{\hat 1}=(\cos k_h z,\sin k_h z,0)$, $\mathfrak{e}^k_{\hat 2}=(-\sin k_h z,\cos k_h z,0)$, $\mathfrak{e}^k_{\hat 3}=(0,0,1)$, with constant wavenumber k_h. The left rotation $\psi\mapsto U\psi$ with $U=\cos(k_h z/2)+e_3\sin(k_h z/2)$ gives $U^{-1}\mathsf{X}(U\psi)=\sum_{\hat a,k}e_{\hat a}\,\mathfrak{e}^k_{\hat a}\,D_k\psi-\frac12 k_h\,\psi$, so $\mathcal{S}=-k_h/2$. The triad $\mathfrak{e}^k_{\hat a}=\delta^k_{\hat a}$ of the same space gives $\mathcal{S}=0$, so $\mathcal{S}$ also depends on the triad. With $\alpha_n=1$ and the constant shift $\mathbf{A}_g=u_{\rm sh}\,\hat{\mathbf{z}}$ of speed $u_{\rm sh}$, for which $\mathbf{B}_g=\mathbf{0}$, the same rotation gives $b_{\hat 3}=-u_{\rm sh}k_h/(2c)$.

The proper volume element of a slice is $(\det h)^{1/2}d^3x$, so the divergence term inside the bracket of the normal derivative, which carries the factor $1/(c\alpha_n)$ of Eq. (39), is $-\frac12(\det h)^{-1/2}\nabla\cdot\big((\det h)^{1/2}\mathbf{A}_g\big)\psi_o$. The identity before Eq. (39) holds with any positive function in place of W^3, so with $(\det h)^{1/2}$ it puts the transport and divergence terms of $\mathring D_n\psi$ into their flat-slice form. Since $e_{\hat a}\psi=(\sigma_{\hat a}\psi)\,e_3$, on a stationary state Eq. (80) holds with

$$\hat V\equiv V-\hat\Pi_g-\hbar c\,\alpha_n\sum_{\hat a}\sigma_{\hat a}\,b_{\hat a},$$

which reduces to the $\hat V$ of Eq. (79) on conformally flat slices. The operator $\hbar\mathsf{X}_h$ is $(\det h)^{1/4}$ times the spatial Dirac operator of the slice times $(\det h)^{-1/4}$. The spatial Dirac operator is symmetric in the inner product $\int\phi^\dagger\chi\,(\det h)^{1/2}d^3x$ of the slice, so $\hbar\mathsf{X}_h$ is symmetric in $\int\phi^\dagger\chi\,d^3x$, and $\mathsf{P}\equiv\alpha_n\hbar\mathsf{X}_h$ is Hermitian in the weighted product $\int\phi^\dagger\chi\,\alpha_n^{-1}d^3x$. With this $\hat V$ and this P, Eqs. (81), (82) and (83) hold in every stationary metric whose positive-frequency subspace is a graph $(u,\mathsf{R}u)$. The existence and uniqueness proof of Section IV B applies under its condition $\epsilon<1$. In weak fields outside the proved cases the graph remains an assumption, as it is on conformally flat slices. On those slices the frequency equation is Eq. (92), and when P, G and $\hat V$ commute it reduces to Eq. (90).

At order $\hbar^0$ the symbol of P is $\alpha_n\sum_{\hat a,k}\sigma_{\hat a}\mathfrak{e}^k_{\hat a}\pi_k$, whose square is $\alpha_n^2 h^{jk}\pi_j\pi_k$. The term $\hbar c\,\alpha_n\sum_{\hat a}\sigma_{\hat a}b_{\hat a}$ of $\hat V$ is of order $\hbar$. So Eqs. (94), (95) and (96) hold with $h^{jk}\pi_j\pi_k$ in place of $W^{-2}|\boldsymbol{\pi}|^2$, with $h^{jk}\pi_k$ in place of the component $W^{-2}\pi_j$ in Eq. (95), and with $h_{jk}A^j_gA^k_g$ in place of $W^2|\mathbf{A}_g|^2$ in Eq. (36). The rest momentum is then $\pi_j=\mathcal{H}h_{jk}A^k_g/(\alpha_n^2c^2)$. Under a stationary triad rotation $U(\mathbf{x})$ the frequency operator becomes $U^{-1}\hat h\,U$, so its spectrum does not depend on the triad, although $b_{\hat a}$ and $\mathcal{S}$ do.

D. Non-Abelian color and weak gauge fields

For a quark, whose field has three color components [58, 108], each chiral component is a column of three quaternions, $\boldsymbol{\psi}=(\psi^{(1)},\psi^{(2)},\psi^{(3)})^T$ in $\mathbb{H}^3$ (12 real components), with norm squared $|\boldsymbol{\psi}|^2=\sum_{j=1}^3\psi^{(j)}\bar\psi^{(j)}$. On $\boldsymbol{\psi}$, i acts as right multiplication of every entry by $-e_3$ and a spatial rotation $\mathcal{R}\in SU(2)$ multiplies every entry on the left. Let λ_A $(A=1,\dots,8)$ be the eight Hermitian traceless 3×3 Gell-Mann matrices obeying $\mathrm{Tr}(\lambda_A\lambda_B)=2\delta_{AB}$ and $[\frac12\lambda_B,\frac12\lambda_C]=if_{BCA}\,\frac12\lambda_A$. Because i is right multiplication by $-e_3$, any complex 3×3 color matrix $M=M_{\rm re}+iM_{\rm im}$ acts on $\boldsymbol{\psi}$ by $M\boldsymbol{\psi}\equiv M_{\rm re}\boldsymbol{\psi}-(M_{\rm im}\boldsymbol{\psi})e_3$, which commutes with the left spatial action of $\mathbf{e}$ and $\mathcal{R}$ by associativity. With the eight gluon four-potentials $(-\Phi^A_c/c,\mathbf{G}^A)$ and strong coupling q_c, the gauge-covariant derivative of Eqs. (29)

and (30) extends to

$$D_t\,\psi = \partial_t\psi - \frac{1}{\hbar}\Big(V + q_c\sum_{A=1}^{8}\tfrac{1}{2}\lambda_A\,\Phi_c^A\Big)\psi\,e_3,$$
$$D_k\,\psi = \partial_k\psi + \frac{1}{\hbar}\Big(qA_k + q_c\sum_{A=1}^{8}\tfrac{1}{2}\lambda_A\,G_k^A\Big)\psi\,e_3. \tag{150}$$

In the commutator $[D_j, D_k]\psi$, the quadratic gluon term applies two right factors e_3 ($e_3^2 = -1$) to the Lie bracket $[\frac{1}{2}\lambda_B, \frac{1}{2}\lambda_C] = if_{BCA}\,\frac{1}{2}\lambda_A$, and since i is right multiplication by $-e_3$, the product $(-1)(-e_3) = +e_3$ restores the single right factor e_3:

$$[D_j, D_k]\,\psi = \frac{1}{\hbar}\,\epsilon_{jkl}\Big(qB_l + q_c\sum_{A=1}^{8}\tfrac{1}{2}\lambda_A\,B_{c,l}^A\Big)\psi\,e_3,$$
$$[D_t, D_k]\,\psi = -\frac{1}{\hbar}\Big(qE_k + q_c\sum_{A=1}^{8}\tfrac{1}{2}\lambda_A\,E_{c,k}^A\Big)\psi\,e_3, \tag{151}$$

with the non-Abelian chromomagnetic field $\mathbf{B}_c^A = \nabla\times\mathbf{G}^A + (q_c/2\hbar)\sum_{B,C} f_{ABC}\,\mathbf{G}^B\times\mathbf{G}^C$ and chromoelectric field $\mathbf{E}_c^A = -\nabla\Phi_c^A - \partial_t\mathbf{G}^A + (q_c/\hbar)\sum_{B,C} f_{ABC}\,\Phi_c^B\mathbf{G}^C$. Because $\frac{1}{2}\lambda_A$ commutes with $\mathbf{e}$ and with the metric coefficients, the gluon potentials enter the wave equation through the covariant derivatives (150), and the color fields enter it through these curvatures, in the places where $q\mathbf{B}$ and $q\mathbf{E}$ enter in Section III. The color matrices do not commute with each other, so the color charge $\mathfrak{Q}_A \equiv \langle\frac{1}{2}\lambda_A\rangle$, a real number for each A, in general changes along a trajectory. At lowest order in $1/c^2$ and first order in q_c, the color-dependent part of the frequency operator is $q_c\sum_B \frac{1}{2}\lambda_B\,(\Phi_c^B - \mathbf{v}\cdot\mathbf{G}^B)$: the first term from D_t in Eq. (150), the second from the cross term of $\hat{\boldsymbol{\pi}}^2/2m$ with $\hat{\boldsymbol{\pi}} = \hat{\mathbf{p}} - q\mathbf{A} - q_c\sum_B \frac{1}{2}\lambda_B\mathbf{G}^B$, where $\mathbf{v}$ is the velocity. The Heisenberg equation $d\mathfrak{Q}_A/dt = (i/\hbar)\langle[\hat{h}, \frac{1}{2}\lambda_A]\rangle$ and $[\frac{1}{2}\lambda_B, \frac{1}{2}\lambda_A] = if_{BAC}\,\frac{1}{2}\lambda_C$ then give

$$\frac{d\mathfrak{Q}_A}{dt} = \frac{q_c}{\hbar}\sum_{B,C=1}^{8} f_{ABC}\,\big(\Phi_c^B - \mathbf{v}\cdot\mathbf{G}^B\big)\,\mathfrak{Q}_C, \tag{152}$$

a rotation of the eight-vector $(\mathfrak{Q}_1, \ldots, \mathfrak{Q}_8)$ in the color field. For a wave packet narrower than the scale on which the gluon fields vary, and whose color state is uncorrelated with its position, $\langle\frac{1}{2}\lambda_A\,\mathbf{E}_c^A\rangle = \mathfrak{Q}_A\,\mathbf{E}_c^A$ and $\langle\frac{1}{2}\lambda_A\,\mathbf{B}_c^A\rangle = \mathfrak{Q}_A\,\mathbf{B}_c^A$, with the fields at the center of the packet. In the force law the Abelian $q\mathbf{E} + q\mathbf{v}\times\mathbf{B}$ then gains the term

$$\mathbf{F}_c = q_c\sum_{A=1}^{8}\mathfrak{Q}_A\,\big(\mathbf{E}_c^A + \mathbf{v}\times\mathbf{B}_c^A\big), \tag{153}$$

which with Eq. (152) is the classical limit of Wong [109]. Whether Eq. (92) and its expansion carry over term by term with non-commuting potentials is not checked here. A chiral $SU(2)_L$ weak gauge field acts by the same right-multiplication rule on a left-handed quaternion doublet $\boldsymbol{\varphi}_L \in \mathbb{H}^2$.

VIII. CONCLUSION

Written in quaternions, the Dirac equation of a particle of nonzero mass is a pair of first-order equations, one for each chiral component, and their product is one second-order wave equation that is equivalent to it. The operator of that equation sorts the fields by the quaternion units they multiply: the minimally coupled Klein–Gordon operator on the diagonal of the unit table, the electric and gravitational field strengths on its border, and the magnetic and spatial-curvature spin couplings in its interior. The gravitomagnetic field enters all three. The Jacobi identity of the covariant derivatives gives the homogeneous Maxwell equations. In flat spacetime all four of Maxwell's equations are the four parts of one biquaternion equation [24]. Gauss's law and the Ampère–Maxwell law in it are the field's own equations, and the conserved current of a Dirac particle is a source that the equation admits. Where the positive-frequency subspace is a graph $(u, \mathsf{R}u)$, a Riccati equation reduces the wave equation to the exact frequency equation of the positive-frequency states, and at zeroth order in $\hbar$ the frequency operator of that equation gives a particle at rest the frequency $(V + mc^2\alpha)/\hbar$. The time-dilated rest frequency, which the companion paper takes from local position invariance, is therefore a consequence of the Dirac equation.

Expanded in powers of $1/c^2$ through order c^{-2}, the frequency operator is a sum of eight terms that contain the frequencies of the interactions of a slow particle of spin $\frac{1}{2}$, with a bound on the error of the truncated series when the norms of $c\mathsf{P}$, G and $\hat{V}$ are small compared with mc^2. The force-frequency identity of the companion paper gives a force law from each of these frequencies. Gravity and inertia enter through the lapse, the shift, and the scale factor W, and electromagnetism through the potentials. The spin couplings, the Darwin term, and the spin-orbit and geodetic couplings come from the quaternion units and from one dressed product. The expansion in $\hbar$ holds at all speeds and field strengths and requires that the fields vary slowly on the scale of the wavelength. It gives the spin precession and, for fields that vary along one direction, energy levels proved to approximate the Dirac levels with an error of order $\hbar^2$ and exact in uniform magnetic and crossed fields and in the nonuniform crossed fields of the profile $qA_y = \mathfrak{a}\tanh(x/L)$.

The derivation goes from the Dirac equation to the frequency and from the frequency to the force. It takes two inputs that it does not derive: the time-dilation factor of a given mass distribution, and the field equations that relate the potentials to their sources.

Appendix A: Derivations for the wave equation and the frequency operator

1. The Lorentz-scalar potential

A *Lorentz-scalar potential* $V_{\rm sc}$ adds to the rest energy itself, unlike Φ, which is the time component of the four-potential (Remark 1). It therefore makes the mass depend on position, $m \to m(\mathbf{x}) = m_0 + V_{\rm sc}(\mathbf{x})/c^2$ (Sections IV and V). The Higgs field gives a fermion its mass in this way, and the scalar mean field of nuclear matter shifts the nucleon mass in the same way [41]. Semiconductors have band-gap-dependent effective masses [42].

For a state at rest the scalar potential adds $V_{\rm sc}(\mathbf{x})/\hbar = [m(\mathbf{x}) - m_0]\,c^2/\hbar$ to the frequency, as an additive potential does (Eq. (109)). In motion the free-space energy is $\hbar\omega = \sqrt{\mathbf{p}^2c^2 + m(\mathbf{x})^2c^4}$. Its negative gradient at fixed momentum is the force $-m(\mathbf{x})\,c^4\,\nabla m/(\hbar\omega) = -\nabla V_{\rm sc}/\gamma$, where $\gamma = \hbar\omega/(m(\mathbf{x})\,c^2)$ is the Lorentz factor. At order c^{-2} this force is $-\nabla V_{\rm sc} + (v^2/2c^2)\nabla V_{\rm sc}$ (Eqs. (109) and (114)). The scalar potential enters the dressed potential $f_W = \hat{V} + \mathsf{G} - V_{\rm sc} - 2m_0c^2\ln W$ of Eq. (102) with the opposite sign to V, so in Eq. (122) the spin-orbit coupling of an attractive scalar potential adds to that of a repulsive vector potential. When the potential changes every mass by the same amount $V_{\rm sc}/c^2$, the force at a given velocity is the same for every species, so the acceleration depends on m_0.

A Lorentz-scalar potential differs from an additive potential in what it does to clocks: it changes the masses on which transition frequencies depend, so a uniform change of mass changes the rate of a clock. Suppose the scalar potential changes every mass by the same fraction, so that the ratio $m(\mathbf{x})/m_0$ is one positive function of position for every particle. Then the energy $\sqrt{\mathbf{p}^2c^2 + m(\mathbf{x})^2c^4}$ is that of a particle of mass m_0 in the conformally flat metric $[m(\mathbf{x})/m_0]^2$ times the Minkowski metric, so the potential belongs to the geometric class of Table I. The frequency $c\,|\mathbf{k}|$ of light in that metric does not depend on the conformal factor, so the potential bends no light. A Lorentz-scalar potential that changes different masses by different fractions lies outside both classes.

2. The first-order pair: derivation, rates, and uniqueness

Symbols introduced in this subsection are local to it. In particular, χ is the second quaternion of the pair, not the wave function of Eq. (83); Γ^μ are the maps of a first-order system, not a spin connection; and A_ν and $F_{\mu\nu}$ are the components of a massive vector field, not the electromagnetic potentials.

Equation (47) is of second order in time, so the future of a solution is fixed only by ψ and $\partial_t\psi$ together at one time, whereas the Schrödinger equation is of first order in time and is fixed by ψ alone. A boost mixes ∂_0 with the spatial derivative along the motion, since the four derivatives ∂_μ transform as the four-gradient of Eq. (40). A wave equation of first order in time that keeps its form under boosts is therefore also of first order in space. For a free particle in flat spacetime, write such an equation as a pair: an operator L_1 of first order (a differential operator, not a Lagrangian) takes a quaternion φ to κ times a quaternion χ, and an operator L_2 of first order takes χ back to κ times φ,

$$\begin{aligned} \mathsf{L}_1\,\varphi &= \kappa\,\chi, \\ \mathsf{L}_2\,\chi &= \kappa\,\varphi, \end{aligned} \tag{A1}$$

where χ may be φ itself; a constant factor on the right side of either line can be absorbed into χ, so both lines contain the same κ. Applying L_2 to the first line and substituting the second gives $\mathsf{L}_2\mathsf{L}_1\varphi = \kappa^2\varphi$, which is Eq. (47) for φ exactly when

$$\mathsf{L}_2\,\mathsf{L}_1 = -\partial_0^2 - (\mathbf{e}\cdot\nabla)^2, \tag{A2}$$

where $\partial_0 = c^{-1}\partial_t$, as in Eq. (40).

Free flat spacetime is the same at every point and at every time. Requiring the equation to share this *translation invariance* makes the coefficients of the derivatives constant, and each operator L_a, $a = 1, 2$, is $\mathsf{L}_a\psi = \mathcal{M}_0(\partial_0\psi) + \sum_k \mathcal{M}_k(\partial_k\psi)$, with four constant real-linear maps $\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ of the quaternions, together with a possible constant term. With $e_0 \equiv 1$, every real-linear map $\mathcal{M}$ of the quaternions can be written in exactly one way as

$$\mathcal{M}\psi = \sum_{a,b=0}^{3} y_{ab}\, e_a\, \psi\, e_b, \tag{A3}$$

with sixteen real coefficients y_{ab}, since the sixteen maps $\psi \mapsto e_a\psi e_b$ are linearly independent and the real-linear maps of a space of four dimensions form a space of sixteen. Write $y_{\mu ab}$ for the coefficients of $\mathcal{M}_\mu$, $\mu = 0, 1, 2, 3$.

Rotation covariance requires the equation to take the same form in rotated axes: rotating φ and χ together and then applying L_1 or L_2 gives the same result as applying the operator first and rotating after. A rotation acts on each quaternion by left multiplication by the unit quaternion $\mathcal{R}$ of Section II C, the same $\mathcal{R}$ on both, and on the three ∂_k as on the components of a vector. Applying $\bar{\mathcal{R}} = \mathcal{R}^{-1}$ on the left to undo the rotation of the result maps a term of Eq. (A3) to $y_{ab}\,(\bar{\mathcal{R}}\, e_a \mathcal{R})\,\psi\, e_b$. The rotation acts on the left index alone: it fixes e_0 and turns e_1, e_2, e_3 as the components of a vector. Since $\bar{\mathcal{R}}\, e_a \mathcal{R}$ is unchanged when $\mathcal{R}$ is replaced by $-\mathcal{R}$, the two unit quaternions that represent one rotation impose the same condition. The rotated operator must agree with the original on every ψ, and $\psi = x^\mu\psi_0$ with a constant ψ_0 has $\partial_\nu\psi = \delta^\mu_\nu\,\psi_0$, so the coefficient map of $\partial_0\psi$ and those of the $\partial_k\psi$ must agree separately. The map $\mathcal{M}_0$ multiplies $\partial_0\psi$, a scalar under rotations, so the rotated $\mathcal{M}_0$ must be $\mathcal{M}_0$ again. The three $\mathcal{M}_k$ multiply the three components $\partial_k\psi$ of a vector, so the rotation turns the index k as well, and the rotated three maps must be the same three. For each b, the three coefficients y_{0jb}, $j = 1, 2, 3$, therefore form a vector that every rotation leaves unchanged, and so do the three coefficients y_{k0b}, $k = 1, 2, 3$. Such a vector is zero: the half turn about the first axis reverses its second and third components, and the half turn about the second axis reverses its first. Hence $\mathcal{M}_0\psi = \psi\,\mathsf{p}$, with the constant quaternion $\mathsf{p} = \sum_b y_{00b}\, e_b$.

For each b the nine coefficients y_{kjb}, $k, j = 1, 2, 3$, form a tensor of two indices unchanged by every rotation. The half turn about the first axis reverses the second and third axes and so the four entries $y_{12b}, y_{13b}, y_{21b}, y_{31b}$, which must therefore vanish, and the half turn about the third axis removes y_{23b} and y_{32b} in the same way. The quarter turn about the third axis takes the first axis to the second, so $y_{11b} = y_{22b}$, and the quarter turn about the first axis gives $y_{22b} = y_{33b}$. So $y_{kjb} = y_{11b}\,\delta_{kj}$ for each b. Only rotations were used, not reflections. Conversely $y_{11b}\,\delta_{kj}$ is unchanged by every rotation, so these four turns already impose every condition that the whole group of rotations imposes, and the form they leave is covariant under every rotation. Hence $\mathcal{M}_k\psi = e_k\,\psi\,\mathsf{q}$, with one constant quaternion $\mathsf{q} = \sum_b y_{11b}\, e_b$ common to the three.

Phase invariance requires the operators to commute with a constant phase. Multiplying φ and χ by the same constant phase (right multiplication by $\exp(e_3\phi)$ with a constant angle ϕ, Section II C) and then applying L_1 or L_2 gives the same result as applying the operator first and multiplying after. This also holds for the Schrödinger equation, whose operator commutes with a constant phase. The phase changes neither the density nor the spin density, and the local form of this invariance introduces the potentials (Remark 1). The term $\psi\,\mathsf{p}$ gives $\psi\exp(e_3\phi)\,\mathsf{p}$, which equals $\psi\,\mathsf{p}\exp(e_3\phi)$ for every ϕ only when p commutes with e_3, so p is a combination of 1 and e_3 (Section II D). The term $e_k\,\psi\,\mathsf{q}$ gives $e_k\,\psi\exp(e_3\phi)\,\mathsf{q}$ in the same way, so q is also a combination of 1 and e_3. Therefore

$$\mathsf{L}_a\,\psi = (\partial_0\psi)\,\mathsf{p} + (\mathbf{e}\cdot\nabla\psi)\,\mathsf{q}, \tag{A4}$$

with p and q in the plane of 1 and e_3. Quaternions in that plane multiply as the complex numbers do, with e_3 as the imaginary unit, so they commute with one another and each nonzero one has an inverse in the plane.

Write L_1 with the coefficients $\mathsf{p}_1, \mathsf{q}_1$ and L_2 with $\mathsf{p}_2, \mathsf{q}_2$. In $\mathsf{L}_2(\mathsf{L}_1\psi)$ the derivatives pass through the constant right factors of L_1, which then stand to the left of those of L_2, so

$$\mathsf{L}_2\mathsf{L}_1\psi = (\partial_0^2\psi)\,\mathsf{p}_1\mathsf{p}_2 + (\mathbf{e}\cdot\nabla\,\partial_0\psi)\,(\mathsf{q}_1\mathsf{p}_2 + \mathsf{p}_1\mathsf{q}_2) + \big((\mathbf{e}\cdot\nabla)^2\psi\big)\,\mathsf{q}_1\mathsf{q}_2. \tag{A5}$$

Equating the coefficients of the three second derivatives with those of Eq. (A2) gives

$$\begin{aligned} \mathsf{p}_1\,\mathsf{p}_2 &= -1, \\ \mathsf{q}_1\,\mathsf{q}_2 &= -1, \\ \mathsf{q}_1\,\mathsf{p}_2 + \mathsf{p}_1\,\mathsf{q}_2 &= 0. \end{aligned} \tag{A6}$$

By the second line q_1 has an inverse. Multiplying the first line of Eq. (A1) on the right by q_1^{-1} and writing the second line for $\chi = (\chi\,\mathsf{q}_1^{-1})\,\mathsf{q}_1$ gives a pair for φ and $\chi\,\mathsf{q}_1^{-1}$ with the coefficients $\mathsf{p}_1\mathsf{q}_1^{-1}$, 1, $\mathsf{q}_1\mathsf{p}_2$, $\mathsf{q}_1\mathsf{q}_2$, which lie in

the same plane. Renaming $\chi\,\mathsf{q}_1^{-1}$ as χ therefore sets $\mathsf{q}_1 = 1$. The conditions then read $\mathsf{p}_1\mathsf{p}_2 = -1$, $\mathsf{q}_2 = -1$, and $\mathsf{p}_2 = \mathsf{p}_1$, so $\mathsf{p}_1^2 = -1$, and since the square roots of -1 in the plane are $\pm e_3$, $\mathsf{p}_1 = \mathsf{p}_2 = \pm e_3$. With the upper sign, $\mathsf{L}_1\psi = (\partial_0\psi)\,e_3 + \mathbf{e}\cdot\nabla\psi$ and $\mathsf{L}_2\psi = (\partial_0\psi)\,e_3 - \mathbf{e}\cdot\nabla\psi$. The lower sign gives the negatives of the same two operators with their places exchanged, which after $\chi \to -\chi$ is the upper sign with the names φ and χ exchanged.

A constant term $\psi\,\mathsf{r}_a$ beside the derivatives of L_a, for $a = 1, 2$, is allowed by translation invariance, rotation covariance, and phase invariance when r_a lies in the plane of 1 and e_3. It adds to $\mathsf{L}_2\mathsf{L}_1\psi$ the terms of first order $(\partial_0\psi)\,(\mathsf{p}_2\mathsf{r}_1 + \mathsf{p}_1\mathsf{r}_2) + (\mathbf{e}\cdot\nabla\psi)\,(\mathsf{q}_2\mathsf{r}_1 + \mathsf{q}_1\mathsf{r}_2)$, which Eq. (A2) does not contain, and the term $\psi\,\mathsf{r}_1\mathsf{r}_2$ of order zero, which would add to the mass term. With $\mathsf{q}_1 = 1$ and the upper sign the two first-order coefficients are $e_3\,(\mathsf{r}_1 + \mathsf{r}_2)$ and $\mathsf{r}_2 - \mathsf{r}_1$. Both vanish only for $\mathsf{r}_1 = \mathsf{r}_2 = 0$, so the term of order zero vanishes with them. The mass therefore cannot enter as a constant term that maps a quaternion to itself. With $\mathsf{q}_1 = 1$ and the upper sign, the pair of Eq. (A1) is

$$\begin{aligned}(\partial_0\varphi)\,e_3 + \mathbf{e}\cdot\nabla\varphi &= \kappa\,\chi,\\ (\partial_0\chi)\,e_3 - \mathbf{e}\cdot\nabla\chi &= \kappa\,\varphi.\end{aligned} \tag{A7}$$

With $\kappa = mc/\hbar$ (Eq. (43)), dividing each line by κ gives χ in terms of the first derivatives of φ, and φ in terms of the first derivatives of χ:

$$\begin{aligned}\chi &= \kappa^{-1}\big[(\partial_0\varphi)\,e_3 + \mathbf{e}\cdot\nabla\varphi\big],\\ \varphi &= \kappa^{-1}\big[(\partial_0\chi)\,e_3 - \mathbf{e}\cdot\nabla\chi\big].\end{aligned} \tag{A8}$$

The factor $\kappa^{-1} = \hbar/mc$ is the reduced Compton wavelength. Substituting the first line into the second gives $\mathsf{L}_2\mathsf{L}_1\varphi = \kappa^2\varphi$, Eq. (47), as after Eq. (A1). Conversely, for any solution φ of Eq. (47), the χ of the first line satisfies the second, since $\mathsf{L}_2\chi = \kappa^{-1}\mathsf{L}_2\mathsf{L}_1\varphi = \kappa\varphi$. A solution of the pair is thus a quaternion φ whose four real components each obey the Klein–Gordon equation, together with $\chi = \kappa^{-1}\mathsf{L}_1\varphi$, and the mass enters the relation between χ and φ only through the factor κ^{-1}. The spin enters through the derivative L_1: its spatial part $\mathbf{e}\cdot\nabla$ contains the units e_k, which act on the spin (Section II C), so χ depends on the direction of the spin of φ relative to its momentum.

For a particle at rest, $\nabla\varphi = \mathbf{0}$, and the plane wave of Eq. (42) with $\mathbf{k} = \mathbf{0}$ and positive frequency $\omega = \kappa c$ has $\partial_0\varphi = \kappa\,\varphi\,e_3$, so $(\partial_0\varphi)\,e_3 = \kappa\,\varphi\,e_3^2 = -\kappa\,\varphi$. The first line of Eq. (A7) then reads $-\kappa\,\varphi = \kappa\,\chi$, so $\chi = -\varphi$: the two quaternions of a particle at rest are equal in magnitude.

