Basic material for Cosmology

Vincent Icke

icke@strw.LeidenUniv.nl

www.strw.LeidenUniv.nl/~icke/

27 January 2005

The following is a summary of the most essential material in the Cosmology lecture series. It is not complete and in some places fairly simplified to show the basic ingredients most clearly. For more details consult the literature, such as the books by Achterberg, Coles & Lucchin, Narlikar, Padmanabhan, Peacock, Peebles or Weinberg. I have tried to be precise, but these pages may contain typing errors.

Equation numbers of important material have been indicated in the text with an ace-of-spades. It is not mandatory to learn these by heart, but it would not be a bad idea, as in

$$E = \gamma mc^2 \tag{0.1}$$ ♠

1. Observations of Cosmic Structure

The most successful cosmological description of the Universe at large, the Big Bang Model, assumes that the Universe has no structure at all, not even atoms. However, it is obvious that this is only an approximation. To what degree does it apply?

People often react to astrophysics with the platitude ‘We are so small, and the Universe is so big.’ Let us see what this means quantitatively. The mean distance between stars is of the order of 10^{16} times bigger than a human. The mass of a typical star is 10^{29} times larger than that of a human, and the average human life span is 200 million times shorter than the age of the Universe. These figures are so extreme, that the expression ‘the Universe is so big’ holds even for galaxies. This leads to the question: on what scales of length, mass, and time can we begin to speak about ‘the properties of the Universe’, if indeed we can do this at all?

It is by no means self-evident that there is such a thing as *the* properties of the Universe. Geometrically, it might be that the Universe has a (multi)fractal structure, such that there is no length scale, be it ever so large, on which we can safely speak about the average properties of the cosmic structure. Physically, it need not even be true that the same laws hold everywhere. Suppose that the Universe has a finite age, say Y years. Then if we look (almost) Y light years in one direction, we can see a patch P of the Universe at a time of almost Y years in the past. In the diametrically opposed direction, we can do the same, and observe a patch Q. Yet P and Q are approximately $2Y$ apart from each other, and therefore cannot yet have communicated in the age Y of the Universe. How, then, could the laws of physics be the same in both?

Problems of this kind are not to be solved *a priori*, but we must *observe* how the Universe in fact behaves. It is a remarkable fact that the laws of physics seem to be the same everywhere (we can see this, among other things, from the behaviour of spectral line frequency ratios and similar dimensionless physical phenomena). Furthermore, it appears that *the farther we go, the less structure the Universe has*, in the following sense. Suppose that we average the mass distribution in the observed Universe over a length scale L, for example by covering space with a tiling of cubes with edge L. Then we can compute the variance $\sigma(L)$ of the distribution $M_i(L)$ of masses in the boxes $i = 1, \cdots, \infty$. It turns out that $\sigma(L)$ *decreases with increasing* L, in such a way that for L larger than about 100 Mpc, the variance in mass density is less than 1% or so.

Furthermore, the Universe becomes more and more *isotropic* for increasing L. Unless we occupy a very privileged place in space and time, we must assume that this holds everywhere in the Universe, from which it follows that the Universe is also *homogeneous*; that is, the large-scale properties of the Universe at a given time are the same at all points in space. This leads to the approximation which is the basis of the Big Bang Model: in describing its overall behaviour, we assume that *the Universe has no structure at all, not even atoms.*

Exercise.

Prove that a static Universe that is isotropic with respect to at least two different points must be spatially homogeneous.

However, the Universe does have a definite structure. On very small scales there is the structure of particles and atoms. On larger scales, the density contrast of stars, galaxies, and groups of galaxies is considerable. Quantitatively, these deviations from homogeneity can be studied by means of various statistical tests, such as the `two point correlation function`, which measures the excess probability for finding a galaxy near another one, and the `minimal spanning tree`, which is a measure for the coherence of structures over larger distances. Please consult the literature, in particular the books by Coles and Peebles, about quantitative determination of these measures.

2. The Oort Constants of the Universe

The motion of a universe that is homogeneous and isotropic is extremely restricted. Due to the severe limitations imposed by this symmetry, only a uniform change of scale is possible.

