M1 Cosmology - 6 - Scalar Fields as the Dark Matter

Nick Kaiser

Contents

List of Figures

1 Introduction

In the last lecture we studied the role of scalar fields in driving inflation; either in the early or late universe. In this lecture we will consider another application of scalar fields; the idea that the dark matter (DM) we see may be a scalar field. Most physicists, if asked to guess what the DM consists of, would probably opt for some kind of fermionic 'weakly interacting massive particle' (WIMP), and would perhaps cite the 'WIMP miracle' as motivation. Candidates abound, and most DM detection experiments are targeted at this kind of DM. But such searches have, thus far, proved fruitless. An alternative is that the DM is some kind of weakly interacting scalar matter, and there is one well motivated candidate, which is the Peccei-Quinn axion.

The mass of the axion is not precisely predicted, but is thought to be, in very rough order of magnitude, to be around $mc^2 \sim 10^{-6}$eV. If that is correct, then, in order to be the DM we see in galaxies and in clusters, it must have a extremely large occupation numbers, which leads on to suspect that it may be well described as a classical field.

The argument for the high occupation number – originated by Scott Tremaine and Jim Gunn – goes as follows: Masses of such structures can be estimated from the velocity dispersion of luminous 'tracers': $\sigma_v^2 \sim GM/R$, so $M \sim \sigma_v^2 R/G$ and the number of DM particles of mass m must be $N_\mathrm{p} \sim \sigma_v^2 R/Gm$. The volume of momentum space is $\sim |\mathbf{p}|^3 \sim m^3\sigma_v^3$, so the volume of phase-space is $\sim R^3|\mathbf{p}|^3 \sim m^3 R^3 \sigma_v^3$. The density of states is, roughly speaking, one per phase-space volume $\hbar^3$, so it follows that the total number of states that have momentum that can be confined within these potential wells is $N_\mathrm{s} \sim m^3 R^3 \sigma_v^3/\hbar^3$, and that the typical occupation number must be $n \simeq N_\mathrm{p}/N_\mathrm{s} \simeq \hbar^3/GR^2\sigma_v m^4$. For particles with a thermal distribution, the typical occupation number is $n \sim 1$. It is possible that 'phase-mixing' could reduce the 'coarse grained' occupation number, but if originally thermal, the occupation number can't substantially exceed unity. Using the observed properties of the densest parts of clusters of galaxies, Tremaine and Gunn came up with a lower bound on the mass for any fermionic DM candidate which was $mc^2 \sim 30$eV[1]. Scalar fields have excitations which are bosonic and so can avoid this bound. But the TG argument says that if the actual mass is $\sim 10^{-6}$eV, the occupation numbers – scaling as $1/m^4$ – must be astronomically huge.

Having a high occupation number doesn't automatically imply that the field be well described by a coherent state (and therefore being essentially classical)[2]. However, the axion is usually considered to have been 'born' in such a state, and its interactions are so weak that it would have remained so. More generally, with strong enough interactions, a field would thermalise and either occupy states with large p – and so not be confinable within astronomical structures – or have a thermal (Bose-Einstein) component with low p (and negligible density) plus a Bose condensate of zero momentum. And the latter, in a potential well, would behave classically.

This leads to the interesting question of how it can be, if the DM is really behaving like a classical scalar *field*, that it behaves 'just like particles' in being confined by gravity. The answer is that there is a duality between classical wave-mechanics and particle dynamics; if the wavelength – and it is the de Broglie wavelength that is relevant here – is very small compared to the size of the system, then waves of the field behave indistinguishably from particles. This follows from the fact that the KG equation for a relativistic scalar field becomes, in the limit appropriate for Newtonian potential wells, the Schrödinger equation, and the arguments leading to the quantum correspondence principle apply equally to classical fields. The other question – not relevant for a $\sim 1^{-6}$eV axion, but possibly relevant for a field of much lower mass – is whether wave-mechanical effects could lead to *different* predictions from particle DM. These are the questions we address below.

We start, in §2, by considering the behaviour of *spatially homogeneous* expanding scalar radiation fields; analogous to the nearly homogeneous and expanding EM radiation that is the CMB, but with mass. In §2.1 we recall the form of the Klein-Gordon (KG) equation in FLRW coordinates and show, in §2.2, how this admits solutions which are damped standing waves whose energy density may redshift. Then, §2.3 we show how this behaviour can be understood from the energy continuity, and, in §2.4 we show how these results are in direct correspondence with the behaviour of a gas of particles (either non- or highly-relativisitic) in a state of expansion.

In §3 we explore how it may be that the DM seen in galaxies and clusters may be waves of a scalar

[1]This is well below the mass of most WIMP candidates, but at the time there were some claims for neutrino masses of this size. Measurements of mixing of neutrinos suggests that there masses are much lower than this, in which case they can't be the DM

[2]A counter example is the Raleigh-Jeans region of the black body spectrum, for which states, the density matrix is diagonal and the states are incoherent, with vanishing expectation value.

field. We start by reviewing how such structures can be described by weak-field gravity, and describing the candidate scalar fields that have been considered. In §3.1 we obtain the KG equation in the weak-field coordinate system; we obtain the local dispersion relation in §3.2 from which we obtain the phase- and group-velocities in §3.3. We show in §3.4 how scalar wave packets can be visualised in space-time and discuss in §3.5 the close analogy between scalar waves in gravitational potential wells and EM waves in a cold plasma and the refractive effects that are common to the two different systems. In §3.6 we develop the mathematics behind this analogy; we formulate the KG equation in terms of the log-amplitude and phase in §3.6.1 and then use this to demonstrate, in §3.6.2, how, in the geometric optics limit, the wave vector obeys the geodesic equation and, in §3.6.3, how the amplitude varies as one would expect if one has a diverging or converging beam or packet. In §3.7 we summarise the above findings which all illustrate a close correspondence between the behaviour of classical scalar fields and particles. In §3.8 we show how one can understand the way in which an initially uniform field can 'fall' into a potential well and how, in what for particles would be the multi-stream region, interference takes place, resulting in a 'speckly' energy density field. In §3.9 we describe how scalar matter waves can be modelled using the Schrödinger equation. In §3.9.1 we show how the KG equation for a real field can be replaced by the Schrödinger equation for a complex field and, in §3.9.2, how a 5th conservation law emerges; in addition to conservation of energy and the three components of momentum, there is conservation of total density (or number of particles). As described in §3.9.3 the Schrödinger approach sheds light on the speckly nature of scalar field DM and the phase-vortices that accompany the speckles. Finally, in §3.9.4, we describe how, with yet another transformation, the Schrödinger equation can be replaced by the Madelung fluid equations, providing yet another way to model scalar matter.

2 Homogeneous Expanding scalar radiation fields

We will now study the evolution of a expanding scalar waves, such as might exist in a homogeneous expanding universe. We consider in §2.1 a spatially flat FLRW model and obtain the KG equation in $\vec{x} \longrightarrow (\tau, x, y, z)$ coordinates where $\mathbf{r} \longrightarrow (x, y, z)$ are dimensionless comoving Cartesian coordinates. Then, in §2.2, we show that these admit solutions that are waves in comoving coordinates: $\phi = \phi(\tau)\cos(\mathbf{k}\cdot\mathbf{r} - \omega_{\mathbf{k}}\tau + \psi)$ with a time varying amplitude. We show that for waves whose frequency is much greater than the expansion rate, the energy density varies as $\mathcal{E} \propto a^{-3}$ if the physical wavelength $2\pi a/k$ is much larger than the Compton wavelength – i.e. just like non relativistic massive particles (which have $\lambda_{\rm dB} = (c/v)\lambda_{\rm C} \gg \lambda_{\rm C}$) – whereas in the opposite limit $2\pi a/k \ll \lambda_{\rm C}$ the density varies as $\mathcal{E} \propto a^{-4}$; i.e. just like highly relativistic particles.

In §2.3 we will show that the same behaviour follows from the energy continuity equation when applied to a field that is a 'sea' of statistically isotropic random waves. This exercise will introduce some tools that are very useful for dealing with the kind of statistically homogeneous random field that are used broadly in cosmology. In §2.4 we develop the correspondence with an expanding gas of particles or photons and show how the power-spectrum of the waves $P_\phi(\mathbf{k})$ corresponds to $f(\mathbf{p})/E(\mathbf{p})$ where $f(\mathbf{p})$ is the phase-space density for particles.

2.1 The Klein-Gordon equation in expanding coordinates

The line element in spatially flat FLRW (τ, x, y, z) coordinates is

$$ds^2 = -c^2 d\tau^2 + a^2(\tau)(dx^2 + dy^2 + dz^2). \tag{1}$$

One can obtain the equation of motion in these coordinates most simply by noting that, in locally inertial – and therefore non-expanding – coordinates $(\tau, \mathbf{r}) = (\tau, a(\tau)\mathbf{x})$, an element of the action is $dS = d\tau d^3r\mathcal{L} = d\tau d^3r\frac{1}{2}(\dot{\phi}^2/c^2 - |\nabla_{\mathbf{r}}\phi|^2 - V(\phi))$. With $d^3r = a^3 d^3x$ and $\nabla_{\mathbf{r}}\phi = \nabla\phi/a(\tau)$, where $\nabla \longrightarrow (\partial_x, \partial_y, \partial_x)$, this is $dS = d\tau d^3x\mathcal{L}'$, identical in form to the original action element, but with an effective Lagrangian density

$$\mathcal{L}'(\phi, \dot{\phi}, \nabla\phi, \tau) = \tfrac{1}{2}a^3(\dot{\phi}^2/c^2 - |\nabla\phi|^2/a^2 - V(\phi)) \tag{2}$$

that has acquired an explicit time dependence through the scale factor $a(\tau)$. The equation of motion is $\partial_\tau(\partial\mathcal{L}'/\partial\dot{\phi}) + \nabla\cdot(\partial\mathcal{L}'/\partial\nabla\phi) = \partial\mathcal{L}'/\partial\phi$, in which the only significant effect of the time dependence is in the first term, in which $\partial\mathcal{L}'/\partial\dot{\phi} = \partial(a^3\dot{\phi}^2/2c^2)/\partial\dot{\phi} = a^3\dot{\phi}/c^2$ so $\partial_\tau(\partial\mathcal{L}'/\partial\dot{\phi}) = \partial_\tau(a^3\dot{\phi}/c^2) = (a^3/c^2)(\ddot{\phi}+3(\dot{a}/a)\dot{\phi})$. Dividing the equation of motion through by a^3/c^2 we obtain

$$\ddot{\phi} + 3H\dot{\phi} - \frac{c^2}{a^2}\nabla^2\phi + dV/d\phi = 0 \tag{3}$$

where $H \equiv \dot{a}/a$. This is identical to the field equation in locally inertial coordinates but with $\nabla_{\mathbf{r}} \Rightarrow \nabla/a$ and with $\ddot{\phi} \Rightarrow \ddot{\phi} + 3H\dot{\phi}$, where we see what is called the *Hubble damping term* $3H\dot{\phi}$, which looks like a 'friction' term in the equation of motion.

The same result can be obtained using the machinery established previously to transform the KG equation – which, in locally inertial coordinates is $\phi^{;\mu}{}_{;\mu} - dV/d\phi = 0$ (commas being equivalent to semi-colons in such coordinates) using the generally covariant form of the d'Alembertian $\phi^{;\mu}{}_{;\mu} = \sqrt{g}^{-1}\partial_\mu(\sqrt{g}g^{\mu\nu}\partial_\nu\phi)$.

The covariant metric components are $g_{\mu\nu} = \mathrm{diag}(-c^2, a^2, a^2, a^2)$ and the Jacobian factor $\sqrt{g} = \sqrt{-|\mathbf{g}|}$ is simply $\sqrt{g} = ca^3$. The inverse metric components are just $g^{\mu\nu} = \mathrm{diag}(-c^{-2}, a^{-2}, a^{-2}, a^{-2})$ so $g^{\mu\nu}\partial_\nu\phi = (-c^{-2}\dot{\phi}, \nabla\phi/a^2)$, and so the d'Alembertian operator applied to ϕ is then

$$\begin{aligned} \phi^{;\mu}{}_{;\mu} = \sqrt{g}^{-1}\partial_\mu(\sqrt{g}g^{\mu\nu}\partial_\nu\phi) &= \frac{1}{ca^3}\partial_\tau\left(-\frac{a^3}{c}\dot{\phi}\right) + \frac{1}{a^2}\nabla^2\phi \\ &= -\frac{1}{c^2}\left(\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi}\right) + \frac{1}{a^2}\nabla^2\phi \end{aligned} \tag{4}$$

which results in (5). Yet another route would be to compute the Christoffel symbols for this metric and use these in $\phi^{;\mu}{}_{;\mu} = \phi^{,\mu}{}_{;\mu} = \phi^{,\mu}{}_{,\mu} + \Gamma^{\mu}{}_{\nu\mu}\phi^{,\nu}$.

Below, we will be mostly interested in free fields, for which $V(\phi) = \frac{1}{2}\mu^2\phi^2$, giving us the KG equation in flat FLRW coordinates:

$$\boxed{\ddot{\phi} + 3H\dot{\phi} - \frac{c^2}{a^2}\nabla^2\phi + \mu^2\phi = 0.} \tag{5}$$

2.2 Damped standing waves in FLRW coordinates

Equation (5) admits solutions that are standing plane waves[3] – or a sum of such waves – in the $\mathbf{r}$-coordinates: $\phi = \phi_{\mathbf{k}}(\tau)\cos(\mathbf{k}\cdot\mathbf{r} + \psi_{\mathbf{k}})$ where $\psi_{\mathbf{k}}$ is the phase and $\mathbf{k}$ is the dimensionless comoving wave-number.

The equation of motion for the amplitude $\phi_{\mathbf{k}}(\tau)$ is then the *damped simple harmonic oscillator*

$$\ddot{\phi}_{\mathbf{k}} + 3H\dot{\phi}_{\mathbf{k}} + \omega^2_{\mathbf{k}}(\tau)\phi_{\mathbf{k}} = 0 \tag{6}$$

with time-dependent frequency

$$\omega_{\mathbf{k}}(\tau) = c\sqrt{k^2/a^2(\tau) + \mu^2}. \tag{7}$$

To elucidate the effect of the damping term it is useful to make a change of variable and set $\phi_{\mathbf{k}} = \varphi_{\mathbf{k}}a^\alpha$ The time derivates are then $\dot{\phi}_{\mathbf{k}} = a^\alpha(\dot{\varphi}_{\mathbf{k}} + \alpha H\varphi_{\mathbf{k}})$ and $\ddot{\phi}_{\mathbf{k}} = a^\alpha(\ddot{\varphi}_{\mathbf{k}} + 2\alpha H\dot{\varphi}_{\mathbf{k}} + \alpha(\ddot{a}/a + (\alpha-1)H^2)\varphi_{\mathbf{k}})$, so the equation of motion for φ is

trick

$$\ddot{\varphi}_{\mathbf{k}} + (3 + 2\alpha)H\dot{\varphi}_{\mathbf{k}} + (\omega^2_{\mathbf{k}}(\tau) + \alpha(\ddot{a}/a + (\alpha-1)H^2)\varphi_{\mathbf{k}} = 0. \tag{8}$$

Thus if we choose $\alpha = -3/2$ the damping term gets cancelled. Moreover, if we restrict attention to waves with period much less than the inverse expansion rate H^{-1} (i.e. for which $\omega^2_{\mathbf{k}}$ is much bigger than H^2 or $\ddot{a}/a$) we have

$$\ddot{\varphi}_{\mathbf{k}} + \omega^2_{\mathbf{k}}(\tau)\varphi_{\mathbf{k}} = 0 \tag{9}$$

which is a simple harmonic oscillator with a time varying frequency.

Since the frequency is varying steadily we can invoke *adiabatic invariance*: the field amplitude will vary sinusoidally $\varphi_{\mathbf{k}}(\tau) \simeq \varphi_{\mathbf{k}0}(\tau)\cos(\omega_{\mathbf{k}}\tau)$ with slowly varying *envelope* $\varphi_{\mathbf{k}0}(\tau)$ changing so that the 'energy' $\langle\dot{\varphi}^2_{\mathbf{k}}\rangle \simeq \omega^2_{\mathbf{k}}|\varphi_{\mathbf{k}0}|^2$ varies in proportion to the frequency. This is illustrated in figure 1. This means that $\varphi_{\mathbf{k}0} \propto 1/\sqrt{\omega_{\mathbf{k}}(\tau)}$ and therefore, since $\phi_{\mathbf{k}} = \varphi_{\mathbf{k}}/a^{3/2}$, that the field is $\phi_{\mathbf{k}}(\tau) \simeq \phi_{\mathbf{k}0}(\tau)\cos(\omega_{\mathbf{k}}\tau)$ with envelope

$$\boxed{\phi_{\mathbf{k}0}(\tau) \propto 1/\sqrt{a^3(\tau)\omega_{\mathbf{k}}(\tau)}.} \tag{10}$$

For waves with physical wavenumber $2\pi/\lambda = k/a \ll \mu$, or equivalently $\lambda \gg \lambda_{\mathrm{C}}$, the Compton wavelength, $k^2/a^2 \ll \mu^2$ and the frequency $\omega_{\mathbf{k}} \simeq c\mu$ and is independent of time. So the field amplitude decays as $\langle\phi^2\rangle \propto 1/a^3$.

[3] We could consider travelling waves also. A pair of oppositely directed travelling waves gives a standing wave. We are interested here in how the amplitude and energy of such waves varies with time, and these are well illustrated by standing waves.

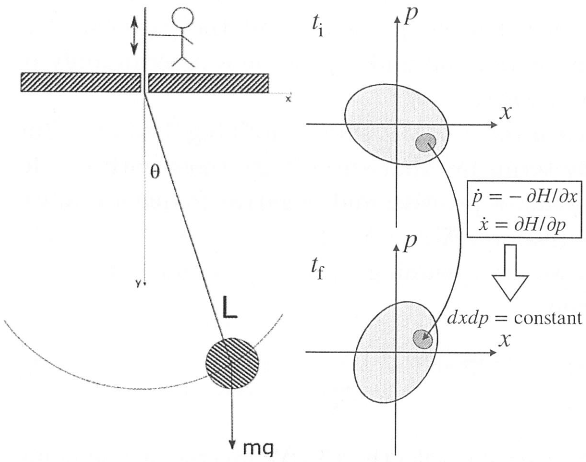

Figure 1: A simple model illustrating adiabatic invariance is a pendulum with the bob rotating in a circle in the horizontal plane with displacement $\mathbf{r}$ from the vertical axis. The angular frequency of rotation is $\omega = \sqrt{g/l}$, so the speed of the bob is $v = |\dot{\mathbf{r}}| = \omega r$. If we steadily shorten the string, the frequency will rise, but there is no torque on the bob, so its angular momentum $J = mrv = mv^2/\omega$ remains constant. So $mv^2 \propto \omega$. Adiabatic invariance can also be understood from the fact that Hamilton's equations imply the phase-space volume of a small cloud of particles remains constant. If the system is changing slowly then while the shape of the light grey region will change, its area remains constant. This implies $\int p\,dq = \int dq\,p$ = constant for the boundary orbit.