For $\kappa = 0$ the two lines decouple into $\mathsf{L}_1\varphi = 0$ and $\mathsf{L}_2\chi = 0$. For the plane wave of Eq. (42), $\partial_0\varphi = (\omega/c)\,\varphi\,e_3$ and $\partial_j\varphi = -k_j\,\varphi\,e_3$. Since $e_3^2 = -1$, $(\partial_0\varphi)\,e_3 = -(\omega/c)\,\varphi$. With $\sigma_k\psi = e_k\,\psi\,e_3^{-1}$ (Eq. (24)) and $e_3^{-1} = -e_3$, $\mathbf{e}\cdot\nabla\varphi = -\sum_j k_j\,e_j\,\varphi\,e_3 = (\mathbf{k}\cdot\boldsymbol{\sigma})\,\varphi$. For $\kappa = 0$ a wave of positive frequency has $\omega = c\,|\mathbf{k}|$, and with the unit vector $\hat{\mathbf{k}} = \mathbf{k}/|\mathbf{k}|$ along the momentum the two lines become

$$\begin{aligned}(\hat{\mathbf{k}}\cdot\boldsymbol{\sigma})\,\varphi &= \varphi,\\ (\hat{\mathbf{k}}\cdot\boldsymbol{\sigma})\,\chi &= -\chi.\end{aligned} \tag{A9}$$

The operator $\hat{\mathbf{k}}\cdot\boldsymbol{\sigma}$ is the Pauli operator along the momentum, whose eigenvalue $+1$ or -1 is the sign of the spin component along it, so a massless φ has its spin along its momentum and a massless χ has its spin against it. A spinor that obeys the first line of Eq. (A9) when its mass is zero is called *right-handed*, and one that obeys the second *left-handed*; φ is the right-handed quaternion and χ the left-handed one. The mass couples a right-handed quaternion to a left-handed one, as the mass term of the Dirac equation contracts a left-handed spinor with a right-handed one [110]. A massive particle, whose two quaternions are equal in magnitude at rest, therefore has no fixed handedness. Each of φ and χ is a real quaternion, four real numbers, and the imaginary unit i, right multiplication by $-e_3$, acts on both alike. The pair therefore has eight real components, as many as the four complex components of a Dirac spinor.

A single quaternion, $\chi = \varphi$ with $\mathsf{L}_2 = \mathsf{L}_1$, would need $\mathsf{p}_1^2 = -1$, $\mathsf{q}_1^2 = -1$, and $2\mathsf{p}_1\mathsf{q}_1 = 0$ by Eq. (A6); the last makes p_1 or q_1 zero, which contradicts the first two, so a massive wave whose equation is invariant under a constant change of phase needs two quaternions, coupled by the mass.

When each derivative is replaced with its covariant derivative, as in Eq. (48), the two operators $\psi \mapsto (\partial_0\psi)\,e_3 \pm \mathbf{e}\cdot\nabla\psi$ of Eq. (A7) become $\mathsf{N} \pm \mathsf{X}_W$, and Eqs. (54) and (55) are Eq. (A1) with $\mathsf{L}_1 = \mathsf{N} + \mathsf{X}_W$, $\mathsf{L}_2 = \mathsf{N} - \mathsf{X}_W$, $\varphi = \varphi_R$, and $\chi = \varphi_L$. The translation of the Dirac equation therefore identifies φ, the quaternion on which $\mathsf{N} + \mathsf{X}_W$ acts, as the right-handed component, in agreement with Eq. (A9). The pair of Eqs. (54) and (55) thus follows from Eq. (47), the requirement of first order in time and space, translation invariance, rotation covariance, phase invariance, and minimal coupling, with no input from the Dirac equation.

Moving the spatial derivatives of Eq. (A7) to the right side of each line and multiplying on the right by $c\,e_3^{-1} = -c\,e_3$ gives, with $\partial_t = c\,\partial_0$ and the rest frequency $\omega_0 = \kappa c$ of Eq. (2),

$$\begin{aligned}\partial_t\varphi &= c\,(\mathbf{e}\cdot\nabla\varphi)\,e_3 - \omega_0\,\chi\,e_3,\\ \partial_t\chi &= -c\,(\mathbf{e}\cdot\nabla\chi)\,e_3 - \omega_0\,\varphi\,e_3.\end{aligned} \tag{A10}$$

For a particle at rest, $\nabla\varphi = \nabla\chi = \mathbf{0}$, and subtracting and adding the two lines gives $\partial_t(\varphi - \chi) = \omega_0\,(\varphi - \chi)\,e_3$ and $\partial_t(\varphi + \chi) = -\omega_0\,(\varphi + \chi)\,e_3$. The plane wave of Eq. (42) with $\mathbf{k} = \mathbf{0}$ has $\partial_t\varphi = \omega\,\varphi\,e_3$, so $\varphi - \chi$ turns at the frequency $+\omega_0$ and $\varphi + \chi$ at $-\omega_0$. A solution of positive frequency has $\varphi + \chi = 0$, the rest state $\chi = -\varphi$ that the first line of Eq. (A7) gives, in which $\partial_t\varphi = -\omega_0\,\chi\,e_3 = \omega_0\,\varphi\,e_3$ and $\partial_t\chi = -\omega_0\,\varphi\,e_3 = \omega_0\,\chi\,e_3$.

Take a plane wave of positive frequency whose φ has its spin along the momentum, $(\hat{\mathbf{k}}\cdot\boldsymbol{\sigma})\,\varphi = \varphi$. By $(\partial_0\varphi)\,e_3 = -(\omega/c)\,\varphi$ and $\mathbf{e}\cdot\nabla\varphi = (\mathbf{k}\cdot\boldsymbol{\sigma})\,\varphi = |\mathbf{k}|\,\varphi$, the first line of Eq. (A7) gives $(|\mathbf{k}| - \omega/c)\,\varphi = \kappa\,\chi$. Multiplying the numerator and the denominator of $(|\mathbf{k}| - \omega/c)/\kappa$ by $\omega/c + |\mathbf{k}|$, with $\omega^2/c^2 - |\mathbf{k}|^2 = \kappa^2$ (Eq. (41)), gives

$$\chi = -\frac{\kappa}{\omega/c + |\mathbf{k}|}\,\varphi, \tag{A11}$$

and the second line then holds as well. The factor is $-mc^2/(\hbar\omega + \hbar c\,|\mathbf{k}|)$, minus the rest energy divided by the sum of the energy and c times the momentum: -1 at rest, $-mc^2/(2\hbar\omega)$ once $\hbar\omega \gg mc^2$, and 0 at $m = 0$, where Eq. (A9) holds.

Consider a first-order system of any number N of quaternions that has translation invariance and phase invariance and a nonzero mass, $\kappa \neq 0$, and whose solutions all obey Eq. (47). Collect the N quaternions into a column $\boldsymbol{\Psi}$, which differs from the four-component Dirac wave function Ψ of Eq. (50). Let i act on each component by right multiplication by $-e_3$, so $i^2 = -1$. By translation invariance and phase invariance the system is $\sum_\mu \tilde{\Gamma}^\mu\partial_\mu\boldsymbol{\Psi} + \mathcal{G}\boldsymbol{\Psi} = 0$, with constant real-linear maps $\tilde{\Gamma}^0, \ldots, \tilde{\Gamma}^3$ and $\mathcal{G}$ of the column that commute with i, where $\mathcal{G}$ is the map of the mass term. The map $\mathcal{G}$ is invertible, because a constant column $\boldsymbol{\Psi}_0$ with $\mathcal{G}\boldsymbol{\Psi}_0 = 0$ would be a solution, and a constant solves Eq. (47) only if $\kappa^2\boldsymbol{\Psi}_0 = 0$. Multiplying the system by $-\kappa\mathcal{G}^{-1}$, which also commutes with i, and writing $\Gamma^\mu \equiv -\kappa\mathcal{G}^{-1}\tilde{\Gamma}^\mu$ gives $\Gamma(\partial)\,\boldsymbol{\Psi} = \kappa\,\boldsymbol{\Psi}$, where $\Gamma(\partial) \equiv \sum_\mu \Gamma^\mu\partial_\mu$. For the pair, $N = 2$, $\boldsymbol{\Psi} = (\varphi, \chi)$, and $\Gamma(\partial)(\varphi, \chi) = (\mathsf{L}_2\chi, \mathsf{L}_1\varphi)$, so that $\Gamma^0(\varphi, \chi) = (\chi\,e_3, \varphi\,e_3)$ and $\Gamma^k(\varphi, \chi) = (-e_k\chi, e_k\varphi)$.

The system is taken to be an evolution equation: Γ^0 is invertible, as it is for the pair, so the system can be solved for the time derivative. With $\mathsf{M} \equiv (\Gamma^0)^{-1}$ and the operator $\mathcal{T} \equiv \mathsf{M}\,(\kappa - \sum_k \Gamma^k\partial_k)$, the system is $\partial_0\boldsymbol{\Psi} = \mathcal{T}\boldsymbol{\Psi}$. The operator $\mathcal{T}$ has constant coefficients, so it commutes with ∂_0, and every solution obeys $\partial_0^2\boldsymbol{\Psi} = \mathcal{T}^2\boldsymbol{\Psi}$. Equation (47) for every component is $\partial_0^2\boldsymbol{\Psi} = (\nabla^2 - \kappa^2)\,\boldsymbol{\Psi}$, since $(\mathbf{e}\cdot\nabla)^2 = -\nabla^2$, so it holds for every solution exactly when $(\mathcal{T}^2 - \nabla^2 + \kappa^2)\,\boldsymbol{\Psi} = 0$ for every solution. Every column whose components are polynomials in $\mathbf{x}$ of degree at most two is the value at $x^0 = 0$ of a solution, $\exp(x^0\mathcal{T})\,\boldsymbol{\Psi}(0)$, because $\mathcal{T}$ maps such columns to such columns, a space of finite dimension on which the exponential series converges. The operator $\mathcal{T}^2 - \nabla^2 + \kappa^2$ has constant coefficients and order at most two, and it gives zero on every such column exactly when its coefficient of order zero, its coefficients of order one, and the parts of its coefficients of order two that are symmetric in the two indices all vanish. Expanding $\mathcal{T}^2$, with $\partial_j\partial_k = \partial_k\partial_j$ and the maps kept in their order, gives

$$\mathcal{T}^2 = \kappa^2\mathsf{M}^2 - \kappa\sum_k \left(\mathsf{M}^2\,\Gamma^k + \mathsf{M}\,\Gamma^k\mathsf{M}\right)\partial_k + \sum_{j,k} \mathsf{M}\,\Gamma^j\mathsf{M}\,\Gamma^k\,\partial_j\partial_k. \tag{A12}$$

Since $\kappa \neq 0$, order zero gives $\mathsf{M}^2 = -1$, so $\Gamma^0\Gamma^0 = -1$. Order one then gives $\mathsf{M}\,\Gamma^k\mathsf{M} = \Gamma^k$, and multiplying by Γ^0 on the left and on the right gives $\Gamma^k = \Gamma^0\,\Gamma^k\,\Gamma^0$, so $\Gamma^0\Gamma^k = \Gamma^0\Gamma^0\,\Gamma^k\,\Gamma^0 = -\Gamma^k\Gamma^0$. With $\mathsf{M}\,\Gamma^j\mathsf{M} = \Gamma^j$ the coefficient of order two is $\Gamma^j\Gamma^k$, and its symmetric part equals that of ∇^2 exactly when $\Gamma^j\Gamma^k + \Gamma^k\Gamma^j = 2\,\delta_{jk}$. Together,

$$\begin{aligned} &\Gamma^0\,\Gamma^0 = -1, \\ &\Gamma^k\,\Gamma^k = 1, \\ &\Gamma^\mu\,\Gamma^\nu = -\Gamma^\nu\,\Gamma^\mu \quad (\mu \neq \nu), \end{aligned} \tag{A13}$$

for $k = 1, 2, 3$ and $\mu, \nu = 0, \ldots, 3$, the relations of the Dirac matrices in the signature $(-, +, +, +)$. Conversely, Eq. (A13) gives $\Gamma(\partial)\,\Gamma(\partial) = -\partial_0^2 + \nabla^2$, so applying $\Gamma(\partial)$ twice gives Eq. (47) for every component.

Without the invertibility of Γ^0 the relations need not hold. The equations $\partial_\mu A_\nu - \partial_\nu A_\mu = \kappa\,F_{\mu\nu}$ and $\sum_\mu \partial^\mu F_{\mu\nu} = \kappa\,A_\nu$ of a massive vector field, where A_ν are four complex fields, $F_{\mu\nu} = -F_{\nu\mu}$ are six more, $\partial^0 = -\partial_0$, and $\partial^k = \partial_k$, are a first-order system of ten complex components, five quaternions, with translation invariance and phase invariance. Applying $\sum_\nu \partial^\nu$ to the second equation gives zero on the left, since $F_{\mu\nu}$ is antisymmetric, so $\sum_\nu \partial^\nu A_\nu = 0$; applying $\sum_\mu \partial^\mu$ to the first equation then gives $(-\partial_0^2 + \nabla^2)\,A_\nu = \kappa^2 A_\nu$, and every component, $F_{\mu\nu}$ included, obeys Eq. (47). But ∂_0 enters only the six equations for F_{0k} and A_k, so Γ^0 has rank six of ten, and the four equations without ∂_0 give A_0 and F_{jk} at each time from the spatial derivatives of F_{0k} and A_k.

Define $C_{12} \equiv i\,\Gamma^1\Gamma^2$ and $C_{03} \equiv \Gamma^0\Gamma^3$. By Eq. (A13), with i commuting with every Γ^μ, $C_{12}^2 = i^2\,\Gamma^1\Gamma^2\Gamma^1\Gamma^2 = -(-\Gamma^1\Gamma^1\Gamma^2\Gamma^2) = 1$ and $C_{03}^2 = -\Gamma^0\Gamma^0\Gamma^3\Gamma^3 = 1$, and C_{12} and C_{03} commute, since moving $\Gamma^0\Gamma^3$ past $\Gamma^1\Gamma^2$ reverses the sign four times. The map Γ^1 anticommutes with C_{12} and commutes with C_{03}, and Γ^0 commutes with C_{12} and

anticommutes with C_{03}, by the same counting of sign changes. Two commuting maps whose squares are 1 split the column space into four subspaces $\mathcal{W}_{\pm\pm}$, on which $C_{12} = \pm 1$ (first sign) and $C_{03} = \pm 1$ (second sign). The four products $\frac{1}{4}(1 \pm C_{12})(1 \pm C_{03})$ project onto these subspaces and add to 1. Each $\mathcal{W}_{\pm\pm}$ is mapped into itself by i, which commutes with C_{12} and C_{03}, so its real dimension is even. The map Γ^1, which is its own inverse, flips the first sign, mapping $\mathcal{W}_{+\pm}$ onto $\mathcal{W}_{-\pm}$, and Γ^0, whose inverse is $-\Gamma^0$, flips the second, mapping $\mathcal{W}_{\pm+}$ onto $\mathcal{W}_{\pm-}$, so the four subspaces have the same real dimension, which is even. The column therefore has $4N$ real components, a multiple of 8, so N is even and at least 2. The massive vector field has $N = 5$, so this conclusion needs the invertibility of Γ^0.

For $N = 2$ each $\mathcal{W}_{\pm\pm}$ has real dimension two. For the pair, i is right multiplication by $-e_3$ and $\sigma_3\psi = -e_3\,\psi\,e_3$ by Eq. (24), so $C_{12}(\varphi, \chi) = (-\sigma_3\varphi, -\sigma_3\chi)$ and $C_{03}(\varphi, \chi) = (-\sigma_3\varphi, \sigma_3\chi)$, and $C_{12}C_{03}$ is 1 on φ and -1 on χ. The subspaces $\mathcal{W}_{++}$ and $\mathcal{W}_{--}$ therefore contain the right-handed quaternion φ, with $\chi = 0$, and $\mathcal{W}_{+-}$ and $\mathcal{W}_{-+}$ contain the left-handed quaternion χ, with $\varphi = 0$; in each, the first sign is minus the sign of the spin component along the third axis, and the product of the two signs is $+1$ for right-handed and -1 for left-handed.

Choose a nonzero column $\mathbf{\Psi}_0$ in $\mathcal{W}_{++}$; then $\mathbf{\Psi}_0$, $\Gamma^1\mathbf{\Psi}_0$, $\Gamma^0\mathbf{\Psi}_0$, and $\Gamma^0\Gamma^1\mathbf{\Psi}_0$ lie in $\mathcal{W}_{++}$, $\mathcal{W}_{-+}$, $\mathcal{W}_{+-}$, and $\mathcal{W}_{--}$, and with their images under i they form a basis of the column space. Equation (A13) fixes the action of Γ^0 and Γ^1 on this basis. For instance, $\Gamma^1(\Gamma^0\mathbf{\Psi}_0) = -\Gamma^0\Gamma^1\mathbf{\Psi}_0$ and $\Gamma^0(\Gamma^0\mathbf{\Psi}_0) = -\mathbf{\Psi}_0$, and the other two maps follow from $\Gamma^2 = -i\,\Gamma^1 C_{12}$ and $\Gamma^3 = -\Gamma^0 C_{03}$. Thus every Γ^μ has the same matrix in such a basis for every system of two quaternions. The map that takes the basis built for one system to the basis built for another commutes with i, because each basis consists of vectors and their images under i, and it maps the first system into the second. The map is therefore a change of the quaternion variables that preserves the imaginary unit and hence the phase. This is the theorem that Pauli [59] states and proves for 4×4 matrices, obtained here from Eq. (A13) and the commutation of every Γ^μ with i. Pauli wrote the relations with every square equal to 1, and with Γ^0 multiplied by i his relations are Eq. (A13).

For any even N the same construction, started from $N/2$ columns of $\mathcal{W}_{++}$ independent over the complex numbers, splits the column into $N/2$ copies of the pair. The pair is therefore, up to a change of variables that preserves i, the only first-order evolution equation for two quaternions that has translation invariance and phase invariance and whose components all obey Eq. (47) with $\kappa \neq 0$. A conserved count, the number of particles minus the number of antiparticles, is the charge of a constant phase symmetry [110], so, among first-order equations of this kind, the pair is the free equation of every spin-$\frac{1}{2}$ particle whose number is conserved, charged or neutral.

The argument does not use rotation covariance, so it holds whether or not a rotation acts alike on the two quaternions. The imaginary unit enters twice: in C_{12}, whose square is 1 because $i^2 = -1$, and in the even real dimension of each $\mathcal{W}_{\pm\pm}$.

Without phase invariance a single quaternion suffices. Then i need not commute with the Γ^μ, and the count fails: p_1 and q_1 need not lie in the plane of 1 and e_3, and $\mathsf{p}_1 = e_3$ and $\mathsf{q}_1 = e_1$, which square to -1 and anticommute, satisfy Eq. (A6) with $\mathsf{p}_2 = \mathsf{p}_1$ and $\mathsf{q}_2 = \mathsf{q}_1$. The one equation

$$(\partial_0\psi)\,e_3 + (\mathbf{e}\cdot\nabla\psi)\,e_1 = \kappa\,\psi \tag{A14}$$

then gives Eq. (47) when its operator is applied twice, but the right factor e_1 anticommutes with e_3, so right multiplication by $\exp(e_3\phi)$ does not map solutions to solutions, and the equation has no phase symmetry.

3. Charge conjugation, the Majorana equation, and the biquaternion form

Right multiplication by e_2 anticommutes with e_3, so it reverses the signs of q and V in Eqs. (29) and (30) and commutes with every other term of D_n and X_W and with the mass term, since each of these terms is built from derivatives, multiplications by real functions and left multiplications. The operator $\mathsf{N}\psi = (D_n\psi)\,e_3$ ends with the right factor e_3, so N with q and V reversed maps ψe_2 to $-(\mathsf{N}\psi)\,e_2$, while X_W with q reversed maps ψe_2 to $(\mathsf{X}_W\psi)\,e_2$. So if (φ_R, φ_L) solves Eqs. (54) and (55), then $(\varphi_L e_2, -\varphi_R e_2)$ solves them with q and V reversed, which is the exchange of the chiral components stated in Section III A. In terms of $\psi = \psi_\uparrow + e_2\psi_\downarrow$, $\psi e_2 = -\bar{\psi}_\downarrow + e_2\bar{\psi}_\uparrow$, which is the standard charge conjugate $-i\sigma_2\psi^*$.

In Eq. (A14), the maps $\psi \mapsto \psi e_3$ and $\psi \mapsto e_k\psi e_1$ generate the full 16-dimensional algebra of real 4×4 matrices, whose commutant is the real scalars $\mathbb{R}$. The particle of Eq. (A14) (equivalent in flat spacetime to Eq. (56) via $\varphi = \psi\mathcal{Q}$ with $\mathcal{Q} = (e_2 - e_3)/\sqrt{2}$) is therefore a real field. A real field is its own charge conjugate, a neutral Majorana fermion [60, 61]. The neutron is not one, since the antineutron is distinct from it [111]; neutrinoless double beta decay occurs only if neutrinos are Majorana fermions [110].

For a charged particle in the Dirac representation $\varphi_\pm = (\varphi_R \pm \varphi_L)/\sqrt{2}$, the sum and difference of Eqs. (54) and (55) are

$$\mathsf{N}\,\varphi_+ + \mathsf{X}_W\,\varphi_- = \kappa\,\varphi_+, \tag{A15}$$

$$\mathsf{N}\,\varphi_- + \mathsf{X}_W\,\varphi_+ = -\kappa\,\varphi_-. \tag{A16}$$

With the second imaginary unit i of Section III D, let $\Xi = \varphi_+ + \mathsf{i}\,\varphi_- e_3$ be a biquaternion and $\Xi^* = \varphi_+ - \mathsf{i}\,\varphi_- e_3$ its i-conjugate. Equations (A15) and (A16) combine into

$$\mathsf{N}\,\Xi + \mathsf{i}\,(\mathsf{X}_W\,\Xi)\,e_3 = \kappa\,\Xi^*, \tag{A17}$$

which in flat spacetime is the Dirac–Lanczos biquaternion equation [14, 106] and the even-subalgebra spacetime-algebra equation of Hestenes [15, 104]. By contrast, the chiral pair of Eqs. (54)–(55) and the second-order wave equation (57) use only the four-dimensional real division algebra $\mathbb{H}$ of quaternions. A quaternion has the four real components of a spin-$\frac{1}{2}$ state, and a normalized state is a point of the unit sphere S^3 [112]. Unit quaternions act on the state as the half-angle rotors of Eq. (27).

4. The commutator with a shift

Let $W = 1$, as in a uniformly moving or rotating frame in flat spacetime, the lapse α_n and the potential V be constant, $\mathbf{A}$ static, and the shift a rigid flow, a constant velocity $\mathbf{u}$ plus the rotation $\frac{1}{2}\,\mathbf{B}_g \times \mathbf{x}$. Then $\mathsf{N}\psi = (D_n\psi)\,e_3$ with $c\alpha_n D_n\psi = D_t\psi - \mathbf{A}_g\cdot\mathbf{D}\psi + \frac{1}{4}\,(\mathbf{e}\cdot\mathbf{B}_g)\,\psi$ by Eqs. (39) and (37); the divergence term $-\frac{1}{2}(\nabla\cdot\mathbf{A}_g)\,\psi$ of Eq. (39) is absent, because a rigid flow has $\nabla\cdot(\mathbf{u} + \frac{1}{2}\,\mathbf{B}_g\times\mathbf{x}) = 0$. In $\mathsf{N}(\mathsf{X}\psi)$ the operators D_t and $\mathbf{A}_g\cdot\mathbf{D}$ pass through the left factors e_k of $\mathsf{X}\psi = \sum_k e_k\,D_k\psi$, while the rotation term multiplies them from the left; in $\mathsf{X}(\mathsf{N}\psi)$ each D_k passes through the constant $1/(c\alpha_n)$, the right factor e_3, and the uniform left factor $\mathbf{e}\cdot\mathbf{B}_g$. Subtracting $\mathsf{X}(\mathsf{N}\psi)$ from $\mathsf{N}(\mathsf{X}\psi)$ term by term gives

$$[\mathsf{N},\mathsf{X}]\,\psi = \frac{1}{c\alpha_n}\sum_k \Big(e_k\,[D_t, D_k]\,\psi - e_k\,[\mathbf{A}_g\cdot\mathbf{D}, D_k]\,\psi + \tfrac{1}{4}\,\big[\mathbf{e}\cdot\mathbf{B}_g, e_k\big]\,D_k\psi\Big)\,e_3, \tag{A18}$$

where $[\mathbf{A}_g\cdot\mathbf{D}, D_k]\,\psi = \mathbf{A}_g\cdot\mathbf{D}(D_k\psi) - D_k(\mathbf{A}_g\cdot\mathbf{D}\psi)$ and $[\mathbf{e}\cdot\mathbf{B}_g, e_k] = (\mathbf{e}\cdot\mathbf{B}_g)\,e_k - e_k\,(\mathbf{e}\cdot\mathbf{B}_g)$.

The first term is Eq. (64). It vanishes, since V is constant and $\mathbf{A}$ static.

In the second and third terms the indices j, k, l run over $1, 2, 3$, and a repeated index is summed. The components A_{gj} and B_{gj} and their derivatives are real numbers, so they commute with every quaternion. By the product rule, $D_k(A_{gj}\,D_j\psi) = (\partial_k A_{gj})\,D_j\psi + A_{gj}\,D_kD_j\psi$, so $[\mathbf{A}_g\cdot\mathbf{D}, D_k]\,\psi = A_{gj}\,[D_j, D_k]\,\psi - (\partial_k A_{gj})\,D_j\psi$. By Eq. (61) and its cyclic relabelings, $[D_j, D_k]\,\psi = \frac{q}{\hbar}\,\epsilon_{jkl}B_l\,\psi\,e_3$, and $\epsilon_{jkl}B_lA_{gj} = \epsilon_{klj}B_lA_{gj} = (\mathbf{B}\times\mathbf{A}_g)_k$, so the first part is $\frac{q}{\hbar}\,(\mathbf{B}\times\mathbf{A}_g)_k\,\psi\,e_3$. In the second part the constant $\mathbf{u}$ of the rigid flow drops out, and $\mathbf{A}_g = \frac{1}{2}\,\mathbf{B}_g\times\mathbf{x}$ gives $\partial_k A_{gj} = \frac{1}{2}\,\epsilon_{jlk}B_{gl}$. The second term of Eq. (A18), $-e_k\,[\mathbf{A}_g\cdot\mathbf{D}, D_k]\,\psi$, is therefore $-\frac{q}{\hbar}\,\mathbf{e}\cdot(\mathbf{B}\times\mathbf{A}_g)\,\psi\,e_3 + \frac{1}{2}\,\epsilon_{jlk}B_{gl}\,e_k\,D_j\psi$, a magnetic part and a derivative part.

In the third term, Eq. (19) gives $e_le_j - e_je_l = 2\,\epsilon_{ljk}\,e_k$: the parts $-\delta_{lj}$ cancel and the parts $\epsilon_{ljk}\,e_k$ add, since $\epsilon_{jlk} = -\epsilon_{ljk}$. With the index of $\mathbf{B}_g$ written l and the summed index of e_k renamed j, the third term is $\frac{1}{4}\,B_{gl}\,(e_le_j - e_je_l)\,D_j\psi = \frac{1}{2}\,\epsilon_{ljk}B_{gl}\,e_k\,D_j\psi$. Since $\epsilon_{ljk} = -\epsilon_{jlk}$, it cancels the derivative part of the second term. The magnetic part remains. Its right factor e_3 multiplies the e_3 of Eq. (A18), and since $e_3e_3 = -1$ this gives

$$[\mathsf{N},\mathsf{X}]\,\psi = \frac{q}{\hbar c\alpha_n}\,\mathbf{e}\cdot(\mathbf{B}\times\mathbf{A}_g)\,\psi. \tag{A19}$$

The vector $\mathbf{B}\times\mathbf{A}_g$ is the electric field $\mathbf{v}\times\mathbf{B}$ that an observer moving at the velocity $\mathbf{v} = -\mathbf{A}_g$, the normal observer, sees in the magnetic field $\mathbf{B}$ to first order in $\mathbf{v}$. With $\mathbf{A} = \mathbf{0}$, case (b) of Eq. (92), the commutator vanishes.