The kinematic properties of the motion of galaxies can be found by defining the `deformation tensor` of the velocity field v_i of the galaxies in space as

$$\mathcal{Q}_{ij} \equiv \frac{\partial v_i}{\partial x^j} \tag{2.1}$$

These kinematic numbers are, so to speak, the Oort Constants of the Universe. The velocity field v_i carries a galaxy in the time interval δt from the position r_i to

$$r_i(t+\delta t) = r_i(t) + \mathcal{Q}_{ij} r^j\,\delta t \equiv r_i(t) + \sum_j \mathcal{Q}_{ij} r^j\,\delta t \tag{2.2}$$

Note the use of the `summation convention`, which will be used throughout these notes. The deformation tensor can be split into three components:

$$\mathcal{Q}_{ii} \qquad \text{(diagonal)} \tag{2.3}$$

$$\frac{1}{2}(\mathcal{Q}_{ij} - \mathcal{Q}_{ji}) \qquad \text{(antisymmetric)} \tag{2.4}$$

$$\frac{1}{2}(\mathcal{Q}_{ij} + \mathcal{Q}_{ji}) \qquad \text{(symmetric)} \tag{2.5}$$

The third one of these still depends on the first, so to make them truly irreducible we write

$$D \equiv \mathcal{Q}_k^k \qquad (\textit{divergence}) \tag{2.6}$$

$$R_{ij} \equiv \frac{1}{2}(\mathcal{Q}_{ij} - \mathcal{Q}_{ji}) \qquad (\textit{rotation}) \tag{2.7}$$

$$S_{ij} \equiv \frac{1}{2}(\mathcal{Q}_{ij} + \mathcal{Q}_{ji}) - \frac{1}{3} D\delta_{ij} \qquad (\textit{shear}) \tag{2.8}$$

In this way, deformations due to R and S keep the volume unchanged; volume changes are entirely due to D. The second tensor has only three independent components, which can be written as a vector:

$$R^i \equiv \epsilon^{ijk} \mathcal{Q}_{jk} \tag{2.9}$$

where ϵ is zero if two of its indices are the same, +1 if $\{ijk\}$ is an even permutation and -1 if it is odd.

In a homogeneous Universe the dynamical quantities do not depend on position. Therefore $\mathcal{Q}_{ij}$ must be constant in space (it may depend on time, though). In a Newtonian universe we can then integrate to obtain

$$\mathcal{Q}_{ij} = C_{ij} \quad \Rightarrow \quad v_i = \int C_{ij}\, dx^j \tag{2.10}$$

which has the solution

$$v_i = C_{ij}\, x^j \tag{2.11}$$

A possible integration constant can be removed by a global Galilei-Huygens transformation, which is allowed in Newtonian dynamics.

On large scales (beyond a gigaparsec or so) the Universe is isotropic at the 1% level. If this is true for all observers, the Universe is *homogeneous* too. Thus we must have

$$S_{ij} = 0 \quad ; \qquad R^i = 0 \quad ; \qquad D = \text{constant in space} \tag{2.12}$$

and so C_{ij} is a multiple of the identity. Consequently

$$v_i = H(t) x_i \tag{2.13}$$ ♠

which is `Hubble's Law`. This is the main underpinning of the Big Bang Model. Our trust in this model does *not* depend on the particular value of H!

So far we have looked at the kinematics of the cosmic motion only. But we can also consider the dynamics. Using the deformation tensor, we can write the equation of motion as

$$\frac{dv^i}{dt} = \frac{\partial v^i}{\partial t} + v_k \frac{\partial v^i}{\partial x_k} = \frac{\partial v^i}{\partial t} + v_k \mathcal{Q}^{ik} \tag{2.14}$$

This can be used in the equation of motion in an acceleration field A_k. The divergence D is the "K term" of stellar dynamics, and the Hubble parameter in cosmology. In Oort's stellar case, D was negligible, in the Universe it is dominant!

$$\frac{dD}{dt} = \frac{\partial A_k}{\partial x_k} - S^2 - \frac{1}{3}D^2 + \frac{1}{2}R^2 = -\frac{1}{\rho}\frac{d\rho}{dt} \tag{2.15}$$

The equation for the rotation is

$$\frac{dR^i}{dt} = \epsilon^{ijk} \frac{\partial A_k}{\partial x^j} + \left(S^{ik} - \frac{2}{3} D \delta^{ik} \right) R_k \tag{2.16}$$

If the matter moves under the influence of a potential Φ only, then the vector R_i does not depend on the acceleration $\vec{A} = -\vec{\nabla}\Phi$:

$$\frac{dR_i}{dt} = \left(S^{ik} - \frac{2}{3} D \delta^{ik} \right) R_k \tag{2.17}$$

so that $R = 0$ is a good solution; if initially $R \neq 0$, then the cosmic expansion in D makes it negligible. *Potential flow is therefore a good approximation in studying megaparsec structure.*