What about the energy density? In a locally inertial frame the energy density measured by a stationary observer is $\mathcal{E} = -T^0{}_0 = \frac{1}{2}(\dot\phi^2/c^2 + |\nabla_{\mathbf{r}}\phi|^2 + \mu^2\phi^2)$. The fundamental observers in FLRW cosmology are inertial, so the energy density they measure is the same[4]. Or, in terms of comoving spatial derivatives, $\mathcal{E} = \frac{1}{2}(\dot\phi^2/c^2 + |\nabla\phi|^2/a^2 + \mu^2\phi^2)$. We will now explore how this behaves for long and short waves.

For $\lambda \gg \lambda_C$ (or $k^2 \ll \mu^2 a^2$) we can neglect the $|\nabla\phi|^2/a^2$ term compared to the other two, and, neglecting the time variation of the envelope $\phi_0(\tau)$, we have

$$\mathcal{E} \simeq \tfrac{1}{2}(\dot\phi^2/c^2 + \mu^2\phi^2) = \tfrac{1}{2}\phi_0^2(\tau)\left(\frac{\omega_{\mathbf{k}}^2}{c^2}\sin^2\omega_{\mathbf{k}}\tau + \mu^2\cos^2\omega_{\mathbf{k}}\tau\right)\cos^2\mathbf{k}\cdot\mathbf{r} = \frac{1}{2}\frac{\omega_{\mathbf{k}}^2}{c^2}\phi_0^2\cos^2\mathbf{k}\cdot\mathbf{r} \tag{11}$$

where we have used $\omega_{\mathbf{k}} \simeq \mu c$. This has no time variation, but fluctuates spatially. These fluctuations are 'beats', arising because we are summing two waves travelling in opposite directions. Had we considered only one travelling wave, the energy density would be spatially uniform in this regime.

The spatial average is $\langle\mathcal{E}\rangle = \frac{1}{4}(\omega_{\mathbf{k}}/c)^2\phi_0^2$, and, in this regime the frequency is constant and $\phi_0 \propto 1/a^{3/2}$, so the mean energy decreases as

$$\boxed{\langle\mathcal{E}\rangle = \tfrac{1}{4}(\omega_{\mathbf{k}}/c)^2\phi_0^2 \propto 1/a^3(\tau) \quad \text{for} \quad \lambda \gg \lambda_C.} \tag{12}$$

In the opposite limit $\lambda \ll \lambda_C$ (or $k^2 \gg \mu^2 a^2$) we can neglect the $\mu^2\phi^2/2$ term, and use

$$\mathcal{E} \simeq \tfrac{1}{2}(\dot\phi^2/c^2 + |\nabla\phi|^2/a^2) = \tfrac{1}{2}\phi_0^2(\tau)\left(\frac{\omega_{\mathbf{k}}^2}{c^2}\sin^2\omega_{\mathbf{k}}\tau\cos^2\mathbf{k}\cdot\mathbf{r} + \frac{|\mathbf{k}|^2}{a^2}\cos^2\omega_{\mathbf{k}}\tau\sin^2\mathbf{k}\cdot\mathbf{r}\right) \tag{13}$$

which fluctuates in both time and space. The dispersion relation again tells us that the two terms have the same amplitude, so, at $\tau = 0$ we have, for a wave directed along the z-axis, $\mathcal{E} \propto \sin^2 kz$ with nodes at $z = n\lambda/2$. But a quarter cycle later in time, $\mathcal{E} \propto \cos^2 kz$ with nodes at $z = (n + 1/2)\lambda/2$. The peaks have become nodes and *vice versa*. There is obviously an energy flux density associated with these fluctuations.

Taking an average over one period in time and one wavelength in space, we have $\langle\mathcal{E}\rangle = \frac{1}{4}\omega_{\mathbf{k}}^2\phi_0^2$ as for long waves. But now the frequency varies as $\omega_{\mathbf{k}} \propto 1/a$ while $\phi_0(\tau) \propto 1/a$ and the energy density varies as

$$\boxed{\langle\mathcal{E}\rangle = \tfrac{1}{4}(\omega_{\mathbf{k}}/c)^2\phi_0^2 \propto 1/a^4 \quad \text{for} \quad \lambda \ll \lambda_C.} \tag{14}$$

2.3 Continuity of energy for a sea of waves

It is instructive to look at this using the energy continuity equation. Let's consider a field that is a sum of waves:

$$\phi(\mathbf{x}, \tau) = \sum_{\mathbf{k}} \tfrac{1}{2}(\phi_{\mathbf{k}}(\tau)e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}}\tau)} + \text{c.c.}) \tag{15}$$

[4] In general, the energy density measured by an observer with 4-velocity $\vec{U}$ is $\mathcal{E} = \vec{U}\cdot\mathbf{T}\cdot\vec{U}/c^2 = U_\mu T^\mu{}_\nu U^\nu/c^2$, since, in a locally inertial frame, and for a stationary observer with $U^\mu = dx^\mu/d\tau = (c, 0, 0, 0)$ (and therefore $U_\mu = \eta_{\mu\nu}U^\nu = (-c, 0, 0, 0)$) this gives $\mathcal{E} = -T^0{}_0 = T^{00}$. Here we want the energy density measured by a fundamental observer, with $U^\mu = (1, 0, 0, 0)$ and $U_\mu = g_{\mu\nu}U^\nu = (-c^2, 0, 0, 0)$. The (mixed components of the) stress energy tensor are, with $\mathcal{L} = -\frac{1}{2}(\phi^{;\mu}\phi_{;\mu} + \mu^2\phi^2) = -\frac{1}{2}(\phi^{,\mu}\phi_{,\mu} + \mu^2\phi^2)$, given by $T^\mu{}_\nu = -\phi_{,\nu}\partial\mathcal{L}/\partial\phi_{,\mu} + \delta^\mu_\nu\mathcal{L} = \phi_{,\nu}\phi^{,\mu} + \delta^\mu_\nu\mathcal{L}$. So $\mathcal{E} = U_\mu T^\mu{}_\nu U^\nu/c^2 = -T^0{}_0$ here also, and this is $\mathcal{E} = \frac{1}{2}(\dot\phi^2/c^2 + |\nabla\phi|^2/a^2 + \mu^2\phi^2)$.

where $\phi_{\mathbf{k}}(\tau)$ is a slowly varying complex amplitude – which encodes both the real amplitude and the phase – and c.c. denotes complex conjugation. These are now travelling waves, and the frequency[5] is $\omega_{\mathbf{k}}(\tau) = c\sqrt{|\mathbf{k}|^2/a(\tau)^2 + \mu^2}$ as before. If we choose the phases at random and $|\phi_{\mathbf{k}}|$ to be a function only of $|\mathbf{k}|$ this gives us a statistically isotropic gaussian random sea of waves.

The squared field is a double sum (over $\mathbf{k}$ and $\mathbf{k}'$) with each term in the sum containing 4 terms. But the average over space of nearly all of these vanishes; the only terms are that survive are those with $\mathbf{k} = \mathbf{k}'$ and of the 4 terms the only two that survive are those involving a positive and negative frequency wave. Similarly for the space average of the square of the spatial derivative: $\boldsymbol{\nabla}\phi = \sum(i\mathbf{k}\phi_{\mathbf{k}}e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}\tau)} + \text{c.c.})$ and the time derivative $\dot{\phi} = \sum(-i\omega_{\mathbf{k}}\phi_{\mathbf{k}}e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}\tau)} + \text{c.c.})$, where we are assuming $\omega_{\mathbf{k}}$ is huge compared to the rate of change of the amplitude. These space averages are then

$$\mu^2\langle\phi^2\rangle = \tfrac{1}{2}\sum_{\mathbf{k}}\mu^2\phi_{\mathbf{k}}\phi^\star_{\mathbf{k}} \qquad \langle\dot{\phi}^2\rangle = \tfrac{1}{2}\sum_{\mathbf{k}}\omega^2_{\mathbf{k}}\phi_{\mathbf{k}}\phi^\star_{\mathbf{k}} \qquad \langle|\boldsymbol{\nabla}\phi|^2\rangle = \tfrac{1}{2}\sum_{\mathbf{k}}|\mathbf{k}^2|\phi_{\mathbf{k}}\phi^\star_{\mathbf{k}}. \tag{16}$$

The space average of the Lagrangian density $\mathcal{L} = -\frac{1}{2}(g^{\mu\nu}\phi_{,\mu}\phi_{\nu} + \mu^2\phi^2)$, with the FLRW inverse metric being $g^{\alpha\gamma} = \text{diag}(-c^{-2}, a^{-2}, a^{-2}, a^{-2})$, is

$$\langle\mathcal{L}\rangle = \tfrac{1}{2}\sum_{\mathbf{k}}(\omega^2_{\mathbf{k}}/c^2 - |\mathbf{k}|^2/a^2 - \mu^2)\phi_{\mathbf{k}}\phi^\star_{\mathbf{k}} \tag{17}$$

which vanishes by virtue of the dispersion relation: $\langle\mathcal{L}\rangle = 0$.

The average of the stress-energy tensor (Lecture 5, equation 87) is just

$$\langle T^\alpha{}_\beta\rangle = -\langle\phi_{,\beta}\partial\mathcal{L}/\partial\phi_{,\alpha}\rangle = g^{\alpha\gamma}\langle\phi_{,\beta}\phi_{\gamma}\rangle \tag{18}$$

whose non-vanishing components (with $\tau = x^0$) are

$$\langle T^0{}_0\rangle = -\langle\dot{\phi}^2\rangle/c^2 \qquad \text{and} \qquad \langle T^i{}_j\rangle = \delta^{im}\langle\phi_{,m}\phi_{,j}\rangle/a^2 \tag{19}$$

but, if the power in the waves $P_\phi(\mathbf{k}) \equiv \phi_{\mathbf{k}}\phi^\star_{\mathbf{k}}$ is isotropic, $\langle\phi_{,m}\phi_{,j}\rangle = \frac{1}{2}\sum_{\mathbf{k}} k_m k_j \phi_{\mathbf{k}}\phi^\star_{\mathbf{k}} = \delta_{mj}\langle|\boldsymbol{\nabla}\phi|^2\rangle/3$ and therefore $\langle T^i{}_j\rangle = \delta^i_j\langle|\boldsymbol{\nabla}\phi|^2\rangle/3$ we have

$$\langle T^\alpha{}_\beta\rangle = \text{diag}(-\mathcal{E}, P, P, P) \tag{20}$$

with

$$\mathcal{E} = \langle\dot{\phi}^2\rangle/c^2 \qquad \text{and} \qquad P = \langle|\boldsymbol{\nabla}\phi|^2\rangle/3a^2 \tag{21}$$

The time component of the continuity equation (Lecture-5 86) is

$$\partial_0(\sqrt{g}T^0{}_0) = \partial_0(\sqrt{g}\mathcal{L}) \tag{22}$$

where, for the FLRW metric, the Jacobian is $\sqrt{g} = ca^3(\tau)$. Since $\langle\mathcal{L}\rangle = 0$ we might be tempted to assume that the right hand side has no effect, and we would have $\partial_0(a^3\mathcal{E}) = 0$ which would lead us to $\dot{\mathcal{E}} = -3(\dot{a}/a)\mathcal{E}$ with solution $\mathcal{E} \propto 1/a^3$. This is correct only for waves with $\lambda \gg \lambda_C$, so evidently we need to look more closely at $\partial_0(\sqrt{g}\mathcal{L})$. Recalling that, unlike on the left hand side, where $T^0{}_0$ is considered a function of $\vec{x}$, so ∂_0 means differentiating w.r.t. τ holding the spatial variables constant, on the right hand side, when operating on $\mathcal{L}$, the operator ∂_0 is the τ-derivative holding ϕ and $\phi_{,\alpha}$ fixed. So the right hand side is $\mathcal{L}\partial_0\sqrt{g} - \frac{1}{2}\sqrt{g}\phi_{,\mu}\phi_{,\nu}\partial_0 g^{\mu\nu}$. The first term here averages to zero, and so has no secular effect, but not so the second, since the spatial components of the FLRW inverse metric $g^{\mu\nu}$ are time varying: $\partial_0 g^{ij} = -2\delta^{ij}\dot{a}/a^3$. Thus the average of the right hand side is

$$\langle\partial_0(\sqrt{g}\mathcal{L})\rangle = -\tfrac{1}{2}\sqrt{g}\langle\phi_{,\mu}\phi_{,\nu}\rangle\partial_0 g^{\mu\nu} = c\dot{a}\delta^{ij}\langle\phi_{,i}\phi_{,j}\rangle = c\dot{a}\langle|\boldsymbol{\nabla}\phi|^2\rangle = 3a^2\dot{a}P \tag{23}$$

so, finally, the spatial average of the energy continuity equation above becomes $\partial_0(-ca^3\mathcal{E}) = 3a^2\dot{a}P$, or

$$\dot{\mathcal{E}} = -3(\dot{a}/a)(\mathcal{E} + P) \tag{24}$$

which, for high frequency waves with $\lambda \ll \lambda_C$, for which $P = \mathcal{E}/3$, gives $\mathcal{E} \propto 1/a^4$.

[5] Note that while these have positive and negative frequency components, that does not imply negative energy for the latter; the total energy for a $\mathbf{k}$-mode is positive.

2.4 Correspondence with an expanding gas of particles

We see here a precise correspondence with the behaviour of particles and ideal fluids. A sea of scalar waves that is set up to have, initially, statistically homogeneous properties on a constant-τ hypersurface in FLRW space-time has an energy density that dilutes just like that of a gas of particles. If the particles are non-relativisitic (so they have $\lambda \gg \lambda_C$) their pressure is negligible compared to $\mathcal{E}$, and the energy density is, to all intents an purposes, just equal to the rest-mass energy density of the particles times c^2, and conservation of particles implies $\mathcal{E} = \rho c^2 \propto 1/a^3$. But if they are rapidly moving, they transport significant momentum, this momentum flux being pressure, and, as the volume per particle increases, the pressure "does work in expanding". That terminology seems sensible if we think of a collisional gas of particles, but if the particles stream freely it makes more sense to say that there is an expanding network of "observers" – those who see no flux density of particles or energy – who measure the travelling particles to have steadily decreasing energies. Nothing happens to a particle as it moves from one such observer to another, infinitesimally distant, one. That's because space-time is locally flat; the central tenet of GR. But the latter measures the particle's energy in a frame that is slightly boosted frame as compared to that of the former, and so measures it to have a lower energy. Either view leads to the same conclusion.

We also see a direct correspondence with what happens for expanding ideal fluids. For an isotropic superposition of high frequency waves the 3-stress for the scalar field is diagonal with pressure $P = \mathcal{E}/3$ just as for a plasma or gas consisting of relativistic particles and/or photons. And the continuity equation for energy density contains, on the right hand side, just as for particles or a fluid, the enthalpy $\mathcal{E} + P$.

The stress energy tensor obtained here can be expressed in term of the power spectrum of the waves $P_\phi(\mathbf{k}) \propto \langle \phi_\mathbf{k} \phi^\star_\mathbf{k} \rangle$ as follows. The spatial average

$$\langle T^\alpha{}_\beta \rangle \equiv \frac{1}{L^3} \int d^3r \phi^2 = g^{\alpha\gamma} \langle \phi_{,\beta} \phi_{,\gamma} \rangle = \tfrac{1}{2} g^{\alpha\gamma} \sum_\mathbf{k} k_\gamma k_\beta \phi_\mathbf{k} \phi^\star_\mathbf{k} = \tfrac{1}{2} \sum_\mathbf{k} k^\alpha k_\beta \phi_\mathbf{k} \phi^\star_\mathbf{k} \tag{25}$$

where $k_\mu = (\omega/c, \mathbf{k})$ and we are taking the field to be periodic within a comoving box of size L so the Fourier modes $\mathbf{k}$ live on a lattice with spacing $\Delta k = 2\pi/L$. The number of modes in a volume d^3k is $d^3k/(\Delta k)^3 = L^3 d^3k/(2\pi)^3$ so we can replace $\sum_\mathbf{k} \ldots \Rightarrow (L/2\pi)^3 \int d^3k \ldots$ and write

$$\langle T^\alpha{}_\beta \rangle = \tfrac{1}{2} \int \frac{d^3k}{(2\pi)^3} k^\alpha k_\beta P_\phi(\mathbf{k}) \tag{26}$$

where $P_\phi(\mathbf{k}) = L^3 \langle \phi_\mathbf{k} \phi^\star_\mathbf{k} \rangle$ and where the averaging here is not a spatial average but an average per mode[6].

This is highly reminiscent of the expression we obtained for the stress-energy tensor in terms of the phase-space density $f(\mathbf{p})$ for a gas of particles:

$$T^\alpha{}_\beta = \int \frac{d^3p}{p^0} p^\alpha p_\beta f(\mathbf{p}) \tag{27}$$

thus there is a direct correspondence between the particle phase-space density $f(\mathbf{p})$ and $k^0 P_\phi(\mathbf{k})$ evaluated[7] at $\mathbf{k} = \mathbf{p}/\hbar$.

And, as with particles and/or ideal fluids, we see the peculiar phenomenon that each *comoving volume* is losing energy as it expands. Where is the energy going? Energy obeys continuity, but doesn't seem to be conserved here[8]. Well, perhaps that is no surprise; the Lagrangian we are working with has an explicit cosmic time dependence, so there is no reason for there to be a conserved Hamiltonian.

Of course, there is nothing particular about the scalar field here. The same thing happens to e.g. the electromagnetic field in cosmology where Maxwell's equations have a damping term that is precisely analogous to the $3H\dot{\phi}$ term in the KG equation, and to which is attributed the red-shifting and consequent

---

[6] The idea here is that we are dealing with a field which whose fluctuations form a statistically homogeneous random process, and for which the power spectrum is some smooth function of $\mathbf{k}$ so, if we let the periodicity scale L become huge, the density of states becomes similarly huge, so there is a sensibly defined local average $\langle \phi_\mathbf{k} \phi^\star_\mathbf{k} \rangle$.