5. The commutator in a static field

For the commutator of Eq. (58), let α_n, V, and $\mathbf{A}$ be static functions of position and $\mathbf{A}_g = \mathbf{0}$. Then $D_n = (c\alpha_n)^{-1}D_t$ and $\mathsf{N}\psi = (c\alpha_n)^{-1}(D_t\psi)\,e_3$. In $\mathsf{N}(\mathsf{X}\psi) = (c\alpha_n)^{-1}\sum_k e_k\,(D_tD_k\psi)\,e_3$ the static factor $1/\alpha_n$ passes through D_t, while in $\mathsf{X}(\mathsf{N}\psi) = \sum_k e_k\,D_k\big((c\alpha_n)^{-1}(D_t\psi)\,e_3\big)$ the derivative D_k differentiates it by the product rule, so

$$[\mathsf{N},\mathsf{X}]\,\psi = \frac{1}{c\alpha_n}\sum_k e_k\,\big([D_t, D_k]\,\psi\big)\,e_3 - \sum_k e_k\,\partial_k\Big(\frac{1}{c\alpha_n}\Big)\,(D_t\psi)\,e_3, \tag{A20}$$

where each e_k multiplies from the left and each e_3 from the right. By Eq. (64) with $\partial_t A_k = 0$, and $e_3e_3 = -1$, the first sum is $-(\hbar c\alpha_n)^{-1}\,(\mathbf{e}\cdot\nabla V)\,\psi$, which for $V = q\Phi$, with $\nabla V = -q\mathbf{E}$ for static fields, is $(q/\hbar c\alpha_n)\,(\mathbf{e}\cdot\mathbf{E})\,\psi$. In the

second, $\partial_k(1/(c\alpha_n)) = -(\partial_k\alpha_n)/(c\alpha_n^2)$, and $(D_t\psi)\,e_3 = (\partial_t\psi)\,e_3 + (V/\hbar)\,\psi$ by $D_t\psi = \partial_t\psi - (V/\hbar)\,\psi\,e_3$. For any state, with $V = q\Phi$,

$$[\mathsf{N},\mathsf{X}]\,\psi = \frac{q}{\hbar c\alpha_n}\,(\mathbf{e}\cdot\mathbf{E})\,\psi + \frac{1}{c\alpha_n^2}\,(\mathbf{e}\cdot\nabla\alpha_n)\left[(\partial_t\psi)\,e_3 + \frac{V}{\hbar}\,\psi\right]. \tag{A21}$$

The electric term contains the factor $1/\alpha_n$ because $\mathbf{E}$ is the coordinate field and the normal observer measures $\mathbf{E}/\alpha_n$. The bracket in the lapse term is $(D_t\psi)\,e_3 = -\hat{\mathcal{E}}\psi/\hbar$, where $\hat{\mathcal{E}}$ is the energy operator of Section III.

On a stationary state of frequency ω, Eq. (75) gives $(\partial_t\psi)\,e_3 = \psi\,e_3\,\omega\,e_3 = -\omega\,\psi$, so the bracket is $-\big[(\hbar\omega - V)/\hbar\big]\psi$, where $\hbar\omega - V$ is a function of position. Define the function of position $\mathcal{E} \equiv (\hbar\omega - V)/\alpha_n$, the value that $\hat{\mathcal{E}}$ takes on the stationary state divided by the lapse. With ∇V written back for $-q\mathbf{E}$,

$$[\mathsf{N},\mathsf{X}]\,\psi = -\frac{1}{\hbar c\alpha_n}\,\mathbf{e}\cdot\big(\nabla V + \mathcal{E}\,\nabla\alpha_n\big)\,\psi. \tag{A22}$$

The factor $\mathbf{e}\cdot(\nabla V + \mathcal{E}\,\nabla\alpha_n)$ is the vector in the bracket with each component multiplied by its unit, a quaternion that multiplies ψ from the left. Equation (A22) is exact. With $\mathbf{A}_g = \mathbf{0}$, N commutes with multiplication by the static function $W^{-1/2}$, so $[\mathsf{N},\mathsf{X}_W]\psi = W^{-1/2}[\mathsf{N},\mathsf{X}](W^{-1/2}\psi)$, and Eq. (A22) applied to $W^{-1/2}\psi$ gives

$$[\mathsf{N},\mathsf{X}_W]\,\psi = -\frac{1}{\hbar c\alpha_n W}\,\mathbf{e}\cdot\big(\nabla V + \mathcal{E}\,\nabla\alpha_n\big)\,\psi,$$

which is Eq. (130) in quaternion form. To leading order in $\hbar$, with each operator replaced by its Weyl symbol (Section IV), $\mathcal{E}$ with $\mathbf{A}_g = \mathbf{0}$ is $\sqrt{(mc^2)^2 + c^2W^{-2}\boldsymbol{\pi}^2}$, the symbol of the root of Eq. (92) at zeroth order in $\hbar$. To that order the vector in the bracket is the gradient of the symbol $V + \alpha_n\sqrt{(mc^2)^2 + c^2W^{-2}\boldsymbol{\pi}^2}$ taken through V and α_n alone, so by Eq. (8) its negative is the force $\mathbf{F}$ that the potential and the lapse exert, and $[\mathsf{N},\mathsf{X}]\psi$ in Eq. (A22) is $(\hbar c\alpha_n)^{-1}\,(\mathbf{e}\cdot\mathbf{F})\,\psi$, that force with each component multiplied by its unit.

For a particle held at rest the velocity $\nabla_{\mathbf{p}}(\hbar\omega)$ of Hamilton's canonical equations (Section II A) vanishes, where $\nabla_{\mathbf{p}}$ is the gradient with respect to the canonical momentum $\mathbf{p} = \hbar\mathbf{k}$. With $\mathbf{A}_g = \mathbf{0}$ this gradient of the symbol is $\alpha_n c^2 W^{-2}\boldsymbol{\pi}/\mathcal{E}$, so $\boldsymbol{\pi} = \mathbf{0}$ and the root is $\mathcal{E} = mc^2$. Write $\mathbf{g}_n \equiv -c^2\nabla\ln\alpha_n$, the force of Eq. (6), with α_n in place of α, divided by $m\alpha_n$; for α_n near 1 it is the Newtonian field $\mathbf{g}$ of Eq. (5). Then the term $mc^2\,\nabla\alpha_n = mc^2\alpha_n\,\nabla\ln\alpha_n$ is $-m\alpha_n\,\mathbf{g}_n$, and since $\nabla V = -q\mathbf{E}$ the commutator is

$$[\mathsf{N},\mathsf{X}]\,\psi = \frac{1}{\hbar c}\left(\frac{q}{\alpha_n}\,\mathbf{e}\cdot\mathbf{E} + m\,\mathbf{e}\cdot\mathbf{g}_n\right)\psi. \tag{A23}$$

As in Eq. (65), $\mathbf{e}\cdot\mathbf{E}$ and $\mathbf{e}\cdot\mathbf{g}_n$ multiply ψ from the left; q and m are the charge and mass of the particle. Divided by the term κ^2 of Eq. (58), the gravitational term is $|\mathbf{g}_n|/(\kappa c^2) = |\nabla\ln\alpha_n|/\kappa$, and the electric term is $|\nabla V|/(\kappa\,\alpha_n mc^2)$: each is the reduced Compton wavelength $1/\kappa$ divided by the distance over which the potential changes by the rest energy. Both are small for a potential that varies slowly on the scale $1/\kappa$, and they enter Eq. (103) through its spin-orbit term ω_{so}.

6. The higher terms of the commutator term

The higher terms of C follow from Eq. (97) and the binomial series of the root of Eq. (92), with $\alpha_n = 1 + \mathsf{G}/mc^2$ and $W = 1 + O(c^{-2})$. Each is a polynomial in $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$, W, G, and $\hat{V}$, and it vanishes when they are numbers, since C then vanishes. Writing the factors of each product in a fixed order changes the polynomial only by products that contain a commutator of two factors, and the reordered polynomial is the value at numbers, which is zero. So every term of C is a sum of products each of which contains a commutator.

Both C and $[\mathsf{N},\mathsf{X}_W]$ are built from commutators, those of $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$, W, G and $\hat{V}$ for C, and those of $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$, α_n and $\hat{V}$ for $[\mathsf{N},\mathsf{X}_W]$; but $[\mathsf{N},\mathsf{X}_W]$ does not determine C. Replace V by $V + \delta\,\alpha_n$ for a number δ. Since $\mathsf{N}\psi = -(\hbar\omega - \hat{V})\,\psi/(\hbar c\alpha_n)$ (Section IV), this adds the number $\delta/\hbar c$ to N and leaves $[\mathsf{N},\mathsf{X}_W]$ unchanged at every frequency. The change $\delta\,\alpha_n = \delta + \delta\,\mathsf{G}/mc^2$ has two parts. The number δ adds δ to H and to $\hat{h}$ and leaves C unchanged. The part $\delta\,\mathsf{G}/mc^2$ is of order c^{-2} and changes the terms of C of order c^{-4} and beyond only at order c^{-6}. Since the term of order c^{-2} of Eq. (99) is linear in f, this part changes the leading term of C (Eq. (104)) by the terms of order c^{-2} of Eq. (107) with $f = \delta\,\mathsf{G}/mc^2$. The spin part of these terms is $(\delta\hbar/4m^3c^4)\,\boldsymbol{\sigma}\cdot(\nabla\mathsf{G}\times\hat{\boldsymbol{\pi}})$ and is not zero wherever the lapse varies. For $\delta \neq 0$ and a lapse that varies, the potentials V and $V + \delta\,\alpha_n$ therefore give the same $[\mathsf{N},\mathsf{X}_W]$ at every frequency and different commutator terms.

7. Fine structure and geodetic precession from the expansion

In the Taylor expansion of the master frequency equation (Eq. (103)), the relativistic term $\hbar\,\omega_{\rm rel}$ (Eq. (120)) and the Darwin and spin-orbit commutator term $\hbar\,\omega_{\rm so}$ (Eq. (122)) give the $O(c^{-2})$ fine structure of atomic levels and the geodetic precession of a spin in a gravitational field. Without the commutator $[\mathsf{N},\mathsf{X}_W]$ of Eq. (58), the Darwin and spin-orbit terms of Eq. (122) are absent and only the relativistic kinetic correction $-\hat{\mathbf{p}}^4/(8m_0^3c^2)$ of Eq. (120) remains.

For hydrogen-like atoms in flat spacetime with $\mathbf{A}=\mathbf{0}$, $V=-Z\alpha_{\rm fs}\hbar c/r$ for a nucleus of charge Ze, where e is the elementary charge, ε_0 is the vacuum permittivity, and $\alpha_{\rm fs}=e^2/(4\pi\varepsilon_0\hbar c)$ is the fine-structure constant. On the state of principal quantum number n, orbital quantum number l, and total angular momentum quantum number j, the three $O(c^{-2})$ terms of Eqs. (120) and (122) have the following expectation values [72], in units of $(Z\alpha_{\rm fs})^4 m_0c^2$. The relativistic kinetic term $-\hat{\mathbf{p}}^4/(8m_0^3c^2)$ gives $-(1/2n^4)\big(n/(l+\frac12)-\frac34\big)$. The Darwin term $(\hbar^2/8m_0^2c^2)\nabla^2V$ gives $1/(2n^3)$ for $l=0$ and zero otherwise. The spin-orbit term $(2m_0^2c^2r)^{-1}(dV/dr)\,\mathbf{S}\cdot\mathbf{L}$ gives $[j(j{+}1){-}l(l{+}1){-}\frac34]/\big(4n^3\,l\,(l{+}\frac12)(l{+}1)\big)$ for $l\geq1$ and zero for $l=0$. For $j=l\pm\frac12$ the three add to

$$\Delta E_{nj}=-\frac{(Z\alpha_{\rm fs})^4\,m_0c^2}{2n^4}\left(\frac{n}{j+\frac12}-\frac34\right), \tag{A24}$$

the fine structure of the Dirac equation at this order, which depends on j alone. The kinetic term alone is the same expression with l in place of j, the Klein–Gordon result, which separates the $2S$ level from the $2P$ levels and leaves $2P_{1/2}$ and $2P_{3/2}$ together. The measured spectrum has the pattern of Eq. (A24), with $2S_{1/2}$ and $2P_{1/2}$ together apart from the much smaller Lamb shift [113].

For a particle in the Schwarzschild metric in Painlevé–Gullstrand coordinates ($\alpha_n=W=1$, $\mathbf{A}_g=v_{\rm ff}\,\hat{\mathbf{r}}$ with $v_{\rm ff}=\sqrt{2GM/r}$, Section V), the strain force of Eq. (108) is $\mathbf{F}_\Sigma=(v_{\rm ff}/r)\,\mathbf{p}-\frac32(v_{\rm ff}/r)\,(\mathbf{p}\cdot\hat{\mathbf{r}})\,\hat{\mathbf{r}}$, whose first term is parallel to $\mathbf{p}$ and so drops out of $\mathbf{p}\times\mathbf{F}_\Sigma$. With $\mathbf{p}=m_0(\mathbf{v}+\mathbf{A}_g)$, the radial momentum is $\mathbf{p}\cdot\hat{\mathbf{r}}=m_0(v_r+v_{\rm ff})$, where $v_r=\mathbf{v}\cdot\hat{\mathbf{r}}$ is the radial velocity. Also $\mathbf{x}\times\mathbf{p}=m_0\,\mathbf{x}\times\mathbf{v}$. The spin-orbit precession rate $\boldsymbol{\Omega}_{\rm spin}=(\mathbf{p}\times\mathbf{F}_\Sigma)/(2m_0^2c^2)$ of Section V on a general (non-circular) orbit is

$$\boldsymbol{\Omega}_{\rm spin}=\frac{3v_{\rm ff}\,(v_{\rm ff}+v_r)}{4c^2r^2}\,(\mathbf{x}\times\mathbf{v}). \tag{A25}$$

The part in $v_{\rm ff}^2=2GM/r$ is

$$\boldsymbol{\Omega}_{\rm geo}=\frac{3GM}{2c^2r^3}\,(\mathbf{x}\times\mathbf{v}), \tag{A26}$$

the geodetic precession of general relativity [77, 87, 88]. The part in $v_{\rm ff}\,v_r$ is a time derivative: $v_{\rm ff}\,v_r/r^2=\sqrt{2GM}\,r^{-5/2}\,dr/dt=-\frac23\sqrt{2GM}\,d(r^{-3/2})/dt$, and $\mathbf{x}\times\mathbf{v}$ is constant along a Newtonian orbit, so that part is the time derivative of the bounded vector $-\sqrt{2GM}\,r^{-3/2}\,(\mathbf{x}\times\mathbf{v})/(2c^2)$. That part turns the spin back and forth by a bounded angle, its mean over a bound orbit vanishes, and on a circular orbit, where $v_r=0$, it vanishes at every instant.

On a circular orbit of radius r and angular frequency $\omega_{\rm orb}=\sqrt{GM/r^3}$, $|\mathbf{x}\times\mathbf{v}|=\omega_{\rm orb}r^2$ and Eq. (A26) is $\frac32(GM/c^2r)\,\omega_{\rm orb}$. Gravity Probe B measured this drift in a polar orbit 642 km above the Earth as 6,601.8±18.3 mas/yr, against the 6,606.1 mas/yr that general relativity gives for its orbit [77]; the circular-orbit expression of Eq. (A26) gives 6.6×10^3 mas/yr at that altitude. On Painlevé–Gullstrand slices ($\alpha_n=W=1$) the whole geodetic precession comes from the shift through the commutator term $\omega_{\rm so}$; in isotropic coordinates ($\mathbf{A}_g=\mathbf{0}$, $\alpha_n\approx1-GM/rc^2$, $W\approx1+GM/rc^2$), one-third comes from the lapse $\nabla\mathsf{G}$ and two-thirds from the spatial curvature $-2m_0c^2\nabla\ln W$ in f_W (Eq. (122)) [9, 10].

Appendix B: The level theorems and their proofs

Definition 1 and the notation of Sections VI and A apply throughout this section. Every quantity in this section is measured in units in which m, c and one fixed length are 1, so $\hbar$, the energies and all constants are numbers, and $\hbar$ small means $\hbar$ small against mc times that length. Where a statement does not contain m, any fixed positive mass may stand in for m. Symbols introduced in this section are local to it: the letters used for Olver's functions, for the matching constants and for the variables of the scalar equation have other meanings elsewhere in the paper. In particular τ is the independent variable of Olver's equation and then of the scalar equation in its Liouville form, not a proper time, and ζ is Olver's Langer variable, not a rapidity. A prime on a function is its derivative, in the variable named where the prime is introduced ($d/d\tau$ or d/dx); a prime on a variable, as in E', J' or the integration variable x', marks a second value of that variable.

Lemma 1 (From the action to the energy)**:** *Let J be a compact interval, and for each $E \in J$ let the level set $h = E$ of a smooth real function h on the phase plane be a closed curve with no critical point that bounds the set where $h < E$, a well; let $\mathcal{A}(E)$ be the area of that set and $T(E)$ the period of the curve under the flow of h. Let $\varphi_\hbar(E)$ be continuously differentiable on J with $|d\varphi_\hbar/dE| \le C_1$, let $\mathcal{I}_1 \equiv \mathcal{A} - \hbar\varphi_\hbar$, and let $\mathcal{I}_\hbar$ be continuous on J with $|\mathcal{I}_\hbar - \mathcal{I}_1 - \pi\hbar| \le C_\mathcal{I}\hbar^2$, where C_1 and $C_\mathcal{I}$ do not depend on $\hbar$. Let $Z_\hbar$ be the set of $E \in J$ with $\mathcal{I}_\hbar(E) \in 2\pi\hbar\mathbb{Z}$, let E_n^0 be the solutions of $\mathcal{I}_1(E) = 2\pi\hbar(n+\frac12)$, and let $S_\hbar$ be a set of real numbers such that, for every compact J_1 inside the interior of J and $\hbar$ small, each point of $S_\hbar \cap J_1$ lies within $\varepsilon_\hbar$ of $Z_\hbar$, and each point of $Z_\hbar \cap J_1$ within $\varepsilon_\hbar$ of $S_\hbar \cap J$, with $\varepsilon_\hbar \le \hbar^3$. Then $S_\hbar$ is a two-way $O(\hbar^2)$ spectral approximation of the E_n^0 on J (Definition 1, with $S_\hbar$ in place of the eigenvalues), and for every compact J' inside the interior of J and $\hbar$ small, each point of $S_\hbar$ in J' lies within $C\hbar^2$ of E_n^0 for exactly one n. If in addition $|d(\mathcal{I}_\hbar - \mathcal{I}_1)/dE| \le C_\mathcal{I}\hbar^2$, or $\mathcal{I}_\hbar$ takes each value in $2\pi\hbar\mathbb{Z}$ at most once on J, exactly one point of $Z_\hbar$ lies within $2C_\mathcal{I}\hbar^2/\min_J T$ of each E_n^0 in J'.*

Proof. By the implicit function theorem the curves depend smoothly on E, so T is continuous and positive on J; let $T_{\min} > 0$ and $T_{\max}$ be its least and largest values. Since $dx\,d\xi = dE\,dt$ and the set where $h < E$ grows with E, $d\mathcal{A}/dE = T$, so $T_{\min}/2 \le d\mathcal{I}_1/dE \le T_{\max} + \hbar C_1$ for $\hbar C_1 \le T_{\min}/2$. This lower bound on the rate uses T and $\varphi_\hbar$ only, no derivative of the remainder $\mathcal{I}_\hbar - \mathcal{I}_1 - \pi\hbar$. Fix $J' \subset J_1$, compact intervals with J' inside the interior of J_1 and J_1 inside the interior of J, and let $\delta_0 \equiv 2C_\mathcal{I}\hbar^2/T_{\min}$. The function $G_n(E) \equiv \mathcal{I}_\hbar(E) - 2\pi\hbar(n+1)$ is $\mathcal{I}_1(E) - \mathcal{I}_1(E_n^0)$ plus the remainder. (1) If $E_n^0 \in J'$, the rate bound gives $G_n \le 0$ at $E_n^0 - \delta_0$ and $G_n \ge 0$ at $E_n^0 + \delta_0$, both points in J_1 for $\hbar$ small, so G_n has a zero, a point of $Z_\hbar \cap J_1$, within δ_0 of E_n^0, and a point of $S_\hbar$ lies within $\delta_0 + \varepsilon_\hbar$ of E_n^0. (2) If $E \in S_\hbar \cap J'$, a point $E' \in Z_\hbar$ lies within $\varepsilon_\hbar$ of E, in J_1 for $\hbar$ small, and $\mathcal{I}_\hbar(E') = 2\pi\hbar(n+1)$ for some n, so $|\mathcal{I}_1(E') - 2\pi\hbar(n+\frac12)| \le C_\mathcal{I}\hbar^2$. Since $\mathcal{I}_1$ increases at a rate of at least $T_{\min}/2$ on J and E' is a fixed distance from the ends of J, for $\hbar$ small that value is taken at a point $E_n^0 \in J$ with $|E' - E_n^0| \le \delta_0$, and $|E - E_n^0| \le \delta_0 + \varepsilon_\hbar$. (3) Consecutive E_n^0 differ in $\mathcal{I}_1$ by $2\pi\hbar$, so they are at least $2\pi\hbar/(T_{\max} + \hbar C_1)$ apart, more than $2(\delta_0 + \varepsilon_\hbar)$ for $\hbar$ small, and the n in (2) is unique. So $C\hbar^2 = \delta_0 + \hbar^3$ serves, with C independent of n and $\hbar$. If $|d(\mathcal{I}_\hbar - \mathcal{I}_1)/dE| \le C_\mathcal{I}\hbar^2$, G_n increases on $[E_n^0 - \delta_0, E_n^0 + \delta_0]$ for $\hbar$ small, and its zero there is unique. If $\mathcal{I}_\hbar$ takes each value in $2\pi\hbar\mathbb{Z}$ at most once, let $E' \in Z_\hbar$ lie within δ_0 of $E_n^0 \in J'$; by (2) and (3), $\mathcal{I}_\hbar(E') = 2\pi\hbar(k+1)$ with E' within δ_0 of E_k^0, so $k = n$, and E' is the one point at which $\mathcal{I}_\hbar = 2\pi\hbar(n+1)$. □

The separation in step (3) fixes the integer n attached to each point of $S_\hbar$, but it does not exclude two points near one E_n^0; the last statement of the lemma needs one of its two additional hypotheses for that. For the theorem of Ifa *et al.* [94], $h = \mathfrak{p}_0$, $\varphi_\hbar = \oint \mathfrak{p}_1\,dt_\mathfrak{p}$, smooth in E because the curves are, and $S_\hbar = Z_\hbar$, the eigenvalues in J, with $\varepsilon_\hbar = 0$.

1. The energy-dependent operator

Write $E \equiv \hbar\omega$, let $p_z = 0$, and let $\mathsf{K}(E)$ be the operator K of the electric reduction at the energy E, with $g = 1/(E - V + mc^2)$. In the representation with diagonal mass term the matrix $\mathrm{diag}(\sigma_z, -\sigma_z)$ commutes with the Dirac operator at $p_z = 0$, since $\boldsymbol{\sigma}\cdot\hat{\boldsymbol{\pi}}$ then contains only σ_x and σ_y, which anticommute with σ_z. Its eigenvalue s labels the invariant subspaces, with upper component along $\check\chi_s$ and lower component along $\check\chi_{-s}$, and the Dirac operator on the subspace of eigenvalue s at fixed p_y is the *block* of sector s, acting on the pair (u, d) of upper and lower amplitudes. This s is a conserved label of the block; it is not the eigenvalue of the Dirac spin matrix $\mathrm{diag}(\sigma_z, \sigma_z)$, which is not conserved here. Since $Q_s^\dagger g Q_s = \hat p\, g\, \hat p + g\varrho^2 - s\hbar\,(g\varrho)'$ and the Weyl symbol of $\hat p\, g\, \hat p$ is $g\,\xi^2 + \frac14\hbar^2 g''$, the symbol of $\mathsf{K}(E)$ is $\mathsf{K}_0 + \hbar\mathsf{K}_1 + \hbar^2\mathsf{K}_2$, with the three terms given in the electric reduction. Since $(g\varrho)' = g\,qB + g^2\varrho V'$ and $dt = 2g\,(E - V)\,dt_0$ on the curve $\mathsf{K}_0 = 0$, the relation $\mathsf{K}_1\,dt_0 = \frac12 s\,\Omega_z\,dt$ follows, where t_0 is the time of the flow of K_0.

Lemma 2 (Levels of the energy-dependent operator)**:** *Let V be bounded with a continuous derivative, let B be continuous, and let J be a compact interval of energies on which $E - V + mc^2 \ge E_{gap} > 0$. Then: (i) the operators $\mathsf{K}(E)$, $E \in J$, are self-adjoint with the common form domain $\mathrm{dom}\,Q_s$; (ii) for $E < E'$ in J, $\mathsf{K}(E') \le \mathsf{K}(E) - (E' - E)$ as forms on $\mathrm{dom}\,Q_s$, so each eigenvalue $\mathsf{k}_j(E)$ of $\mathsf{K}(E)$ below its essential spectrum, counted from the bottom, decreases with E at a rate of at least 1 and has at most one zero in J; (iii) if 0 lies below the essential spectrum of $\mathsf{K}(E)$ for every $E \in J$, the Dirac levels in J with upper component $F\check\chi_s$ are exactly the zeros of the k_j, and each is simple.*

Proof. (i) On J, g lies between two positive constants, so the form $c^2\|g^{1/2}Q_sF\|^2$ is comparable to $\|Q_sF\|^2$ and closed on $\mathrm{dom}\,Q_s$, and $\mathsf{K}(E)$ is the operator of that form minus a bounded function. (ii) $g(E) - g(E') = (E' - E)\,g(E)g(E') \ge 0$, and $-(E - V - mc^2)$ decreases by $E' - E$. By the minimax principle, the form inequality gives $\mathsf{k}_j(E') \le \mathsf{k}_j(E) - (E' - E)$ for each j. (iii) A Dirac level is an E with $\mathsf{K}(E)F = 0$ for some $0 \ne F \in \mathrm{dom}\,Q_s$, an isolated eigenvalue below the essential spectrum and so one of the $\mathsf{k}_j(E)$. The equation sets $\hbar^2(gF')'$ equal to F times the continuous function $g\varrho^2 - s\hbar(g\varrho)' - (E - V - mc^2)/c^2$, so F and gF' are continuously differentiable. For two solutions the Wronskian $g\,(F_1F_2' - F_2F_1')$ therefore has derivative $F_1(gF_2')' - F_2(gF_1')' = 0$ and is constant. Since $Q_s = \hat p - is\varrho$,

$F' = (i/\hbar)\,Q_s F - (s\varrho/\hbar)\,F$, and the terms in ϱ cancel: $F_1F_2' - F_2F_1' = (i/\hbar)\,(F_1\,Q_sF_2 - F_2\,Q_sF_1)$, a sum of products of square-integrable functions for $F_1, F_2 \in \operatorname{dom} Q_s$. The Wronskian is thus constant and integrable, so it is zero and the level is simple. □

Lemma 3 (The essential spectrum, uniformly in the energy)**:** *Let $qB = \varrho'$ be bounded, let J be a compact interval of energies, and for $E \in J$ let $\mathsf{K}(E) \equiv c^2\,Q_s^\dagger\,g\,Q_s - \mathcal{U}$ on the form domain $\operatorname{dom} Q_s$, where $g(x,E)$ lies between two positive constants, $(\ln g)'$ and $\mathcal{U}(x,E)$ are bounded, all uniformly in $E \in J$, and $c^2g\varrho^2 - \mathcal{U}$ decreases with E at each x. Let $\Lambda_\infty(E)$ be the lower limit of $c^2g\varrho^2 - \mathcal{U}$ as $|x| \to \infty$, and let $\Lambda_\infty(\max J) > 0$. Then there are $\hbar_0 > 0$ and $\Lambda_{ess} > 0$ such that for $\hbar < \hbar_0$ the essential spectrum of $\mathsf{K}(E)$ lies in $[\Lambda_{ess}, \infty)$ for every $E \in J$.*

Proof. Since $c^2g\varrho^2 - \mathcal{U}$ decreases with E, $\Lambda_\infty(E) \geq \Lambda_\infty(\max J)$ on J, and any finite Λ_* with $0 < \Lambda_* \leq \Lambda_\infty(\max J)$ is a lower bound for $\Lambda_\infty(E)$ on J, also where ϱ is unbounded and $\Lambda_\infty = +\infty$. Fix one such Λ_*. Three steps give a lower bound on the essential spectrum. (1) Since $(g\varrho)' = g\,qB + g\varrho\,(\ln g)'$, with g between two positive constants and $2c\sqrt{g\Lambda_*}\,|\varrho| \leq \Lambda_* + c^2g\varrho^2$, there is a C with $\hbar c^2|(g\varrho)'| \leq C\hbar\,(\Lambda_* + c^2g\varrho^2)$. For $F \in \operatorname{dom} Q_s$, $\langle F, \mathsf{K}F\rangle = c^2\|g^{1/2}\hat{p}F\|^2 + \int \big(c^2g\varrho^2 - s\hbar c^2(g\varrho)' - \mathcal{U}\big)|F|^2dx$, and the first term is nonnegative, so $\langle F, \mathsf{K}F\rangle \geq \int \big((1-C\hbar)\,(c^2g\varrho^2 - \mathcal{U}) - C_{\mathcal{U}}\hbar\big)|F|^2dx$, with $C_{\mathcal{U}} \equiv C\,(\Lambda_* + \sup|\mathcal{U}|)$. In particular $\mathsf{K}(E) \geq -C_K$ for a constant C_K. Let $\Lambda_{\text{cut}} \equiv (1 - C\hbar)\Lambda_*/2 - C_{\mathcal{U}}\hbar$. There is an x_{cut} such that $\langle F, \mathsf{K}F\rangle \geq \Lambda_{\text{cut}}\|F\|^2$ for every $F \in \operatorname{dom} Q_s$ supported in $|x| > x_{\text{cut}}$. The point x_{cut} is chosen at $E = \max J$, and the bound then holds with the same x_{cut} for every $E \in J$ because $c^2g\varrho^2 - \mathcal{U}$ decreases with E. (2) Let $\mathfrak{c}_1, \mathfrak{c}_2$ be smooth cutoff functions with $\mathfrak{c}_1^2 + \mathfrak{c}_2^2 = 1$, $\mathfrak{c}_1 = 1$ on $|x| \leq x_{\text{cut}}$, $\mathfrak{c}_1 = 0$ on $|x| \geq 2x_{\text{cut}}$, and $|\mathfrak{c}_j'| \leq 2/x_{\text{cut}}$. Expanding $|(\mathfrak{c}_jF)'|^2$ gives, on $\operatorname{dom} Q_s$, $\langle F, \mathsf{K}F\rangle = \sum_j \langle \mathfrak{c}_jF, \mathsf{K}\mathfrak{c}_jF\rangle - \hbar^2c^2\int g\,(\mathfrak{c}_1'^2 + \mathfrak{c}_2'^2)\,|F|^2dx$. With $C_{\text{cut}} \equiv C_K + \Lambda_{\text{cut}}$, steps (1) and (2) give $\langle F, (\mathsf{K} + C_{\text{cut}}\mathfrak{c}_1^2)F\rangle \geq (\Lambda_{\text{cut}} - 8\hbar^2c^2\sup g/x_{\text{cut}}^2)\|F\|^2$, where $\sup g$ is taken over x and $E \in J$ and is finite by hypothesis. (3) Functions in $\operatorname{dom} Q_s$ restricted to $|x| < 2x_{\text{cut}}$ lie in H^1 of that interval, which embeds compactly in L^2 (Rellich), so the multiplication by $C_{\text{cut}}\mathfrak{c}_1^2$ is compact relative to the form of K, and by Weyl's theorem K and $\mathsf{K} + C_{\text{cut}}\mathfrak{c}_1^2$ have the same essential spectrum. So, for every $E \in J$, the essential spectrum of $\mathsf{K}(E)$ lies in $[\Lambda_{\text{cut}} - 8\hbar^2c^2\sup g/x_{\text{cut}}^2, \infty)$; this bound does not depend on E, and for $\hbar$ small it is at least $\Lambda_{\text{ess}} \equiv \Lambda_*/4$. □

For the operator of the electric reduction, $\mathcal{U} = E - V - mc^2$ and $(\ln g)' = gV'$; both terms of $c^2g\varrho^2 - \mathcal{U}$ decrease with E, and $\Lambda_\infty(E)$ is the lower limit of K_0 at infinity, since K_0 is least at $\xi = 0$. So Lemma 3 applies whenever B and V' are bounded and 0 lies below the lower limit of K_0 at infinity at $E = \max J$.