The third equation is the one for the shear:

$$\frac{dS_{kj}}{dt} = \frac{1}{2}\left(\frac{\partial A_j}{\partial x^k} + \frac{\partial A_k}{\partial x^j} \right) - \frac{1}{3}\frac{\partial A_m}{\partial x_m}\delta_{kj} + \frac{1}{3}\left(S^2 + \frac{1}{4}R^2 \right) - S_j^m S_{mk} - \frac{2}{3} D S_{kj} - \frac{1}{4} R_k R_j \tag{2.18}$$

Assuming potential flow, this simplifies to

$$\frac{dS_{kj}}{dt} = -\frac{\partial^2\Phi}{\partial x^k \partial x^j} + \frac{1}{3}(\Delta\Phi + S^2)\delta_{kj} - S_{mk}S_j^m - \frac{2}{3}DS_{kj} \tag{2.19}$$

Using Poisson's Equation, we finally get

$$-\frac{dD}{dt} = 4\pi G\rho + S^2 + \frac{1}{3}D^2 = \frac{1}{\rho}\frac{d\rho}{dt} \tag{2.20}$$

$$\frac{dS_{kj}}{dt} = -\frac{\partial^2\Phi}{\partial x^k \partial x^j} + \frac{1}{3}(4\pi G\rho + S^2)\delta_{kj} - S_{mk}S_j^m - \frac{2}{3}DS_{kj} \tag{2.21}$$

Clearly, both D (expansion) and S_{ij} (shear) can be used to map the cosmic density. Traditionally, one used $S = 0$ (Hubble flow) only.

3. Symmetrie en klassieke mechanica

De regels van de mechanica zijn van zeer groot belang in de sterrenkunde. Het is dus essentieel om te weten waar die regels vandaan komen. Het blijkt dat de 'wetten' van de beweging voortkomen uit bepaalde regelmatigheden in de Natuur, 'symmetrie' genoemd. Dat zijn: homogeniteit van tijd en ruimte, en relativiteit van de snelheid. Uit deze symmetrieën leidt men de bewegingsvergelijkingen af. Bij deze afleiding blijkt dat wiskundige analyse niet gemist kan worden.

Om erachter te komen welke vergelijking(en) gebruikt kunnen worden om beweging van deeltjes te beschrijven moeten we eerst zien welke fysica moet worden ingebouwd. Op het eerste gezicht zou je denken dat de vergelijking voor de baan van een deeltje een *algebraïsche* vorm is, bv. de vergelijking van de parabool

$$y = a + bt + ct^2 \tag{3.1}$$

waarin y zoiets als de hoogte van een deeltje en t de tijd. Maar het blijkt dat het niet zo gaat. Dat komt doordat de absolute positie en de absolute snelheid van deeltjes in ons Heelal blijkbaar geen meetbare grootheden zijn: aan niets is af te lezen wat de ruimtelijke coördinaten van een deeltje zijn. Dat is geen 'principe' of zo; in de studeerkamer, afgesloten van de werkelijkheid, zouden we best een heelal kunnen verzinnen waarin deeltjes een soort inwendige kilometerteller hebben waarop je de plaats ervan kunt aflezen. Maar het Heelal waarin wij wonen werkt niet zo. In onze natuur geldt de `homogeniteit van de ruimte`: als je bij de positie van een deeltje een willekeurig vast getal optelt, verandert er niets. Dit komt neer op een globale verschuiving van het Heelal. Met andere woorden, er geldt blijkbaar een soort 'relativiteit van de ruimte': de waarde van een coördinaat als zodanig is niet waarneembaar. In formule luidt deze invariantie

$$\vec{r} \quad \Longrightarrow \quad \vec{r} + \vec{a} \tag{3.2}$$ ♠

Dus is niet de absolute positie van een deeltje van belang, maar de *relatieve* plaats, in het bijzonder de verandering van de plaats in de loop van de tijd: de snelheid. Dus alleen het verschil in positie telt, vandaar dat de bewegingsvergelijking geen algebraïsche vergelijking voor de positie $\vec{r}$ is maar een *differentiaalvergelijking*[*1] voor de verandering van de plaats,

[*1] Zie bijvoorbeeld W.T. van Horssen, *Differentiaalvergelijkingen*, Epsilon Uitgaven, Utrecht 1993; J. Grasman, *Wiskundige methoden toegepast*, Epsilon Uitgaven, Utrecht 1992.