[7] One can understand the reason for the k^0 here by thinking about the contribution to the variance $d\langle \phi^2 \rangle$ from the modes in some small volume d^3k around $\mathbf{k} = 0$ (or $k_\mu = (\mu, \mathbf{0})$) (there always being some observer who sees $\mathbf{k} = 0$). In another frame, the boosted modes occupy a volume $d^3k' = \gamma d^3k = (k^{0\prime}/\mu) d^3k$. Thus d^3k/k^0 is a Lorentz invariant (just like d^3p/p^0 for particles). But the contribution $d\langle \phi^2 \rangle = d^3k P_\phi(\mathbf{k})/(2\pi)^3$ from these modes must be Lorentz invariant so $P_\phi(\mathbf{k})$ itself cannot be. But $k^0 P_\phi(\mathbf{k})$ is, and it is that that we can identify with the similarly Lorentz invariant phase-space density $f(\mathbf{p})$.

[8] In a closed universe, we can integrate the total energy, and it is finite. And, if there is non-negligible pressure it is also varying with time.

loss of energy of the radiation field. This term is often said to arise because of the *coupling of the EM – or here scalar – field to the gravitational field of an expanding universe.* It is also often said that "the expansion of the universe" stretches the wavelengths of light, reducing their energy, and the same thing is happening here if the waves have significant momentum.

As I have complained vociferously, this is misleading. We would have found the same behaviour if, instead of a flat FLRW universe, we had considered an empty universe – i.e. the Milne model – since we are considering here waves that have wavelength that is small compared to the curvature scale. But we would then have been simply considering scalar waves in Minkowski space, where there is no gravitational field, but viewed from an expanding coordinate system. The lesson from the red-shifting radiation field in the 1D Milne model is that what we have here is a field of waves of the scalar field that is – *of itself* – in a state of expansion. That expansion is traceable to its initial conditions; the field was set up in such a way that its statistical properties were homogeneous on an initial surface of constant τ. And, since we are thinking about a sea of waves with large-scale homogeneity, we can imagine a network of observers who measure vanishing energy flux density (averaged over a suitable volume), and these observers are expanding away from one another, just like the fundamental observers of the FLRW model, but it is the radiation itself that is defining that expansion.

3 A scalar waves as the dark matter in galaxies and clusters

Structures like galaxies and clusters of galaxies have velocity dispersions (orbital velocities) $\sigma_v \sim 100 - 1000$km/s (or in the range $(0.3 - 3) \times 10^{-3}c$). The virial theorem says $\sigma_v^2 \sim GM/R$, with M and R the mass and size of the system. So the dimensionless gravitational potential is $\Phi \equiv GM/Rc^2$ and ranges from about 10^{-7} for a small galaxy to 10^{-5} for a very massive cluster. For all such systems $\Phi \ll 1$ to an extremely good approximation and, moreover, they are approximately static, in the sense that any time-variation of the density is much less than (c times) the space-variation.

Such systems are therefore well described by *weak-field gravity* in which we write $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ with metric perturbations $|h_{\mu\nu}| \ll 1$. If one adopts the *Lorenz gauge* (which is also called the de Donder gauge) in which the trace reversed metric perturbations $\overline{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$ (where $h = \eta^{\mu\nu}h_{\mu\nu}$) are taken to be divergence-free: $\overline{h}^{\alpha\beta}{}_{,\alpha} = 0$, then the Einstein equations become

$$\Box\overline{h}_{\mu\nu} = -16\pi(G/c^4)T_{\mu\nu} \tag{28}$$

and, for slowly moving sources, as here, only the $T_{00} = \rho c^2$ component is significant and it is slowly varying, so this equation permits solutions where the time variation of the metric is also small (so $\Box \Rightarrow \nabla^2$). Thus we need to solve $\nabla^2\overline{h}_{\mu\nu} = -16\pi G\rho\delta^0_\mu\delta^0_\nu/c^2$, whose solution we can take to be $\overline{h}_{\mu\nu} = -4\Phi\delta^0_\mu\delta^0_\nu$ where

$$\nabla^2\Phi = 4\pi G\rho/c^2 \tag{29}$$

i.e. Φ the Newtonian potential divided by c^2 and, if we are dealing with an isolated galaxy or cluster, we can require, as boundary conditions, that $\Phi \to 0$ far away from the object.

Trace-reversing $\overline{h}_{\mu\nu}$, we recover the metric perturbations $h_{\mu\nu} = -2\Phi(\mathbf{r})\delta_{\mu\nu}$, so the line element is

$$ds^2 = -(1 + 2\Phi)c^2dt^2 + (1 - 2\Phi)(dx^2 + dy^2 + z^2). \tag{30}$$

As discussed earlier, a well motivated candidate for the dark matter is the *Peccei-Quinn axion*; a scalar field with mc^2 thought to be on the order of 10^{-6}eV (though with considerable slop in the precise value), corresponding to a Compton wavelength on the order of a metre, so tiny compared to the size of a galaxy, and, if they were to behave like particles with this mass and have velocity of say $v \sim 300$km/s, would have a de-Broglie wavelength $\lambda_{\rm dB} = (c/v)\lambda_{\rm C}$ of about a km, still tiny. Less well motivated, but still widely considered, is the idea that the DM is an *ultra-light axion-like field*. It the mass is such that $mc^2 \sim 10^{-22}$eV then the de Broglie scale becomes comparable to the size of the inner parts of galaxies[9]. This is known as the *fuzzy dark matter* scenario.

The question we will address here is: how would such matter behave in e.g. galaxies? We know that *space-time tells particles how to move* through the requirement that their action, being $S = -mc^2\int d\tau = -mc^2\int d\lambda\sqrt{-g_{\mu\nu}p^\mu p^\mu}$, be extremised. How does this work if the DM is scalar waves?

[9]So one might expect to see wave-mechanical effects in the centres of galaxies, where there have been claims that the standard ΛCDM model produces 'cores' that are too 'cuspy'.

3.1 The Klein-Gordon equation in weak-field gravity

In our analysis of expanding scalar wave fields in FLRW space-times we used the KG equation (Lecture 5, equation 78) obtained by extremising the transformed action. Here it is more convenient to use the alternative form (Lecture 5, equation 82) obtained by writing out covariant derivative in the kinetic term ${\phi^{;\mu}}_{;\nu} = g^{\mu\nu}\phi_{;\mu;\nu} = g^{\mu\nu}\phi_{,\mu;\nu}$ in terms of Christoffel symbols to obtain:

$$g^{\mu\nu}(\phi_{,\mu\nu} - {\Gamma^{\alpha}}_{\mu\nu}\phi_{,\alpha}) - \mu^2\phi = 0. \tag{31}$$

The reason for using this is that, for weak field gravity, and working, as we are, in the Lorenz gauge, the Christoffel symbol term actually vanishes. This is because, working to first order in the metric perturbations, we have

$$\begin{aligned} g^{\mu\nu}{\Gamma^{\alpha}}_{\mu\nu} &= \tfrac{1}{2}g^{\mu\nu}g^{\alpha\gamma}(g_{\gamma\mu,\nu} + g_{\gamma\nu,\mu} - g_{\mu\nu,\gamma}) \\ &= \tfrac{1}{2}\eta^{\mu\nu}\eta^{\alpha\gamma}(h_{\gamma\mu,\nu} + h_{\gamma\nu,\mu} - h_{\mu\nu,\gamma}) \\ &= \eta^{\alpha\gamma}\left({h_{\gamma\mu}}^{,\mu} - \tfrac{1}{2}h_{,\gamma}\right) \\ &= \eta^{\alpha\gamma}\left(h_{\gamma\mu} - \tfrac{1}{2}\eta_{\gamma\mu}h\right)^{,\mu}. \end{aligned} \tag{32}$$

In the first step here we have used the fact that, with $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, derivatives of components of the metric like $g_{\gamma\mu,\nu}$ are the same as $h_{\gamma\mu,\nu}$, which is a first order quantity, and so we can ignore the metric perturbations in $g^{\mu\nu}g^{\alpha\gamma}$ and replace this by $\eta^{\mu\nu}\eta^{\alpha\gamma}$. And in the second, we have used the symmetry of $\eta^{\mu\nu}$ and the definition of the trace $h \equiv \eta^{\mu\nu}h_{\mu\nu}$.

But we recognise the quantity in parentheses as our friend the trace-reversed metric perturbation, $\overline{h}_{\gamma\mu} \equiv h_{\gamma\mu} - \frac{1}{2}\eta_{\gamma\mu}h$, so $g^{\mu\nu}{\Gamma^{\alpha}}_{\mu\nu} = \eta^{\alpha\gamma}{\overline{h}_{\gamma\mu}}^{,\mu}$ and, if we work in the Lorenz (or de Donder) gauge, this vanishes. Put another way, with this choice of gauge,

$$g^{\mu\nu}\phi_{;\mu;\nu} \Rightarrow g^{\mu\nu}\phi_{,\mu,\nu} \tag{33}$$

with corrections that are only second order in the metric perturbations (so roughly a million times smaller than any 1st order effects).

Thus the Lorenz gauge proves useful, not just for providing the simplified Einstein's equations (28), but it also banishes any connection terms from the generalised KG equation (Lecture 5, equation 82).

So it appears that, in this instance, and in the Lorenz gauge, that the 'comma becomes semi-colon' rule has no effect, which might lead one to think that there is no coupling of gravity to a scalar field. But not quite. We still have the metric perturbation in $g^{\mu\nu}$. This is the inverse of $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. Writing $g^{\mu\nu} = \eta^{\mu\nu} + h^{\mu\nu}$, we have

$$\delta^{\mu}_{\nu} = g^{\mu\gamma}g_{\gamma\nu} = (\eta^{\mu\gamma} + h^{\mu\gamma})(\eta_{\gamma\nu} + h_{\gamma\nu}) = \underbrace{\eta^{\mu\gamma}\eta_{\mu\gamma}}_{\delta^{\mu}_{\nu}} + \eta^{\mu\gamma}h_{\gamma\nu} + h^{\mu\gamma}\eta_{\gamma\nu} + \mathcal{O}(h^2) \tag{34}$$

so, at 1st order, $\eta^{\mu\gamma}h_{\gamma\nu} + h^{\mu\gamma}\eta_{\mu\gamma} = 0$, implying[10] $h^{\mu\nu} = -\eta^{\mu\alpha}\eta^{\mu\beta}h_{\alpha\beta}$. Thus with $h_{\alpha\beta} = -2\Phi\delta_{\alpha\beta}$, $h^{\mu\nu} = +2\Phi\delta^{\alpha\beta}$ and hence the kinetic term is $g^{\mu\nu}\phi_{,\mu,\nu} = (\eta^{\mu\nu} + h^{\mu\nu})\phi_{,\mu\nu} = {\phi^{,\mu}}_{,\mu} + 2\Phi\delta^{\mu\nu}\phi_{,\mu\nu}$ and the KG equation becomes

$${\phi^{,\mu}}_{,\mu} + 2\Phi\delta^{\mu\nu}\phi_{,\mu\nu} - \mu^2\phi = 0. \tag{35}$$

This is a non-covariant equation[11], but is valid nonetheless, but only in a coordinate system compatible with our choice of gauge. In 3+1 notation, this is

$$\boxed{-(1 - 2\Phi)\ddot{\phi}/c^2 + (1 + 2\Phi)\nabla^2\phi - \mu^2\phi = 0.} \tag{36}$$

This is quite general, and could be used to describe the propagation of scalar waves of arbitrary momentum (i.e. wave-number). If, however, the field gained its 3-momentum by 'falling' into a Newtonian potential well with $\Phi \ll 1$, one would expect that its wavelength would be, to order of magnitude, the same as the de-Broglie wavelength for a particle in the same potential well; so larger than the Compton wavelength μ^{-1} by a factor $c/\sigma_v \sim \Phi^{-1/2} \gg 1$. Or, equivalently, that the group velocity $v_g = d\omega/dk = (k/\omega)d\omega^2/dk^2 \simeq c^2k/\omega \simeq ck/\mu = c\lambda_C/\lambda$ be approximately equal to σ_v. This means that

[10] Note that the inverse metric perturbations $h^{\mu\nu}$ are *not* obtained simply by raising indices on $h_{\mu\nu}$; the sign is different.

[11] To avoid possible confusion, $\delta^{\nu\mu}$ here has components diag(1, 1, 1, 1), and is quite different from the index-raised version of the (covariant) $\delta^{\mu}_{\nu} = \text{diag}(1, 1, 1, 1)$ (which is $\eta^{\nu\alpha}\delta^{\mu}_{\alpha} = \text{diag}(-1, 1, 1, 1)$).

$\nabla^2\phi \sim k^2\phi \sim (\sigma_v^2/c^2)\mu^2\phi \sim \Phi\mu^2\phi$. So $\nabla^2\phi$ is of 1st order in Φ, so we can ignore the term $2\Phi\nabla^2\phi$ (it being second order in Φ). To zeroth order in Φ we have $\ddot{\phi} = -\mu^2 c^2 \phi$, so we can replace, again at 1st order, $2\Phi\ddot{\phi}/c^2 \Rightarrow -2\mu^2\Phi\phi$. Moving this over to the right hand side, and multiplying by c^2 the KG equation is

$$\boxed{-\ddot{\phi} + c^2\nabla^2\phi - {\omega_C}^2(\mathbf{x})\phi = 0} \tag{37}$$

where what was the (angular) Compton frequency $\omega_C = \mu/c = mc^2/\hbar$ has become position dependent:

$$\omega_C \Rightarrow \omega_C(\mathbf{x}) \equiv (1 + \Phi(\mathbf{x}))\omega_C. \tag{38}$$

or, equivalently, that the mass m has become position dependent:

$$m \Rightarrow m(\mathbf{x}) = (1 + \Phi(\mathbf{x}))m. \tag{39}$$

This is both simple and rather pleasing and intuitively reasonable. The metric here is static, so distances between constant-$\mathbf{x}$ observers is not changing. But they are in a potential well with, in general, a non-vanishing tidal field, so they must be accelerated[12] in order to maintain constant $\mathbf{x}$. And, as we know, in an accelerating frame, time is 'warped'; clocks at the nose of an accelerating rocket run faster than those at the tail, and clocks higher up in the Earth's potential well, for instance, run faster than those lower down[13]. The angular frequency of scalar field oscillations (of long spatial wavelength) in locally inertial coordinates is ω_C and constitutes a clock. And like any other good clock, it will run slow if lowered into a gravitational potential well (where, we recall, $\Phi < 0$).

But we need to be clear what is meant by this physically; $\omega_C(\mathbf{x})$ is frequency in *coordinate time*. This is *not* the frequency that would be measured by a local constant-$\mathbf{x}$ observer: According to the metric, for such an observer, an interval dt of coordinate time corresponds to an interval of proper time – as he or she would measure with a clock – $d\tau = \sqrt{1+2\Phi}dt \simeq (1+\Phi)dt$, so the coordinate time for one oscillation: $dt = 2\pi/\omega_C(\mathbf{x}) = 2\pi/(1+\Phi)\omega_C$ corresponds to the same proper time $d\tau = 2\pi/\omega_C$ as in the absence of gravity. But that is as it should be; in GR physics is locally the same everywhere[14]. Nonetheless, the varying coordinate time frequency results in analogous refractive effects, just as in an inhomogeneous plasma. In the coordinate system here, coordinate time corresponds to proper time for an observer outside the potential well where $\Phi = 0$. So if an observer in the potential at $\mathbf{x}$ sends a light signal once every cycle of the scalar field, the paths the light signals take being time invariant, the external observer will receive these signals with intervals dilated by a factor $1 - \Phi(\mathbf{x})$.

3.2 The local dispersion relation for scalar waves

If the gravitational potential $\Phi(\mathbf{x})$ is varying on a sufficiently large scale, the KG equation (37) will admit locally monochromatic solutions where $\phi(\vec{x}) = a\cos(k_\mu x^\mu + \Psi_0)$ where a is a constant amplitude, Ψ_0 is the phase, and the components k_μ of the wave-number are to be determined. Equivalently we can write this as

$$\phi(\vec{x}) = \tfrac{1}{2}(\phi_0 e^{ik_\mu x^\mu} + \text{c.c.}). \tag{40}$$

with constant complex amplitude $\phi_0 = ae^{i\Psi_0}$ that encodes both the phase and the amplitude.

With this trial solution (and writing $k_\alpha = (-\omega_\mathbf{k}/c, \mathbf{k})$ so, with $x^\mu = (ct, \mathbf{x})$, $k_\mu x^\mu = \mathbf{k}\cdot\mathbf{x} - \omega_\mathbf{k} t$), and considering Φ to be (locally) constant, the KG equation (36) becomes the dispersion relation linking frequency and wave-number:

$$(1 - 2\Phi)\omega_\mathbf{k}^2/c^2 = \mu^2 + (1 + 2\Phi)|\mathbf{k}|^2 \tag{41}$$

where we are dropping the restriction that $|\mathbf{k}|^2 \ll \mu^2$.

So, just as in Minkowski space where the normalisation condition on the 4-wavenumber is $\eta^{\mu\nu}k_\mu k_\nu = -\mu^2$, here the dispersion relation is $g^{\mu\nu}k_\mu k_\nu = -\mu^2$.

[12] Perhaps they are being supported by some kind of rigid lattice. Perhaps they are riding on rockets. Who knows? The important point is that the coordinate system here is distinctly non-inertial.

[13] This means that if you go up the Eiffel tower and spend some time there, when you come back down, your watch will register a longer elapsed time than if you had stayed at the bottom. This may sound a bit like an effect of curvature, but it isn't. It's just flat space-time special relativity in an accelerated frame – and no different in essence from what happens in the well-known 'twin paradox' – and that's why we talk of 'time being warped', here rather than 'space-time being curved'.

[14] One can say loosely that the varying $\omega_C(\mathbf{x})$ is like having a varying plasma frequency $\omega_p(\mathbf{x})$ in an inhomogeneous cold plasma. But the latter is different in the sense that a local observer can measure the changing plasma frequency – the frequency it rings at if some electrons are displaced – whereas here a local observer does not see any change in the Compton frequency.

Specialising to the non-relativistic regime (as appropriate for waves in quasi-Newtonian potential wells), we can replace $(1+2\Phi)|\mathbf{k}|^2 \Rightarrow |\mathbf{k}|^2$, and, dividing through by $(1-2\Phi) \simeq 1/\sqrt{1+\Phi}$, the dispersion relation is

$$\omega_{\mathbf{k}}^2 = \omega_{\rm C}{}^2(\mathbf{x}) + c^2|\mathbf{k}|^2 \tag{42}$$

which we could have obtained from (37).