Why a rule at fixed energy is not enough. A Dirac level is an energy E at which 0 is an eigenvalue of $\mathsf{K}(E)$, so the energy is both the unknown and a parameter of the operator. A Bohr–Sommerfeld theorem for $\mathsf{K}(E)$ at each fixed E locates the eigenvalues of that one operator with a constant that may depend on E. Two things are missing for the levels. (1) To locate the zeros in E to within $C\hbar^2$, the error of the rule must be bounded by one constant for all E in the interval. With a constant that depends on E, the error may grow without bound along a sequence of energies, and nothing is known about the zeros there. (2) To count, an approximation of each $\mathcal{I}_\hbar(E)$ by the classical action does not by itself exclude that $\mathcal{I}_\hbar$ takes a quantized value two or three times near one solution of the rule, since the error may oscillate in E. Lemma 1 needs $\mathcal{I}_\hbar$ continuous, and the count needs it to take each quantized value at most once. Explicit error bounds whose constants depend only on quantities bounded on the whole interval give (1), and a Sturm comparison that uses the monotonicity in E of the coefficients gives (2).

2. The uniform Airy approximation

The explicit error bounds are those of the uniform Airy approximation of Olver [114]. Olver states it for $d^2w/dx^2 = \{u^2f(x) + g(x)\}w$ with a positive parameter u and adds that "Generally the parameter u is to be thought of as being large, but this is inessential in the error analysis, nor need $f(x)$ and $g(x)$ be independent of u" [114, Ch. 11, §3.1, p. 398]. With his x, u, f and g renamed τ, ν_O, f_O and g_O, it reads as follows.

Proposition 1 (Olver's uniform Airy approximation)**:** *Let $\nu_O > 0$, let (τ_a, τ_b) be an interval, and let f_O and g_O be real functions on (τ_a, τ_b), which may depend on ν_O, such that f_O has exactly one zero τ_* in (τ_a, τ_b), $f_O/(\tau - \tau_*)$ is positive and twice continuously differentiable, and g_O is continuous. Let ζ be defined by $\frac{2}{3}\zeta^{3/2} = \int_{\tau_*}^{\tau} f_O^{1/2}d\tau$ for $\tau \geq \tau_*$ and $\frac{2}{3}(-\zeta)^{3/2} = \int_{\tau}^{\tau_*}(-f_O)^{1/2}d\tau$ for $\tau \leq \tau_*$, let $\hat{f} \equiv f_O/\zeta$, and let*

$$\psi_O \equiv \frac{g_O}{\hat{f}} - \hat{f}^{-3/4}\,\frac{d^2\hat{f}^{-1/4}}{d\tau^2}.$$

Let $\mathcal{V}_O(\tau_1, \tau_2)$ be the variation on $[\tau_1, \tau_2]$ of $\mathcal{H}_O \equiv -\int_0^\zeta |t|^{-1/2}\psi_O\,dt$, and let $X_O \equiv \nu_O^{2/3}\zeta$. Then $d^2w/d\tau^2 = (\nu_O^2f_O + g_O)\,w$ has twice continuously differentiable solutions

$$w_1 = \hat{f}^{-1/4}\big(\mathrm{Bi}(X_O) + \epsilon_1\big),$$

$$w_2 = \hat f^{-1/4}\big(\mathrm{Ai}(X_O) + \epsilon_2\big),$$

whose errors satisfy, wherever the variations are finite,

$$\max\Big(\frac{|\epsilon_1|}{M_O(X_O)},\, \frac{|d\epsilon_1/d\tau|}{\nu_O^{2/3}\hat f^{1/2} N_O(X_O)}\Big) \le \frac{E_O(X_O)}{\lambda_O}\Big(e^{\lambda_O \mathcal{V}_O(\tau_a,\tau)/\nu_O} - 1\Big),$$

$$\max\Big(\frac{|\epsilon_2|}{M_O(X_O)},\, \frac{|d\epsilon_2/d\tau|}{\nu_O^{2/3}\hat f^{1/2} N_O(X_O)}\Big) \le \frac{1}{\lambda_O\, E_O(X_O)}\Big(e^{\lambda_O \mathcal{V}_O(\tau,\tau_b)/\nu_O} - 1\Big).$$

Here $\lambda_O \equiv \sup_X \pi|X|^{1/2} M_O(X)^2 = 1.0395\ldots$, and E_O, M_O and N_O are fixed functions of one variable built from Ai *and* Bi *alone: $E_O = 1$ below the negative root c_A of $\mathrm{Ai} = \mathrm{Bi}$ of least absolute value and $E_O = (\mathrm{Bi}/\mathrm{Ai})^{1/2}$ above it, and M_O and N_O are the moduli in $E_O\,\mathrm{Ai} = M_O \sin\theta_O$, $E_O^{-1}\mathrm{Bi} = M_O\cos\theta_O$, $E_O\,\mathrm{Ai}' = N_O\sin\omega_O$ and $E_O^{-1}\mathrm{Bi}' = N_O\cos\omega_O$. In particular, if f_O and g_O depend also on a parameter E, the bounds hold for each (E,ν_O) with the same λ_O, E_O, M_O and N_O, so they are uniform on any set of (E,ν_O) on which $\mathcal{V}_O(\tau_a,\tau_b)/\nu_O$ is bounded.*

This is Olver [114, Ch. 11, Lemma 3.1 and Theorem 3.1], with the auxiliary functions of its §2. Lemma 3.1 there gives that $\zeta/(\tau-\tau_*)$ is positive and twice continuously differentiable and that ψ_{O} is continuous, and ψ_{O} is the function ψ of its Eq. (3.05) written with derivatives in τ. The last sentence holds because the constant and the three functions do not depend on the equation, which enters only through $\mathcal{V}_{\mathrm{O}}/\nu_{\mathrm{O}}$.

Proposition 2 (Uniform Airy matching)**:** *Let J be a compact interval. For $E \in J$ and $\hbar > 0$ consider on the line the equation*

$$\hbar^2 u'' = \mathfrak{q}\, u,$$

with $\mathfrak{q} \equiv \mathfrak{q}_0 + \hbar\mathfrak{q}_1 + \hbar^2\mathfrak{q}_2$, a prime denoting $d/d\tau$, and $\mathfrak{q}_0$, $\mathfrak{q}_1$, $\mathfrak{q}_2$ real functions of (τ,E) that do not depend on $\hbar$, such that:

(U1) $\mathfrak{q}_0$, $\mathfrak{q}_1$ and their first three τ-derivatives, and $\mathfrak{q}_2$ and $\mathfrak{q}_2'$, are continuous in (τ,E);

(U2) for each $E \in J$, $\mathfrak{q}_0$ has exactly two zeros $\tau_L < \tau_R$, both simple, and $\mathfrak{q}_0 < 0$ between them;

(U3) there are $\delta, \mathfrak{q}_* > 0$ with $\mathfrak{q}_0 \ge \mathfrak{q}_*$ for $\tau \ge \tau_R + \delta$ and for $\tau \le \tau_L - \delta$, for every $E \in J$;

(U4) there is C_0 such that on these two half-lines $|\mathfrak{q}_1|$, $|\mathfrak{q}_2|$, $|\mathfrak{q}_0'|$, $|\mathfrak{q}_1'|$ and $|\mathfrak{q}_2'|$ are at most $C_0\mathfrak{q}_0$.

Let $\tau_0 \equiv (\tau_L+\tau_R)/2$, $I_+ \equiv [\tau_0, \tau_R+\delta]$, $I_- \equiv [\tau_L-\delta, \tau_0]$, $\mathfrak{f} \equiv \mathfrak{q}_0 + \hbar\mathfrak{q}_1$ and $c_* \equiv \min_{E\in J}\min\big(\mathfrak{q}_0'(\tau_R), -\mathfrak{q}_0'(\tau_L)\big)$. Then there are $\hbar_0 > 0$, C and $c_{dec} > 0$ such that for every $E \in J$ and $0 < \hbar \le \hbar_0$:

(M0) there is a solution u_+ with no zero on $\tau \ge \tau_R + \delta$, with $\int \mathfrak{q} u_+^2\, d\tau$ finite there, with u_+ and u_+' continuous in (τ,E) and $|\hbar u_+'/u_+| \le 2\mathfrak{q}^{1/2}$ there, and every solution square-integrable on that half-line is a multiple of u_+; the same holds for a solution u_- on $\tau \le \tau_L - \delta$;

(M1) $\mathfrak{f}$ has exactly one zero on each of I_+ and I_-, $\tau_{R,\hbar}$ and $\tau_{L,\hbar}$, each simple with $|\mathfrak{f}'| \ge \frac12 c_*$ there and within $C\hbar$ of τ_R and τ_L, and continuous in $(E,\hbar)$;

(M2) Proposition 1 applies on I_+ and on I_- with $\nu_O = 1/\hbar$, $f_O = \mathfrak{f}$ and $g_O = \mathfrak{q}_2$, and the variation of its error-control function there is at most a constant;

(M3) $u_\pm = \alpha_\pm w_2 + \beta_\pm w_1$ on $I_\pm$, where w_1 and w_2 are the solutions of Proposition 1 on that interval, $\alpha_\pm \ne 0$ and $|\beta_\pm/\alpha_\pm| \le C\hbar\, e^{-c_{dec}/\hbar}$;

(M4) with $a_m \equiv (-\mathfrak{f}(\tau_0))^{1/2}$, $\Phi_+ \equiv \int_{\tau_0}^{\tau_{R,\hbar}} \sqrt{-\mathfrak{f}}\, d\tau$, $\Phi_- \equiv \int_{\tau_{L,\hbar}}^{\tau_0} \sqrt{-\mathfrak{f}}\, d\tau$, and the angles $\vartheta_\pm$ of $u_\pm$ at τ_0, defined by $u_\pm = r_\pm \sin\vartheta_\pm$ and $\hbar u_\pm'/a_m = r_\pm\cos\vartheta_\pm$ with $r_\pm > 0$, $|\vartheta_+ + \Phi_+/\hbar + \pi/4| \le C\hbar$ and $|\vartheta_- - \Phi_-/\hbar - \pi/4| \le C\hbar$ modulo π.

Let M_2 be the largest value of $|\mathfrak{q}_0''|$ for $E \in J$ and $\tau_L - \delta \le \tau \le \tau_R + \delta$, let $\ell_* \equiv \min_{E\in J}(\tau_R - \tau_L)$, let $\delta_{tp} \equiv \min\big(\ell_*/5,\, \delta/2,\, c_*/(8M_2)\big)$, and let μ be the least value of $|\mathfrak{q}_0|$ for $E \in J$, $\tau_L - \delta \le \tau \le \tau_R + \delta$ and τ at distance at least δ_{tp} from τ_L and from τ_R. The constants $\hbar_0$, C and c_{dec} depend on the equation only through c_*, through ℓ_*, μ, δ, $\mathfrak{q}_*$ and C_0, and through the bounds of the functions of (U1) for $E \in J$ and $\tau_L - \delta \le \tau \le \tau_R + \delta$; the numbers c_*, ℓ_* and μ are positive, and none of these quantities depends on $E \in J$. The upper bounds of (U1) alone do not bound μ from below, since μ measures how close $\mathfrak{q}_0$ comes to a third zero.

Proof. Every constant in this proof is a least or largest value, over a compact set, of a function continuous in E and the other variables, so none depends on $E \in J$ or on $\hbar$. Let $\tau_1 \equiv \tau_R + \delta$ and $\tau_{-1} \equiv \tau_L - \delta$. Because $\mathfrak{q}_0$ is continuous in (τ, E) with a τ-derivative continuous in (τ, E), and τ_L and τ_R are its only zeros, both simple, the implicit function theorem makes them continuous in E; hence $c_* > 0$ and $\ell_* > 0$. The number $\delta_{\rm tp}$ of the statement is less than $(\tau_R - \tau_L)/4$ for every $E \in J$, and since $|\mathfrak{q}_0''| \le M_2$, for every $E \in J$, $\mathfrak{q}_0' \ge \frac{3}{4}c_*$ within $2\delta_{\rm tp}$ of τ_R and $\mathfrak{q}_0' \le -\frac{3}{4}c_*$ within $2\delta_{\rm tp}$ of τ_L. The number μ of the statement is positive, as the least value of a continuous function on a compact set of (τ, E) on which $\mathfrak{q}_0 \ne 0$. The functions of (U1) are bounded on the compact set of (τ, E) with $E \in J$ and $\tau_{-1} \le \tau \le \tau_1$. By (U3) and (U4), for $\hbar \le \min(1, 1/(4C_0))$, $\mathfrak{q} \ge \mathfrak{q}_0/2 \ge \mathfrak{q}_*/2$ and $|\mathfrak{q}'| \le C_T\mathfrak{q}$ with $C_T \equiv 6C_0$, for $\tau \ge \tau_1$ and for $\tau \le \tau_{-1}$.

(a) *The decaying solutions, (M0).* On $\tau \ge \tau_1$ let $\mathfrak{v} = \hbar u'/u$, so that $\hbar\mathfrak{v}' = \mathfrak{q} - \mathfrak{v}^2$, and $\mathfrak{v} = -\sqrt{\mathfrak{q}} + z$, so that

$$\hbar z' = z\,(2\sqrt{\mathfrak{q}} - z) + \hbar\,(\sqrt{\mathfrak{q}})',$$

with $|(\sqrt{\mathfrak{q}})'| \le \frac{1}{2}C_T\sqrt{\mathfrak{q}}$. Let $b_T \equiv \hbar C_T$, at most $\sqrt{\mathfrak{q}_*}/4$ for $\hbar$ small, so that $b_T \le \frac{1}{2}\sqrt{\mathfrak{q}}$. At $z = b_T$, $\hbar z' \ge b_T\sqrt{\mathfrak{q}}\,(2 - \frac{1}{2} - \frac{1}{2}) > 0$, and at $z = -b_T$, $\hbar z' \le -b_T\sqrt{\mathfrak{q}}\,(2 - \frac{1}{2}) < 0$, so a solution with $z(\mathcal{C}) = 0$, followed toward smaller τ, stays in $|z| \le b_T$ on $[\tau_1, \mathcal{C}]$.

For two solutions $z, \tilde z$ in that band, $\Delta z \equiv z - \tilde z$ satisfies $\hbar\,\Delta z' = \Delta z\,(2\sqrt{\mathfrak{q}} - z - \tilde z)$, whose factor is at least $\sqrt{\mathfrak{q}} \ge \sqrt{\mathfrak{q}_*/2}$, so $|\Delta z(\tau)| \le 2b_T\,e^{-\sqrt{\mathfrak{q}_*/2}\,(\mathcal{C}-\tau)/\hbar}$ for $\tau \le \mathcal{C}$. Let $z_\mathcal{C}$ be the solution with $z_\mathcal{C}(\mathcal{C}) = 0$, defined on $[\tau_1, \mathcal{C}]$. For $\mathcal{C}_0 < \mathcal{C} < \mathcal{C}_1$, the bound with $z = z_{\mathcal{C}_1}$ and $\tilde z = z_\mathcal{C}$ gives $|z_{\mathcal{C}_1} - z_\mathcal{C}| \le 2b_T\,e^{-\sqrt{\mathfrak{q}_*/2}\,(\mathcal{C}-\mathcal{C}_0)/\hbar}$ on $[\tau_1, \mathcal{C}_0]$, uniformly in $E \in J$. So as $\mathcal{C} \to \infty$ the $z_\mathcal{C}$ converge uniformly on each compact interval $[\tau_1, \mathcal{C}_0]$ and $E \in J$, to a function z on $[\tau_1, \infty)$. The limit satisfies $|z| \le b_T$, since every $z_\mathcal{C}$ does. It solves the equation for z, since the solutions of that equation through the values $z_\mathcal{C}(\tau_1)$ converge on each $[\tau_1, \mathcal{C}_0]$ to the solution through $z(\tau_1)$. It is continuous in (τ, E), as a uniform limit on $[\tau_1, \mathcal{C}_0] \times J$ of continuous functions.

Let u_+ be the solution with $\hbar u_+'/u_+ = -\sqrt{\mathfrak{q}} + z$ and $u_+(\tau_1) = 1$. Since $-2\sqrt{\mathfrak{q}} \le \hbar u_+'/u_+ \le -\frac{1}{2}\sqrt{\mathfrak{q}}$ and $\hbar C_T \le \sqrt{\mathfrak{q}}$,

$$\bigl(\sqrt{\mathfrak{q}}\,u_+^2\bigr)' \le -\frac{\mathfrak{q}\,u_+^2}{2\hbar},$$

so $\int_{\tau_1}^\infty \mathfrak{q}u_+^2\,d\tau \le 2\hbar\sqrt{\mathfrak{q}(\tau_1)}$, and u_+ is square-integrable on the tail since $\mathfrak{q} \ge \mathfrak{q}_*/2$. Every solution u square-integrable on the tail is a multiple of u_+. On the tail $(u^2)'' = 2u'^2 + 2\mathfrak{q}u^2/\hbar^2 \ge 0$, so the square of every solution is convex there, and a convex function integrable on the tail is nonincreasing. The combination of u and u_+ that vanishes at a point of the tail is a solution square-integrable on the tail, so its square vanishes on the rest of the tail, and the combination is zero by the uniqueness of the solution with given value and derivative at one point. The same holds on $\tau \le \tau_{-1}$ for a solution u_-, with τ replaced by $-\tau$.

(b) *The perturbed turning point, (M1).* Let $\hbar_0$ be small enough that $\hbar \sup|\mathfrak{q}_1| < \min(\mu/2, \frac{3}{4}c_*\delta_{\rm tp})$ and $\hbar \sup|\mathfrak{q}_1'| \le \frac{1}{4}c_*$, the suprema taken over the compact set on which the functions of (U1) are bounded. Then $\mathfrak{f}' \ge \frac{1}{2}c_*$ within $2\delta_{\rm tp}$ of τ_R; $\mathfrak{f} < 0$ at $\tau_R - \delta_{\rm tp}$ and $\mathfrak{f} > 0$ at $\tau_R + \delta_{\rm tp}$; and at the points of I_+ farther than $\delta_{\rm tp}$ from τ_R, which are at least $2\delta_{\rm tp}$ from τ_L because τ_0 is, $|\mathfrak{f}| \ge \mu/2$ with the sign of $\mathfrak{q}_0$. So on I_+, $\mathfrak{f}$ has exactly one zero $\tau_{R,\hbar}$, which is simple, with $|\tau_{R,\hbar} - \tau_R| \le 2\hbar \sup|\mathfrak{q}_1|/c_*$, and $\tau_{R,\hbar}$ is continuous in $(E, \hbar)$ on $J \times [0, \hbar_0]$. The lower bound $\mathfrak{f}' \ge \frac{1}{2}c_*$, uniform in E and $\hbar$, keeps the turning point simple. In particular $\mathfrak{f}(\tau_0) \le -\mu/2$, so $a_{\rm m}$ is defined.

(c) *Olver's hypotheses, (M2).* Proposition 1 is applied on (τ_0, τ_1) with $\nu_{\rm O} = 1/\hbar$, $f_{\rm O} = \mathfrak{f}$, $g_{\rm O} = \mathfrak{q}_2$ and $\tau_* = \tau_{R,\hbar}$. Since $\mathfrak{f}$ and $\mathfrak{q}_2$ are continuous on the closed interval $[\tau_0, \tau_1]$ and the variations are finite, w_j, ϵ_j and their derivatives extend continuously to its ends, and the bounds hold there by taking limits. With $\varsigma \equiv \tau - \tau_{R,\hbar}$,

$$\mathfrak{d}_\varsigma \equiv \frac{\mathfrak{f}}{\varsigma},$$

$$\mathfrak{d}_\varsigma = \int_0^1 \mathfrak{f}'(\tau_{R,\hbar} + t\varsigma)\,dt,$$

so $\mathfrak{d}_\varsigma$ and its derivatives $d\mathfrak{d}_\varsigma/d\varsigma = \int_0^1 t\,\mathfrak{f}''\,dt$ and $d^2\mathfrak{d}_\varsigma/d\varsigma^2 = \int_0^1 t^2\mathfrak{f}'''\,dt$ are continuous in $(\varsigma, E, \hbar)$ and bounded, by (U1). For τ within $\delta_{\rm tp}$ of τ_R the segment from $\tau_{R,\hbar}$ to τ lies within $2\delta_{\rm tp}$ of τ_R, so $\mathfrak{d}_\varsigma \ge \frac{1}{2}c_*$; at the other points of I_+, $\mathfrak{d}_\varsigma \ge \mu/(2L_+)$, where L_+ is the largest length of I_+ on J. So the hypotheses hold. With ζ of the proposition for $f_{\rm O} = \mathfrak{f}$, $\hat{\mathfrak{f}} \equiv \mathfrak{f}/\zeta$, so that $d\zeta/d\tau = \hat{\mathfrak{f}}^{1/2}$, and $X_{\rm O} \equiv \hbar^{-2/3}\zeta$, the proposition gives two solutions $w_1 = \hat{\mathfrak{f}}^{-1/4}(\mathrm{Bi}(X_{\rm O}) + \epsilon_1)$ and $w_2 = \hat{\mathfrak{f}}^{-1/4}(\mathrm{Ai}(X_{\rm O}) + \epsilon_2)$, whose errors satisfy, on I_+,

$$\max\Bigl(\frac{|\epsilon_1|}{M_{\rm O}(X_{\rm O})},\, \frac{|d\epsilon_1/d\tau|}{\hbar^{-2/3}\hat{\mathfrak{f}}^{1/2}N_{\rm O}(X_{\rm O})}\Bigr) \le \frac{E_{\rm O}(X_{\rm O})}{\lambda_{\rm O}}\Bigl(e^{\lambda_{\rm O}\hbar\mathcal{V}_{\rm O}(\tau_0,\tau)} - 1\Bigr),$$

$$\max\Big(\frac{|\epsilon_2|}{M_{\mathrm{O}}(X_{\mathrm{O}})},\,\frac{|d\epsilon_2/d\tau|}{\hbar^{-2/3}\hat{\mathfrak{f}}^{1/2}N_{\mathrm{O}}(X_{\mathrm{O}})}\Big)\le\frac{1}{\lambda_{\mathrm{O}}\,E_{\mathrm{O}}(X_{\mathrm{O}})}\Big(e^{\lambda_{\mathrm{O}}\hbar\mathcal{V}_{\mathrm{O}}(\tau,\tau_1)}-1\Big),$$

with $\psi_{\mathrm{O}}=\mathfrak{q}_2/\hat{\mathfrak{f}}-\hat{\mathfrak{f}}^{-3/4}\,d^2\hat{\mathfrak{f}}^{-1/4}/d\tau^2$ in the error-control function. So the constants of the bounds depend on E and $\hbar$ only through $\mathcal{V}_{\mathrm{O}}(\tau_0,\tau_1)$.

The variation is bounded. As in the proof of Olver's Lemma 3.1, $\zeta/\varsigma=\big(\frac{3}{2}\int_0^1 t^{1/2}\mathfrak{d}_\varsigma(t\varsigma)^{1/2}dt\big)^{2/3}$, where $\mathfrak{d}_\varsigma(t\varsigma)$ is the value of $\mathfrak{d}_\varsigma$ at $\tau_{R,\hbar}+t\varsigma$; so ζ/ς and its first two derivatives in ς are continuous in $(\varsigma,E,\hbar)$, and $\zeta/\varsigma\ge(\min\mathfrak{d}_\varsigma)^{1/3}$. Then $\hat{\mathfrak{f}}=\mathfrak{d}_\varsigma\,\varsigma/\zeta$ is positive, bounded below, and twice continuously differentiable, and ψ_{O} is a continuous function of $\mathfrak{d}_\varsigma$, ζ/ς, their first two derivatives and $\mathfrak{q}_2$. Each of these is continuous on the compact set of $(\varsigma,E,\hbar)$ with $E\in J$, $\hbar\in[0,\hbar_0]$ and $\tau_{R,\hbar}+\varsigma\in I_+$, whose ends move continuously. So on that set $|\psi_{\mathrm{O}}|\le\Psi_{\max}$ and $|\zeta|$ is bounded, both with bounds independent of E and $\hbar$. Hence

$$\mathcal{V}_{\mathrm{O}}(\tau_0,\tau_1)\le\Psi_{\max}\int_{\zeta(\tau_0)}^{\zeta(\tau_1)}|\zeta|^{-1/2}d\zeta,$$

$$\Psi_{\max}\int_{\zeta(\tau_0)}^{\zeta(\tau_1)}|\zeta|^{-1/2}d\zeta=2\Psi_{\max}\big(|\zeta(\tau_0)|^{1/2}+|\zeta(\tau_1)|^{1/2}\big),$$

a bound uniform on $J\times[0,\hbar_0]$, although the turning point $\tau_{R,\hbar}$ moves with E and $\hbar$. The table lists each hypothesis of Proposition 1, its form in this problem, and the step that verifies it.

Hypothesis of Olver	**Here**	**Established in**
The equation is $d^2w/d\tau^2=(\nu_{\mathrm{O}}^2f_{\mathrm{O}}+g_{\mathrm{O}})\,w$, with f_{O} and g_{O} allowed to depend on ν_{O}	$\nu_{\mathrm{O}}=1/\hbar$, $f_{\mathrm{O}}=\mathfrak{f}=\mathfrak{q}_0+\hbar\mathfrak{q}_1$, $g_{\mathrm{O}}=\mathfrak{q}_2$	the equation $\hbar^2u''=\mathfrak{q}u$
f_{O} is real with one zero τ_* in the interval	$\mathfrak{f}$ has the one zero $\tau_{R,\hbar}$ on I_+	(b)
$f_{\mathrm{O}}/(\tau-\tau_*)$ is positive	$\mathfrak{d}_\varsigma\ge\min(\frac{1}{2}c_*,\mu/(2L_+))$	(c), from c_*, δ_{tp} and μ
$f_{\mathrm{O}}/(\tau-\tau_*)$ is twice continuously differentiable	$\mathfrak{d}_\varsigma$, $d\mathfrak{d}_\varsigma/d\varsigma$, $d^2\mathfrak{d}_\varsigma/d\varsigma^2$ continuous and bounded	(c), from (U1)
g_{O} is continuous	$\mathfrak{q}_2$ continuous and bounded	(U1)
The variation of the error-control function $\mathcal{H}_{\mathrm{O}}$ is finite	$\mathcal{V}_{\mathrm{O}}(\tau_0,\tau_1)\le2\Psi_{\max}\big(\lvert\zeta(\tau_0)\rvert^{1/2}+\lvert\zeta(\tau_1)\rvert^{1/2}\big)$	(c), the variation is bounded

The equation enters Olver's bounds only through $\hbar\mathcal{V}_{\mathrm{O}}(\tau_0,\tau_1)$, which is at most a constant times $\hbar$ on $J\times[0,\hbar_0]$.