$d\vec{r}$, in een klein intervalletje van tijd, dt. Die verandering van plaats heeft een naam, de `snelheid`:

$$\vec{v} \equiv \lim_{\Delta t \to 0} \frac{\vec{r}(t+\Delta t) - \vec{r}(t)}{(t+\Delta t) - t} = \frac{d\vec{r}}{dt} \qquad (3.3) ♠$$

Merk op dat hierin niet alleen de homogeniteit van de ruimte is gebruikt (omdat we het *verschil* van twee posities bepalen), maar ook de homogeniteit van de tijd (omdat we door een tijdsverschil delen, en niet door een absolute tijdswaarde).

Nu zou je denken dat dan tenminste de vergelijking voor de snelheid v van een deeltje een algebraïsche vorm zou kunnen zijn, bijvoorbeeld

$$v = a + bt + ct^2 \qquad (3.4)$$

maar ook dat is niet het geval. Dat komt doordat de absolute snelheid van deeltjes in ons Heelal blijkbaar geen meetbare grootheid is: aan niets is af te lezen wat de ruimtelijke snelheid van een deeltje is. Ook dat is geen 'principe'; in de studeerkamer kunnen we best een heelal verzinnen waarin deeltjes een soort inwendig wijzertje hebben waarop je de snelheid van het deeltje kunt aflezen. Maar het Heelal waarin wij wonen werkt niet zo. In onze natuur geldt de `Galilei-Huygens symmetrie`, dit is de invariantie

$$\vec{v} \quad \Longrightarrow \quad \vec{v} + \vec{w} \qquad (3.5) ♠$$

Als je bij de snelheid van een deeltje een willekeurige vaste waarde $\vec{w}$ optelt, verandert er niets. Met andere woorden, er geldt blijkbaar een soort 'relativiteit van de snelheid': de waarde van een snelheid als zodanig is niet waarneembaar. Dus is niet de absolute snelheid van een deeltje van belang, maar de *relatieve* snelheid, in het bijzonder de verandering van de snelheid in de loop van de tijd: de `versnelling`

$$\vec{a} \equiv \lim_{\Delta t \to 0} \frac{\vec{v}(t+\Delta t) - \vec{v}(t)}{(t+\Delta t) - t} = \frac{d\vec{v}}{dt} \qquad (3.6)$$

Omdat alleen het verschil in snelheid telt, is de bewegingsvergelijking geen algebraïsche vergelijking voor de snelheid $\vec{v}$, maar een *differentiaalvergelijking* voor de verandering van de snelheid, $d\vec{v}$.

Hogere symmetrieën zijn er blijkbaar niet, want we kunnen een versnelling wèl absoluut meten (experiment met draaiende emmer). Ook dat is iets wat ons door de Natuur wordt voorgeschoteld, en waarvan we nog niet weten wat de diepere achtergrond is. Omdat er blijkbaar niet zoiets bestaat als een 'relativiteit van versnelling' is de bewegingsvergelijking een tweede-orde differentiaalvergelijking, de klassieke `bewegingsvergelijking`

$$\frac{d^2\vec{r}}{dt^2} = \frac{d\vec{v}}{dt} = \vec{a} = \text{“iets”} \qquad (3.7) ♠$$

waarin "iets" de versnelling ten gevolge van uitwendige invloeden. Dat kan van alles zijn; in dit college zullen wij zeer zwaar leunen op het voorschrift van de Newtonse zwaartekracht.

Om deze tweede-orde differentiaalvergelijking te kunnen oplossen, moeten wij dus zeven getallen invoeren: de beginpositie (drie stuks), de beginsnelheid (ook 3), en de begintijd. Die laatste doet er niet toe als de uitwendige versnelling niet van de tijd afhangt. Pas als we al deze gegevens hebben gekozen, kunnen we (in principe, tenminste) de vergelijkingen oplossen. Dat is dus heel iets anders dan een algebraïsche functie uitrekenen! Toch hebben veel mensen zo'n spoorrails-beeld in hun hoofd als ze aan bewegingen denken. We zien dat aardig terug in rare uitspraken van het type "het ruimtevoertuig is uit zijn baan geraakt."