An alternative, and somewhat illuminating, way to look at this is to write

$$\phi(\vec{x}) = \tfrac{1}{2}(\phi_0 e^{i\Psi(\vec{x})} + \text{c.c.}) \tag{43}$$

which is equivalent to the above if we expand the phase as $\Psi(\vec{x}) = \Psi_0 + \Psi_{,\mu}x^\mu + \ldots$, then the dispersion relation relates the components of the phase-derivative 1-form $\tilde{d}\Psi \rightarrow \Psi_{,\mu}$, and is then

$$(1-2\Phi)(-\partial\Psi/\partial t)^2/c^2 = \mu^2 + (1+2\Phi)(\boldsymbol{\nabla}\Psi)^2. \tag{44}$$

This reminds us of the Hamilton-Jacobi equations. If we have a beam of (possibly relativistic) particles emanating from a common starting point with a range of 3-momenta, the HJ equations are $H = -\partial S/\partial t$ and $\mathbf{p} = \boldsymbol{\nabla} S$ where $S(\vec{x}) = -mc^2\int d\tau$ is the action. Equivalently these say $\tilde{p} \longrightarrow p_\mu = (-H/c, \mathbf{p}) = \partial_\mu S$. The normalisation condition on the covariant momentum components is $g^{\mu\nu}p_\mu p_\mu = -m^2c^2$, or

$$(1-2\Phi)(-\partial S/\partial t)^2/c^2 = m^2c^2 + (1+2\Phi)(\boldsymbol{\nabla} S)^2. \tag{45}$$

But $\mu = mc/\hbar$, so the above two equations are equivalent if the phase $\Psi(\vec{x}) = S(\vec{x})/\hbar$.

This is reminiscent of the Dirac-Feynman prescription, according to which the wave-function ψ for a system that has a classical action S is equal to $\exp(iS/\hbar)$. And the KG equation was originally conceived of by Schrödinger as an equation for a single particle wave function, and its incarnation (36) in weak gravitational fields is the operator equation $(g^{\mu\nu}p_\mu p_\nu + m^2c^2)\phi = 0$ with $p_\mu = (-H/c, \mathbf{p})$ replaced by the operators $H \Rightarrow i\hbar\partial_t$ and $\mathbf{p} \Rightarrow -i\hbar\boldsymbol{\nabla}$, or $p_\mu \Rightarrow -i\hbar\partial_\mu$ and with the parameter $\mu \Rightarrow mc/\hbar$.

But the field $\phi(\vec{x})$ here is a purely classical one, and is certainly not a single particle wave function; here we use the KG equation to describe fields for which the occupation number is typically enormous. If $\hbar$ appears here, it is really just because we are writing the parameter μ as $mc/\hbar$. As discussed earlier, the field can be considered to be a collection of harmonic oscillators – one for each Fourier mode $\mathbf{k}$ of the field, with amplitude $\phi_{\mathbf{k}}$ and frequency $\omega_{\mathbf{k}}$. These can be quantised in the usual way, by asserting that there is a wave function $\psi(\phi_{\mathbf{k}})$ that obeys the (non-relativistic) Schrödinger equation. This is called 'second-quantisation'. So there isn't a single wave-function here; there's a wave-function for each mode and, while the energy eigenstates $|n;\mathbf{k}\rangle$ all have vanishing $\overline{\phi}_{\mathbf{k}} = \langle n;\mathbf{k}|\,\phi_{\mathbf{k}}\,|n;\mathbf{k}\rangle = 0$, if we assume each mode is in a coherent state $|\alpha;\mathbf{k}\rangle \propto \sum_n(\alpha^n/\sqrt{n!})\,|n;\mathbf{k}\rangle$ the expectation value $\overline{\phi}_{\mathbf{k}}$ behaves classically – and, for large $|\alpha|^2$ – the fluctuations about this are tiny. And the KG equation describes the behaviour of $\phi(\mathbf{x},t)$ as synthesised from these states.

3.3 The phase- and group-velocities for scalar waves

The phase of a locally planar wave is $\Psi(\vec{x}) = k_\mu x^\mu = \mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}}t$. What speed are the wave-fronts moving? In 1D it is unambiguous, but in 3D is is more subtle. One could say it is the answer to the question: what is the speed of an observer who sees no change in the phase? But that has no unique answer; it depends on the direction the observer is moving. One natural choice is the velocity of an observer moving along the path with (contravariant) $x^i(t)$ proportional to the contravariant components of the wave-vector $k^i = g^{ij}k_j$ (we're assuming here $g^{0i} = 0$). The phase-velocity then has components $v_{\rm p}^i = \omega_{\mathbf{k}}k^i/(k^jk_j)$. In the Newtonian limit weak field metric $g^{ij} = (1+2\Phi)\,\text{diag}(1,1,1)$ and so, in terms of the (covariant) components $v_{\rm p}^i = \omega_{\mathbf{k}}k_i/(k_jk_j)$, or

$$\mathbf{v}_{\rm p} = \frac{d\mathbf{x}(t)}{dt} = \hat{\mathbf{k}}\frac{\omega_{\mathbf{k}}}{|\mathbf{k}|}. \tag{46}$$

According to the dispersion relation, and for $\Phi = 0$, $|\mathbf{v}_{\rm p}|$ is always greater than c. In the non-relativistic regime $\omega_{\mathbf{k}} \Rightarrow \omega_{\rm C}$ and $\mathbf{v}_{\rm p} = \hat{\mathbf{k}}\omega_{\rm C}/|\mathbf{k}|$ diverges as $\mathbf{k} \rightarrow 0$.

The speed with which 'groups' of waves or wave-packets – or 'beats' in a wave-train constructed from a superposition of 2 waves of slightly different frequencies – travel is called the 'group-velocity' and is, in contrast, unambiguously defined: $\mathbf{v}_{\rm g} = d\mathbf{x}/dt \longrightarrow (dx^i/dt)$ is

$$\mathbf{v}_{\rm g} = \frac{d\omega_{\mathbf{k}}}{d|\mathbf{k}|} = \frac{\mathbf{k}}{\omega_{\mathbf{k}}}\frac{d\omega_{\mathbf{k}}^2}{d|\mathbf{k}|^2} = c^2\frac{1+2\Phi}{1-2\Phi}\frac{\mathbf{k}}{\omega_{\mathbf{k}}} \tag{47}$$

and, for $\Phi = 0$, is always less than (or, in the limit $|\mathbf{k}| \gg \mu = \omega_C/c$ equal to) c. This is also the speed at which information can be propagated. Put another way,

$$v_g^i = c\frac{k^i}{k^0}. \tag{48}$$

These expressions are quite analogous to the 3-velocity of a particle. If the particle has 4-momentum with contravariant components $p^\mu = dx^\mu/d\lambda$ where $d\lambda = d\tau/m$ (or the limit of this as $m \to 0$) then the coordinate velocity $\mathbf{v} = d\mathbf{x}/dt = cd\mathbf{x}/dx^0$ has components $v^i = cp^i/p^0$. But $p^\mu = g^{\mu\nu}p_\nu$ and therefore, the weak-field metric being diagonal, $p^0 = -(1-2\Phi)p_0$ and $p^i = (1+2\Phi)p_i$ so, if we write, for the covariant components[15] $p_\mu = (-H/c, \mathbf{p})$, the coordinate velocity is

$$\mathbf{v} = c^2\frac{1+2\Phi}{1-2\Phi}\frac{\mathbf{p}}{H} \tag{49}$$

just like the group velocity above.

Thus the group velocity is the same as the coordinate 3-velocity of a particle with $p_\mu = \text{constant} \times k_\mu$. To obtain the constant of proportionality, we may note that an observer moving a distance $d\mathbf{x} = \mathbf{v}_g dt$ in coordinate time dt would record a squared proper time increment

$$d\tau^2 = (1+2\Phi)dt^2 - (1-2\Phi)|d\mathbf{x}|^2/c^2 = (1+2\Phi)dt^2\left(1 - \frac{1-2\Phi}{1+2\Phi}\frac{|\mathbf{v}_g|^2}{c^2}\right). \tag{50}$$

But the last factor is, from the dispersion relation simply $\mu^2c^2/\omega_\mathbf{k}^2$, from which

$$\begin{aligned} dt &= (1-2\Phi)\frac{\omega_\mathbf{k}}{c\mu}d\tau \\ d\mathbf{x} &= (1+2\Phi)\frac{\mathbf{k}}{\mu}d\tau \end{aligned} \tag{51}$$

But, since $k_\mu = (-\omega_\mathbf{k}/c, \mathbf{k})$, these say that the 4-velocity of this observer is

$$\boxed{\frac{dx^\mu}{d\tau} = \frac{c}{\mu}g^{\mu\nu}k_\mu} \tag{52}$$

and that a particle of mass m moving at $\mathbf{v}_g$ has 4-momentum

$$\boxed{\tilde{p} = \frac{mc}{\mu}\tilde{k} = \hbar\tilde{k}.} \tag{53}$$

And the change of phase $\Psi = k_\mu x^\mu$ as perceived by an observer moving at $\mathbf{v}_g$ is

$$\begin{aligned} d\Psi &= -\omega_\mathbf{k}dt + \mathbf{k}\cdot d\mathbf{x} \\ &= (1-2\Phi)(-\omega_\mathbf{k}^2 + \mathbf{k}\cdot\mathbf{v}_g)d\tau/c\mu \\ &= (1-2\Phi)\underbrace{\left(-\omega_\mathbf{k}^2 + c^2|\mathbf{k}|^2\frac{1+2\Phi}{1-2\Phi}\right)}_{-\mu^2c^2/(1-2\Phi)}d\tau/c\mu = -\mu c d\tau \end{aligned} \tag{54}$$

(again invoking the dispersion relation) so

$$\boxed{d\Psi = -\frac{mc^2}{\hbar}d\tau = \frac{1}{\hbar}dS} \tag{55}$$

Thus, an observer moving along with a wave-packet (or with a beam of waves, at a velocity such that he or she sees no energy flux or momentum density) would see the phase rotating at the Compton frequency $\omega_C = mc^2/\hbar$.

[15]Where, as usual, $\mathbf{p} \equiv \partial L/\partial\dot{\mathbf{x}}$ with Lagrangian $L(\mathbf{x}, \dot{\mathbf{x}}, t) = -mc^2\sqrt{-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}$ and $H = \dot{\mathbf{x}}\mathbf{p} - L$.

3.4 Visualising wave-packets in space-time

Packets of scalar waves behave a lot like particles. Figure 6 in Lecture 5 showed a picture of a 2D wave-packet in space. What does such a packet look like in space-time? This is shown – for one spatial dimension – in figure 2. The left panel shows a packet with vanishing 3-momentum. In reality, owing to its finite extent in space, the packet will spread with time. But for a nearly monochromatic packet as shown here this takes a long time. We can see what the phase-fronts look like for a $\mathbf{p} \neq 0$ packet by boosting the stationary packet to make the centre panel. Note that the phase-gradient 1-form $\tilde{d}\Psi$ points 'back in time'. Note how the wave-fronts move through the packet from the back to the front; and they do so at a super-luminal velocity[16].

On the right is what happens as the boost velocity $\Rightarrow c$. This is a wave-packet that corresponds to a highly relativistic particle. Note that the scalar product $\tilde{p}(\vec{p})$ – the number of iso-Ψ surfaces pierced by the arrow representing $\vec{p}$ – vanishes.

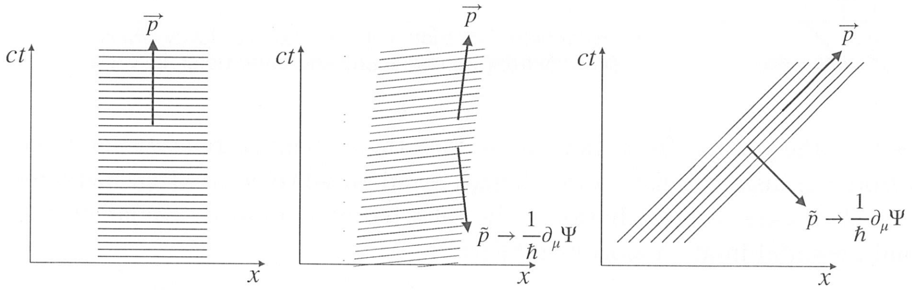

Figure 2: At left is shown a scalar wave-packet that has vanishing 3-momentum. Middle is the same thing in a boosted frame. Note that the t-component of $\tilde{p}$ is opposite to that of $\vec{p}$. Right is a highly relativistic packet.

3.5 The analogy with EM waves in a plasma

The dispersion relation for non-relativistic scalar waves in a gravitational potential (42) is identical to that for EM waves in a cold plasma

$$\omega^2 = \omega_{\rm p}^2 + c^2|\mathbf{k}|^2 \tag{56}$$

with $(1+\Phi)mc$ playing the role of the plasma frequency $\omega_{\rm p}$. The physics of this is illustrated in the left-hand panel of figure 3. The density of electrons defines a frequency $\omega_{\rm p}$ which is the frequency that the plasma would 'ring' at if the electrons in some region were displaced (the resulting charge imbalance creating a restoring force). Maxwell's equations – here in integral form – show that travelling wave solutions are only allowed at frequencies above $\omega_{\rm p}$. Here the analogous frequency $\omega = cm(1+\Phi)$ arises for completely different reasons, but, as for EM waves in an inhomogeneous plasma, it varies with position, and this results in interesting refractive effects on the propagation of waves.

If we turn the dispersion relation around and use it to determine the wave-number $\mathbf{k}$ for a wave of a specified frequency ω we get $|\mathbf{k}|^2 = (\omega^2 - \omega_{\rm p}^2)/c^2$. This has real solutions – corresponding to travelling waves – only for $\omega > \omega_{\rm p}$. For $\omega < \omega_{\rm p}$, the wave-number $\mathbf{k}$ is imaginary and any fluctuations of the field at these frequencies are evanescent.

As one enters the ionosphere, the density of electrons rises at first and then decreases, so the plasma frequency $\omega_{\rm p} = \sqrt{nq^2/\epsilon_0 m_e}$ has a maximum, which, it turns out, is at $\nu_{\rm p} = \omega_{\rm p}/2\pi \simeq 30$MHz.

Terrestrial EM waves of frequency less than this get trapped and reflected as illustrated in figure 3. For a given temporal frequency, the spatial frequency decreases with altitude, so the wavelength increases, and it is this stretching of the wavelengths with height that causes the refraction of such waves[17]. This is how

[16]The same phenomenon is seen for de Broglie waves in a particle in a flat-bottomed potential well $V = -|V| =$ constant. There, Schrödinger's equation $-i\hbar\dot{\psi} = -\hbar^2\nabla^2\psi/2m - |V|\psi$ gives, for a travelling wave $\psi \propto e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$, the (inverse) dispersion relation for $k = |\mathbf{k}|$ as a function of ω: $k(\omega) = (\sqrt{2m}/\hbar)\sqrt{|V| - \hbar\omega}$, which is real – i.e. the wave is actually travelling rather then evanescent, with $\psi \sim e^{-kx}$ – only if $\hbar\omega < |V|$. But it implies that $v_{\rm p} = \omega/k(\omega)$ increases without limit as $\hbar\omega \Rightarrow |V|$ from below (i.e. as $k \to 0$). Similarly, for EM waves in a cold plasma, the inverse dispersion relation is $ck(\omega) = \sqrt{\omega^2 - \omega_{\rm p}^2}$. So again $v_{\rm p} = \omega/k(\omega) = c\omega/\sqrt{\omega^2 - \omega_{\rm p}^2}$ blows up as $\omega \Rightarrow \omega_{\rm p}$ (now from above) and, for that matter, is superluminal for all $\omega > \omega_{\rm p}$. But it was realised early on that this is not in any way unphysical. The speed at which energy or momentum (or information for that matter) can be carried (and this is seen most clearly by considering a wave-packet) is the *group-velocity* $v_{\rm g} = d\omega/dk$. For de Broglie waves this is $v_{\rm g} = \hbar k/m$ and tends to zero as $k \to 0$ (in contrast to $v_{\rm p}$). And, for EM waves in a plasma, $v_{\rm g} = c^2 k/\omega$ and again is perfectly regular as $k \to 0$ and is sub-luminal for all k.

[17]Note that, for such waves, the group velocity *decreases* with altitude. What matters for refraction is the phase-velocity, which increases with altitude, and one can think of the (wave-crests in the) upper part of of a beam from a transmitter as out-running those in the lower parts and thus causing the beam to turn.

'short-wave' radio transmissions can be detected around the world.

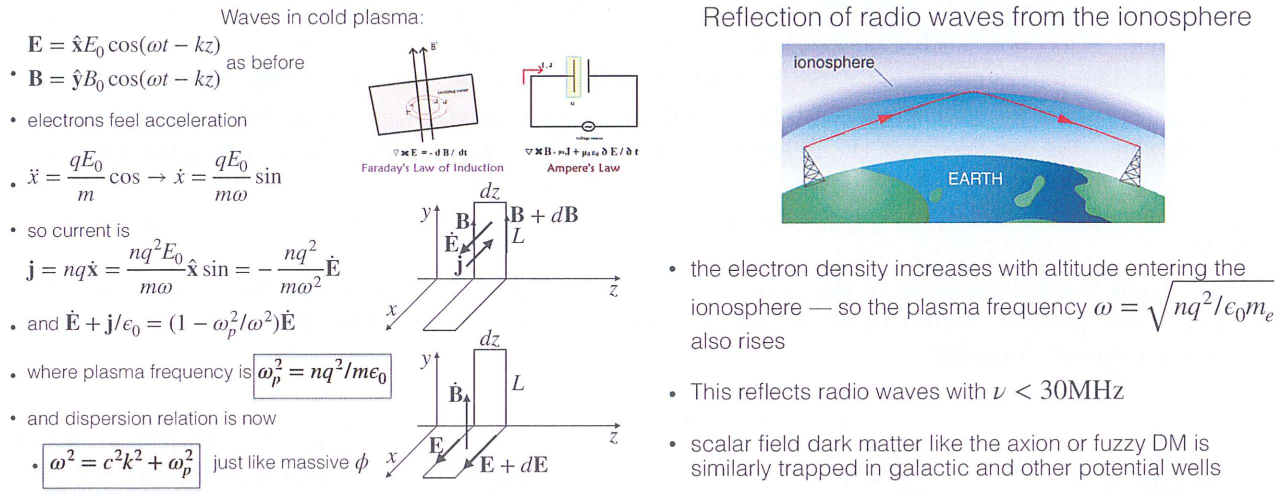

Figure 3: The left panel describes how the plasma frequency enters into the dispersion relation for radio waves. Waves of a given frequency from a transmitter have a wavelength – and phase velocity – that increases with height in the lower ionosphere. This causes the radiation to be reflected and trapped. Scalar matter waves are trapped in a gravitational potential in an analogous manner.