(d) *Matching to the decaying tail, (M3).* At τ_1 the variation $\mathcal{V}_{\mathrm{O}}(\tau_1,\tau_1)$ is 0, so ϵ_2 and $d\epsilon_2/d\tau$ vanish and $w_2=\hat{\mathfrak{f}}^{-1/4}\mathrm{Ai}(X_{\mathrm{O}})$ there. Also $\zeta(\tau_1)$ is bounded below by a positive constant, so X_{O} is at least a positive constant times $\hbar^{-2/3}$ at τ_1, where $\mathrm{Ai}'/\mathrm{Ai}=-X_{\mathrm{O}}^{1/2}(1+O(X_{\mathrm{O}}^{-3/2}))$, $M_{\mathrm{O}}/E_{\mathrm{O}}=\sqrt{2}\,\mathrm{Ai}$, $M_{\mathrm{O}}E_{\mathrm{O}}=\sqrt{2}\,\mathrm{Bi}$, and $N_{\mathrm{O}}E_{\mathrm{O}}$ is of the order of Bi'. Since $\hbar^{1/3}\hat{\mathfrak{f}}^{1/2}X_{\mathrm{O}}^{1/2}=\sqrt{\mathfrak{f}}$, $X_{\mathrm{O}}^{-3/2}=O(\hbar)$ and $\hbar\,d\hat{\mathfrak{f}}/d\tau=O(\hbar)$, the logarithmic derivatives $\mathfrak{y}_j\equiv\hbar w_j'/w_j$ at τ_1 are

$$\mathfrak{y}_2=-\sqrt{\mathfrak{f}}\,\big(1+O(\hbar)\big),$$

$$\mathfrak{y}_1=\sqrt{\mathfrak{f}}\,\big(1+O(\hbar)\big),$$

while by (a), $\mathfrak{y}_+\equiv\hbar u_+'/u_+=-\sqrt{\mathfrak{q}}+z$, with $|z|\le\hbar C_T$ and $\sqrt{\mathfrak{q}}=\sqrt{\mathfrak{f}}\,(1+O(\hbar^2))$. Write $u_+=\alpha_+w_2+\beta_+w_1$. At τ_1, $\alpha_+w_2(\mathfrak{y}_+-\mathfrak{y}_2)=\beta_+w_1(\mathfrak{y}_1-\mathfrak{y}_+)$, and $\mathfrak{y}_1-\mathfrak{y}_+\ge\sqrt{\mathfrak{f}}\ge\sqrt{\mathfrak{q}_*/2}$; so $\alpha_+\ne0$, and

$$\Big|\frac{\beta_+\,w_1(\tau_1)}{\alpha_+\,w_2(\tau_1)}\Big|\le C\hbar.$$

Since $w_1/w_2=(\mathrm{Bi}/\mathrm{Ai})(1+O(\hbar))$ at τ_1 and $\mathrm{Ai}/\mathrm{Bi}\le e^{-4X_{\mathrm{O}}^{3/2}/3}$ for every $X_{\mathrm{O}}\ge0$, $|\beta_+/\alpha_+|\le C\hbar\,e^{-c_{\mathrm{dec}}/\hbar}$. Here $c_{\mathrm{dec}}=\frac{4}{3}\min\zeta(\tau_1)^{3/2}$, with the minimum taken over $J\times[0,\hbar_0]$. It is positive because $\zeta(\tau_1)$ is positive and continuous in $(E,\hbar)$ on that compact set, although τ_R, $\tau_{R,\hbar}$ and τ_1 move with E. Every constant here is uniform on J because $\mathcal{V}_{\mathrm{O}}(\tau_0,\tau_1)$, c_*, μ, $\mathfrak{q}_*$ and C_T are.

(e) *Phases at the center, (M4).* At τ_0, $X_{\rm A} \equiv -\hbar^{-2/3}\zeta = -X_{\rm O}$ is at least a constant times $\hbar^{-2/3}$, and Ai and Ai$'$ at $X_{\rm O} = -X_{\rm A}$ are $\pi^{-1/2}X_{\rm A}^{-1/4}\big(\sin\psi_{\rm A} + O(X_{\rm A}^{-3/2})\big)$ and $-\pi^{-1/2}X_{\rm A}^{1/4}\big(\cos\psi_{\rm A} + O(X_{\rm A}^{-3/2})\big)$, with $\psi_{\rm A} = \frac{2}{3}X_{\rm A}^{3/2} + \frac{\pi}{4}$. Here $\frac{2}{3}X_{\rm A}^{3/2} = \Phi_+/\hbar$ and $\hbar\cdot\hbar^{-2/3}\hat{\mathfrak{f}}^{1/2}X_{\rm A}^{1/2} = a_{\rm m}$. At $X_{\rm O} = -X_{\rm A}$ the bounds of (c) with $E_{\rm O} = 1$, $M_{\rm O} \le CX_{\rm A}^{-1/4}$ and $N_{\rm O} \le CX_{\rm A}^{1/4}$ make ϵ_2 and $\hbar\,(d\epsilon_2/d\tau)/a_{\rm m}$ at most $C\hbar X_{\rm A}^{-1/4}$, so, with these two relations, the vector $(w_2, \hbar w_2'/a_{\rm m})$ is $\pi^{-1/2}X_{\rm A}^{-1/4}\hat{\mathfrak{f}}^{-1/4}$ times $(\sin\psi_{\rm A}, -\cos\psi_{\rm A}) + O(\hbar)$. The same bounds make the norm of $(w_1, \hbar w_1'/a_{\rm m})$ at most $CX_{\rm A}^{-1/4}\hat{\mathfrak{f}}^{-1/4}$, so by (d) the part $\beta_+ w_1$ of u_+ changes this vector by a relative amount at most $C\hbar\, e^{-c_{\rm dec}/\hbar}$. So $\vartheta_+ \equiv -\Phi_+/\hbar - \pi/4 + O(\hbar)$ modulo π. The same argument on I_-, with τ replaced by $-\tau$, gives $\vartheta_- \equiv \Phi_-/\hbar + \pi/4 + O(\hbar)$ modulo π. □

Lemma 4 (From the phase error to the action)**:** *Let J be a compact interval, and for $E \in J$ and $0 < \hbar \le \hbar_0$ let Φ_+ and Φ_- be real and continuous in E, and let ϑ_+ and ϑ_- be angles defined modulo π, continuous in E, with $|\vartheta_+ + \Phi_+/\hbar + \pi/4| \le C\hbar$ and $|\vartheta_- - \Phi_-/\hbar - \pi/4| \le C\hbar$ modulo π, where $4C\hbar_0 < \pi$. Let $\Phi_\hbar \equiv \Phi_- + \Phi_+$. Then there is a function $\mathcal{I}_\hbar$, continuous on J, such that $\vartheta_- - \vartheta_+ \in \pi\mathbb{Z}$ if and only if $\mathcal{I}_\hbar \in 2\pi\hbar\mathbb{Z}$, and*

$$|\mathcal{I}_\hbar - 2\Phi_\hbar - \pi\hbar| \le 4C\hbar^2.$$

Proof. Modulo π, $\vartheta_- - \vartheta_+ - \Phi_\hbar/\hbar - \pi/2$ is the sum of the two quantities bounded in the hypothesis, so it has a representative $\delta_\vartheta(E)$ with $|\delta_\vartheta| \le 2C\hbar < \pi/2$. It is the only representative in $(-\pi/2, \pi/2)$, so it is continuous in E. Let $\mathcal{I}_\hbar \equiv 2\Phi_\hbar + \pi\hbar + 2\hbar\delta_\vartheta$. Then $\vartheta_- - \vartheta_+$ is congruent to $\mathcal{I}_\hbar/(2\hbar)$ modulo π, which gives the first statement, and $|\mathcal{I}_\hbar - 2\Phi_\hbar - \pi\hbar| = 2\hbar|\delta_\vartheta| \le 4C\hbar^2$. □

The action is $2\hbar$ times the phase, so a phase error of order $\hbar$ is an action error of order $\hbar^2$. From the action error, Lemma 1 bounds the energy error, using only that $\mathcal{I}_\hbar$ is continuous and that the period of the orbit is bounded below. Neither step uses a derivative in E of the error $\mathcal{I}_\hbar - 2\Phi_\hbar - \pi\hbar$.

Lemma 5 (Sturm comparison)**:** *Let J be a compact interval and $\hbar > 0$. For $E \in J$ let $g(x,E) > 0$ and $k_S(x,E)$ be real and continuous in (x,E), and consider for real F, with F and gF' continuously differentiable and a prime denoting d/dx, the equation*

$$(\hbar g F')' = \frac{k_S}{\hbar}\, F.$$

Assume:

(S1) for $E < E'$ in J and every x, $g(x,E') \le g(x,E)$ and $k_S(x,E') < k_S(x,E)$;

(S2) there is $x_* > 0$, independent of E, and for each $E \in J$ there are solutions F_+, positive for $x \ge x_*$, and F_-, positive for $x \le -x_*$, with $\int_{x_*}^{\infty} gF_+^2\,dx$ and $\int_{-\infty}^{-x_*} gF_-^2\,dx$ finite;

(S3) gF_+'/F_+ at x_* and gF_-'/F_- at $-x_*$ are continuous in E.

For a real solution $F \ne 0$ write $F = r_P \sin\theta$ and $\hbar g F' = r_P\cos\theta$, with $r_P > 0$ and θ continuous in x, and let $\theta_\pm$ be the angle of $F_\pm$ with $\theta_\pm(\pm x_*) \in (0,\pi)$. Then $\Delta_\theta(E) \equiv \theta_-(0,E) - \theta_+(0,E)$ is continuous and strictly increasing on J.

Proof. The angle of a solution satisfies

$$\theta' = \frac{\cos^2\theta}{\hbar g} - \frac{k_{\rm S}}{\hbar}\sin^2\theta.$$

Comparison. Let $E < E'$, let θ and $\tilde\theta$ solve this equation at E and E' on an interval, and let $\Delta\theta \equiv \tilde\theta - \theta$. Then $\Delta\theta' = a_{\rm S}\,\Delta\theta + b_{\rm S}$, where $a_{\rm S}$ is the mean over the segment from θ to $\tilde\theta$ of the θ-derivative of the right side at E', bounded on bounded intervals, and $b_{\rm S}$ is the right side at (θ, E') minus that at (θ, E), so by (S1)

$$b_{\rm S} \ge \frac{k_{\rm S}(x,E) - k_{\rm S}(x,E')}{\hbar}\,\sin^2\theta,$$

$$\frac{k_{\rm S}(x,E) - k_{\rm S}(x,E')}{\hbar}\,\sin^2\theta \ge 0.$$

So $\Delta\theta\, e^{-\int a_{\rm S}}$ is nondecreasing. It increases strictly on every interval, since $b_{\rm S} > 0$ wherever $\sin\theta \ne 0$, and the zeros of $\sin\theta$ are isolated because $\theta' = 1/(\hbar g) > 0$ there. *The left tail.* For $\mathcal{C} > x_*$ the solution $F_{\mathcal{C}} \equiv F_-(x)\int_{-\mathcal{C}}^{x} dx'/(gF_-^2)$

vanishes at $-\mathcal{C}$, has $\hbar g F_{\mathcal{C}}' = \hbar/F_-(-\mathcal{C}) > 0$ there, and is positive on $(-\mathcal{C}, -x_*]$; so its angle $\theta_{\mathcal{C}}$ with $\theta_{\mathcal{C}}(-\mathcal{C}) = 0$ has $\theta_{\mathcal{C}}(-x_*) \in (0,\pi)$, and

$$\frac{\hbar g F_{\mathcal{C}}'}{F_{\mathcal{C}}} = \frac{\hbar g F_-'}{F_-} + \frac{\hbar}{F_-^2 \int_{-\mathcal{C}}^x dx'/(gF_-^2)}.$$

By the Cauchy–Schwarz inequality, $(x+\mathcal{C})^2 \le \int_{-\mathcal{C}}^x g F_-^2\, dx' \int_{-\mathcal{C}}^x dx'/(gF_-^2)$, and $\int g F_-^2$ is finite by (S2), so the last term at $x = -x_*$ tends to 0 as $\mathcal{C} \to \infty$. Since cot maps $(0,\pi)$ onto $\mathbb{R}$ one to one, $\theta_{\mathcal{C}}(-x_*) \to \theta_-(-x_*)$. At $E < E'$ the comparison on $[-\mathcal{C}, -x_*]$, from $\Delta\theta(-\mathcal{C}) = 0$, gives $\theta_{\mathcal{C}}(-x_*, E') > \theta_{\mathcal{C}}(-x_*, E)$; in the limit $\theta_-(-x_*, E') \ge \theta_-(-x_*, E)$, and the comparison on $[-x_*, 0]$ gives

$$\theta_-(0, E') > \theta_-(0, E).$$

The right tail. In the same way the solution $F_+(x) \int_x^{\mathcal{C}} dx'/(gF_+^2)$ vanishes at $\mathcal{C}$ with $\hbar g F' < 0$ there, so its angle with the value π at $\mathcal{C}$ lies in $(0,\pi)$ at x_* and tends to $\theta_+(x_*)$ as $\mathcal{C} \to \infty$. The comparison on $[x_*, \mathcal{C}]$ from $\Delta\theta(\mathcal{C}) = 0$, with $\Delta\theta\, e^{-\int a_{\mathrm{S}}}$ nondecreasing, gives $\Delta\theta(x_*) < 0$; in the limit $\theta_+(x_*, E') \le \theta_+(x_*, E)$, and the comparison on $[0, x_*]$ gives $\theta_+(0, E') < \theta_+(0, E)$. *Continuity.* $\theta_\pm(\pm x_*)$ is the value in $(0,\pi)$ of the inverse cotangent of $\hbar g F_\pm'/F_\pm$ at $\pm x_*$, continuous in E by (S3), and the solution of the angle equation on $[-x_*, x_*]$ depends continuously on its initial value and on E. So Δ_θ is continuous, and by the two tails it is strictly increasing. □

Hypothesis (S1) says that the coefficient k_{S}, the potential minus the energy, decreases at every x as E increases. The proof shows that $\theta_-(0, E)$ then increases with E and $\theta_+(0, E)$ decreases, so their difference passes each multiple of π at most once on J.

The angles of the Sturm comparison are measured at $x = 0$. The angles $\vartheta_\pm$ of the matching are measured at τ_0, in other coordinates, and a linear map that depends on E relates the two. An angle of a nonzero vector $\mathfrak{u} \in \mathbb{R}^2$ is a θ with $\mathfrak{u} = |\mathfrak{u}|(\sin\theta, \cos\theta)$, as in Lemma 5.

Lemma 6 (Zeros, Sturm angles and integer labels)**:** *Let J be a compact interval and $\hbar > 0$. For $E \in J$ let $\mathcal{G}(E)$ be a real 2×2 matrix with $\det \mathcal{G}(E) > 0$, continuous in E, and let $\theta_+(E)$ and $\theta_-(E)$ be real and continuous on J, with $\Delta_\theta \equiv \theta_- - \theta_+$ strictly increasing. Let $\vartheta_\pm(E)$ be an angle, defined modulo π, of $\mathcal{G}(E)(\sin\theta_\pm, \cos\theta_\pm)$, and let $\mathcal{I}_\hbar$ be continuous on J with $\mathcal{I}_\hbar/(2\hbar)$ congruent to $\vartheta_- - \vartheta_+$ modulo π. Then:*

(a) there is a function $\theta_{\mathcal{G}}(\theta, E)$, continuous on $\mathbb{R}\times J$, such that $\theta_{\mathcal{G}}(\theta, E)$ is an angle of $\mathcal{G}(E)(\sin\theta, \cos\theta)$, $\theta_{\mathcal{G}}$ is strictly increasing in θ, and $\theta_{\mathcal{G}}(\theta+\pi, E) = \theta_{\mathcal{G}}(\theta, E) + \pi$;

(b) there is an integer m_L, the same for every $E \in J$, such that for every integer k, $\mathcal{I}_\hbar(E) = 2\pi\hbar k$ if and only if $\Delta_\theta(E) = (k - m_L)\pi$;

(c) $\vartheta_- - \vartheta_+ \in \pi\mathbb{Z}$ exactly when $\Delta_\theta \in \pi\mathbb{Z}$; for each integer k there is at most one $E \in J$ with $\mathcal{I}_\hbar(E) = 2\pi\hbar k$; and two such energies with no other between them have consecutive integers k.

Proof. (a) The vector $\mathfrak{u}(\theta, E) \equiv \mathcal{G}(E)(\sin\theta, \cos\theta)$ is nonzero, since $\det \mathcal{G}(E) \ne 0$, and continuous on $\mathbb{R}\times J$. This set is convex, hence simply connected, so by the lifting property of the covering of the unit circle by the line, $\mathfrak{u}/|\mathfrak{u}|$ has a continuous angle $\theta_{\mathcal{G}}$ on $\mathbb{R}\times J$. The derivative of $\theta_{\mathcal{G}}$ in θ is $\det \mathcal{G}(E)$ divided by $|\mathfrak{u}|^2$, so $\theta_{\mathcal{G}}$ increases strictly in θ. Since $\mathfrak{u}(\theta+\pi, E) = -\mathfrak{u}(\theta, E)$, the difference $\theta_{\mathcal{G}}(\theta+\pi, E) - \theta_{\mathcal{G}}(\theta, E)$ is an odd multiple of π; it is continuous on the connected set $\mathbb{R}\times J$, so it is a constant $j\pi$, and $j \ge 1$ since $\theta_{\mathcal{G}}$ increases. For $\theta < \theta_1 < \theta + \pi$ the vectors $\mathfrak{u}(\theta_1, E)$ and $\mathfrak{u}(\theta, E)$ are not parallel, since $\mathcal{G}(E)$ is invertible and $(\sin\theta_1, \cos\theta_1)$ is not parallel to $(\sin\theta, \cos\theta)$; so $\theta_{\mathcal{G}}(\theta_1, E) - \theta_{\mathcal{G}}(\theta, E)$ is not a multiple of π there, and by continuity it stays in $(0,\pi)$. Hence $j = 1$.

(b) Since $\vartheta_\pm$ is congruent to $\theta_{\mathcal{G}}(\theta_\pm(E), E)$ modulo π, the function

$$h_{\mathcal{G}}(E) \equiv \frac{\mathcal{I}_\hbar(E)}{2\hbar} - \theta_{\mathcal{G}}(\theta_-(E), E) + \theta_{\mathcal{G}}(\theta_+(E), E)$$

takes values in $\pi\mathbb{Z}$. It is continuous on J, since $\mathcal{I}_\hbar$, $\theta_\pm$ and $\theta_{\mathcal{G}}$ are, and J is connected, so $h_{\mathcal{G}}$ is a constant $m_{\mathrm{L}}\pi$. So $\mathcal{I}_\hbar(E) = 2\pi\hbar k$ exactly when $\theta_{\mathcal{G}}(\theta_-, E) - \theta_{\mathcal{G}}(\theta_+, E) = (k - m_{\mathrm{L}})\pi$. By (a), for each integer j, $\theta_{\mathcal{G}}(\theta_+ + j\pi, E) = \theta_{\mathcal{G}}(\theta_+, E) + j\pi$, and $\theta_{\mathcal{G}}$ is strictly increasing in θ; so $\theta_{\mathcal{G}}(\theta_-, E) = \theta_{\mathcal{G}}(\theta_+, E) + j\pi$ exactly when $\theta_- = \theta_+ + j\pi$. With $j = k - m_{\mathrm{L}}$ this is (b).

(c) The first statement is (b) for all k together. Since Δ_θ is continuous and strictly increasing, it takes the value $(k - m_{\mathrm{L}})\pi$ at most once, and between the energies at which it equals $j\pi$ and $(j+1)\pi$ it takes no other multiple of π, which with (b) gives the other two statements. □

Proposition 2 and Lemmas 4, 5 and 6 combine to give the rule for the energy-dependent operator.

Lemma 7 (The rule for the energy-dependent operator)**:** *Let J be a compact interval and $s=\pm1$. For $E\in J$ let $\varrho(x,E)$, $g(x,E)$ and $\mathcal{U}(x,E)$ be real, continuous in (x,E) and smooth in x, with g, $\mathcal{U}$ and each of their x-derivatives, and each x-derivative of ϱ, continuous in (x,E) and bounded on $\mathbb{R}\times J$, and let g lie between two positive constants. Let $Q_s=\hat p-is\varrho$ and $\mathsf{K}(E)\equiv c^2\,Q_s^\dagger\,g\,Q_s-\mathcal{U}$ on the form domain $\operatorname{dom}Q_s$, $\mathsf{K}_0\equiv c^2g\,(\xi^2+\varrho^2)-\mathcal{U}$ and $\mathsf{K}_1\equiv -s\,c^2\,(g\varrho)'$, and let t_0 be the time of the flow of K_0. Let (i) for each $E\in J$ the level set $\mathsf{K}_0=0$ be a closed curve with no critical point of K_0, where $\mathcal{A}(E)$ denotes the area of the set where $\mathsf{K}_0<0$, and (ii) let there be x_∞ and $\varpi_\infty>0$ with $c^2g\varrho^2-\mathcal{U}\ge\varpi_\infty$ for $|x|\ge x_\infty$ and $E\in J$. Then there are C and $\hbar_0>0$ and, for $\hbar<\hbar_0$, a function $\mathcal{I}_\hbar$ continuous on J, such that 0 is an eigenvalue of $\mathsf{K}(E)$ if and only if $\mathcal{I}_\hbar(E)\in2\pi\hbar\mathbb{Z}$, and on J*

$$\left|\mathcal{I}_\hbar-\mathcal{A}+\hbar\oint\mathsf{K}_1\,dt_0-\pi\hbar\right|\le C\hbar^2.$$

Let in addition (iii) for $\hbar<\hbar_0$, $E<E'$ in J and every x, $g(x,E')\le g(x,E)$ and $k_{S,\hbar}(x,E')<k_{S,\hbar}(x,E)$, where $k_{S,\hbar}\equiv g\varrho^2-\mathcal{U}/c^2-s\hbar\,(g\varrho)'$. Then $\mathcal{I}_\hbar$ takes each value in $2\pi\hbar\mathbb{Z}$ at most once on J.

Proof. Write $\mathfrak{q}_0\equiv g\varrho^2-\mathcal{U}/c^2$ and $\mathfrak{q}_1\equiv\mathsf{K}_1/c^2$, so that $\mathsf{K}_0=c^2(g\xi^2+\mathfrak{q}_0)$. Since $Q_s^\dagger gQ_s=\hat p\,g\,\hat p+g\varrho^2-s\hbar\,(g\varrho)'$, the equation $\mathsf{K}F=0$ is

$$-\hbar^2(gF')'+(\mathfrak{q}_0+\hbar\mathfrak{q}_1)\,F=0,$$

and its solutions have F and gF' continuously differentiable, as in Lemma 2(iii). The coefficients are real, so the real and imaginary parts of a solution are solutions, and it suffices to treat real solutions.

(1) *Liouville form.* Let $\tau\equiv\int_0^x g^{-1/2}\,dx$, $w_{\mathrm L}\equiv g^{1/2}$ and $F=w_{\mathrm L}^{-1/2}u$, and write a prime for $d/d\tau$ from here on. With $\mathfrak{q}_2\equiv\frac12 w_{\mathrm L}^{-1/2}\big(w_{\mathrm L}^{-1/2}w_{\mathrm L}'\big)'$ and $\mathfrak{q}\equiv\mathfrak{q}_0+\hbar\mathfrak{q}_1+\hbar^2\mathfrak{q}_2$, the equation becomes

$$\hbar^2u''=\mathfrak{q}\,u,$$

and $\int|F|^2dx=\int|u|^2d\tau$. The map $(x,\xi)\mapsto(\tau,p_\tau)$ with $p_\tau=g^{1/2}\xi$ preserves area, and takes K_0/c^2 to $p_\tau^2+\mathfrak{q}_0$. Since τ is continuous in (x,E) and increasing in x, x is continuous in (τ,E), and so is every function in this proof that is continuous in (x,E).

(2) *The hypotheses of Proposition 2.* (U2): the projection of the curve $\mathsf{K}_0=0$ on the τ axis is the set where $\mathfrak{q}_0\le0$; it is connected, so it is an interval $[\tau_L,\tau_R]$. At a zero τ_* of $\mathfrak{q}_0$ the point $(\tau_*,0)$ lies on the curve, so $\mathfrak{q}_0'(\tau_*)\ne0$ by (i); a zero inside the interval would be a maximum of $\mathfrak{q}_0\le0$, where $\mathfrak{q}_0'=0$. So $\mathfrak{q}_0<0$ inside, and the zeros at the ends are simple. (U3): by (ii) and the compactness of J there are $\delta,\mathfrak{q}_*>0$ with $\mathfrak{q}_0\ge\mathfrak{q}_*$ for $\tau\ge\tau_R+\delta$ and for $\tau\le\tau_L-\delta$. (U1) and (U4): each of $\mathfrak{q}_0$, $\mathfrak{q}_1$ and $\mathfrak{q}_2$, and each of their τ-derivatives, with $d/d\tau=g^{1/2}d/dx$, is a sum of terms ϱ^2, ϱ or 1 times products of x-derivatives of ϱ, $g^{\pm1/2}$ and $\mathcal{U}$, which are bounded and continuous in (x,E), so these functions are continuous in (τ,E). The term ϱ^2 occurs only in $\mathfrak{q}_0$ and its derivatives, and ϱ does not occur in $\mathfrak{q}_2$. Where $\mathfrak{q}_0\ge\mathfrak{q}_*$, $\varrho^2\le(\mathfrak{q}_0+\sup|\mathcal{U}|/c^2)/\inf g\le C_{\mathfrak q}\mathfrak{q}_0$ for a constant $C_{\mathfrak q}$, and $1\le\mathfrak{q}_0/\mathfrak{q}_*$, so $|\mathfrak{q}_1|$, $|\mathfrak{q}_2|$, $|\mathfrak{q}_0'|$, $|\mathfrak{q}_1'|$ and $|\mathfrak{q}_2'|$ are at most $C_0\mathfrak{q}_0$ there. Proposition 2 therefore gives (M0)–(M4), with constants independent of $E\in J$.

(3) *The decaying solutions lie in the form domain.* By (M0), $\int\mathfrak{q}u_+^2\,d\tau$ is finite on $\tau\ge\tau_R+\delta$; with $|u_+'|\le2\sqrt{\mathfrak q}\,|u_+|/\hbar$ and ϱ^2 at most a constant times $\mathfrak q$ there, F and Q_sF are square-integrable on the tail for $F=w_{\mathrm L}^{-1/2}u_+$, and in the same way for u_- on the other tail.

(4) *The criterion.* Let $\mathcal{W}\equiv u_-u_+'-u_+u_-'$, which is constant. If $\mathsf{K}(E)F=0$ with $0\ne F\in\operatorname{dom}Q_s$, $u=w_{\mathrm L}^{1/2}F$ is square-integrable on both tails, so it is a multiple of u_+ and of u_-, and $\mathcal{W}=0$. If $\mathcal{W}=0$, u_+ is a multiple of u_-, so $F=w_{\mathrm L}^{-1/2}u_+$ lies in $\operatorname{dom}Q_s$. For $\tilde F\in\operatorname{dom}Q_s$, integration by parts on $[-X,X]$ writes $c^2\langle Q_s\tilde F,gQ_sF\rangle-\langle\tilde F,\mathcal{U}F\rangle$ as the limit of the boundary terms $i\hbar c^2\,\bar{\tilde F}\,g\,Q_sF$ at $\pm X$. These boundary terms are products of square-integrable functions, so they tend to 0 along a sequence $X_k\to\infty$. The form of $\mathsf{K}(E)$ therefore vanishes at $(\tilde F,F)$ for every $\tilde F$, so $\mathsf{K}(E)F=0$. So 0 is an eigenvalue if and only if $\mathcal{W}=0$.