Een stel eenvoudige gevallen kunnen wij meteen oplossen. De meest voor de hand liggende is 'iets is niets':

$$\frac{d^2\vec{r}}{dt^2} = 0 \qquad (3.8)$$

met de oplossing

$$\vec{v} = \text{constant} = \vec{v}_0 \;; \qquad \vec{r} = \vec{r}_0 + \vec{v}_0 t \qquad (3.9)$$

Met andere woorden, *een deeltje waarop geen kracht werkt, beweegt met constante snelheid.* Let op: snelheid is een vector, dus zowel de richting als de grootte van de snelheid blijven constant! Dit noemt men wel de `wet van de traagheid`. Die komt er dus uitrollen dankzij de symmetrieën die we in de bewegingsvergelijking hebben ingebouwd.

Exercise.

Van de klassieke oudheid tot de Middeleeuwen vond men niet de rechtlijnige beweging, maar de cirkelbeweging de 'meest ideale'. Als je dit opvat in de zin van "een deeltje waarop geen kracht werkt, beweegt met constante snelheid *op een cirkel*" dan zegt dit iets over een soort vervanger van de Galilei-Huygens symmetrie. Probeer daarvoor eens een formulering te vinden.

De volgende vorm van de bewegingsvergelijking veronderstelt dat de uitwendige versnelling constant is:

$$\frac{d^2\vec{r}}{dt^2} = \vec{g} \tag{3.10}$$

De oplossing hiervan is

$$\vec{v} = \vec{v}_0 + \vec{g}t \tag{3.11}$$

Met andere woorden: bij constante versnelling neemt de snelheid vanaf een gegeven beginwaarde $\vec{v}_0$ lineair toe (of af) met de tijd t. Dit is een `eenparig versnelde beweging`. De vergelijking voor $\vec{v}$ is weer een differentiaalvergelijking voor $\vec{x}$, die we oplossen als

$$\vec{r} = \vec{r}_0 + \vec{v}_0 t + \frac{1}{2}\vec{g}t^2 \tag{3.12}$$

Als wij even aannemen dat $\vec{g}$ langs de y-as gericht is, vinden we de beroemde paraboolbaan

$$y - y_0 = \frac{v_y}{v_x}(x - x_0) + \frac{g}{2v_x^2}(x - x_0)^2 \tag{3.13}$$

Galilei heeft daar zeer veel experimenteel onderzoek naar gedaan.

Wij zien hier het optreden van de `integratieconstanten` $\vec{v}_0$ en $\vec{r}_0$. Die komen er in omdat we zagen dat de ruimte homogeen is, dus de absolute positie van een voorwerp doet er niet toe in de bewegingsvergelijking; de ijking $\vec{x} = \vec{x}_0$ op tijd $t = 0$ moeten we dus achteraf vastleggen. Op soortgelijke manier moet ook de beginsnelheid $\vec{v}_0$ worden gegeven, omdat Galilei-Huygens symmetrie zegt dat absolute snelheden er niet toe doen. Dus hebben we een tweede-orde differentiaalvergelijking, die dan ook twee (vector-)constanten nodig heeft voor de oplossing.

Wanneer je een bewegingsvergelijking opschrijft die uitgaat van een versnelling, krijg je (als boven) een vorm van het type

$$\frac{d^2\vec{r}}{dt^2} = \vec{a} \tag{3.14}$$

waarin $\vec{a}$ een of andere van buiten opgelegde 'oorzaak' van de versnelling $d^2\vec{r}/dt^2$. In de klassieke mechanica wordt de 'kracht' $\vec{F}$ aangemerkt als verantwoordelijk voor de waargenomen versnelling, en omdat experimenteeol blijkt dat niet elk deeltje hetzelfde reageert op zo'n kracht, wordt de massa m ingevoerd volgens

$$\vec{a} \equiv \frac{1}{m}\vec{F} \tag{3.15}$$

In zekere zin is dit een cirkelredenering, en opmerkelijk genoeg komt in moderne theorieën van wisselwerkingen het begrip 'kracht' dan ook niet voor (zie Wilczek 2004a en 2004b).

4. Newton-kosmologie

Op zeer grote afstand, boven ongeveer een miljard jaar, is het Heelal isotroop: het ziet er in alle richtingen hetzelfde uit. Tenzij wij een zeer bijzondere plaats innemen in ruimte of tijd, betekent dit dat het Heelal ook homogeen is: op een gegeven tijdstip is de toestand overal in het Heelal gemiddeld hetzelfde. Daarom is een bol met een voldoende grote straal een goede afspiegeling van het Heelal als geheel. Hieruit volgt, dat het Heelal maar op één manier kan bewegen: door uniforme verandering van de ruimtelijke schaal. Zodoende kunnen wij onze oude bekende vergelijkingen voor beweging onder invloed van een centrale kracht ook hier gebruiken. De oplossingen zijn van hetzelfde type als tevoren, analoog aan de Kepler-oplossingen, maar nu voor rechtlijnige banen.