If the dark matter is the axion or an ultra-light axion-like field then it is trapped in the gravitational potential wells of galaxies and clusters etc. in much the same manner. The situation is somewhat different from short-wave radio waves in that whereas in the ionosphere the plasma frequency increases from zero to its maximum value and then drops again, in the galaxy the effective plasma frequency is everywhere very nearly constant, being equal to cm at infinity and with only a small suppression within bound systems (of at most about 1 part in 10^5 – this being in clusters of galaxies). So the situation we have is that, within a cluster of galaxies say, we have waves just above the local plasma frequency but just below its asymptotic value at infinity.

3.6 Refraction and focussing of scalar beams and wave-packets

In §3.3 and §3.2 we considered a scalar matter wave in a region of space within which the gravitational potential was considered to be constant. In §3.5 we argued, by analogy with EM waves in a cold plasma, that, if the potential is spatially varying, the waves would get refracted.

Here we will elaborate on this, and show how, in the 'geometric optics limit' – i.e. in the limit that the wavelength is very small compared to the scale over which the gravitational potential or field amplitude varies – the wave-vector $\tilde{k}$ of a beam or a wave packet obeys the geodesic equation. So a nearly monochromatic wave-packet falls under gravity in the same way as a material test particle, and if we were to launch ourselves to be initially moving with a packet, we would remain co-moving with it.

3.6.1 Klein Gordon equation in terms of log-amplitude and phase

We write the field as

$$\phi(\vec{x}) = \tfrac{1}{2} e^{a(\vec{x}) + i\Psi(\vec{x})} + \text{c.c.} \tag{57}$$

where $a(\vec{x})$ – the logarithm of the amplitude for a locally planar wave – is now position dependent, and $\Psi(\vec{x})$ is the phase, as before, but now we are allowing that the wave-vector $k_\mu = \Psi_{,\mu}$ may vary with position. There is clearly some freedom in how to do this as one cannot uniquely determine both $a(\vec{x})$ and $\Psi(\vec{x})$ from a single real function $\phi(\vec{x})$.

Taking the partial derivative with respect to x^μ gives $\phi_{,\mu} = \frac{1}{2}(i\Psi_{,\mu} + a_{,\mu})e^{a+i\Psi} + \text{c.c.}$, and a further derivative gives

$$\phi_{,\mu\nu} = \tfrac{1}{2}\{(i\Psi_{,\mu} + a_{,\mu})(i\Psi_{,\nu} + a_{,\nu}) + i\Psi_{,\mu\nu} + a_{,\mu\nu}\}e^{a+i\Psi} + \text{c.c.} \tag{58}$$

which, in the KG equation $g^{\mu\nu}\phi_{,\mu\nu} - \mu^2\phi = 0$, gives us

$$\{g^{\mu\nu}(-\Psi_{,\mu}\Psi_{,\nu} + 2i\Psi_{,\mu}a_{,\nu} + a_{,\mu}a_{,\nu} + a_{,\mu\nu} + i\Psi_{,\mu\nu}) - \mu^2\}e^{a+i\Psi} + \text{c.c.} = 0. \tag{59}$$

This may be solved if $\{\dots\}$ (as well as its conjugate) vanishes, as this gives us two equations for the two fields. Moreover, since $e^{+i\Psi}$ and $e^{-i\Psi}$ are rapidly varying functions, if we want the log-amplitude and phase to be slowly varying functions, we are forced to impose this. The KG equation then becomes the pair of coupled equations

$$\boxed{g^{\mu\nu}(-\Psi_{,\mu}\Psi_{,\nu} + a_{,\mu}a_{,\nu} + a_{,\mu\nu}) - \mu^2 = 0} \tag{60}$$

and

$$\boxed{g^{\mu\nu}(2\Psi_{,\mu}a_{,\nu} + \Psi_{,\mu\nu}) = 0.} \tag{61}$$

3.6.2 Refraction of scalar waves in the geometric optics limit

While somewhat complicated, the first of these simplifies in the limit that the wavelength is very small compared with the width of the beam or the size of the wave-packet L (and also compared to the scale over which the properties of the packet are changing), as the terms $a_{,\mu}a_{,\nu}$ and $a_{,\mu\nu}$ are on the order of $1/L^2$ whereas $g^{\mu\nu}\Psi_{,\mu}\Psi_{,\nu} + \mu^2$ is on the order of $1/\lambda^2$.

Dropping the terms involving a in (60) decouples it from (61) and gives

$$g^{\mu\nu}\Psi_{,\mu}\Psi_{,\nu} + \mu^2 = 0. \tag{62}$$

This doesn't look very useful as it seems simply to express the normalisation $k_\mu k^\mu = -\mu^2$. To obtain a differential equation for the rate of change of $k_\mu = \Psi_{,\mu}$ we can choose some point $\vec{x}$ as the origin of coordinates, and perform a Taylor expansion, letting $\Psi_{,\mu}(\vec{x}) = \overline{\Psi}_{,\mu} + x^\gamma\overline{\Psi}_{,\mu\gamma} + \dots$ where $\overline{\Psi}_{,\mu}$ and $\overline{\Psi}_{,\mu\gamma}$ denotes values at $\vec{x} = 0$, and $\dots$ denotes terms involving second or higher powers of x^μ. Similarly letting $g^{\mu\nu}(\vec{x}) = \overline{g}^{\mu\nu} + x^\gamma\overline{g}^{\mu\nu}{}_{,\gamma} + \dots$, we get, collecting terms that are either of zeroth or first order in x^μ,

$$(\overline{g}^{\mu\nu}\overline{\Psi}_{,\mu}\overline{\Psi}_{,\nu} + \mu^2) + x^\gamma(\overline{g}^{\mu\nu}{}_{,\gamma}\overline{\Psi}_{,\mu}\overline{\Psi}_{,\nu} + 2\overline{g}^{\mu\nu}\overline{\Psi}_{,\mu\gamma}\overline{\Psi}_{,\nu}) + \dots = 0. \tag{63}$$

The second term should vanish for arbitrary small x^γ, so we must have, using $\overline{\Psi}^{,\nu} = \overline{g}^{\mu\nu}\overline{\Psi}_{,\nu}$,

$$\overline{\Psi}^{,\mu}\overline{\Psi}_{\mu,\gamma} = -\tfrac{1}{2}\overline{g}^{\mu\nu}{}_{,\gamma}\overline{\Psi}_{,\mu}\overline{\Psi}_{,\nu} \tag{64}$$

or, with $\overline{\Psi}_{,\nu} = k_\nu$ and $\overline{\Psi}_{,\nu\gamma} = k_{\nu,\gamma}$, and, dropping the overline on the metric and using $g^{\mu\nu}{}_{,\gamma} = -g^{\mu\alpha}g^{\nu\beta}g_{\alpha\beta,\gamma}$,

$$\boxed{k^\mu k_{\gamma,\mu} = \tfrac{1}{2}g_{\mu\nu,\gamma}k^\mu k^\nu} \tag{65}$$

which we see is the (covariant form of the) geodesic equation.

Now since $k^\mu = p^\mu/\hbar = (m/\hbar)dx^\mu/d\tau$ (the τ being proper time as measured by the comoving particle) this says

$$\frac{dk_\gamma}{d\tau} = \frac{\hbar}{2m}g_{\mu\nu,\gamma}k^\mu k^\nu \tag{66}$$

which can be integrated to obtain the wave-vector along the packet path (and hence also the path itself).

Note that if the potential Φ is static, as assumed here, then $g_{\mu\nu,0} = 0$, and the time component k_0 – which is rate of change of phase Ψ with respect to coordinate time – is unchanging. This is fairly obvious in the plasma analogy; if the transmitter is broadcasting at some frequency ν then, if the plasma is static, the radiation everywhere will have the same frequency.

3.6.3 Focussing of beams and wave-packets

The second part of the KG equation (61) allows us, given a solution k_μ of the geodesic equation, to determine the rate of change of the log-amplitude along the beam or packet path. This says $g^{\mu\nu}k_\mu a_{,\nu} = k^\mu a_{,\mu} = -\frac{1}{2}g^{\mu\nu}k_{\mu,\nu} = -\frac{1}{2}g^{\mu\nu}k_{\mu;\nu}$, since, in the Lorenz gauge, $g^{\mu\nu}\Gamma^\alpha{}_{\mu\nu} = 0$. Thus

$$\frac{da}{d\tau} = -\frac{\hbar}{2m}g^{\mu\nu}k_{\mu;\nu} \tag{67}$$

which can be integrated to give $a(\tau)$.

The meaning of this is clearest in the inertial frame that is comoving with the beam (i.e. for which $v_{\rm g}^i = ck^i/k^0 = ck^i/\mu = 0$). In that frame this says $da/d\tau = -(c/2\mu)k^\mu{}_{,\mu}$ but $k^\mu{}_{,\mu} = k^0{}_{,0} + k^i{}_{,i} = k^i{}_{,i}$, since, again, $k^0 = \mu$ in that frame. And, in that frame, the divergence of the group velocity is $\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g} = c(k^i/k^0)_{,i} = ck^i{}_{,i}/k^0 - ck^ik^0{}_{,i}/(k^0)^2 = ck^i{}_{,i}/\mu$ (since $k^i = 0$). So a comoving observer would measure

$$\boxed{\frac{da}{d\tau} = -\tfrac{1}{2}\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g}} \tag{68}$$

and if we have a diverging beam, for example, the log-amplitude will decrease with time as measured by a comoving observer. This makes perfect sense; the energy density is $\mathcal{E} = \frac{1}{2}(\dot\phi^2/c^2 + \mu^2\phi^2)$ and $\phi = e^a\cos(\Psi)$ so, since we are assuming $\dot a \ll \dot\Psi = -\mu c$ this is simply $\mathcal{E} = \frac{1}{2}\mu^2 e^{2a}$ so $d\log\mathcal{E}/d\tau = \mathcal{E}^{-1}d\mathcal{E}/d\tau = 2da/d\tau$ and therefore

$$\frac{d\mathcal{E}}{d\tau} = -\mathcal{E}\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g} \tag{69}$$

which is the usual law of energy conservation[18].

Put yet another way, the rate of change of a small volume V bounded by a set of comoving observers – those who see no energy flux and are therefore moving at the local group velocity $\mathbf{v}_{\rm g}(\mathbf{x})$ – is given by the integral over the surface of $\hat{\mathbf{n}}\cdot\mathbf{v}_{\rm g}$ where $\hat{\mathbf{n}}$ is the outward normal. And by the divergence theorem, this is

$$\dot V = \int dA\,\hat{\mathbf{n}}\cdot\mathbf{v}_{\rm g} = \int d^3r\,\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g} \Rightarrow V\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g} \tag{70}$$

if we take the volume to be small so the variation of the divergence (and of $\mathcal{E}$) can be ignored. So (69) is saying $\dot{\mathcal{E}}/\mathcal{E} = -\dot V/V$, and therefore that $\mathcal{E}V = \text{constant}$[19]. And for a steady beam – i.e. not varying with time – there can be no variation of $\mathbf{v}_{\rm g}$ along the beam, and the area of the beam obeys $\dot A = A\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g}$ and $\mathcal{E}A = \text{constant}$.

3.7 Classical wave-packet – particle duality

We started with the Lagrangian density $\mathcal{L}$ for a massive scalar field characterised by a single parameter μ. This is a field for which the (bosonic) quantum excitations have mass $m = \hbar\mu/c$. Extremising the action gave us the Klein Gordon equation, which allows, as solutions, nearly monochromatic wave-packets, $\phi \sim e^{ik_\mu x^\mu}$, the wave-vectors for which obey the normalisation $k^\mu k_\mu = -\mu^2$.

Such packets have total energy and momentum $P^\mu = \int d^3r\,T^{0\mu} = \frac{1}{2}(L/2\pi)^3\int d^3k\,k^0k^\mu P_\phi(\mathbf{k})$ which we can write as $P^\mu = N\hbar k^\mu$, where $N = \frac{1}{2}\hbar^{-1}(L/2\pi)^3\int d^3k\,k^0 P_\phi(\mathbf{k})$ is the effective number of particles and, for free fields, is constant. Since $k^\mu k_\mu = -\mu^2$, the total 4-momentum obeys $P^\mu P_\mu = -M^2c^2$ where $M = Nm$.

And we have just seen that, in a gravitational field, a packet moves with velocity $\mathbf{v} \longrightarrow ck^i/k^0 = cP^i/P^0$, just like a massive particle with 4-momentum $\vec P$. And, in the geometric optics limit, which means taking μ – or equivalently the particle mass m – to be large: $\mu \gg 1/L$, the wave-vector $\vec k$ (and therefore also the total 4-momentum $\vec P$) obeys the geodesic equation[20]. So a massive test particle, if it starts with $\vec p$ such that $p^i/p^0 = v^i = k^i/k^0$, will follow the same trajectory as the packet. But of course it had to be that way by virtue of the equivalence principle; if we're in free fall, we shouldn't see a zero-momentum wave-packet in our vicinity start to move.

So there is a very close correspondence – or duality – between the behaviour of these classical waves and massive particles.

And, at the risk of anthropomorphising, it looks like the wave-packet is 'trying to find the path' that extremises some action. You might have been thinking that this would be the time integral of the Lagrangian $L = \int d^3x\,\mathcal{L}$. But that can't be right, as this actually vanishes for a packet in the monochromatic limit. That's quite different from the Lagrangian for a relativistic particle $L = -mc^2 d\tau/dt = -mc^2\sqrt{-g_{\mu\nu}\dot x^\mu\dot x^\mu}$ (where $\dot x^\mu = dx^\mu/dt = c(1, p^i/p^0)$). There is no way to write $\int d^3x\,\mathcal{L}$ as some function of cP^i/P^0. So what is the thing that the actual path of a wave-packet is extremising? It must be proportional to the integrated proper time measured by the comoving observer. And, as we saw, that observer sees the field oscillating at the proper frequency $\omega = \mu c = \omega_{\rm C}$. So the effective action for the packet is simply proportional the integrated phase; the paths that a packet can take are those that extremise this. The way this comes about

[18] If you were thinking this should be $\dot{\mathcal{E}} = -\boldsymbol{\nabla}\cdot(\mathcal{E}\mathbf{v}_{\rm g})$, which looks different, you'd be right. That is $\dot{\mathcal{E}} = -\mathcal{E}\boldsymbol{\nabla}\cdot\mathbf{v}_{\rm g} - \mathbf{v}_{\rm g}\cdot\boldsymbol{\nabla}\mathcal{E}$, but our observer has $\mathbf{v}_{\rm g} = 0$, so they're the same.

[19] Note that there is no PdV term here; an expanding – and possibly stretching – beam being analogous to a beam of particles with no velocity dispersion.

[20] And if we take a complex field coupled to the EM field with $\partial_\mu \Rightarrow D_\mu = \partial_\mu + i(q/\hbar)A_\mu$ it gets deflected just like a particle of charge q.

– in the picture developed here – is that if we have a beam of waves, for instance, propagating through a region with some static but spatially varying potential, then its coordinate time frequency remains constant but the coordinate wavelength changes and the beam is forced to change direction.

The opposite side of this is that particles, is seems, are behaving just like KG wave packets or beams of waves. The reason, in the Feynman and Dirac view, is that the 'classical system' is a particle that started at some place and time $(\mathbf{x}_0, t_0)$ and ends up somewhere else at a different time: $(\mathbf{x}, t)$. The classical action is $S(\mathbf{x}, t) = \int dt L = -mc^2 \int dt \sqrt{-g_{\mu\nu}\dot{x}^\mu \dot{x}^\mu}$ and the quantum mechanical amplitude is $\psi(\mathbf{x}, t) \sim e^{iS/\hbar}$. But $S(\mathbf{x}, t)$ obeys the classical Hamilton-Jacobi relations $S_{,\mu} = p_\mu$ where $p_\mu = (-H/c, \mathbf{p})$ with, as usual, $\mathbf{p} \equiv \partial L/\partial \dot{\mathbf{x}}$ and $H \equiv \dot{\mathbf{x}}\mathbf{p} - L$, and which, for this L, enforces the normalisation $g^{\mu\nu} p_\mu p_\nu = g^{\mu\nu} S_{,\mu} S_{,\nu} = -m^2c^2$. These relations force the QM amplitude to obey the Klein-Gordon equation.

3.8 Speckly nature of scalar DM in the multi-streaming regime

If we have a potential well – that of a spherical mass concentration say – and we have test particles released from rest then they will fall into the potential. A scalar wave that starts off spatially homogeneous – corresponding to zero momentum particles – will, initially, start to develop infall momentum in exactly the same way. As illustrated in figure 4, the wave will oscillate at a lower frequency (as a function of coordinate time, that is) deeper into the potential. Different parts of the wave will get out of phase with each other. So the initially spatially homogeneous field will develop ripples. Unsurprisingly these give the field an inward directed momentum density and, guess what? it is just that that the corresponding beam of particles would develop.

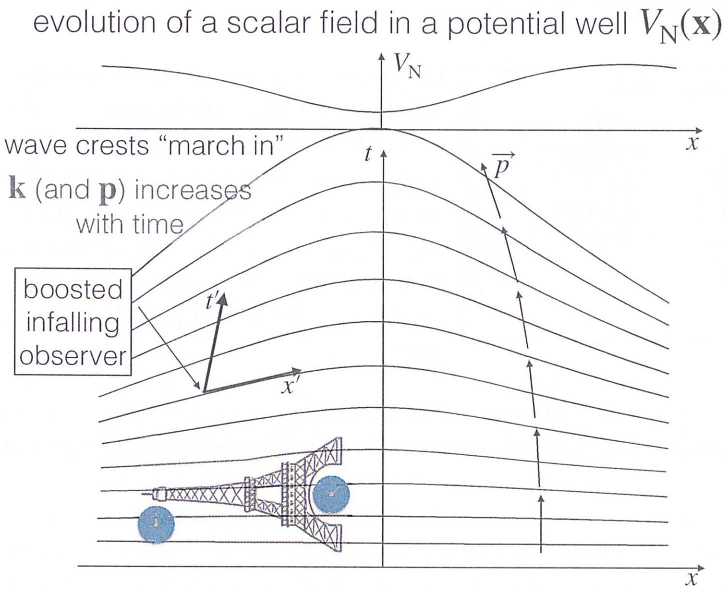

Figure 4: Illustration of how a scalar field 'falls' into a potential well. The upper plot shows the gravitational potential we have in mind, and the Eiffel tower at the bottom is there to remind us which way is up. Also shown are two clocks that, being at rest in our coordinate system and therefore being accelerated, are drifting out of synch. The curves in the space-time diagram are iso-Ψ surfaces. Deep in the potential, the Compton frequency $\omega_C(\mathbf{x})$ is reduced, so the iso-phase surfaces become advanced. The field, which initially had $\nabla\phi = 0$ will develop spatial ripples with the wave-fronts moving towards the centre; the field is developing in-fall momentum. An in-falling observer, will see $\nabla\phi = 0$ in his frame.