(5) *The phase condition.* At τ_0, $\mathcal{W}=(a_{\mathrm m}/\hbar)\,r_-r_+\sin(\vartheta_--\vartheta_+)$, so $\mathcal{W}=0$ if and only if $\vartheta_--\vartheta_+\in\pi\mathbb{Z}$. The angles $\vartheta_\pm$ are continuous in E, since $u_\pm$ and $u_\pm'$ are continuous in E on the tails by (M0) and the solutions depend continuously on E, and so are $\Phi_\pm$, by (M1). By (M4) and Lemma 4 there is a continuous $\mathcal{I}_\hbar$ with $\mathcal{W}=0$ if and only if $\mathcal{I}_\hbar\in2\pi\hbar\mathbb{Z}$, and $|\mathcal{I}_\hbar-2\Phi_\hbar-\pi\hbar|\le C\hbar^2$.

(6) *The area.* The number $2\Phi_\hbar=\int_{\tau_{L,\hbar}}^{\tau_{R,\hbar}}2\sqrt{-\mathfrak f}\,d\tau$ is the area of the set where $p_\tau^2+\mathfrak{q}_0+\hbar\mathfrak{q}_1<0$, which by (1) is the area of the set where $\mathsf{K}_0+\hbar\mathsf{K}_1<0$. For $|\varepsilon|\le\hbar_0$, step (b) of the proof of Proposition 2, applied with ε in place of $\hbar$ on I_+ and on I_-, shows that $\mathfrak f_\varepsilon\equiv\mathfrak{q}_0+\varepsilon\mathfrak{q}_1$ has exactly two zeros $\tau_L(\varepsilon)<\tau_R(\varepsilon)$ on $[\tau_{-1},\tau_1]$, that both are simple and continuously differentiable in ε, and that $\mathfrak f_\varepsilon$ is positive outside them. So the curves $\mathsf{K}_0+\varepsilon\mathsf{K}_1=0$ are closed with

no critical point and lie in a fixed compact set. By the coarea formula the derivative in ε of the area of the set where $\mathsf{K}_0+\varepsilon\mathsf{K}_1<0$ is $-\oint \mathsf{K}_1\,dt_\varepsilon$, where t_ε is the time of the flow of $\mathsf{K}_0+\varepsilon\mathsf{K}_1$. In the variables of (1), the factors c^2 cancel and $\oint \mathsf{K}_1\,dt_\varepsilon=\int_{\tau_L(\varepsilon)}^{\tau_R(\varepsilon)}\mathfrak{q}_1(-\mathfrak{f}_\varepsilon)^{-1/2}\,d\tau$.

Its derivative in ε cannot be taken under this integral, whose derivative diverges at the turning points. Write instead $-\mathfrak{f}_\varepsilon=(\tau-\tau_L(\varepsilon))(\tau_R(\varepsilon)-\tau)\,r_\varepsilon$. The factor r_ε is a divided difference of $\mathfrak{f}_\varepsilon$, which has three continuous derivatives in τ; r_ε is positive and continuously differentiable in (τ,ε). Substitute $\tau=\tau_L+(\tau_R-\tau_L)(1-\cos\phi)/2$. The integral becomes $\int_0^\pi \mathfrak{q}_1\,r_\varepsilon^{-1/2}\,d\phi$, whose integrand is continuously differentiable in ε and continuous in E, so its derivative in ε is bounded on the compact set of (ε,E). So $2\Phi_\hbar=\mathcal{A}-\hbar\oint \mathsf{K}_1\,dt_0+O(\hbar^2)$, and with (4) and (5) this is the first statement.

(7) *Each value once.* Assume (iii). This step works in x: in it a prime denotes d/dx, and derivatives in τ are written $d/d\tau$. The equation of $\mathsf{K}F=0$ is $(\hbar g F')'=(k_{\mathrm{S},\hbar}/\hbar)\,F$, the equation of Lemma 5 with $k_\mathrm{S}=k_{\mathrm{S},\hbar}$. Its hypotheses hold: (S1) is (iii). Choose x_* such that $|x|\ge x_*$ lies in the tails $\tau\ge\tau_R+\delta$ and $\tau\le\tau_L-\delta$ for every $E\in J$, which is possible since these ends are continuous in E. By (M0), $F_\pm=w_\mathrm{L}^{-1/2}u_\pm$ has no zero on its tail, $F_\pm>0$ there for the right sign of $u_\pm$, $\int gF_\pm^2\,dx$ is finite there by (3), and $gF_\pm'/F_\pm$ at $\pm x_*$ is continuous in E. Together these give (S2) and (S3). So $\Delta_\theta\equiv\theta_--\theta_+$ at $x=0$ is continuous and strictly increasing on J, by Lemma 5.

Back to $\mathcal{I}_\hbar$. The map from $(F,\hbar gF')$ at $x=0$ to $(u,\hbar\,(du/d\tau)/a_\mathrm{m})$ at τ_0 is the solution flow from 0 to the point τ_0, whose determinant is 1, followed by

$$\begin{pmatrix} w_\mathrm{L}^{1/2} & 0 \\ \hbar\,g^{1/2}(w_\mathrm{L}^{1/2})_x/a_\mathrm{m} & w_\mathrm{L}^{1/2}g^{-1/2}/a_\mathrm{m}\end{pmatrix},$$

whose determinant $w_\mathrm{L}\,g^{-1/2}/a_\mathrm{m}$ is positive. Call the product $\mathcal{G}(E)$. The quantities τ_0, a_m, w_L and g are continuous in E, as are the solutions, so $\mathcal{G}(E)$ is continuous in E. Also $\det\mathcal{G}(E)>0$. The vector $(F_\pm,\hbar gF_\pm')$ at $x=0$ has the angle $\theta_\pm(0,E)$, and $\mathcal{G}(E)$ maps it to $(u_\pm,\hbar\,(du_\pm/d\tau)/a_\mathrm{m})$ at τ_0, whose angle is $\vartheta_\pm$ modulo π (the sign of $F_\pm$ changes an angle by π). By (5), $\mathcal{I}_\hbar$ is continuous and $\mathcal{I}_\hbar/(2\hbar)$ is congruent to $\vartheta_--\vartheta_+$ modulo π. So Lemma 6 applies with $\theta_\pm(E)=\theta_\pm(0,E)$, and by its part (c) $\mathcal{I}_\hbar$ takes each value $2\pi\hbar k$ at most once on J. □

Lemma 7 concerns the class of operators $c^2Q_s^\dagger gQ_s-\mathcal{U}$ and uses nothing about the Dirac equation. Its hypotheses are global and sufficient: the regularity stated first, with g between two positive constants; (i), one regular closed level curve at each E; (ii), a lower bound at infinity, uniform in E; and, for the count only, (iii), monotonicity in E at every x. The Dirac equation enters in two places only: the exact reduction of a block to $\mathsf{K}(E)$, and the check of these hypotheses for the coefficients of $\mathsf{K}(E)$, made in the proof of each theorem. Four steps make the Airy argument uniform in E. (1) *The decaying solutions and their logarithmic derivatives.* By (ii) and the regularity, (U3) and (U4) hold with $\mathfrak{q}_*$ and C_0 independent of E. In step (a) of the proof of Proposition 2 the logarithmic derivative $\hbar u'/u=-\sqrt{\mathfrak{q}}+z$ obeys a Riccati equation for which the band $|z|\le\hbar C_T$ is invariant toward smaller τ, and two solutions in the band approach each other at the rate $\sqrt{\mathfrak{q}_*/2}/\hbar$; the tails beyond $\tau_{\pm1}$ enter only there. (2) *Turning-point separation and lower bounds.* All constants of the matching are largest values of functions made continuous by (U1), or least values of positive continuous functions, on a compact set of $(\tau,E,\hbar)$ whose ends move continuously with E. The lower bounds are the slope c_* at the turning points, which keeps them simple; their separation ℓ_*, which keeps τ_0 at a distance at least $2\delta_\mathrm{tp}$ from both; the bound μ on $|\mathfrak{q}_0|$ away from them, which excludes a third near zero; and $\mathfrak{q}_*$ on the tails. As any of c_*, ℓ_* or μ tends to 0 the constants may grow without bound; hypothesis (i), a closed curve with no critical point at every E of the compact J, keeps all three positive. (3) *Matching to the square-integrable tails.* At τ_1 Olver's ϵ_2 vanishes, and since every solution square-integrable on the tail is a multiple of u_+ (M0), the eigenvalue condition is imposed on the solution matched to the Airy function. (4) *From the local phase to a continuous global action.* Lemma 4 takes the one representative of the phase error in $(-\pi/2,\pi/2)$ and so defines $\mathcal{I}_\hbar$ continuous on all of J, with no derivative in E. For the count, Lemma 6 transfers the labels from the Sturm angle at $x=0$ to $\mathcal{I}_\hbar$ through the change of coordinates $\mathcal{G}(E)$, so distinct crossings have distinct labels, and none is skipped.

In the applications, hypothesis (iii) takes one form, $k_{\mathrm{S},\hbar}(x,E')-k_{\mathrm{S},\hbar}(x,E)=-(E'-E)\,\beta_\mathrm{mon}(x,E,E')$, where β_mon is a positive function plus a term of order $\hbar$ that grows with $|\varrho|$.

Lemma 8 (Monotonicity in the energy)**:** *Let J be a compact interval, $s=\pm1$, and for $x\in\mathbb{R}$ and $E,E'\in J$ let $b_{mon}(x,E,E')$ and $a_{mon}(x,E,E')$ be real, with a_{mon} continuously differentiable in x, and let $\varrho(x)$ be real and continuously differentiable. Let $b_{mon}\ge b_*>0$, $a_{mon}\ge a_*>0$, $|\partial_x a_{mon}|\le M_{mon}$ and $a_{mon}\,|\varrho'|\le M_{mon}$ on $\mathbb{R}\times J\times J$, with $M_{mon}>0$, and let*

$$\beta_{mon}\equiv b_{mon}+a_{mon}\varrho^2-s\hbar\,(a_{mon}\varrho)'.$$

Then for $0<\hbar\le\hbar_{mon}\equiv\min\big(b_*/(4M_{mon}),\,\sqrt{a_*b_*}/M_{mon}\big)$, $\beta_{mon}\ge b_*/2$ at every x and every $E,E'\in J$.

Proof. Since $(a_{\rm mon}\varrho)' = (\partial_x a_{\rm mon})\,\varrho + a_{\rm mon}\varrho'$, $|s\hbar\,(a_{\rm mon}\varrho)'| \le \hbar M_{\rm mon}\,(1+|\varrho|)$, so

$$\beta_{\rm mon} \ge b_* - \hbar M_{\rm mon} + \big(a_*\varrho^2 - \hbar M_{\rm mon}|\varrho|\big).$$

The quadratic $a_*t^2 - \hbar M_{\rm mon}t$ has its least value $-\hbar^2 M_{\rm mon}^2/(4a_*)$ at $t = \hbar M_{\rm mon}/(2a_*)$, so

$$\beta_{\rm mon} \ge b_* - \hbar M_{\rm mon} - \frac{\hbar^2 M_{\rm mon}^2}{4a_*}.$$

For $\hbar \le b_*/(4M_{\rm mon})$ the second term is at most $b_*/4$, and for $\hbar \le \sqrt{a_*b_*}/M_{\rm mon}$ the third is at most $b_*/4$. □

3. Levels in a magnetic field along one direction

Theorem 1 (Levels in a magnetic field along one direction)**:** *In flat spacetime let V be constant and $\mathbf{B} = B(x)\,\hat{\mathbf{z}}$, with B smooth and bounded with bounded derivatives and $qB > 0$, and let p_y, p_z and s be fixed, with ϱ tending to $\varrho_- < 0$ as $x \to -\infty$ and to $\varrho_+ > 0$ as $x \to +\infty$, and $\Lambda_c \equiv \min(\varrho_-^2, \varrho_+^2)$. The operator $\mathcal{P}_s$ on $L^2(\mathbb{R})$ is self-adjoint. For every compact interval $I \subset (0, \Lambda_c)$ there are C and $\hbar_0 > 0$ such that for $\hbar < \hbar_0$ the eigenvalues of $\mathcal{P}_s$ in I are simple and are a one-to-one $O(\hbar^2)$ spectral approximation (Definition 1) of the solutions E_n^0 of Eq. (141) without its remainder. For every profile, 0 is a simple eigenvalue of $\mathcal{P}_{+1}$ and not of $\mathcal{P}_{-1}$, and the positive eigenvalues of $\mathcal{P}_{+1}$ and $\mathcal{P}_{-1}$ coincide.*

Proof. For $\mathcal{P}_s$, $\mathsf{K}(\Lambda) \equiv c^2(\mathcal{P}_s - \Lambda)$ has the form of Lemma 7 with $g = 1$, $\mathcal{U} = c^2\Lambda$ and ϱ independent of Λ, so $\mathsf{K}_0 = c^2(\mathfrak{p}_0 - \Lambda)$, $\mathsf{K}_1 = c^2\mathfrak{p}_1$, and $k_{{\rm S},\hbar} = \varrho^2 - s\hbar\varrho' - \Lambda$ decreases strictly in Λ, which is hypothesis (iii). On a compact $I \subset (0, \Lambda_c)$ hypothesis (i) holds because $x \mapsto \varrho$ is increasing, so that each level set $\xi^2 + \varrho^2 = \Lambda$ is one closed curve with no critical point, and hypothesis (ii) holds with $\varpi_\infty = c^2(\Lambda_c - \max I)/2$, since ϱ^2 tends to at least Λ_c at both ends. Self-adjointness of $\mathcal{P}_s$ is that of $Q^\dagger Q$ and $QQ^\dagger$ for the closed operator Q. The eigenvalues of $\mathcal{P}_s$ in I are the Λ at which 0 is an eigenvalue of $\mathsf{K}(\Lambda)$, the set $Z_\hbar$ of Lemma 1; each is simple, since by step (4) of the proof of Lemma 7 the square-integrable solutions at one Λ are multiples of one function. Lemma 7 bounds $\mathcal{I}_\hbar - \mathcal{A} + \hbar\oint \mathsf{K}_1\,dt_0 - \pi\hbar$ by $C\hbar^2$. The flow of K_0 is that of $\mathfrak{p}_0$ with $t_0 = t_\mathfrak{p}/c^2$, and $\mathsf{K}_1 = c^2\mathfrak{p}_1$. Therefore $\oint \mathsf{K}_1\,dt_0 = \oint \mathfrak{p}_1\,dt_\mathfrak{p} = -s\pi$, since $\oint qB\,dt_\mathfrak{p} = \pi$. By hypothesis (iii) of Lemma 7, $\mathcal{I}_\hbar$ takes each value in $2\pi\hbar\mathbb{Z}$ at most once. Lemma 1 with this $\mathcal{I}_\hbar$, $h = \mathfrak{p}_0$, $\varphi_\hbar = \oint \mathfrak{p}_1\,dt_\mathfrak{p}$ and $S_\hbar = Z_\hbar$ gives the two-way and the one-to-one statements; the theorem of Ifa *et al.* [94] gives the two-way statement independently. For the zero mode and the pairing, $QF_0 = 0$ is solved by $F_0 = \exp(-\int^x \varrho\,dx'/\hbar)$, which is square-integrable because $\varrho \to \varrho_\pm$ with $\varrho_- < 0 < \varrho_+$, while $Q^\dagger F = 0$ gives $\exp(+\int^x \varrho\,dx'/\hbar)$, which is not; if $\mathcal{P}_{+1}F = \Lambda F$ with $\Lambda > 0$, then $\mathcal{P}_{-1}(QF) = Q\,Q^\dagger QF = \Lambda\,QF$ with $QF \ne 0$, and conversely with $Q^\dagger$. □

4. Levels with an electric potential

In flat spacetime let B, p_y and s be as in Theorem 1, with ϱ tending to limits $\varrho_\pm$, finite or infinite, as $x \to \pm\infty$, and $\varrho_- < 0 < \varrho_+$. Let $V(x)$ be smooth and bounded with bounded derivatives, $m > 0$ and $p_z = 0$. Let J be a compact interval of energies $E = \hbar\omega$ on which $E - V + mc^2 \ge E_{\rm gap} > 0$, the level set $\mathsf{K}_0 = 0$ is a closed curve with no critical point of K_0, and 0 lies below the lower limit of K_0 at infinity. These are the hypotheses of this subsection.

Let $\mathcal{D}_s$ be the block of sector s, the map $(\mathsf{u}, \mathsf{d}) \mapsto \big((V+mc^2)\mathsf{u} + cQ_s^\dagger\mathsf{d},\ cQ_s\mathsf{u} + (V-mc^2)\mathsf{d}\big)$ on $\operatorname{dom}Q_s \oplus \operatorname{dom}Q_s^\dagger$, whose Weyl symbol is $H_s \equiv V\mathbf{1} + c\xi\,\sigma_1 - sc\varrho\,\sigma_2 + mc^2\sigma_3$, since $Q_s = \hat{p} - is\varrho$ has the symbol $\xi - is\varrho$. It is self-adjoint there: its off-diagonal part is self-adjoint because Q_s is closed, so that $(Q_s^\dagger)^\dagger = Q_s$, and $V\mathbf{1} + mc^2\sigma_3$ is bounded. The eigenvalues of H_s are $h_\pm = V \pm |\mathbf{P}|$, where $\mathbf{P} \equiv (c\xi,\ -sc\varrho,\ mc^2)$, and

$$\mathsf{K}_0 = g\,(h_+ - E)(E - h_-),$$

with $E - h_- \ge E - V + mc^2 \ge E_{\rm gap}$. So the set $h_+ = E$ is the curve $\mathsf{K}_0 = 0$, on which $\nabla\mathsf{K}_0 = g\,(E - h_-)\,\nabla h_+$, and the set where $\mathsf{K}_0 < 0$ is the set where $H_0 = h_+ < E$.

Lemma 7 applies to $\mathsf{K}(E)$ with $g = 1/(E - V + mc^2)$ and $\mathcal{U} = E - V - mc^2$, whose x-derivatives are continuous in (x, E) and bounded on $\mathbb{R} \times J$ because V has bounded derivatives and $E - V + mc^2 \ge E_{\rm gap}$. Its hypothesis (i) is the hypothesis on the curve $\mathsf{K}_0 = 0$. Its hypothesis (ii) holds because 0 lies below the lower limit Λ_∞ of K_0 at infinity at $E = \max J$. For a finite ϖ_∞ with $0 < \varpi_\infty < \Lambda_\infty$ there is then x_∞ with $c^2 g\varrho^2 - \mathcal{U} \ge \varpi_\infty$ for $|x| \ge x_\infty$ at $E = \max J$, and this bound holds for every $E \in J$, since $c^2 g\varrho^2 - \mathcal{U}$ decreases with E. Its hypothesis (iii) holds: $g(E') - g(E) = -(E' - E)\,g(E)g(E')$, and ϱ does not depend on E, so

$$k_{{\rm S},\hbar}(E') - k_{{\rm S},\hbar}(E) = -(E' - E)\Big[\frac{1}{c^2} + g(E)g(E')\varrho^2 - s\hbar\big(g(E)g(E')\varrho\big)'\Big],$$

which is Lemma 8 with $b_{\rm mon} = 1/c^2$, $a_{\rm mon} = g(E)g(E')$ and $\varrho' = qB$. Since $E - V + mc^2 \geq E_{\rm gap}$, $g \leq 1/E_{\rm gap}$ and $g \geq a_g \equiv 1/(\max J - \inf V + mc^2)$, and $\partial_x g = g^2 V'$, so the lemma applies with $b_* = 1/c^2$, $a_* = a_g^2$ and $M_{\rm mon} = (2\sup|V'|/E_{\rm gap} + \sup|qB|)/E_{\rm gap}^2$, and the bracket is at least $1/(2c^2)$ at every x for $\hbar \leq \hbar_{\rm mon}$. By the relation $\mathsf{K}_1\, dt_0 = \frac{1}{2} s\, \Omega_z\, dt$, the first-order term is $\oint \mathsf{K}_1\, dt_0 = \frac{1}{2} s\, \Theta_{\rm spin}$, smooth in E because the orbits are.

Proposition 3 (The chain of equivalences)**:** *Let J be a compact interval on which $E - V + mc^2 \geq E_{gap} > 0$ and on which $\mathsf{K}(E)$ meets hypotheses (i) and (ii) of Lemma 7, and let $\hbar < \hbar_0$ of that lemma. For $E \in J$ the following are equivalent:*

(a) E is an eigenvalue of $\mathcal{D}_s$;

(b) 0 is an eigenvalue of $\mathsf{K}(E)$;

(c) the Wronskian $\mathcal{W}(E)$ of the solutions u_+ and u_- of Proposition 2, decaying at $+\infty$ and at $-\infty$, is zero;

(d) $\mathcal{I}_\hbar(E) \in 2\pi\hbar\mathbb{Z}$, with $\mathcal{I}_\hbar$ of Lemma 7.

The map $F \mapsto (F,\, c\, g\, Q_s F)$ is a linear bijection from $\ker \mathsf{K}(E)$ onto $\ker(\mathcal{D}_s - E)$, with inverse $(\mathsf{u}, \mathsf{d}) \mapsto \mathsf{u}$. When (c) holds, $\ker \mathsf{K}(E)$ is spanned by $w_L^{-1/2} u_+$, with w_L of step (1) of the proof of Lemma 7. Both kernels have dimension at most 1.

Proof. (b) implies (a). Let $F \in \ker \mathsf{K}(E)$, so $F \in \operatorname{dom} Q_s$ and, since $\mathsf{K}(E)$ is the operator of its form (Lemma 2(i)), $c^2\langle Q_s\tilde{F}, g Q_s F\rangle = \langle \tilde{F}, (E - V - mc^2)F\rangle$ for every $\tilde{F} \in \operatorname{dom} Q_s$. Let $\mathsf{d}_F \equiv c\, g\, Q_s F$, which is square-integrable since $g \leq 1/E_{\rm gap}$. The identity says $\langle Q_s\tilde{F}, \mathsf{d}_F\rangle = \langle \tilde{F}, (E - V - mc^2)F/c\rangle$ for every $\tilde{F} \in \operatorname{dom} Q_s$, which is the definition of $\mathsf{d}_F \in \operatorname{dom} Q_s^\dagger$ with $cQ_s^\dagger \mathsf{d}_F = (E - V - mc^2)F$, since $Q_s^\dagger$ is the adjoint of Q_s. So (F, d_F) lies in the domain of $\mathcal{D}_s$, its first row is $(V + mc^2)F + (E - V - mc^2)F = EF$, and its second is $\mathsf{d}_F/g + (V - mc^2)\mathsf{d}_F = E\mathsf{d}_F$. (a) implies (b). Let $\mathcal{D}_s(\mathsf{u}, \mathsf{d}) = E(\mathsf{u}, \mathsf{d}) \neq 0$ in the domain of $\mathcal{D}_s$. The second row gives $\mathsf{d} = c\, g\, Q_s \mathsf{u}$, so $\mathsf{u} \neq 0$, and the first gives $cQ_s^\dagger \mathsf{d} = (E - V - mc^2)\mathsf{u}$. For $\tilde{F} \in \operatorname{dom} Q_s$, $c^2\langle Q_s\tilde{F}, g Q_s \mathsf{u}\rangle = c\langle \tilde{F}, Q_s^\dagger \mathsf{d}\rangle = \langle \tilde{F}, (E - V - mc^2)\mathsf{u}\rangle$, so the form of $\mathsf{K}(E)$ vanishes on $(\tilde{F}, \mathsf{u})$ for every $\tilde{F}$ in the form domain, and $\mathsf{u} \in \ker \mathsf{K}(E)$. The two maps are linear and inverse to each other, which is the bijection. (b) is equivalent to (c) by step (4) of the proof of Lemma 7, where also every $F \in \ker \mathsf{K}(E)$ is $w_{\rm L}^{-1/2}$ times a multiple of u_+, so the dimension is at most 1; and (c) is equivalent to (d) by step (5) of that proof. □

The domain of the block is part of the statement: an eigenvector of $\mathcal{D}_s$ has $\mathsf{u} \in \operatorname{dom} Q_s$, which is the form domain of $\mathsf{K}(E)$, and $\mathsf{d} \in \operatorname{dom} Q_s^\dagger$, which the elimination produces from the vanishing of the form rather than assumes. The correspondence is a bijection of kernels at each E and preserves their dimensions; it is not a unitary equivalence between the block and the family $\mathsf{K}(E)$, which depends on E.

Theorem 2 (Levels with an electric potential)**:** *Under the hypotheses of this subsection, for $\hbar$ small the Dirac levels in J of the block of fixed p_y and s are simple within the block, no essential spectrum of the block lies in J, and the levels are a one-to-one $O(\hbar^2)$ spectral approximation (Definition 1) of the solutions E_n^0 of Eq. (143) without its remainder.*

Proof. Simplicity is Lemma 2(iii), whose hypothesis is that 0 lies below the essential spectrum of $\mathsf{K}(E)$ for every $E \in J$. That hypothesis holds by Lemma 3, which applies as stated after it, with a $\Lambda_{\rm ess} > 0$ below the essential spectrum of every $\mathsf{K}(E)$, $E \in J$. The bijection of Proposition 3 gives simplicity for the block, so the eigenspace of the block at a level has dimension 1.

No essential spectrum of the block lies in J. Let $E \in J$, and suppose that $\Psi_k = (\mathsf{u}_k, \mathsf{d}_k)$ in the domain of the block is a singular Weyl sequence: $\|\Psi_k\| = 1$, $\Psi_k \to 0$ weakly, and the residuals $\mathfrak{r}_{k,1} \equiv (V + mc^2 - E)\,\mathsf{u}_k + c\, Q_s^\dagger \mathsf{d}_k$ and $\mathfrak{r}_{k,2} \equiv c\, Q_s \mathsf{u}_k + (V - E - mc^2)\,\mathsf{d}_k$ tend to 0. Since $E - V + mc^2 \geq E_{\rm gap}$, $\mathsf{d}_k = c\, g\, Q_s \mathsf{u}_k - g\, \mathfrak{r}_{k,2}$, so $\|Q_s \mathsf{u}_k\|$ is bounded, and the inner product of u_k with $\mathfrak{r}_{k,1}$ gives, for the form $\mathfrak{k}$ of $\mathsf{K}(E)$,

$$\mathfrak{k}(\mathsf{u}_k) = \langle \mathsf{u}_k, \mathfrak{r}_{k,1}\rangle + c\,\langle Q_s \mathsf{u}_k, g\, \mathfrak{r}_{k,2}\rangle,$$

which tends to 0. If $\|\mathsf{u}_k\| \to 0$ on a subsequence, then $\|Q_s \mathsf{u}_k\| \to 0$, since $c^2 \inf g\, \|Q_s \mathsf{u}_k\|^2 \leq \mathfrak{k}(\mathsf{u}_k) + \sup|E - V - mc^2|\, \|\mathsf{u}_k\|^2$, and then $\|\mathsf{d}_k\| \to 0$, against $\|\Psi_k\| = 1$; so $\|\mathsf{u}_k\| \geq \delta_{\mathsf{u}} > 0$ for k large. The spectral projection Π_P of $\mathsf{K}(E)$ on $(-\infty, \Lambda_{\rm ess}/2)$ has finite rank, so $\Pi_P \mathsf{u}_k \to 0$ because $\mathsf{u}_k \to 0$ weakly, and $\mathfrak{k}(\mathsf{u}_k) \geq (\Lambda_{\rm ess}/2)\, \|(1 - \Pi_P)\mathsf{u}_k\|^2 + \langle \Pi_P \mathsf{u}_k, \mathsf{K}\Pi_P \mathsf{u}_k\rangle$ has limit inferior at least $\Lambda_{\rm ess}\delta_{\mathsf{u}}^2/2 > 0$, a contradiction. By Weyl's criterion E is not in the essential spectrum of the block, so its spectrum in J consists of isolated eigenvalues.

For the bound in both directions, $E \in J$ is a level of the block exactly when 0 is an eigenvalue of $\mathsf{K}(E)$, by Proposition 3. So, by Lemma 7, whose hypotheses were checked at the start of this subsection, the levels in J are

the set $Z_\hbar$ of Lemma 1, with $h = H_0$, $\varphi_\hbar = \oint \mathsf{K}_1\, dt_0 = \frac{1}{2} s\, \Theta_{\mathrm{spin}}$, $S_\hbar = Z_\hbar$ and $\varepsilon_\hbar = 0$. That lemma gives the two-way statement, and since hypothesis (iii) of Lemma 7 holds, $\mathcal{I}_\hbar$ takes each value in $2\pi\hbar\mathbb{Z}$ at most once on J, so the same lemma gives the one-to-one statement. □

The constants. The proof of Lemma 1 gives the distance $\delta_0 + \varepsilon_\hbar \le (2C_{\mathcal{I}}/T_{\min} + \hbar)\,\hbar^2$, so $C = 2C_{\mathcal{I}}/T_{\min} + \hbar_0$ for $\hbar \le \hbar_0$, a constant independent of $\hbar$. It needs $\hbar$ small enough for three conditions: $\hbar C_1 \le T_{\min}/2$; $\delta_0 + \varepsilon_\hbar$ less than the distance from J' to the ends of J_1 and from J_1 to the ends of J; and $2\pi\hbar/(T_{\max} + \hbar C_1) > 2(\delta_0 + \varepsilon_\hbar)$. Here $T_{\min}$ and $T_{\max}$ are the least and largest periods of the orbits $H_0 = E$ on J, C_1 is the largest absolute value on J of the derivative of $\frac{1}{2} s\, \Theta_{\mathrm{spin}}$ in E, and $\varepsilon_\hbar = 0$. The constant $C_{\mathcal{I}}$ is the constant C of Lemma 7, built from the bounds of g, $\mathcal{U}$ and ϱ' and their derivatives and, in the proof of Proposition 2, from the numbers c_*, ℓ_*, δ_{tp}, μ, δ, $\mathfrak{q}_*$ and C_T and the bound $\Psi_{\max}$ of Olver's function ψ_{O}. Each of these numbers is a least or largest value over the compact J of a function continuous in E; none depends on n. This is the sense of the uniformity on J in Theorem 2.

TABLE IV. The steps of the proof of Theorem 2.

Step	Statement	Objects	Kind	Source
1	Eliminating d makes a level a zero of an eigenvalue of $\mathsf{K}(E)$; the block $\mathcal{D}_s$ is self-adjoint with Weyl symbol exactly H_s	$\mathsf{K}(E)$, H_s	exact	Lemma 2, Proposition 3
2	The curve $\mathsf{K}_0 = 0$ is the orbit $H_0 = E$; $\mathsf{K}(E)$ meets hypotheses (i)–(iii) of Lemma 7 on J	K_0, $h_\pm$, H_0	exact	Lemma 8
3	0 is an eigenvalue of $\mathsf{K}(E)$ exactly when $\mathcal{I}_\hbar(E) \in 2\pi\hbar\mathbb{Z}$, with $\mathcal{I}_\hbar$ equal to the area minus $\hbar \oint \mathsf{K}_1\, dt_0$ plus $\pi\hbar$, up to $C\hbar^2$ uniformly on J; $\mathcal{I}_\hbar$ takes each value in $2\pi\hbar\mathbb{Z}$ at most once	$\mathsf{K}(E)$, $\mathcal{I}_\hbar$	semiclassical; the count exact	Lemma 7, Proposition 2, Lemmas 4, 5 and 6
4	The first-order term is spin transport, $\oint \mathsf{K}_1\, dt_0 = \frac{1}{2} s \Theta_{\mathrm{spin}}$ on $H_0 = E$	K_1, Ω_z	exact	$\mathsf{K}_1\, dt_0 = \frac{1}{2} s\, \Omega_z\, dt$
5	The remainder $O(\hbar^2)$ of the action becomes a one-to-one bound $C\hbar^2$ in the energy, with no derivative of the remainder	$\mathcal{I}_\hbar$, E_n^0	exact estimate	Lemma 1
6	The levels are simple and no essential spectrum of the block lies in J	$\mathsf{K}(E)$, $\mathcal{D}_s$	exact	Lemmas 2 and 3, Weyl's criterion

The steps of the argument are listed in Table IV, with the objects each step uses and whether the step is exact or semiclassical; an exact step makes no approximation in $\hbar$.

The count and the minimax principle. The one-to-one count can also be phrased through the eigenvalues of $\mathsf{K}(E)$. By Lemma 2(ii) and (iii) each k_j decreases in E and the levels are simple, so the number of levels in $[E_a, E_b) \subset J$ is $N_-(E_b) - N_-(E_a)$, where $N_-(E)$ is the number of negative eigenvalues of $\mathsf{K}(E)$, finite by Lemma 3. The count therefore says that N_- rises by exactly one across each solution of the rule. A semiclassical counting function, $N_-(E) = \mathcal{I}_\hbar(E)/(2\pi\hbar) + O(1)$, does not give this: an error of order one allows a window that contains two levels and a neighboring one that contains none, which the two-way statement of Lemma 1 does not exclude either. The count needs $N_-(E)$ exactly. In one dimension the oscillation theorem of Sturm–Liouville theory gives it exactly, as the number of zeros of the solution that decays at $+\infty$. Counting those zeros through the turning points requires its phase there and the continuity of the label, which Lemmas 5 and 6 give at $x = 0$. The Sturm angle is therefore one constructive way to the exact count, and that oscillation theorem would be another; both need an exact scalar equation on the whole line. The exact reduction gives the ordinary differential equation $\mathsf{K}(E)F = 0$ on the whole line for every $\hbar$ and E, so comparison and oscillation arguments apply to it directly. Reducing a 2×2 system to a scalar operator near the curve (the semiclassical normal form) yields the rule but does not by itself give such an equation on the whole line.

5. Levels in crossed fields

Theorem 3 (Levels in crossed fields)**:** *In flat spacetime let B be smooth and bounded with bounded derivatives and $qB > 0$, let $V = \beta_d c\, q A_y$ with $0 < \beta_d < 1$, and let $m > 0$, p_y and s be fixed and $p_z = 0$. Call an orbit admissible if it is a closed curve of $c^2(\xi^2 + \varrho^2) = \mathcal{E}^2 - (mc^2)^2$ on which $\mathcal{E} \ge mc^2$, and call the levels with $E_d > 0$ the* electronic branch.

(a) On every admissible orbit $\Theta_{spin} = -2\pi$.

(b) Let J be a compact interval of $E_d > 0$ on which every orbit is admissible, with $0 < \Lambda < \Lambda_c(b_d)$, where $\Lambda_c(b_d)$ is the threshold of the continuum of $\mathcal{P}_s$ for the field B/γ_d at the transverse momentum b_d. At fixed laboratory p_y and for $\hbar$ small, the levels of the electronic branch with $E_d \in J$ are simple and are a two-way $O(\hbar^2)$ spectral approximation (Definition 1, with E_d as the energy) of the solutions of Eq. (143) without its remainder.

(c) For uniform qB, Eq. (143) without its remainder gives Eq. (145) with $E_d > 0$, which are all the exact Dirac levels of the electronic branch for every integer $n \geq 0$ and $s = \pm 1$.

(d) For $qA_y = \mathfrak{a}\tanh(x/L)$ with $\mathfrak{a}, L > 0$, Eq. (143) without its remainder gives Eq. (146) for the integers $n \geq 0$ with $\mathcal{N}_N > \sqrt{\mathfrak{a}_d|b_d|}$, and these are all the exact Dirac levels of the electronic branch.

Proof. (a) With $V = \beta_{\rm d} c\, qA_y$ and $\mathcal{E} = E_{\rm d} - \beta_{\rm d} c\, \varrho$, the level set $c^2(\xi^2 + \varrho^2) = \mathcal{E}^2 - (mc^2)^2$ is the ellipse $\xi^2 + (1-\beta_{\rm d}^2)\,(\varrho - \varrho_{\rm e})^2 = \xi_{\max}^2$ in the plane of (ϱ, ξ), with $\varrho_{\rm e} \equiv -\beta_{\rm d} E_{\rm d}/[c\,(1-\beta_{\rm d}^2)]$ and $\xi_{\max}^2 \equiv E_{\rm d}^2/[c^2(1-\beta_{\rm d}^2)] - m^2c^2$, the same for every profile. With $dt = \mathcal{E}\, d\varrho/(c^2 qB\, \xi)$, $V' = \beta_{\rm d} c\, qB$ and $\mathcal{E} + mc^2 = \beta_{\rm d} c\,(\varrho_{\rm b} - \varrho)$, $\varrho_{\rm b} \equiv (E_{\rm d} + mc^2)/(\beta_{\rm d} c)$, the expression for Ω_z gives

$$\Omega_z\, dt = -\frac{\varrho_{\rm b}\, d\varrho}{(\varrho_{\rm b} - \varrho)\,\xi}.$$

One period runs twice over $[\varrho_1, \varrho_2]$, between the zeros of $\xi^2 = (1-\beta_{\rm d}^2)(\varrho_2 - \varrho)(\varrho - \varrho_1)$, and since $\varrho_{\rm b} > \varrho_2$,

$$\int_{\varrho_1}^{\varrho_2} \frac{d\varrho}{(\varrho_{\rm b} - \varrho)\sqrt{(\varrho_2 - \varrho)(\varrho - \varrho_1)}} = \frac{\pi}{\sqrt{(\varrho_{\rm b} - \varrho_1)(\varrho_{\rm b} - \varrho_2)}}.$$

At $\varrho = \varrho_{\rm b}$, where $\mathcal{E} = -mc^2$, the equation of the ellipse gives $\xi^2 = -\varrho_{\rm b}^2$, so $(1-\beta_{\rm d}^2)(\varrho_{\rm b} - \varrho_1)(\varrho_{\rm b} - \varrho_2) = \varrho_{\rm b}^2$, and

$$\Theta_{\rm spin} = -\frac{2\pi\, \varrho_{\rm b}}{\sqrt{1-\beta_{\rm d}^2}\,\sqrt{(\varrho_{\rm b} - \varrho_1)(\varrho_{\rm b} - \varrho_2)}},$$

which is

$$\Theta_{\rm spin} = -2\pi$$

on every admissible orbit, for every $\beta_{\rm d}$ and every profile $B(x)$.

(b) Since $\|\hbar F'\|^2 + \|\varrho F\|^2 = \|Q_s F\|^2 + s\hbar\langle F, qBF\rangle$, V is bounded relative to the off-diagonal part of the block with relative bound $\beta_{\rm d} < 1$, so the block is self-adjoint by the Kato–Rellich theorem, and the Lorentz boost to the drift frame is exact for the Dirac equation. By the boost, a level of the electronic branch is an $E_{\rm d}$ at which 0 is an eigenvalue of $\mathsf{K}(E_{\rm d}) \equiv c^2\, Q_s^\dagger Q_s - c^2\Lambda$, with $\varrho_{\rm d}$ in place of ϱ in Q_s. This is the operator of Lemma 7 with the parameter $E_{\rm d}$, $g = 1$, $\mathcal{U} = c^2\Lambda$ and $\varrho_{\rm d}$ in place of ϱ; the derivative $qB/\gamma_{\rm d}$ of $\varrho_{\rm d}$ does not depend on $E_{\rm d}$. Hypothesis (i): $\mathsf{K}_0 = c^2(\xi^2 + \varrho_{\rm d}^2 - \Lambda)$ is critical only at $\xi = \varrho_{\rm d} = 0$, inside the curve, and the curve is closed for $0 < \Lambda < \Lambda_c(b_{\rm d})$, as for $\mathcal{P}_s$. Hypothesis (ii): the limits $\varrho_{\rm d\pm}$ of $\varrho_{\rm d}$ as $x \to \pm\infty$ are continuous in $E_{\rm d}$ where finite, $\varpi_0 \equiv \min_J(\Lambda_c - \Lambda) > 0$, and since $\varrho_{\rm d}$ increases with x, a finite cover of J gives one x_∞ with $\varrho_{\rm d}^2 - \Lambda \geq \varpi_0/2$ for $|x| \geq x_\infty$, so $\varpi_\infty = c^2\varpi_0/2$. Along the flow of K_0 the angle of $\xi + i\varrho_{\rm d}$ turns at the rate $2c^2 qB/\gamma_{\rm d}$, so $\oint \mathsf{K}_1\, dt_0 = -s\, c^2 \oint (qB/\gamma_{\rm d})\, dt_0 = -s\pi = \frac{1}{2} s\, \Theta_{\rm spin}$. For Lemma 1 take $h = H_0 - \beta_{\rm d} c\, p_y$ on the laboratory phase plane, so that $h < E_{\rm d}$ where $H_0 < E$. The maximum of $\beta_{\rm d} c\, \varrho - \sqrt{c^2\varrho^2 + (mc^2)^2}$ is $-mc^2/\gamma_{\rm d}$, so $E - h_- \geq E_{\rm d} + mc^2/\gamma_{\rm d} > 0$ on the whole plane for $E_{\rm d} > 0$. By the boost identity, $\mathsf{K}_0 = c^2(\xi^2 + \varrho_{\rm d}^2 - \Lambda) = |\mathbf{P}|^2 - \mathcal{E}^2 = (h_+ - E)(E - h_-)$. Since $E - h_- > 0$, the curve $h = E_{\rm d}$ is the curve $\mathsf{K}_0 = 0$, with no critical point of h, and the set where $h < E_{\rm d}$ is the set where $\mathsf{K}_0 < 0$. The two lemmas, with $\varphi_\hbar = -s\pi$, $S_\hbar = Z_\hbar$ and $\varepsilon_\hbar = 0$, give a bound of order $\hbar^2$ in both directions between the levels and the solutions of Eq. (143) without its remainder. Hypothesis (iii) of Lemma 7 is not available: $k_{{\rm S},\hbar} = \varrho_{\rm d}^2 - \Lambda$ up to the term in $\hbar$, and $\partial(\varrho_{\rm d}^2 - \Lambda)/\partial E_{\rm d} = (2\gamma_{\rm d}/c)(\beta_{\rm d}\varrho_{\rm d} - \gamma_{\rm d} E_{\rm d}/c)$, which is positive where $\beta_{\rm d}\varrho_{\rm d} > \gamma_{\rm d} E_{\rm d}/c$, a set that is not empty when $\varrho_{\rm d}$ is unbounded. Each level is simple, since the boost maps its eigenvectors one to one onto those of $\mathcal{P}_s$ for the field $B/\gamma_{\rm d}$.

(c) For uniform qB the ellipse encloses the area $\oint \xi\, dx = \pi\, \xi_{\max}^2/(qB\sqrt{1-\beta_{\rm d}^2})$, since $dx = d\varrho/(qB)$, and Eq. (143) with $\Theta_{\rm spin} = -2\pi$ and no remainder sets it to $2\pi\hbar\,(n + \frac{1}{2} - \frac{1}{2}s)$, which is Eq. (145). These are the exact Dirac levels: $E_{\rm d}^2$ is $\gamma_{\rm d}^{-2}$ times the squared Dirac level $(mc^2)^2 + (2n+1-s)\,\hbar q B_{\rm d} c^2$ of a particle with $p_z = 0$ in the uniform magnetic field $B_{\rm d} = B/\gamma_{\rm d}$.

(d) Let $qA_y = \mathfrak{a}\tanh(x/L)$, so that $qB = (\mathfrak{a}/L)\,\mathrm{sech}^2(x/L)$ and $V = \beta_{\rm d} c\, \mathfrak{a}\tanh(x/L)$ is bounded. In the drift frame $\varrho_{\rm d} = \mathfrak{a}_{\rm d}\tanh(x/L) - b_{\rm d}$. For integers $j \geq 0$ let $b_{{\rm d},j} \equiv \mathfrak{a}_{\rm d} b_{\rm d}/\mathcal{N}_j$ and $Q_{(j)} \equiv \hat{p} - i\,(\mathcal{N}_j \tanh(x/L) - b_{{\rm d},j})$, so that $\mathcal{N}_0 = \mathfrak{a}_{\rm d}$ and $b_{{\rm d},0} = b_{\rm d}$. Then $\mathcal{P}_{+1} = Q_{(0)}^\dagger Q_{(0)}$ and $\mathcal{P}_{-1} = Q_{(0)} Q_{(0)}^\dagger$, and comparing the coefficients of $\tanh^2$, tanh and 1 gives the shape invariance [98]

$$Q_{(j)} Q_{(j)}^\dagger = Q_{(j+1)}^\dagger Q_{(j+1)} + \mathcal{N}_j^2 + b_{{\rm d},j}^2 - \mathcal{N}_{j+1}^2 - b_{{\rm d},j+1}^2.$$

The zero mode of $Q_{(j)}$ is square-integrable when $|b_{\mathrm{d},j}| < \mathcal{N}_j$, that is when $\mathcal{N}_j > \sqrt{\mathfrak{a}_\mathrm{d}|b_\mathrm{d}|}$, so by iteration the eigenvalues of $\mathcal{P}_{+1}$ are

$$\Lambda_n = \mathfrak{a}_\mathrm{d}^2 + b_\mathrm{d}^2 - \mathcal{N}_n^2 - \frac{\mathfrak{a}_\mathrm{d}^2 b_\mathrm{d}^2}{\mathcal{N}_n^2},$$

for every integer $n \geq 0$ with $\mathcal{N}_n > \sqrt{\mathfrak{a}_\mathrm{d}|b_\mathrm{d}|}$, and for both signs of s the levels are Λ_N, with $N = n + \frac{1}{2}(1-s)$ as in Eq. (146). The limits of ϱ_d at $x \to \pm\infty$ are $\varrho_{\mathrm{d}\pm} = \pm\mathfrak{a}_\mathrm{d} - b_\mathrm{d}$, and $dx = \frac{1}{2}L\big[1/(\varrho_{\mathrm{d}+} - \varrho_\mathrm{d}) + 1/(\varrho_\mathrm{d} - \varrho_{\mathrm{d}-})\big]\, d\varrho_\mathrm{d}$. With $\int_{-\sqrt{\Lambda}}^{\sqrt{\Lambda}} \sqrt{\Lambda - \varrho_\mathrm{d}^2}\, d\varrho_\mathrm{d}/|\varrho_{\mathrm{d}\pm} - \varrho_\mathrm{d}| = \pi\big(|\varrho_{\mathrm{d}\pm}| - \sqrt{\varrho_{\mathrm{d}\pm}^2 - \Lambda}\big)$,

$$\oint \xi\, dx = \pi L\Big[2\mathfrak{a}_\mathrm{d} - \sqrt{(\mathfrak{a}_\mathrm{d} - b_\mathrm{d})^2 - \Lambda} - \sqrt{(\mathfrak{a}_\mathrm{d} + b_\mathrm{d})^2 - \Lambda}\Big]$$

for $0 < \Lambda < \Lambda_c(b_\mathrm{d}) = \min\big((\mathfrak{a}_\mathrm{d} - b_\mathrm{d})^2, (\mathfrak{a}_\mathrm{d} + b_\mathrm{d})^2\big)$. Equation (143) with $\Theta_\mathrm{spin} = -2\pi$ and no remainder sets this to $2\pi\hbar N$, which is

$$\sqrt{(\mathfrak{a}_\mathrm{d} + b_\mathrm{d})^2 - \Lambda} + \sqrt{(\mathfrak{a}_\mathrm{d} - b_\mathrm{d})^2 - \Lambda} = 2\mathcal{N}_N.$$

The difference of the two squares under the roots is $4\mathfrak{a}_\mathrm{d} b_\mathrm{d}$, so $\sqrt{(\mathfrak{a}_\mathrm{d} + b_\mathrm{d})^2 - \Lambda} = \mathcal{N}_N + \mathfrak{a}_\mathrm{d} b_\mathrm{d}/\mathcal{N}_N$, and

$$\Lambda = \Lambda_N,$$

exactly, for every level: the remainder of Eq. (143) is zero in this nonuniform electric and magnetic field, as the lowest-order supersymmetric rule is exact for shape-invariant potentials [98]. The exact laboratory levels are $(\gamma_\mathrm{d} E_\mathrm{d})^2 = (mc^2)^2 + c^2\Lambda_N$, an equation for E_d, which is Eq. (146). □

The uniform and sech^2 profiles of (c) and (d) are exactly solvable checks of the rule.

6. Levels in a static gravitational field

Take the metric, the functions α_n, W, V, B and ϱ, the operator K with $g = 1/[W(\mathcal{E}+mc^2)]$, and $\Upsilon = \ln[W(\mathcal{E}+mc^2)]$, as in Section VI B. With $E = \hbar\omega$, $g = \alpha_n/[W(E - V + \alpha_n mc^2)]$. Let $\mathcal{U} \equiv W(E - V - \alpha_n mc^2)/\alpha_n$, so that $\mathsf{K} = c^2 Q_s^\dagger g Q_s - \mathcal{U}$.

Lemma 9 (Reduction in the static metric)**:** *Let α_n and W be smooth and bounded with bounded derivatives and bounded below by positive constants, let V and B be as in the hypotheses of Section B 4, and let J be a compact interval of energies on which $E - V + \alpha_n mc^2 \geq E_{gap} > 0$. Then $E \in J$ is a Dirac level of the block of fixed p_y and s if and only if 0 is an eigenvalue of $\mathsf{K}(E)$, and statements (i)–(iii) of Lemma 2 hold for this K, with the rate $\inf(W/\alpha_n)$ in place of 1 in (ii).*

Proof. The equivalence is the exact elimination of Section VI B. (i) g lies between two positive constants that depend only on E_gap, $\sup_J |E - V|$ and the bounds of α_n and W, so $\operatorname{dom} Q_s$ is again the common form domain, and $\mathcal{U}$ is a bounded function. (ii) $g(E) - g(E') \geq 0$ for $E < E'$, and $-\mathcal{U}$ decreases by $(W/\alpha_n)(E' - E) \geq \inf(W/\alpha_n)\,(E' - E)$. (iii) The equation again sets $\hbar^2(gF')'$ equal to F times a continuous function, unbounded where ϱ is, and the Wronskian argument of Lemma 2 applies unchanged, since it uses only this continuity and $F_1, F_2 \in \operatorname{dom} Q_s$. □

The symbol of K is $\mathsf{K}_0 + \hbar\,\mathsf{K}_1 + \hbar^2\mathsf{K}_2$ with $\mathsf{K}_0 = c^2 g\,(\xi^2 + \varrho^2) - \mathcal{U}$, $\mathsf{K}_1 = -s\,c^2\,(g\varrho)'$ and $\mathsf{K}_2 = \frac{1}{4}c^2 g''$. The level set $\mathsf{K}_0 = 0$ is $c^2(\xi^2 + \varrho^2) = W^2(\mathcal{E}^2 - (mc^2)^2)$, the trajectory of $H_0 = V + \alpha_n R$ at the energy E, on which $dx/dt_0 = 2c^2 g\,\xi$ and $dx/dt = \alpha_n c^2 \xi/(W^2\mathcal{E})$, so $dt = (2gW^2\mathcal{E}/\alpha_n)\, dt_0$. Since $g = e^{-\Upsilon}$, $(g\varrho)' = g\,(qB - \varrho\,\Upsilon')$, and with $\Omega_z = \alpha_n c^2(\varrho\,\Upsilon' - qB)/(W^2\mathcal{E})$,

$$\mathsf{K}_1\, dt_0 = \tfrac{1}{2}\, s\, \Omega_z\, dt,$$

as in flat spacetime. The spin angle has a closed form in the polar angles $\theta_\mathbf{n}$ and $\phi_\mathbf{n}$ of $\mathbf{n} = (c\xi/W, c\varrho/W, mc^2)/\mathcal{E}$, where $\cos\theta_\mathbf{n} = mc^2/\mathcal{E}$ and $\phi_\mathbf{n} = \arg(\xi + i\varrho)$, as for the flat-spacetime $\mathbf{n}$ of Section VI B.

Lemma 10 (Spin angle in the static metric)**:** *Let $m > 0$ and $\mathcal{E}' = -(V' + \mathcal{E}\,\alpha_n')/\alpha_n$. On every orbit of H_0 at the energy E, $\mathbf{n}$ is a unit vector; at each point of the orbit where $\xi^2 + \varrho^2 > 0$, the spin turns relative to the momentum at the rate*

$$\Omega_z + \dot\phi_{\mathbf{n}} = -\cos\theta_{\mathbf{n}}\,\frac{\alpha_n\,\varrho\,\mathcal{E}'}{\xi^2+\varrho^2},$$

which does not contain W', and

$$\Theta_{spin} = -\oint (1-\cos\theta_{\mathbf{n}})\,d\phi_{\mathbf{n}} + \oint \frac{mc^2\,\alpha_n c^2}{W^2\mathcal{E}^2}\Big(\frac{\varrho\,W'}{W} - qB\Big)\,dt,$$

where the first integral is the solid angle enclosed by the curve that $\mathbf{n}$ traces.

Proof. Hamilton's equations of $H_0 = V + \alpha_n\sqrt{(mc^2)^2 + c^2(\xi^2+\varrho^2)/W^2}$, with $\varrho' = qB$ and $V' + \alpha_n'\mathcal{E} = -\alpha_n\mathcal{E}'$, are

$$\dot x = \frac{\alpha_n c^2\,\xi}{W^2\mathcal{E}},$$

$$\dot\varrho = qB\,\dot x,$$

$$\dot\xi = \alpha_n\mathcal{E}' - \frac{\alpha_n c^2}{W^2\mathcal{E}}\Big(\varrho\,qB - (\xi^2+\varrho^2)\,\frac{W'}{W}\Big).$$

Then $(\xi^2+\varrho^2)\,\dot\phi_{\mathbf{n}} = \xi\dot\varrho - \varrho\dot\xi$ gives

$$\dot\phi_{\mathbf{n}} = \frac{\alpha_n c^2}{W^2\mathcal{E}}\Big(qB - \frac{\varrho\,W'}{W}\Big) - \frac{\alpha_n\,\varrho\,\mathcal{E}'}{\xi^2+\varrho^2}.$$

With $\Upsilon' = W'/W + \mathcal{E}'/(\mathcal{E}+mc^2)$ the expression for Ω_z is

$$\Omega_z = \frac{\alpha_n c^2}{W^2\mathcal{E}}\Big(\frac{\varrho\,W'}{W} - qB\Big) + \frac{\alpha_n c^2\,\varrho\,\mathcal{E}'}{W^2\mathcal{E}(\mathcal{E}+mc^2)}.$$

In the sum $\Omega_z + \dot\phi_{\mathbf{n}}$ the terms in qB and in W' cancel exactly. On the orbit $c^2(\xi^2+\varrho^2) = W^2(\mathcal{E}^2 - (mc^2)^2)$, so $|\mathbf{n}| = 1$, and

$$\frac{c^2}{W^2\mathcal{E}(\mathcal{E}+mc^2)} = \frac{1-\cos\theta_{\mathbf{n}}}{\xi^2+\varrho^2},$$

which gives the rate. Writing $\Omega_z = -(1-\cos\theta_{\mathbf{n}})\,\dot\phi_{\mathbf{n}} - \cos\theta_{\mathbf{n}}\,\dot\phi_{\mathbf{n}} + (\Omega_z + \dot\phi_{\mathbf{n}})$ and inserting $\dot\phi_{\mathbf{n}}$ gives $\Omega_z = -(1-\cos\theta_{\mathbf{n}})\,\dot\phi_{\mathbf{n}} + \cos\theta_{\mathbf{n}}\,\alpha_n c^2(\varrho W'/W - qB)/(W^2\mathcal{E})$, whose integral over one period is the closed form. The integrand $(1-\cos\theta_{\mathbf{n}})\,\dot\phi_{\mathbf{n}}$ is smooth on the whole plane, and $d[(1-\cos\theta_{\mathbf{n}})\,d\phi_{\mathbf{n}}] = d\phi_{\mathbf{n}}\wedge d\cos\theta_{\mathbf{n}}$ is the area form of the unit sphere. □

On an orbit on which $\xi^2+\varrho^2 > 0$, let $w_{\mathbf{n}}$ be the winding number of $\xi + i\varrho$ about 0 in one period, so that $\oint \dot\phi_{\mathbf{n}}\,dt = 2\pi w_{\mathbf{n}}$. Then

$$\Theta_{\text{spin}} = -2\pi w_{\mathbf{n}} - \oint \frac{\cos\theta_{\mathbf{n}}\,\alpha_n\varrho\,\mathcal{E}'}{\xi^2+\varrho^2}\,dt,$$

in which $\cos\theta_{\mathbf{n}} = mc^2/\mathcal{E}$ and $\mathcal{E}'$ contain no W, so W enters only through the orbit and its period. For α_n and V constant, $\mathcal{E}' = 0$, so $\Omega_z = -\dot\phi_{\mathbf{n}}$, and the point $\xi = \varrho = 0$ lies inside the orbit, so $\Theta_{\text{spin}} = -2\pi$ for every profile $W(x)$ and $B(x)$. The lapse enters as the electric potential does, with V' replaced by $\mathcal{E}\,\alpha_n'$, so it couples to the spin through the energy of the particle and not its rest energy.