But the particles will, eventually, reach the centre of the potential, and they will meet up with particles that are coming the other way. We say that a 'multi-streaming' region has developed. What happens to the waves in the same situation?

The answer is that there will be interference. This is illustrated in figure 5 which shows a space-time diagram of the development of a multi-streaming region for particles that have some smooth initial velocity field that is focussing the particles.

Provided only that the velocity field is smooth, this leads to the formation of 'caustics'. These are two dimensional surfaces in space on which the density of particles (or light rays from a point source) is infinite. These are seen on the bottom of a swimming pool on a sunny day. One can easily show (see below) that the density falls off inversely as the square root of the distance from the caustic for these so-called 'fold-catastrophes'. The classification and study of universal nature of such singularities is called 'catastrophe theory' and was developed in the Soviet Union by V.I. Arnold.

We saw that a locally planar wave $\phi \propto \mathrm{Re}(e^{i\Psi})$ solves the KG equation (and that, in the Hamilton-Jacobi set-up, the phase is $\Psi = S/\hbar$ where S is the action). The iso-phase surfaces are orthogonal to the 4-velocities of the particles (in the special relativistic sense).

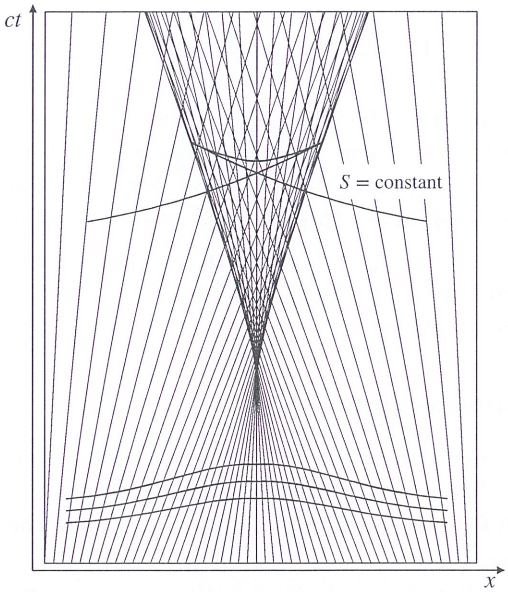

Figure 5: Space-time diagram of trajectories of particles (thin lines) that are focussing shows that, in general, caustics will form. The caustics bound the multi-streaming region and lie at the turning points of the mapping $x_t(x_{t_i})$. Note that the density ρ of trajectories becomes very large – actually infinite as $\rho \propto 1/\sqrt{x}$ – approaching the caustics from inside the multi-stream region. The heavy curves show hypersurfaces of constant action S for these particles. These are orthogonal to the trajectories of the particles in the special relativistic sense. Outside the caustics, there is, locally, a single beam of particles. The corresponding scalar field is, locally, $\phi \propto \text{Re}(e^{iS(\vec{x})/\hbar})$. This oscillates, but the energy density – which is $\mathcal{E} \simeq \frac{1}{2}(\dot{\psi}^2/c^2 + \mu^2\phi^2)$ – is smoothly varying. Inside, in the multi-streaming region, we have a superposition of multiple beams (three here). These beams will interfere wave-mechanically and the density will show interference patterns. If we have many overlapping beams the energy density develops a universal 'speckly' pattern.

So outside multi-streaming region the wave-fronts of the field look like the iso-S surfaces shown at the bottom. Inside it is very different; we have the superposition of multiple – here three – waves, and we can expect that there will be interference. This is illustrated in figure 6.

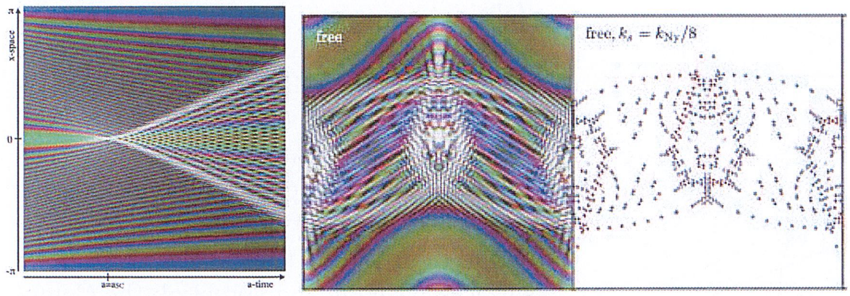

Figure 6: Left panel shows a wave-function closely analogous to the particle caustics in figure 5. The middle plot is for a more complex collapse, and the right panel shows the locations of the phase-vortices (red and blue for vortices and anti-vortices). From Uhlemann *et al.*, 2019.

This is illustrated in figure 8 which shows the result of a numerical simulation of structure formation in a universe dominated by so-called 'fuzzy' dark matter (a classical scalar field with Compton wavelength of order a fraction of a parsec).

What about the universal $\rho \propto 1/\sqrt{x}$ density of rays close to the caustics? How does that arise wave-mechanically? And how do wave-mechanical effects change the behaviour very close to the caustics (which, one feels, must somehow be 'softened')? The latter is seen in the numerical calculations shown in figure 6 but can also be understood qualitatively analytically as follows:

The behaviour of scalar fields in this regard is closely analogous to the illumination pattern seen on the bottom of a swimming pool. In geometric optics, the flux density is proportional to the inverse of the determinant of the tensor $\partial_i\partial_j h$, where $h(\mathbf{r})$ where the vertical deviation of the wavefront. Caustics arise at points $\mathbf{x}$ on the pool bottom that are the images of lines on the incoming wavefront where one of the eigenvalues of $\partial_i\partial_j h$ vanishes. The situation, in a slice along the corresponding direction, is illustrated in figure 7.

At a general point x, the field strength is proportional to the area A of the 'Fresnel zone' within which the phase-difference for paths to the point in question vary by ~ 1 radian or less so there is constructive interference.

This area is $A \sim r_0 r_\perp$ where, as indicated in the figure, $r_0 \sim \sqrt{\lambda/h''}$, ($\lambda$ being the wavelength of the light), and where $r_\perp$ is the analogous distance in the other direction for which $h''_\perp r_\perp^2 \sim \lambda$. So A is proportional to r_0. We are assuming here, for simplicity, that $|h''| \gg 1/D$.

The field strength on the pool bottom is proportional to A, and the energy flux density $\mathcal{E}$, being proportional to the field squared, is proportional to A^2. Close to a turning point, $h'' = 0 + h'''r + \ldots$, where we are now measuring distance from the point r where $h''(r) = 0$. The differential of the mapping is $dx/dr = Dh'' = Dh'''r + \ldots$ (which is also the inverse of the linear magnification $M_\| = dr/dx$; the to-

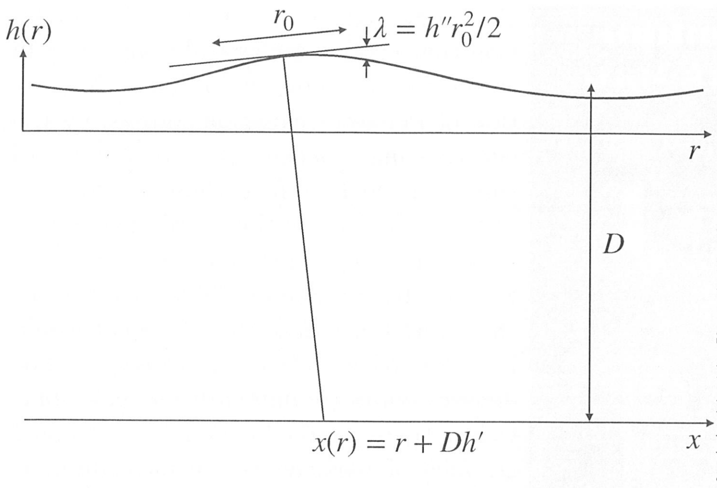

Figure 7: Light deflection in a swimming pool. The curve depicts a wave-front from a distant source. We see here a slice along one eigen-axis of the tensor $\partial_i \partial_j h$. In the geometric optics limit, rays – like the one indicated – are perpendicular to the wavefront and give the mapping $x(r) = r + Dh'(r)$. Points x near caustics are close to turning points (where $dx/dr = Dh''$ vanishes). Wave-mechanically, the field at a point x receives significant contribution only from an area of the incoming wave-front where the phase difference is less than ~ 1 radian. The field strength is proportional to the area A of this 'Fresnel-zone' and the flux density is proportional to A^2.

tal magnification being $M = M_\parallel M_\perp$). Integrating this gives, for small distances from the turning point, $x(r) = \int dr \times dx/dr = \frac{1}{2} Dh''' r^2$ or $r = \sqrt{2x/Dh'''}$, where x is measured from the caustic location.

The energy density varies as $\mathcal{E} \propto A^2 \propto r_0^2 \propto 1/h'' = 1/h'''r = 1/\sqrt{2xh'''/D}$. So this blows up as one approaches the 'classical' (i.e. geometric optics limit) caustic surface as $\mathcal{E} \propto 1/\sqrt{x}$ in just the same manner as the density of particles (and, of course, just like $M_\parallel$ also).

What is happening here is that the 'Fresnel-area' A is blowing up – actually getting stretched out along the singular eigen-direction – as one approaches the classical caustic. We have only considered the energy coming from one such zone. More precisely, what we should do is add up the field from all of the zones and then compute its squared modulus, and we would then have the interference pattern modulating the $\mathcal{E} \propto 1/\sqrt{x}$ fall-off. As one crosses a classical fold caustic, two such zones appear, which, in the geometric optics limit, have infinite magnification and opposite parity. They rapidly move apart and shrink[21].

The $\mathcal{E} \propto 1/\sqrt{x}$ divergence is cut-off in the wave-mechanical treatment as the analysis above is valid only if $r_0 \lesssim r$. For sufficiently small r (and hence correspondingly small x) the two Fresnel zones merge into one. This merging happens when $r_0 \sim \sqrt{\lambda/h''} \sim \sqrt{\lambda/h'''r} \sim r$, or $r \sim (\lambda/h''')^{1/3}$ for which the linear magnification is $M_\parallel = 1/Dh'''r \sim 1/D\lambda^{1/3}(h''')^{2/3}$. So this (times $M_\perp$) is the maximum flux-density magnification, and it occurs over a region of width $x_c \sim Dh'''r^2 \sim D\lambda^{2/3}(h''')^{1/3}$. Both of these are rendered finite if λ is finite.

Much more elaborate, and computationally expensive, calculations have been performed, as illustrated in figure 8 and with the field evolving self-consistently within the potential generated by its energy density. These were performed using the Schrödinger equation, which we turn to now.

3.9 Evolution of scalar fields via the Schrödinger equation

The Klein-Gordon equation was proposed by Schrödinger to describe ψ, the quantum mechanical wave function of a particle. He obtained it by replacing H and $\mathbf{p}$ in the relativistic energy-momentum relation by the operators $i\hbar\partial_t$ and $-i\hbar\nabla$. In response to the problem that the probability density $\rho = \psi\psi^\star$ is not generally conserved he dropped this in favour of what we usually call the Schrödinger equation, which is obtained by taking the non-relativistic limit. Here the KG equation is considered as that obeyed by a classical scalar field ϕ but, if we are considering such a field as the dark matter, we can similarly use the Schrödinger equation. This is useful in numerical simulation as the Schrödinger field evolves less rapidly than the scalar field. It is also useful conceptually as it shows that, in this limit, the field has, in addition to the 4-conserved quantities $\int d^3r T^{0\mu}$ (the total energy and 3-momentum), a 5th conserved quantity that corresponds to particle number. It is also useful as it makes it somewhat simpler to understand phenomenology such as the speckly nature of the DM in the multi-streaming regime.

[21] This leads to two theorems that are useful in gravitational lensing. The first is that the number of images of a source is always odd and the second is that the probability to have high amplification is asymptotically $P(> M) \propto 1/M^2$

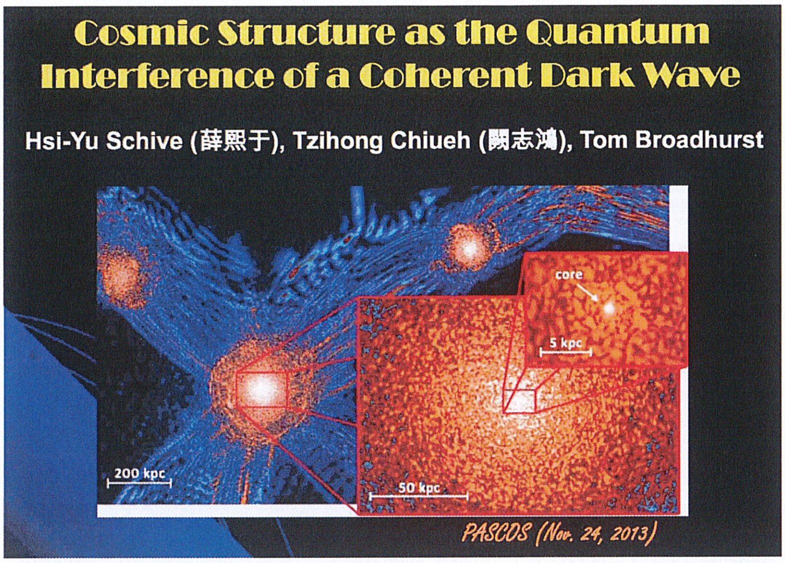

Figure 8: Result of a numerical simulation that evolves a classical scalar field in the gravitational potential that is the solution of Poisson's equation sourced by the effective mass density $\rho = m^2\langle\phi^2\rangle/c^2$ of the scalar field. In fact, these results were obtained by solving the Schrödinger equation, and using $\psi psi^\star$ for the density. In the low density regions (blue) one can see the interference of 3-waves in what would be, for particles, the 3-stream region. The denser regions (orange) are where, for particles, there would be many overlapping streams of particles so we have interference of multiple 'beams' and this is what gives rise to the characteristic speckly pattern.

3.9.1 From the Klein-Gordon equation to the Schrödinger equation

Our starting point is the KG equation in the non-relativistic limit

$$-\ddot{\phi} + c^2\nabla^2\phi - (1 + 2\Phi(\mathbf{x}))\omega_C{}^2\phi = 0. \tag{71}$$

For scalar DM waves in galaxies etc., the terms involving $\nabla^2\phi$ and Φ are very small compared to the others, so, to a first approximation, we get solutions that are $\phi \sim \mathrm{Re}(a(\vec{x})e^{-i\omega_C t})$ with rapid time oscillations and where the amplitude $a(\vec{x})$ is very slowly varying.

We get the Schrödinger equation by 'factoring out' the rapid time variation and, replacing $\omega_C \Rightarrow \omega$ here for clarity, writing

$$\phi(\mathbf{x}, t) = \psi(\mathbf{x}, t)e^{-i\omega t} + \mathrm{c.c.} \tag{72}$$

where the complex field $\psi(\mathbf{x}, t)$ is assumed to be slowly varying and c.c. denotes complex conjugation so, by construction, ϕ is real.

Taking the time derivative gives

$$\dot{\phi} = (-i\omega\psi + \dot{\psi})e^{-i\omega t} + \mathrm{c.c.} \tag{73}$$

in which we expect the $|\dot{\psi}| \ll |\omega\psi|$.

Taking a further time derivative yields

$$\ddot{\phi} = (-\omega^2\psi - 2i\omega\dot{\psi} + \ddot{\psi})e^{-i\omega t} + \mathrm{c.c.} \tag{74}$$

where the terms in parentheses are in order of decreasing magnitude.

The Laplacian of the field is simpler:

$$\nabla^2\phi = \nabla^2\psi e^{-i\omega t} + \mathrm{c.c.} \tag{75}$$

With these substitutions in the KG equation, we see that the leading order term in $\ddot{\phi}$ cancels the $-\omega^2\phi$ term and we have

$$(2i\omega\dot{\psi} - \ddot{\psi} + c^2\nabla^2\psi - \Phi(\mathbf{x})\omega^2\psi)e^{-i\omega t} + \mathrm{c.c.} = 0 \tag{76}$$

Now $|\ddot{\psi}| \ll |2i\omega\dot{\psi}|$, so we can drop $\ddot{\psi}$ above, and since ψ is supposed to be relatively slowly varying compared to $e^{i\omega t}$ this requires that both the coefficient of $e^{i\omega t}$ and of $e^{-i\omega t}$ must vanish, which means that

$$2i\omega\dot{\psi} + c^2\nabla^2\psi - 2\Phi(\mathbf{x})\omega^2\psi = 0. \tag{77}$$

Now $\omega = mc^2/\hbar$ and $\Phi(\mathbf{x})$ is the dimensionless Newtonian potential, so the Newtonian potential energy for a particle of mass m is $V(\mathbf{x}) = c^2 m\Phi(\mathbf{x})$. With these substitutions, and multiplying by $\hbar^2/2mc^2$ gives us

$$\boxed{i\hbar\dot{\psi} = -\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{x})\psi} \tag{78}$$

which we recognise as Schrödinger's equation for a particle of mass m in a potential $V(\mathbf{x})$, which can, alternatively be stated as

$$E\psi = H(\mathbf{p}, \mathbf{x})\psi \tag{79}$$

where

$$E \Rightarrow i\hbar\partial_t \qquad \text{and} \qquad H = |\mathbf{p}|^2/2m + V(\mathbf{x}) \qquad \text{with} \qquad \mathbf{p} \Rightarrow -i\hbar\boldsymbol{\nabla}. \tag{80}$$

It may seem strange that we have been able to replace the KG equation, which is second order in time, and therefore requires that one specify both ϕ and $\dot{\phi}$ as initial conditions to obtain a solution, by one that is first order in time, and therefore only requires that one specify the initial field ψ. But this is quite reasonable when we count degrees of freedom since ψ has both a real and imaginary part. Also, one may note that our starting point (72) does not allow one to determine ψ given the initial field ϕ alone, but, if augmented by $\dot{\phi} = -i\omega\psi e^{-i\omega t} + i\omega\psi^\star e^{i\omega t}$, which we obtain by taking the dominant terms in (73), we have, at $t = 0$, $\psi = (\phi - \dot{\phi}/i\omega)/2$.

While this looks like quantum mechanics, it isn't. We have simply converted the single second order classical KG equation for $\phi(\mathbf{x}, t)$ into a pair of coupled 1st order equations for the equally classical fields $\mathrm{Re}(\psi(\mathbf{x}, t))$ and $\mathrm{Im}(\psi(\mathbf{x}, t))$. Planck's constant only enters here because we have chosen to re-parameterise $\mu = mc/\hbar$.

While this is equivalent – for non-relativistic fields – to the KG equation, it is useful nonetheless. First, one can apply intuition gained in QM courses. Second, in numerical simulations like whose results are illustrated in figure 8, it is more efficient to compute the evolution of the relatively slowly evolving Schrödinger field.

In such simulations, the potential $V(\mathbf{x})$ is assumed to be determined by the energy density of the scalar field. This is $\mathcal{E} = \frac{1}{2}(\dot{\phi}^2/c^2 + |\boldsymbol{\nabla}\phi|^2 + \mu^2\phi^2)$. The middle term is negligible compared to the other two. Taking $\dot{\phi} \simeq -i\omega_C\psi e^{-i\omega_C t} + \text{c.c.}$ (i.e. dropping the relatively negligible term $\dot{\psi} + \text{c.c.}$) and with $\mu = \omega_C/c$ we get $\mathcal{E} = 2\mu^2\psi\psi^\star$ so $\rho = \mathcal{E}/c^2 = 2\omega_C{}^2\psi\psi^\star$. So the equations solved are the *coupled Schrödinger-Poisson system* which are (78) coupled to $\nabla^2\varphi = 4\pi G\rho$ with $V = m\varphi$.

3.9.2 The 5th conservation law: conservation of particle number

A scalar field obeys, in general, 4 conservation laws (or continuity equations); those of energy and the 3 components of spatial momentum. But, for non relativistic particles – where the energy of the particles is too small to create new particles in collisions – there is a 5th conservation law; that of the number of particles.

The corresponding law for a non-relativistic *field* is the law of conservation of total probability (if we think of ψ as a wave-function whose squared modulus gives the probability to find the particle). This is

$$\dot{\rho} + \boldsymbol{\nabla} \cdot \mathbf{j} = 0 \tag{81}$$

where

$$\begin{aligned} \rho &= \psi\psi^\star \\ \mathbf{j} &= \frac{i\hbar}{2m}(\psi^\star\boldsymbol{\nabla}\psi - \psi\boldsymbol{\nabla}\psi^\star) \end{aligned} \tag{82}$$

the latter being the usual Schrödinger 3-current density.

3.9.3 Speckles and phase-vortices from the Schrödinger perspective

The Schrödinger formalism gives an interesting perspective on the structure of the density field as illustrated in figure 9 and described in the caption.

It is not difficult to show that, if we write the field as $\psi = \sqrt{\rho}e^{i\theta}$, the phase θ will wrap by 2π if one follows a loop around one of the lines where ρ vanishes; these lines are *'phase-vortices'*. We will discuss this further below.

One can also show that the Schrödinger current is proportional to ρ times the gradient of θ. That means that if we define the velocity as $\mathbf{v} = \mathbf{j}/\rho$, this is divergent; tending to infinity inversely with distance from the $\rho = 0$ line.

As an aside, the same speckly pattern arises in a seemingly very different context; astronomical optics. Astronomers on Earth observing distant sources are looking through an atmosphere that is introducing

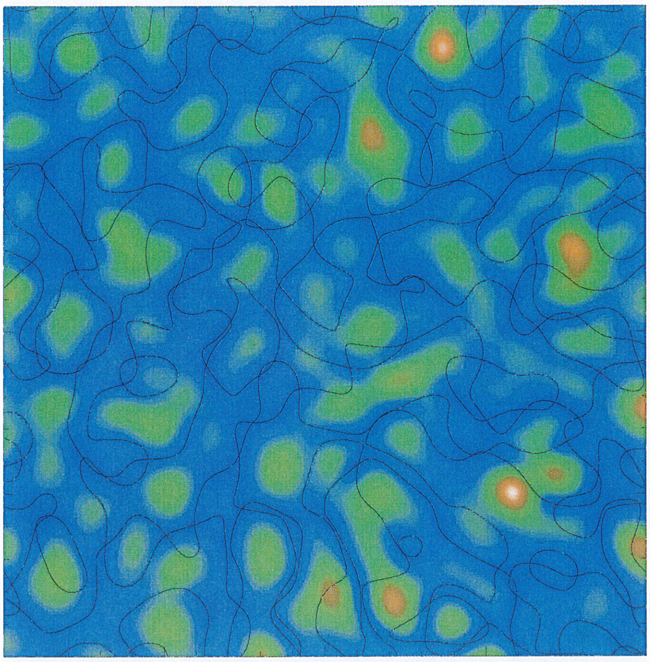

Figure 9: In the strongly multi-streaming regime the Schrödinger field will be the sum of many independent 'beams' coming from various directions. The real and imaginary parts of ψ will then behave – by virtue of the central limit – as Gaussian random fields. The coherence length of these randomly fluctuating field is on the order of the de Broglie wavelength. The real part $\mathrm{Re}\,\psi$ will vanish on one set of 2-dimensional surfaces and $\mathrm{Im}\,\psi$ vanishes on another. In the single stream region these are interleaved so the density $\rho = |\psi|^2$ can never vanish, and is smooth in that region. But in a multi-stream region, $\mathrm{Re}\,\psi$ and $\mathrm{Im}\,\psi$ become effectively statistically independent, and their zero surfaces will cross, which means that ρ will vanish on a set of lines. This is illustrated at left, where the colour image is the density $\rho = |\psi|^2$ on a 2-dimensional slice, and the lines are contours of zero $\mathrm{Re}\,\psi$ and $\mathrm{Im}\,\psi$. Vortices exist at the crossing points, and are lines in three dimensions.

random deformation of the wavefronts. In the so-called para-axial approximation this is modelled by writing the EM field at the entrance aperture of the telescope as the real part of a complex 2-dimensional field $E(\mathbf{x}) \sim E_0 e^{i\psi(\mathbf{x})}$ where the 'phase-error' ψ is $2\pi h(\mathbf{x})/\lambda$ where $h(\mathbf{x})$ is the height of the wavefront deformation. The EM field $E(\mathbf{r})$ on the focal plane is the convolution of the incoming field with a certain kernel – the so-called 'Fresnel kernel' – which, it turns out, is identical in form to the Green's function of the Schrödinger equation. So a telescope can be thought of as an analogue computer that is solving Schrödinger's equation in 2D. And the energy density on the detector is $E(\mathbf{r})E(\mathbf{r})^\star$. The upshot is that if the diameter of the telescope aperture is N times the 'coherence-length' for the phase-fluctuations – this is called the 'Fried length' after David Fried[22] and is the distance $|\mathbf{d}|$ such that $\langle(\psi(\mathbf{x}) - \psi(\mathbf{x}+\mathbf{d}))^2\rangle^{1/2} \simeq 1$ – then the 'point-spread function' will consist not of one sharp diffraction limited peak, but of $\sim N^2$ speckles (these dance around as the atmosphere drifts across the sky, so you need to take a short exposure to see them).

3.9.4 Madelung's formalism

Schrödinger's equation is $E\psi = H\psi$ where $H = H(\mathbf{p}, \mathbf{x}, t)$ is the Hamiltonian operator and $E \Rightarrow i\hbar\partial_t$ and $\mathbf{p} \Rightarrow -i\hbar\boldsymbol{\nabla}$. For a particle in a potential $V(\mathbf{x}, t)$, so $H = |\mathbf{p}|^2/2m + V$, it is, with $\partial_t\psi \Rightarrow \dot{\psi}$,

$$i\hbar\dot{\psi} = -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi. \tag{83}$$

This implies that the time derivative of $\rho \equiv \psi^\star\psi$ is

$$\dot{\rho} = \frac{i\hbar}{2m}\boldsymbol{\nabla}\cdot(\psi^\star\boldsymbol{\nabla}\psi) + \text{c.c.}$$

which, we note, does not contain the potential.

So, defining

$$\mathbf{p}(\mathbf{x}, t) \equiv \frac{i\hbar}{2m}\psi\boldsymbol{\nabla}\psi^\star + \text{c.c.} \tag{84}$$

and

$$\mathbf{v}(\mathbf{x}, t) \equiv \mathbf{p}/\rho \tag{85}$$

we have a continuity equation

$$\dot{\rho} + \boldsymbol{\nabla}\cdot(\rho\mathbf{v}) = 0 \tag{86}$$

[22] Fried was also the first to point out the existence of the phase-vortices of the EM field on the focal plane. These pose a problem for engineers building devices to do 'adaptive optics' where one tries to correct the phase errors by bouncing the light off a 'rubber mirror' with actuators that aim to undo the effect of the atmosphere.

which is just like continuity of mass for a compressible fluid with density ρ and flow velocity $\mathbf{v}$.

Integrating $\dot{\rho}$ over all space implies the conservation law

$$\frac{d}{dt}\int d^3r\rho = \int d^3r\,\dot{\rho} = -\int d^3r\,\nabla\cdot(\rho\mathbf{v}) = 0,$$

which allows the interpretation of ρ as a probability density and $\mathbf{p} = \rho\mathbf{v}$ as the probability flux density (here we will consider them to be mass density and momentum density).

The velocity defined in (85) has some interesting properties. If we write the complex field as

$$\psi(\mathbf{x}, t) = \sqrt{\rho(\mathbf{x}, t)}e^{iS(\mathbf{x},t)/\hbar} \tag{87}$$

and substitute this in (84) we find

$$\mathbf{p} = \frac{i\hbar}{2m}\sqrt{\rho}(\nabla\sqrt{\rho} - \sqrt{\rho}i\nabla S/\hbar) + \text{c.c.} = \frac{\rho}{m}\nabla S$$

(the first term, being purely imaginary, vanishes when added to its complex conjugate) so the velocity is

$$\mathbf{v} = \frac{1}{m}\nabla S = \frac{\hbar}{m}\nabla\theta$$

where we are defining the phase $\theta \equiv S/\hbar$. Thus the velocity defined by (85) is a 'potential flow'. One might be tempted to conclude[23] that the vorticity, and therefore also the circulation $\oint d\mathbf{l}\cdot\mathbf{v}$, vanishes. But that is not correct. As we have discussed, in the multi-stream region we expect to find lines in space where both $\text{Re}(\psi)$ and $\text{Im}(\psi)$ vanish (these lines in space being the intersection of the two dimensional surfaces where either one or other vanishes). The velocity is not defined on these lines and, quite generally, the phase θ will wrap by $\pm 2\pi$ as one traverses a small loop threaded by one of these 'phase vortices'. This is illustrated in figure 10 where we show the phase for a pair of phase vortices[24].

Thus we have for a general loop $\oint d\mathbf{l}\cdot\mathbf{v} = 2\pi n\hbar/m$ where n is the number of vortices minus the number of anti-vortices that thread the loop. For a small circular loop of radius r around a vortex, $|\mathbf{v}| = \pm\hbar/mr$ and diverges as $r \to 0$. And for $r < \hbar/mc$ (i.e. less than the Compton wavelength divided by 2π) the velocity is superluminal. This might seem unreasonable, but there is nothing really physically divergent. The density ρ vanishes on these lines, and tends to zero as $\rho \propto r^2$, so while $\mathbf{v}$ is divergent, the momentum density $\rho\mathbf{v}$ is finite (it is actually zero on the vortex line).

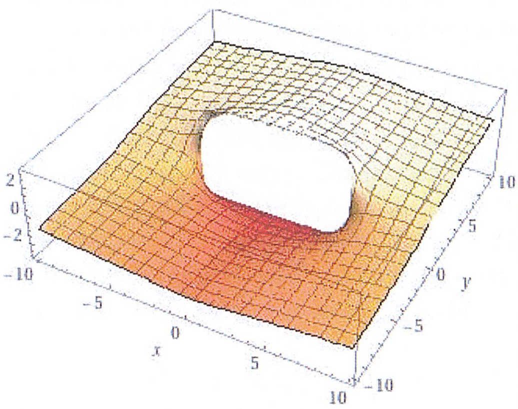

Figure 10: The height of the surface is the phase θ for a pair of parallel phase vortices of opposite winding number. The x and y coordinates are in a plane perpendicular to the vortices which lie at $(x, y) = (\pm 5, 0)$. Phase $-\pi$ and $+\pi$ are to be identified, so the apparent discontinuity on the line joining the vortices is not real; the gradient of the phase $\nabla\theta$ – and therefore also the velocity – is regular there. But close to either vortex the velocity – being $\mathbf{v} = (\hbar/m)\nabla\theta$ – diverges. For a loop that encloses either one of the vortices, the phase 'winds', but in opposite senses for the two vortices. A loop that encloses both vortices (or neither) has zero winding number.

Madelung, in his 1927 paper *'Quantentheorie in Hydrodynamischer Form'*, showed that (83) can be recast in a form very reminiscent of Euler's equation for a fluid. He found that the convective derivative $d\mathbf{v}/dt \equiv \dot{\mathbf{v}} + (\mathbf{v}\cdot\nabla)\mathbf{v}$ is given by

$$\boxed{\frac{d\mathbf{v}}{dt} = -\frac{1}{m}\nabla(V + Q)} \tag{88}$$

[23] The english translation of Madelung's paper defines the 'flux' u to be the phase gradient, so the same as $\mathbf{v}$ here, and says "the flux u is vortex free".

[24] One can show that as the multi-stream region develops, vortices are created in pairs of opposite handedness. They are just like cosmic strings in that respect.

where

$$\boxed{Q \equiv -\frac{\hbar^2}{2m}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}}} \tag{89}$$

is known as the 'quantum potential' (or the Bohm potential), and $-\boldsymbol{\nabla} Q$ is called the 'quantum force'.

We may compare (88) with the Euler equation giving the convective time derivative of the mean streaming velocity of particles in a Newtonian gravitational potential φ (so $V = m\varphi$)

$$\frac{d\mathbf{v}}{dt} = -\boldsymbol{\nabla}\varphi - \frac{1}{\rho}\boldsymbol{\nabla}\cdot\mathbf{P} \tag{90}$$

where $\mathbf{P}$ is the pressure tensor (being isotropic in the limit of an inviscid fluid). It is not difficult to show that the 'quantum acceleration' $-m^{-1}\boldsymbol{\nabla} Q$ in (88) is $1/\rho$ times the gradient of the 'quantum pressure tensor' $\mathbf{P}$ with components

$$P_{ij} = -\left(\frac{\hbar}{2m}\right)^2 \rho\partial_i\partial_j \ln\rho. \tag{91}$$

Madelung was thus able to show that the Schrödinger equation (83) for a particle in a potential $V(\mathbf{x})$ is equivalent to the continuity and Euler equations (86) and (88) for ρ and $\mathbf{v}$, together with the definition of Q in (89). Thus, given some initial wave function $\psi(\mathbf{x}, t_\mathrm{i})$, we could, in principle, calculate the initial density $\rho(\mathbf{x}, t_\mathrm{i}) = |\psi|^2$ and velocity $\mathbf{v}(\mathbf{x}, t_\mathrm{i}) = (\hbar/m)\boldsymbol{\nabla}\arctan(\mathrm{Im}(\psi)/\mathrm{Re}(\psi))$, and then evolve these using the Euler and continuity equations. The resulting final velocity $\mathbf{v}(\mathbf{x}, t_\mathrm{f})$ would furnish us with the gradient of the phase $\boldsymbol{\nabla}\theta(\mathbf{x}, t_\mathrm{f})$, which could be spatially integrated to give $\theta(\mathbf{x}, t_\mathrm{f})$ (up to a constant of integration) and thereby the final Schrödinger wave-function $\psi(\mathbf{x}, t_\mathrm{f}) = \sqrt{\rho}e^{i\theta}$. This should, in principle, be identical to the result of evolving $\psi(\mathbf{x}, t)$ using (83).

This parallel between quantum mechanics and fluid mechanics is an interesting one, but it is also somewhat controversial; some people seem to find it offensive. Perhaps this is because the complex wave-function – which has to be squared to compute probabilities etc. – has been replaced by the real functions ρ and S. Pauli mentioned it in one of his famous reviews of quantum physics in the *Handbuch der Physik*, but dismissed it as being a mathematical curiosity and not of great interest. Aside from its connection to Bohm, it did not play a major role in the development of quantum mechanics.

Here we are interested in the Schrödinger equation as an approximate description – an extremely good one – of classical scalar fields in gravitational potentials, on which the Madelung formalism provides a usefully different perspective. In this regard, we may note that, in a qualitative sense, the form of the 'quantum pressure' is quite reasonable. In the infall region outside of a forming galaxy, where one would be, in the particle picture, in the single stream regime, the corresponding Schrödinger wave would have slowly varying amplitude $\sqrt{\rho}$. The pressure is then negligible and the velocity would be everywhere regular – there being no phase-vortices in this region as the surfaces of vanishing $\mathrm{Re}(\psi)$ and $\mathrm{Im}(\psi)$ are interleaved and are therefore non-intersecting – and the momentum density would increase with time just like that of cold infalling particles. In the multi-stream region, on the other hand, interference will result in rapid variation of $\sqrt{\rho}$. Note that, given the speckly nature of ρ, one would expect there to be an inverse correlation between the amplitude and its Laplacian, so it seems reasonable that Q and the pressure (or momentum flux density) would be tend to be positive. The size of the speckles being on the order of the (de Broglie) wavelength $\lambda_\mathrm{dB} \sim \hbar/m\sigma_v$, in order of magnitude $P \sim \rho(\hbar/m)^2/\lambda_\mathrm{dB}^2 \sim \rho\sigma_v^2$, just like the pressure particles would have in this region.

But going beyond this there are some puzzling features of the Madelung approach. Particles in a gravitating structure – stars in our galaxy, for instance – are described by a distribution function $f(\mathbf{v})$; there is a range of velocities at each point in space. The wave-mechanical model, be it the KG, Schrödinger or Madelung equations, should, in the limit that the mass m is large – i.e. in the geometric optics limit – reproduce this. But the Madelung velocity is a *field*, with a definite value at each point in space, which is quite different. Moreover, it is singular close to the vortices. A consequence of this is that the volume averaged squared velocity $\langle|\mathbf{v}|^2\rangle = V^{-1}\int d^3r\, |\mathbf{v}|^2$ is infinite. It is only a logarithmic divergence: $\langle|\mathbf{v}|^2\rangle \sim \sigma_v^2 \log(\lambda_\mathrm{C}/r_\mathrm{min})$, but it is nonetheless infinite as $r_\mathrm{min} \to 0$. However, the density is very small near the vortices, so few tracer particles are found there (and the particle averaged mean squared velocity is finite). As discussed previously, the thing that corresponds to the coarse-grained $f(\mathbf{v})$ is the spatial power spectrum of the scalar (or Schrödinger) field, which, in the multi-stream region, will have multiple well defined peaks at the $\mathbf{k}$ values equal to the $\mathbf{p}/\hbar = m\mathbf{v}/\hbar$ values for the streams of particles.