Lemma 11 (The block in the static metric)**:** *Let $m > 0$ and let J be as in Lemma 9, with the level set $\mathsf{K}_0 = 0$ a closed curve with no critical point of K_0 for each $E \in J$ and 0 below the lower limit of K_0 at infinity. Then, with $\nu \equiv \alpha_n/W$, the block of fixed p_y and s is unitarily equivalent to a self-adjoint operator whose Weyl symbol is exactly $H_s = V\mathbf{1} + c\nu\xi\,\sigma_1 - sc\nu\varrho\,\sigma_2 + \alpha_n mc^2\sigma_3$, with upper eigenvalue $h_+ = H_0$; the set where $\mathsf{K}_0 < 0$ is the set where $H_0 < E$; and $\mathsf{K}(E)$ meets the hypotheses of Lemma 7 on J, with $\oint \mathsf{K}_1\,dt_0 = \frac{1}{2}s\,\Theta_{spin}$.*

Proof. The operator $V+\alpha_n mc^2\,\sigma_3+c\,\alpha_n\,\mathfrak{B}\,\sigma_1$ on the pair (u,d) is symmetric with the weight α_n^{-1}. Multiplying by $\alpha_n^{-1/2}$ gives off-diagonal entries $(c\nu)^{1/2}Q_s(c\nu)^{1/2}$ and its adjoint, where $c\nu$ is the coordinate speed of light; since $(c\nu)^{1/2}$ is bounded with bounded inverse and derivative, the resulting operator is self-adjoint on $\operatorname{dom}Q_s\oplus\operatorname{dom}Q_s^{\dagger}$. Since $(c\nu)^{1/2}\hat{p}\,(c\nu)^{1/2}$ has the Weyl symbol $c\nu\xi$, the block has the exact symbol H_s, the operator of Obukhov *et al.* [10] restricted to one block. With $\mathbf{P}\equiv(c\nu\xi,\,-sc\nu\varrho,\,\alpha_n mc^2)$, $h_{\pm}=V\pm|\mathbf{P}|$, $|\mathbf{P}|=\alpha_n R$ and

$$\mathsf{K}_0=\frac{gW^2}{\alpha_n^2}\,(h_+-E)(E-h_-),$$

with $E-h_-\geq E_{\mathrm{gap}}$ and $h_+=H_0$. The set $h_+=E$ is the curve $\mathsf{K}_0=0$, on which $\nabla\mathsf{K}_0=(gW^2/\alpha_n^2)(E-h_-)\nabla h_+$, and $\mathsf{K}_0<0$ exactly where $h_+<E$. For Lemma 7, g and $\mathcal{U}$ have x-derivatives continuous in (x,E) and bounded on $\mathbb{R}\times J$, since α_n, W and V have bounded derivatives and $E-V+\alpha_n mc^2\geq E_{\mathrm{gap}}$, and g lies between two positive constants by Lemma 9. Hypothesis (i) is assumed. Hypothesis (ii) follows as in flat spacetime, since $c^2g\varrho^2-\mathcal{U}$ decreases with E by Lemma 9(ii) and its lower limit at infinity at $E=\max J$ is that of K_0, which is positive. Hypothesis (iii) holds as in flat spacetime: with $w_g\equiv W/\alpha_n$, $g(E')-g(E)=-(E'-E)\,w_g\,g(E)g(E')$ and $\mathcal{U}(E')-\mathcal{U}(E)=(E'-E)\,w_g$, so

$$k_{\mathrm{S},\hbar}(E')-k_{\mathrm{S},\hbar}(E)=-(E'-E)\Big[\frac{w_g}{c^2}+w_g\,g(E)g(E')\varrho^2-s\hbar\big(w_g\,g(E)g(E')\varrho\big)'\Big],$$

and Lemma 8 makes the bracket at least $\inf w_g/(2c^2)$ at every x for $\hbar\leq\hbar_{\mathrm{mon}}$. The lemma is applied with $b_{\mathrm{mon}}=w_g/c^2$, $a_{\mathrm{mon}}=w_g\,g(E)g(E')$, $b_*=\inf w_g/c^2$, $a_*=\inf w_g\,(\inf g)^2$, and M_{mon} equal to the bound of $|\partial_x a_{\mathrm{mon}}|+a_{\mathrm{mon}}\,qB$, which is finite since α_n, W, V and qB are bounded with bounded derivatives and g lies between two positive constants. The first-order term is $\oint\mathsf{K}_1\,dt_0=\frac{1}{2}s\,\Theta_{\mathrm{spin}}$ by the relation $\mathsf{K}_1\,dt_0=\frac{1}{2}s\,\Omega_z\,dt$ in this metric. □

The 2×2 rule of Becker *et al.* [26] agrees. The vector $(|\mathbf{P}|+\alpha_n mc^2,\ c\nu(\xi-is\varrho))$ is an eigenvector of H_s for h_+, nonzero for $m>0$, and dividing it by its norm gives a unit eigenvector $\mathbf{e}_+$; the terms with $\{h_+,c\nu\}$ cancel, and on the orbit the Berry integrand is

$$\tfrac{1}{i}\big\langle\{h_+,\mathbf{e}_+\},\mathbf{e}_+\big\rangle=-\frac{sc^2\alpha_n}{2W^2\mathcal{E}}\Big(\frac{qB\,(\mathcal{E}-mc^2)}{\mathcal{E}}-\frac{\varrho\,\mathcal{E}'}{\mathcal{E}+mc^2}-\frac{\varrho\,W'\,(\mathcal{E}-mc^2)}{W\mathcal{E}}\Big).$$

From $\mathbf{P}\cdot(\partial_\xi\mathbf{P}\times\partial_x\mathbf{P})=smc^4(\alpha_n^3/W^2)(\varrho W'/W-qB)$, the Rammal–Wilkinson integrand is

$$\tfrac{1}{2i}\big\langle\{H_s-h_+,\mathbf{e}_+\},\mathbf{e}_+\big\rangle=\frac{sc^2\alpha_n}{2W^2\mathcal{E}}\,\frac{mc^2}{\mathcal{E}}\Big(\frac{\varrho\,W'}{W}-qB\Big).$$

The terms in $mc^2/\mathcal{E}$ cancel in the sum, which is $\frac{1}{2}s\,\Omega_z$, and $h_+=H_0$, so the two phases sum to $\frac{1}{2}s\,\Theta_{\mathrm{spin}}$, the first-order term of Lemma 7 for this block.

Theorem 4 (Levels in a static gravitational field)**:** *In the metric of Eq. (125) let $\alpha_n(x)$ and $W(x)$ be smooth and bounded with bounded derivatives and bounded below by positive constants, and let V, B, m, p_z, p_y, s and ϱ be as in the hypotheses of Section B 4. Let J be a compact interval of energies on which $\hbar\omega-V+\alpha_n mc^2\geq E_{gap}>0$, the level set $\mathsf{K}_0=0$ is a closed curve with no critical point of K_0, and 0 lies below the lower limit of K_0 at infinity. Then for $\hbar$ small the Dirac levels in J of the block of fixed p_y and s are simple within the block, and they are a one-to-one $O(\hbar^2)$ spectral approximation (Definition 1) of the solutions E_n^0 of Eq. (143) without its remainder, where $H_0=V+\alpha_n R$ and Θ_{spin} is the spin angle of Eq. (136). On every orbit Θ_{spin} is minus the solid angle enclosed by the curve that $\mathbf{n}$ traces plus an orbit integral of first order in mc^2, and $\Theta_{spin}=-2\pi$ whenever α_n and V are constant, for every profile $W(x)$ and $B(x)$.*

Proof. By Lemma 9(iii) and the bijection $F\mapsto(W^{1/2}F,\,cW^{1/2}g\,Q_sF)$ from $\ker\mathsf{K}(E)$ onto the eigenspace at E, the levels are simple within the block. This bijection is proved as Proposition 3 is, with the domain of the block of Lemma 11, once 0 is shown to lie below the essential spectrum of $\mathsf{K}(E)$. That step is Lemma 3 with this $\mathcal{U}$: $\Upsilon'=W'/W+\mathcal{E}'/(\mathcal{E}+mc^2)$ is bounded, with $\mathcal{E}+mc^2\geq E_{\mathrm{gap}}/\sup\alpha_n$; $c^2g\varrho^2-\mathcal{U}$ decreases with E; and its lower limit at infinity is that of K_0, positive at $E=\max J$.

The Weyl-sequence argument of the proof of Theorem 2 then shows that no essential spectrum of the block lies in J. It applies to the self-adjoint form of Lemma 11, whose norm is equivalent to the unweighted one since α_n is bounded above and below, with $W^{-1/2}\mathsf{u}_k$ and $W^{-1/2}\mathsf{d}_k$ in place of u_k and d_k and with the g and $\mathcal{U}$ of this subsection, because the two relations of the elimination, divided by α_n and multiplied by $W^{1/2}$, have the form of the flat ones with residuals that still tend to 0.

For the bound in both directions, the hypotheses of Lemmas 9 and 7 hold by Lemma 11, so the levels in J are the set $Z_\hbar$ of Lemma 1, with $h=H_0$, $\varphi_\hbar=\frac{1}{2}s\,\Theta_{\mathrm{spin}}$, $S_\hbar=Z_\hbar$ and $\varepsilon_\hbar=0$. That lemma gives the statement, with the constants listed for the flat block, and the one-to-one statement follows from hypothesis (iii) as in the proof of Theorem 2. The last statement is Lemma 10 with $\mathcal{E}'=0$, for which $\Omega_z=-\dot{\phi}_{\mathbf{n}}$ and the orbit encloses $\xi=\varrho=0$. □

7. What is proved, and the interpretation

Theorem 4 is a statement about a Dirac operator whose lapse and scale factor depend on one Cartesian coordinate and lie between two positive constants; it uses nothing else about the metric. The Schwarzschild metric, whose lapse and scale factor depend on r, and a uniform gravitational field, whose lapse is unbounded, are outside this class. As noted in Section VI, the theorem applies to a weak static gravitational field, where W gives γ_{PPN} and α_n gives $\frac{1}{2}$, as in Eq. (132). It also applies, as noted there, to a gapped two-dimensional Dirac material with position-dependent Fermi velocity $v_{\mathrm{F}}(x) = c/W(x)$ and $\alpha_n = 1$ [103]; there $c\nu = v_{\mathrm{F}}$, and $v_{\mathrm{F}}^{1/2}\hat{p}\,v_{\mathrm{F}}^{1/2} = \frac{1}{2}(v_{\mathrm{F}}\hat{p} + \hat{p}v_{\mathrm{F}})$ maps the block of Lemma 11 to the operator $V\mathbf{1} + \frac{1}{2}\{v_{\mathrm{F}}, \hat{p}\}\sigma_1 - sv_{\mathrm{F}}\varrho\sigma_2 + \Delta_{\mathrm{gap}}\sigma_3$ with $\Delta_{\mathrm{gap}} = mc^2 > 0$. The theorem needs $\Delta_{\mathrm{gap}} > 0$, since $m > 0$ is used in Lemma 11; the massless case and the Dirac point are excluded.

8. What the proofs use

TABLE V. The results used by the proof of each level theorem.

Theorem	Reduction	Rule for $\mathsf{K}(E)$	Energy	Essential spectrum	First-order term
1	$\mathsf{K}(\Lambda) = c^2(\mathcal{P}_s - \Lambda)$	Lemma 7	Lemma 1	threshold Λ_c	$\oint qB\,dt_{\mathfrak{p}} = \pi$
2	Lemma 2, Proposition 3	Lemma 7	Lemma 1	Lemma 3, Weyl's criterion	$\mathsf{K}_1\,dt_0 = \frac{1}{2}s\,\Omega_z\,dt$
3(b)	boost to the drift frame, b_{d} depending on E_{d}	Lemma 7 with the parameter E_{d}, without (iii)	Lemma 1	threshold $\Lambda_c(b_{\mathrm{d}})$	(a): $\Theta_{\mathrm{spin}} = -2\pi$
3(c), (d)	boost to the drift frame	exact levels, no rule	none	none	(a)
4	Lemmas 9 and 11	Lemma 7	Lemma 1	Lemma 3, as for Theorem 2	Lemma 10

The proofs listed in Table V use Lemma 7. That lemma uses Proposition 2 and Lemmas 4, 5 and 6. The only external input to Proposition 2 is the published theorem stated as Proposition 1. Lemma 8 verifies hypothesis (iii) of Lemma 7 for the electric potential and for the static metric.

No theorem uses the preprint of Becker *et al.* [26], but its Theorems 1.1 and 1.2 apply to these blocks. They concern the Weyl quantization of a Hermitian symbol $h_0\mathbf{1} + \mathbf{h}\cdot\boldsymbol{\sigma}$ in a class $S^0(\mu_{\mathrm{o}}^{1/2})$, with eigenvalues $h_\pm = h_0 \pm |\mathbf{h}|$, and a compact interval J of energies with four properties: (H1) each set $h_+ = E$, $E \in J$, is a simple closed curve on which $\nabla h_+ \neq 0$; (H2) $\mathbf{h} \neq 0$ near these curves; (H3) $h_- < \min J$ on the region R_J swept by the curves; and (H4) the symbol minus E is elliptic outside every neighborhood of $\bar{R}_J$, uniformly in $E \in J$. They prove the rule for an operator modified outside a neighborhood of $\bar{R}_J$ (their Theorem 2.6), as a Hausdorff distance $O(\hbar^\infty)$ between its spectrum in J and the solutions of $\mathcal{I}_\hbar(E) \in 2\pi\hbar\mathbb{Z}$, and return to the original operator by their Proposition 2.3. The block symbol H_s of Lemma 11, which is the flat block for $\alpha_n = W = 1$, meets these hypotheses on every J of Theorems 2 and 4, with $\mu_{\mathrm{o}} = 1 + (\xi^2 + \varrho^2)/(mc)^2$ and $\mathbf{h} = \mathbf{P}$. With $c_{\mathrm{h}} \equiv mc^2 \min(\inf\nu, \inf\alpha_n)$ and $c^{\mathrm{h}} \equiv mc^2 \max(\sup\nu, \sup\alpha_n)$, $c_{\mathrm{h}}^2\,\mu_{\mathrm{o}} \le |\mathbf{P}|^2 \le (c^{\mathrm{h}})^2\,\mu_{\mathrm{o}}$, and every derivative of an entry of H_s is a bounded function times ξ, ϱ or 1, so $H_s \in S^0(\mu_{\mathrm{o}}^{1/2})$. (H1) holds because $h_+ = E$ is the curve $\mathsf{K}_0 = 0$, on which $\nabla\mathsf{K}_0$ is a positive multiple of ∇h_+. (H2) holds because the third component of $\mathbf{P}$ is $\alpha_n mc^2 > 0$. (H3) holds on the whole plane, since $h_- \le V - \alpha_n mc^2 \le \min J - E_{\mathrm{gap}}$. For (H4), $\|(H_s - E)^{-1}\| = 1/\min(|h_+ - E|, |h_- - E|)$. With $M_J \equiv \max_{E\in J}\sup_x |E - V|$, where $|\mathbf{P}| \ge 2M_J$ both $|h_\pm - E| \ge |\mathbf{P}|/2 \ge \frac{1}{2}c_{\mathrm{h}}\,\mu_{\mathrm{o}}^{1/2}$. Where $|\mathbf{P}| < 2M_J$, $\mu_{\mathrm{o}} < 4M_J^2/c_{\mathrm{h}}^2$, so ξ is bounded, and $E - h_- \ge E_{\mathrm{gap}}$. There $h_+ - E$ is bounded below by a positive constant for large $|x|$ because 0 lies below the lower limit of K_0 at infinity, and on the remaining compact set outside a neighborhood of $\bar{R}_J$ the continuous function $|h_+ - E|$ has a positive least value over $E \in J$.

Their theorems therefore apply, with the same first-order term (the remark after Lemma 11, whose flat case is $\alpha_n = W = 1$). Combined with Lemma 1, they give the two-way part of Theorems 2 and 4, which are in this sense applications of their rule with an independent proof. A Hausdorff distance does not count eigenvalues, so the one-to-one part is not contained in their statement; whether their reduction to a scalar operator near the curve gives it is not examined here. This paper proves it by the scalar argument (hypothesis (iii) of Lemma 7 and Lemmas 5 and 6). The direct proof also gives the constants explicitly, as least or largest values of stated functions over compact sets,

and evaluates the first-order term as the orbit integral $\frac{1}{2}s\,\Theta_{\mathrm{spin}}$. The two-way part of Theorem 1 also follows from the published theorem of Ifa *et al.* [94].

---

[1] M. Planck, Annalen der Physik **309**, 553 (1901).
[2] A. Einstein, Annalen der Physik **322**, 132 (1905).
[3] L. A. D. Hutchison, arXiv preprint (2026), companion paper.
[4] P. A. M. Dirac, Proceedings of the Royal Society A **117**, 610 (1928).
[5] L. L. Foldy and S. A. Wouthuysen, Physical Review **78**, 29 (1950).
[6] E. Eriksen, Physical Review **111**, 1011 (1958).
[7] F. W. Hehl and W.-T. Ni, Physical Review D **42**, 2045 (1990).
[8] A. J. Silenko, Physical Review A **77**, 012116 (2008), arXiv:0710.4218.
[9] Y. N. Obukhov, A. J. Silenko, and O. V. Teryaev, Phys. Rev. D **80**, 064044 (2009), arXiv:0907.4367 [gr-qc].
[10] Y. N. Obukhov, A. J. Silenko, and O. V. Teryaev, Phys. Rev. D **84**, 024025 (2011), arXiv:1106.0173 [hep-th].
[11] Y. N. Obukhov, A. J. Silenko, and O. V. Teryaev, Phys. Rev. D **88**, 084014 (2013), arXiv:1308.4552 [gr-qc].
[12] J.-L. Heully, I. Lindgren, E. Lindroth, S. Lundqvist, and A.-M. Mårtensson-Pendrill, Journal of Physics B: Atomic and Molecular Physics **19**, 2799 (1986).
[13] W. Kutzelnigg and W. Liu, The Journal of Chemical Physics **123**, 241102 (2005).
[14] C. Lanczos, Z. Phys. **57**, 447 (1929), arXiv:physics/0508002 [physics.hist-ph].
[15] D. Hestenes, Journal of Mathematical Physics **8**, 798 (1967).
[16] S. De Leo and P. Rotelli, Prog. Theor. Phys. **92**, 917 (1994), arXiv:hep-th/9401009 [hep-th].
[17] S. De Leo and W. A. Rodrigues, Jr., Int. J. Theor. Phys. **37**, 1511 (1998), arXiv:hep-th/9806057 [hep-th].
[18] A. Lasenby, C. Doran, and S. Gull, Phil. Trans. Roy. Soc. Lond. A **356**, 487 (1998), arXiv:gr-qc/0405033 [gr-qc].
[19] R. P. Feynman and M. Gell-Mann, Physical Review **109**, 193 (1958).
[20] S. I. Rubinow and J. B. Keller, Physical Review **131**, 2789 (1963).
[21] J. Bolte and S. Keppeler, Annals of Physics **274**, 125 (1999), arXiv:quant-ph/9811025.
[22] G. Panati, H. Spohn, and S. Teufel, Advances in Theoretical and Mathematical Physics **7**, 145 (2003), arXiv:math-ph/0201055.
[23] S. Keppeler, Annals of Physics **304**, 40 (2003), arXiv:quant-ph/0212082.
[24] L. Silberstein, Phil. Mag. 6, **23**, 790 (1912).
[25] I. Bialynicki-Birula, in *Progress in Optics*, Vol. 36, edited by E. Wolf (Elsevier, Amsterdam, 1996) pp. 245–294, arXiv:quant-ph/0508202 [quant-ph].
[26] S. Becker, S. Fujiié, and J. Wittsten, Bohr–Sommerfeld rules for systems, arXiv:2508.21013v2 (2025), version 1 of 28 August 2025; version 2 of 27 April 2026 is the version cited, for Theorems 1.1 and 1.2 and Lemma 3.1, arXiv:2508.21013 [math-ph].
[27] A. Einstein, Annalen der Physik **323**, 639 (1905).
[28] L. de Broglie, Annales de Physique **3**, 22 (1925), phD thesis, 1924; 10th series, tome III.
[29] A. Einstein, Jahrbuch der Radioaktivität und Elektronik **4**, 411 (1907).
[30] I. Newton, *Philosophiæ Naturalis Principia Mathematica* (Royal Society, London, 1687).
[31] W. R. Hamilton, Philosophical Transactions of the Royal Society of London **125**, 95 (1835).
[32] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics*, 3rd ed. (Addison-Wesley, 2002).
[33] P. Ehrenfest, Zeitschrift für Physik **45**, 455 (1927).
[34] E. Madelung, Zeitschrift für Physik **40**, 322 (1927).
[35] D. Bohm, Physical Review **85**, 166 (1952).
[36] H. A. Lorentz, *Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern* (E. J. Brill, Leiden, 1895).
[37] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999).
[38] C.-A. Coulomb, Mémoires de l'Académie Royale des Sciences , 569 (1785), année 1785, published 1788.
[39] W. Ehrenberg and R. E. Siday, Proceedings of the Physical Society. Section B **62**, 8 (1949).
[40] Y. Aharonov and D. Bohm, Physical Review **115**, 485 (1959).
[41] J. D. Walecka, Annals of Physics **83**, 491 (1974).
[42] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*, Monographies de physique (Les Éditions de Physique, Les Ulis, 1988).
[43] H. Weyl, Zeitschrift für Physik **56**, 330 (1929).
[44] C.-N. Yang and R. L. Mills, Physical Review **96**, 191 (1954).
[45] W. R. Hamilton, Philosophical Magazine **25**, 10 (1844).
[46] W. Pauli, Zeitschrift für Physik **43**, 601 (1927).
[47] R. Arnowitt, S. Deser, and C. W. Misner, General Relativity and Gravitation **40**, 1997 (2008), originally in L. Witten (ed.), *Gravitation: An Introduction to Current Research* (Wiley, New York, 1962), chapter 7, pp. 227–265, arXiv:gr-qc/0405109 [gr-qc].
[48] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).

[49] T. W. Baumgarte and S. L. Shapiro, Physics Reports **376**, 41 (2003), arXiv:gr-qc/0211028 [gr-qc].
[50] C. M. Will, Living Reviews in Relativity **17**, 4 (2014).
[51] J. Lense and H. Thirring, Physikalische Zeitschrift **19**, 156 (1918).
[52] P. Painlevé, Comptes rendus hebdomadaires des séances de l'Académie des sciences **173**, 677 (1921), séance du 24 octobre 1921.
[53] K. Martel and E. Poisson, Am. J. Phys. **69**, 476 (2001), arXiv:gr-qc/0001069 [gr-qc].
[54] Y. Kosmann, Annali di Matematica Pura ed Applicata **91**, 317 (1971).
[55] I. Y. Kobzarev and L. B. Okun', Soviet Physics JETP **16**, 1343 (1963).
[56] O. Klein, Zeitschrift für Physik **37**, 895 (1926).
[57] W. Gordon, Zeitschrift für Physik **40**, 117 (1926).
[58] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading, MA, 1995).
[59] W. Pauli, Ann. Inst. Henri Poincaré **6**, 109 (1936).
[60] E. Majorana, Il Nuovo Cimento **14**, 171 (1937), english translation by L. Maiani, Soryushiron Kenkyu 63 (1981) 149, revised in *Ettore Majorana Scientific Papers*, Paper no. 9, pp. 218–233.
[61] H. Arodź and Z. Świerczyński, J. Phys. G **48**, 065001 (2021), arXiv:1910.13920 [quant-ph].
[62] L. Silberstein, Ann. Phys. **327**, 579 (1907).
[63] J. Schwinger, Physical Review **73**, 416 (1948).
[64] L. L. Foldy, Physical Review **87**, 688 (1952).
[65] Y. Aharonov and A. Casher, Physical Review Letters **53**, 319 (1984).
[66] A. Cimmino, G. I. Opat, A. G. Klein, H. Kaiser, S. A. Werner, M. Arif, and R. Clothier, Physical Review Letters **63**, 380 (1989).
[67] T. L. Curtright and C. K. Zachos, Asia Pacific Physics Newsletter **1**, 37 (2012), arXiv:1104.5269.
[68] A. Einstein, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften , 844 (1915).
[69] B. D. Serot and J. D. Walecka, in *Advances in Nuclear Physics*, Vol. 16, edited by J. W. Negele and E. Vogt (Plenum Press, New York, 1986) pp. 1–327.
[70] L. Wolfenstein, Phys. Rev. D **17**, 2369 (1978).
[71] R. Grimm, M. Weidemüller, and Y. B. Ovchinnikov, Adv. At. Mol. Opt. Phys. **42**, 95 (2000).
[72] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, 3rd ed. (Cambridge University Press, 2020).
[73] M. Zhu, Y. Liu, C. Wei, H. Ni, and Q. Wei, J. Chem. Phys. **158**, 164306 (2023).
[74] D. J. BenDaniel and C. B. Duke, Physical Review **152**, 683 (1966).
[75] W. Gerlach and O. Stern, Zeitschrift für Physik **9**, 349 (1922).
[76] B. Mashhoon, Physical Review Letters **61**, 2639 (1988).
[77] C. W. F. Everitt, D. B. DeBra, B. W. Parkinson, *et al.*, Physical Review Letters **106**, 221101 (2011).
[78] B. Demirel, S. Sponar, and Y. Hasegawa, New Journal of Physics **17**, 023065 (2015).
[79] A. Danner, B. Demirel, W. Kersten, H. Lemmel, R. Wagner, S. Sponar, and Y. Hasegawa, npj Quantum Information **6**, 23 (2020).
[80] S. J. Barnett, Physical Review **6**, 239 (1915).
[81] M. Matsuo, J. Ieda, K. Harii, E. Saitoh, and S. Maekawa, Physical Review B **87**, 180402 (2013).
[82] M. V. Berry, Proceedings of the Royal Society A **392**, 45 (1984).
[83] G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999).
[84] C. G. Darwin, Proceedings of the Royal Society of London A **118**, 654 (1928).
[85] L. H. Thomas, Nature **117**, 514 (1926).
[86] J. Sinova, S. O. Valenzuela, J. Wunderlich, C. H. Back, and T. Jungwirth, Reviews of Modern Physics **87**, 1213 (2015).
[87] W. de Sitter, Monthly Notices of the Royal Astronomical Society **77**, 155 (1916).
[88] A. D. Fokker, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen **23**, 729 (1920).
[89] H.-P. Duerr, Physical Review **103**, 469 (1956).
[90] A. J. Silenko and O. V. Teryaev, Phys. Rev. D **71**, 064016 (2005), arXiv:gr-qc/0407015 [gr-qc].
[91] M. A. Oancea and A. Kumar, Physical Review D **107**, 044029 (2023), arXiv:2212.04414.
[92] P. A. Wolff, J. Phys. Chem. Solids **25**, 1057 (1964).
[93] Y. Fuseya, M. Ogata, and H. Fukuyama, J. Phys. Soc. Jpn. **84**, 012001 (2015), arXiv:1407.2179 [cond-mat.mes-hall].
[94] A. Ifa, H. Louati, and M. Rouleux, Journal of Mathematical Sciences, the University of Tokyo **25**, 91 (2018), arXiv:1702.06758 [math-ph].
[95] A. Ifa, H. Louati, and M. Rouleux, Journal of Mathematical Sciences, the University of Tokyo **27**, 81 (2020).
[96] O. Sutton and A. B. Watson, WKB spectral asymptotics for a one-dimensional Dirac operator with a slowly varying mass profile, arXiv:2605.27592 (2026), arXiv:2605.27592 [math-ph].
[97] A. Kormányos, P. Rakyta, L. Oroszlány, and J. Cserti, Physical Review B **78**, 045430 (2008), arXiv:0805.2527 [cond-mat.mes-hall].
[98] F. Cooper, A. Khare, and U. Sukhatme, Physics Reports **251**, 267 (1995), arXiv:hep-th/9405029.
[99] P. Carmier and D. Ullmo, Physical Review B **77**, 245413 (2008), arXiv:0801.4727 [cond-mat.mes-hall].
[100] J. N. Fuchs, F. Piéchon, M. O. Goerbig, and G. Montambaux, The European Physical Journal B **77**, 351 (2010), arXiv:1006.5632 [cond-mat.mes-hall].
[101] V. Lukose, R. Shankar, and G. Baskaran, Physical Review Letters **98**, 116802 (2007), arXiv:cond-mat/0603594.
[102] D. O-Campa, O. Pedraza, L. A. López, and E. Díaz-Bautista, The European Physical Journal Plus **140**, 1246 (2025), arXiv:2509.00401.

[103] F. de Juan, M. Sturla, and M. A. H. Vozmediano, Physical Review Letters **108**, 227205 (2012), arXiv:1201.2656.
[104] D. Hestenes, Foundations of Physics **20**, 1213 (1990).
[105] B. Thaller, *The Dirac Equation*, Texts and Monographs in Physics (Springer-Verlag, Berlin, Heidelberg, 1992).
[106] A. Gsponer and J.-P. Hurni, The physical heritage of sir W.R. hamilton (2002), arXiv:math-ph/0201058 [math-ph].
[107] A. Garat and R. H. Price, Phys. Rev. D **61**, 124011 (2000), arXiv:gr-qc/0002013 [gr-qc].
[108] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, Physics Letters B **47**, 365 (1973).
[109] S. K. Wong, Il Nuovo Cimento A **65**, 689 (1970).
[110] F. T. Avignone, III, S. R. Elliott, and J. Engel, Rev. Mod. Phys. **80**, 481 (2008), arXiv:0708.1033 [nucl-ex].
[111] B. Cork, G. R. Lambertson, O. Piccioni, and W. A. Wenzel, Phys. Rev. **104**, 1193 (1956).
[112] R. Mosseri and R. Dandoloff, J. Phys. A: Math. Gen. **34**, 10243 (2001), arXiv:quant-ph/0108137 [quant-ph].
[113] W. E. Lamb and R. C. Retherford, Physical Review **72**, 241 (1947).
[114] F. W. J. Olver, *Asymptotics and Special Functions* (Academic Press, New York, 1974) reprinted by A K Peters, Wellesley, MA, 1997.