The Madelung formalism also provides an alternative method for computing the evolution of such fields coupled to the Poisson's equation with ρ (or $\delta\rho$ if one is working in expanding coordinates) as the source. The simulation shown earlier was made using the Schrödinger equation, but the Madelung approach has also been used, and the question of which method is best has been debated.

A positive feature of the Madelung approach is that techniques for solving the Euler equation for particles are well developed as there is a long history of doing 'N-body simulations' in cosmology. So Madelung calculations can take advantage of this. Such methods have the advantage over evolving fields that, in following 'tracer' particles, which congregate in highly over-dense regions, they tend to concentrate the calculation where it is most needed. But to compute the quantum potential requires that the density field ρ be well-sampled at the intra-speckle scale, which is computationally expensive (as compared to particle simulations) even for fuzzy dark matter. In pure particle simulations a distribution of velocities is naturally established in dense regions as streams cross. In a Madelung calculation the quantum force is strongly fluctuating, with potentials Q diverging as $1/r^2$ close to the ubiquitous phase-vortices, and giving strong deflections of the tracer particle velocities. So one might expect the velocities of the particles being followed as 'tracers' of the density field would also develop a distribution. But if they did, that would be a numerical error; in reality the velocity has to be a field; but that is not imposed by the structure of the Euler equation.

A Derivation of the quantum potential

Madelung obtained Q by noting that, with the transformation $\psi = ae^{i\theta}$, Schrödinger's equation implies that the phase θ obeys the same equation obeyed by the velocity potential – of which the velocity is the gradient – for a fluid. Here we will take a different approach, which is to derive the equation for continuity of momentum – starting from the Lagrangian density appropriate for the Schrödinger field description of a relativistic scalar ϕ in the Newtonian-gravity limit – and then convert this to an Euler equation.

The Lagrangian density for a scalar field, in the Newtonian limit, is $\mathcal{L} = \frac{1}{2}(\dot\phi^2 - |\nabla\phi|^2 - \mu^2(1+2\Phi(\mathbf{x}))\phi^2)$, where we are working with $t' = ct$ and then dropping the prime. Replacing ϕ by a Schrödinger field ψ such that $\phi = \psi e^{-i\mu t} + \text{c.c.}$, and throwing out all rapidly time-varying terms (those containing factors $e^{\pm 2i\mu t}$) as well as a term $\dot\psi\dot\psi^\star$ which is negligible compared to $\nabla\psi\cdot\nabla\psi^\star$, we obtain an effective Lagrangian density $\mathcal{L}(\psi, \dot\psi, \nabla\psi, \ldots, \mathbf{x})$ where $\ldots$ denotes the conjugate field and its derivatives, given by

$$\boxed{\mathcal{L} = \frac{i\mu}{2}(\dot\psi\psi^\star - \psi\dot\psi^\star) - \tfrac{1}{2}\nabla\psi\cdot\nabla\psi^\star - \Phi(\mathbf{x})\mu^2\psi\psi^\star.} \tag{92}$$

Requiring that the variation of the action $\delta S = \int dt \int d^3x\, \delta\mathcal{L}$ for a variation $\delta\psi^\star$ vanishes gives

$$\partial_t \frac{\partial\mathcal{L}}{\partial\dot\psi^\star} + \nabla\cdot\frac{\partial\mathcal{L}}{\partial\nabla\psi^\star} = \frac{\partial\mathcal{L}}{\partial\psi^\star} \tag{93}$$

which is the general form of the Euler-Lagrange equation. On performing the derivatives, this is

$$i\dot\psi = -\frac{\nabla^2\psi}{2\mu} + \mu\Phi\psi \tag{94}$$

which, recalling that $\mu = mc/\hbar$ and time derivative here is $1/c$ times the derivative w.r.t. physical time and that the dimensionless potential Φ here gives a physical potential $V = mc^2\Phi$, is Schrödinger's equation $E\psi = (|\mathbf{p}|^2/2m + V)\psi$.

Taking the partial derivative of $\mathcal{L}(\mathbf{x}, t)$ with respect to the i^{th} spatial coordinate gives, using the chain rule,

$$\partial_i\mathcal{L}(\mathbf{x}, t) = \frac{\partial\mathcal{L}}{\partial\psi}\psi_{,i} + \frac{\partial\mathcal{L}}{\partial\dot\psi}\dot\psi_{,i} + \frac{\partial\mathcal{L}}{\partial\psi_{,j}}\psi_{,ji} + \text{c.c.} + \partial_i\mathcal{L} \tag{95}$$

where the last term represents the partial spatial derivative holding ψ and $\psi^\star$ and their derivatives constant.

Using (93) and its conjugate to eliminate $\partial\mathcal{L}/\partial\psi$ and $\partial\mathcal{L}/\partial\psi^\star$ then gives

$$\partial_t\left(\frac{\partial\mathcal{L}}{\partial\dot\psi}\psi_{,i} + \text{c.c.}\right) + \partial_j\left(\frac{\partial\mathcal{L}}{\partial\psi_{,j}}\psi_{,i} + \text{c.c.} - \delta_{ij}\mathcal{L}\right) + \partial_i\mathcal{L} = 0 \tag{96}$$

which, on performing the derivatives of (92), flipping the sign, and dividing by μ^2, becomes

$$\underbrace{\partial_t \left(\tfrac{i}{2\mu} \psi \psi^{\star}_{,i} + \text{c.c.} \right)}_{\text{momentum density } p_i} + \underbrace{\partial_j \tfrac{1}{\mu^2} \left(\tfrac{1}{2} \psi^{\star}_{,j} \psi_{,i} + \text{c.c.} + \delta_{ij} \mathcal{L} \right)}_{\text{pressure tensor } P_{ij}} + \underbrace{\Phi_{,i} \psi \psi^{\star}}_{\text{gravity}} = 0 \tag{97}$$

which expresses continuity of momentum.

The analogue of this for a gravitating system of particles is Jeans's equation: $\partial_t(n\overline{\mathbf{v}}) = -\boldsymbol{\nabla} \cdot (n\overline{\mathbf{v}\mathbf{v}}) + n\mathbf{g}$ where n is the particle number density, $\overline{\mathbf{v}}$ is the mean – or 'streaming' – velocity, $n\overline{\mathbf{v}\mathbf{v}}$ is the kinetic pressure tensor and $\mathbf{g} = -c^2 \boldsymbol{\nabla} \Phi$ is the gravity[25]. And, just as for particles, the pressure tensor (the momentum flux density) here is, in general anisotropic, as can be seen by considering a wave $\psi = a e^{i(\mathbf{k}\cdot x - \omega_{\mathbf{k}} t)}$. But it is not the same, in detail, as the 'quantum pressure tensor'. But that is not surprising as this is like *Jeans's* equation, giving the *partial* time derivative $\dot{\mathbf{p}}$ of the momentum density, not the *Euler* equation, which gives the *total* derivative $d\mathbf{v}/dt = \dot{\mathbf{v}} + (\mathbf{v} \cdot \boldsymbol{\nabla})\mathbf{v}$ of $\mathbf{v} = \mathbf{p}/\rho$. For particles, the Euler equation is $d\overline{\mathbf{v}}/dt = -n^{-1} \boldsymbol{\nabla} \cdot (n\boldsymbol{\sigma}) + \mathbf{g}$ where $\boldsymbol{\sigma}$ is the velocity dispersion tensor $\boldsymbol{\sigma} \equiv \overline{(\mathbf{v} - \overline{\mathbf{v}})(\mathbf{v} - \overline{\mathbf{v}})}$, so $n\boldsymbol{\sigma}$ measures the momentum flux density in the frame of an observer moving at the streaming velocity, which is naturally different from $n\overline{\mathbf{v}\mathbf{v}}$ which is the momentum flux density seen by a stationary observer. We would expect therefore an analogous change in the wave-mechanical pressure.

According to Madelung, if we make the appropriate adjustments to (97) to convert it into an expression for $d\mathbf{v}/dt$ we will find that $m d\mathbf{v}/dt$ is simply $-\boldsymbol{\nabla}(V + Q)$. Showing this involves some tedious algebra. First, we need to establish the relation between $\dot{\mathbf{p}} = \rho \mathbf{v}$ and $d\mathbf{v}/dt$. The former is $\dot{\mathbf{p}} = \rho \dot{\mathbf{v}} + \mathbf{v} \dot{\rho} = \rho \dot{\mathbf{v}} - \mathbf{v}(\boldsymbol{\nabla} \cdot \mathbf{p})$, which gives

$$\begin{aligned} \rho \frac{d\mathbf{v}}{dt} &= \overbrace{\qquad \dot{\mathbf{p}} \qquad}^{} + \overbrace{(\boldsymbol{\nabla} \cdot \mathbf{p} + \rho \mathbf{v} \cdot \boldsymbol{\nabla}) \mathbf{v}}^{} \\ &= \mathbf{F}_1 + \mathbf{F}_2 - \rho \boldsymbol{\nabla} \Phi + \tfrac{1}{\mu^2} [a^2 (\boldsymbol{\nabla}\theta \cdot \boldsymbol{\nabla}) \boldsymbol{\nabla}\theta^{\mathrm{b}} + a^2 \nabla^2 \theta \boldsymbol{\nabla} \theta^{\mathrm{d}} + 2a(\boldsymbol{\nabla}\theta \cdot \boldsymbol{\nabla} a) \boldsymbol{\nabla}\theta^{\mathrm{c}}] . \end{aligned} \tag{98}$$

where in the second line we have split the force density – the divergence of the pressure tensor in (97) – into two terms, and have used $\mathbf{v} = \mu^{-1} \boldsymbol{\nabla} \theta$ and have defined the wave-amplitude $a \equiv \sqrt{\rho}$.

Next, we need to compute the first part of the force density $F_{1i} = -\partial_j (\psi^{\star}_{,j} \psi_{,i})/2\mu^2 + c.c.$, which is the i^{th} component of

$$\begin{aligned} \mathbf{F}_1 &= -\tfrac{1}{2\mu^2} [\nabla^2 \psi^{\star} + (\boldsymbol{\nabla} \psi^{\star} \cdot \boldsymbol{\nabla})] \boldsymbol{\nabla} \psi + \text{c.c.} \\ &= -\tfrac{1}{2\mu^2} [\nabla^2 (a e^{-i\theta}) + (\boldsymbol{\nabla}(a e^{-i\theta}) \cdot \boldsymbol{\nabla})] \boldsymbol{\nabla}(a e^{i\theta}) + \text{c.c.} \\ &= -\tfrac{1}{\mu^2} [\nabla^2 a \boldsymbol{\nabla} a + a^2 \nabla^2 \theta \boldsymbol{\nabla} \theta^{\mathrm{d}} + 2a(\boldsymbol{\nabla}\theta \cdot \boldsymbol{\nabla} a) \boldsymbol{\nabla}\theta^{\mathrm{c}} + (\boldsymbol{\nabla} a \cdot \boldsymbol{\nabla}) \boldsymbol{\nabla} a^{\mathrm{a}} + a^2 (\boldsymbol{\nabla}\theta \cdot \boldsymbol{\nabla}) \boldsymbol{\nabla}\theta^{\mathrm{b}}]. \end{aligned} \tag{99}$$

Finally, we need to compute the force density $\mathbf{F}_2$ coming from the isotropic part of the pressure tensor $\delta_{ij} \boldsymbol{\nabla} \mathcal{L}/\mu^2$ also. Using (94) to eliminate the time derivatives in the Lagrangian density gives

$$\begin{aligned} \mathbf{F}_2 &= -\tfrac{1}{\mu^2} \boldsymbol{\nabla} \mathcal{L} = +\tfrac{1}{4\mu^2} \boldsymbol{\nabla} (\psi^{\star} \nabla^2 \psi + \boldsymbol{\nabla} \psi \cdot \boldsymbol{\nabla} \psi^{\star} + \text{c.c.}) \\ &= \tfrac{1}{2\mu^2} \boldsymbol{\nabla} (a \nabla^2 a + \boldsymbol{\nabla} a \cdot \boldsymbol{\nabla} a) \\ &= \tfrac{1}{\mu^2} \left(\tfrac{1}{2} \nabla^2 a \boldsymbol{\nabla} a + (\boldsymbol{\nabla} a \cdot \boldsymbol{\nabla}) \boldsymbol{\nabla} a^{\mathrm{a}} + \tfrac{1}{2} a \boldsymbol{\nabla} \nabla^2 a \right) \end{aligned} \tag{100}$$

As you can see, there is a lot of cancellation. All that remains is

$$\rho \left(\frac{d\mathbf{v}}{dt} + \boldsymbol{\nabla} \Phi \right) = \tfrac{1}{2\mu^2} (a \boldsymbol{\nabla} \nabla^2 a - \nabla^2 a \boldsymbol{\nabla} a) = \tfrac{1}{2\mu^2} a^2 \boldsymbol{\nabla} \tfrac{\nabla^2 a}{a}. \tag{101}$$

Finally, changing back to physical time, so $d\mathbf{v}/dt \Rightarrow c^{-2} d\mathbf{v}/dt$, and replacing μ by $mc/\hbar$ and $\Phi \Rightarrow V/mc^2$ and $a \Rightarrow \sqrt{\rho}$, we have Madelung's result

$$m \frac{d\mathbf{v}}{dt} = -\boldsymbol{\nabla} \left(V - \frac{\hbar^2}{2m} \frac{\nabla^2 \sqrt{\rho}}{\sqrt{\rho}} \right) . \tag{102}$$

[25] Jeans's equation, if multiplied by the particle mass, simply says that the rate of change of momentum density in a volume element is the rate at which particle motions are delivering momentum to that volume plus the rate at which gravity is giving the particles momentum.

B The scalar field in Rindler rocket coordinates

It is quite interesting to look at the the behaviour of a scalar field that is a superposition of random planar waves

$$\phi(\vec{x}) = \sum_{\mathbf{k}} \phi_{\mathbf{k}} e^{i\vec{k}\cdot\vec{x}} \tag{103}$$

where $\vec{k} \longrightarrow k_\mu = (-\omega_{\mathbf{k}}/c, \mathbf{k})$ and with complex amplitudes $\phi_{\mathbf{k}}$ with random phases, and with reality of $\phi(\vec{x})$ being ensured by having $\phi_{-\mathbf{k}} = \phi^\star_{\mathbf{k}}$.

The space-time derivatives of the field are

$$\phi_{,\alpha} = \sum_{\mathbf{k}} ik_\alpha \phi_{\mathbf{k}} e^{i\vec{k}\cdot\vec{x}} \tag{104}$$

and this allows us to express the terms in the stress energy tensor, which are all quadratic in the field or its derivatives, as double sums so, for example,

$$\phi^2(\vec{x}) = \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \phi_{\mathbf{k}} \phi_{\mathbf{k}'} e^{i(\vec{k}+\vec{k}')\cdot\vec{x}}. \tag{105}$$

All but the diagonal terms for which $\mathbf{k}' = -\mathbf{k}$ are oscillating, so if we average over position and time these may be ignored, and we have

$$\langle \phi^2(\vec{x}) \rangle = \sum_{\mathbf{k}} |\phi_{\mathbf{k}}|^2 \tag{106}$$

where $|\phi_{\mathbf{k}}|^2 = \phi_{\mathbf{k}}\phi^\star_{\mathbf{k}}$ is the *power spectrum*, and with analogous results for the mean squared derivatives.

The *expectation value of the stress-energy tensor* is then, from (??),

$$\begin{aligned} T^\alpha{}_\beta &= \langle \phi^{,\alpha}\phi_{,\beta} + \delta^\alpha_\beta(-\tfrac{1}{2}\phi_{,\gamma}\phi^{,\gamma} - \tfrac{1}{2}\mu^2\phi^2) \rangle \\ &= \sum_{\mathbf{k}} |\phi_{\mathbf{k}}|^2 (k^\alpha k_\beta - \tfrac{1}{2}\delta^\alpha_\beta (k_\gamma k^\gamma + \mu^2)). \end{aligned} \tag{107}$$

If we assume that the power spectrum is an isotropic function of the wave-vector $|\phi_{\mathbf{k}}|^2 = P_\phi(k = |\mathbf{k}|)$ – which might come about through some relatively weak self-interaction term (weak in the sense that the contribution to the stress-energy tensor is small) – then this is diagonal:

$$T^\alpha{}_\beta = \mathrm{diag}(-\mathcal{E}, P, P, P) = \sum_{\mathbf{k}} P_\phi(k) \left(-\frac{\omega^2_{\mathbf{k}}}{c^2}, \frac{k^2}{3}, \frac{k^2}{3}, \frac{k^2}{3} \right) \tag{108}$$

so this would be quite analogous to the stress energy for a collisionless gas of particles, but with the phase space density $f(p)$ replaced by by $P_\phi(k)$. Note that for waves of high frequency $|\mathbf{k}| \gg \mu$, so $k^2 = \omega_k^2/c^2$ this would give $P = \mathcal{E}/3$ just as for a gas of relativistic particles or photons.

In Rindler rocket coordinates – that of an observer being accelerated in the x-direction with $d^2x/d\tau^2 = a$ – the metric is $g_{\mu\nu} = \mathrm{diag}(-(1 + ax/c^2)^2, 1, 1, 1)$. If we assume that the stress energy tensor is is locally isotropic like this, but, with energy density and pressure varying with position, then the $\beta = x$ component of (??) – i.e. the equation of continuity for the x-momentum – is

$$\begin{aligned} T^\alpha{}_{x,\alpha} &= T^0{}_{x,0} + T^x{}_{x,x} = \tfrac{1}{2} g^{\mu\sigma}(g_{\sigma\tau,x} T^\tau{}_\mu - g_{\mu\sigma,\tau} T^\tau{}_x) \\ &= \tfrac{1}{2} g^{00} g_{00,x} (T^0{}_0 - T^x{}_x) \\ &= -(1 + ax/c^2)^{-1}(a/c^2)(\mathcal{E} + P) \end{aligned} \tag{109}$$

where we are using the fact that the only variation of the metric is with x, and it is only the time-time component g_{00} that varies.

For x such that $ax \ll c^2$ – which is not very restrictive; for say $a = 9.81\mathrm{m/sec}^2$ this says x must be much less than about a parsec – this equation has static solutions provided the pressure gradient $dP/dx = T^x{}_{x,x}$ satisfies the *relativistic equation of hydrostatic equilibrium*

$$\frac{dP}{dx} = -a(\rho + P/c^2) \tag{110}$$

just as we found for an ideal fluid (where we note that as in that case it is the *enthalpy* that appears on the right hand side).