Een van de meest wonderlijke, en tot dusver onbegrepen, eigenschappen van ons Heelal is dat het er – mits we ons beperken tot voorwerpen op zeer grote afstand, zeg meer dan een miljard jaar ver – in alle richtingen gemiddeld hetzelfde uitziet. Het Heelal is `isotroop`. Dit staat in krasse tegenstelling tot het zeer nabije Heelal: de objecten in ons Zonnestelsel staan altijd binnen een smalle band aan de hemel (de Dierenriem). Evenzo met objecten op middelgrote afstand: de sterren in ons sterrenstelsel staan voor het merendeel binnen de band van de Melkweg. Op grote afstand zien we `clusters van sterrenstelsels` tot een paar honderd megaparsec, maar op afstanden boven een gigaparsec wordt de verdeling van materie steeds gelijkmatiger.

Tenzij wij een uitzonderlijke positie innemen in ruimte of in tijd, moeten we concluderen dat het Heelal isotroop is *op alle tijden en gezien vanuit alle standpunten* in de ruimte. Daaruit volgt dat het Heelal ook `homogeen` is: de massadichtheid is gemiddeld overal hetzelfde. We zien dit direct in door vast te stellen dat een gegeven punt in de ruimte door een passende draaiing om een willekeurig centrum altijd kan worden overgevoerd in een ander punt.

Laten we nu veronderstellen dat deze eigenschap exact geldt. Dat betekent: *het Heelal heeft geen enkele structuur*. Er zijn geen sterrenstelsels, geen sterren, zelfs geen atomen, de materieverdeling is volmaakt glad. We zagen al dat dit op kleinere schaal (onder de 100 megaparsec of zo) beslist onjuist is, en op kleine schaal verwachten we dan ook problemen. Niettemin kunnen we op grote schaal de vraag stellen: *hoe kan een Heelal dat homogeen is en isotroop, bewegen?* Als er bijzondere gebieden aan de hemel waren (zoals Dierenriem of Melkweg) dan wisten we het wel: de bewegingen zouden in een schijf liggen. Maar dat kan nu niet, want een heelal waarin het er aan de hemel overal hetzelfde uitziet moet ook overal hetzelfde bewegen.

Scherper geformuleerd: het snelheidsveld $\vec{v}$ moet zo zijn dat de vervorming die het teweeg brengt overal dezelfde is. Dat betekent het volgende. Laat v_i de component van de snelheid zijn in de richting van de coördinaat x_i. Neem nu een klein stapje Δx_j in de richting van de x_j-coördinaat en kijk hoe $\vec{v}$ verandert. Deze verandering noemen we $\mathcal{Q}$, en omdat zowel $\vec{v} = (v1, v2, v3)$ als $\vec{r} = (x1, x2, x3)$ 3-vectoren zijn is $\mathcal{Q}$ een 3×3 object, de `vervormingstensor`

$$\mathcal{Q}_{ij} = \frac{\Delta v_i}{\Delta x_j} \tag{4.1}$$

In de limiet voor infinitesimale stapjes schrijven we

$$\mathcal{Q}_{ij} = \frac{\partial v_i}{\partial x_j} \tag{4.2}$$

De notatie $\partial f / \partial y$ betekent dat we de afgeleide van f nemen naar de variabele y, terwijl we ervoor zorgen dat alle andere variabelen waar f nog van zou kunnen afhangen hetzelfde blijven (dat heet een `partiële afgeleide`).

In een homogeen isotroop heelal is de vervormingstensor constant, en de vervorming mag geen voorkeursrichtingen hebben. In eerste instantie zou je denken dat de eenvoudigste oplossing voor een heelal dat 'overal hetzelfde beweegt' gegeven wordt door $\vec{v} = 0$. Dat is echter onmogelijk vanwege de zwaartekracht: alle materie en energie trekt elkaar aan, en dan is er geen evenwicht mogelijk. Vervolgens zou je kunnen denken dat dan tenminste $\mathcal{Q} = 0$ zou kunnen zijn. Dat betekent echter wegens Eq.(4.2) dat

$$\vec{v} = \text{constant} \tag{4.3}$$

Maar deze constante snelheid kunnen wij wegtransformeren met behulp van de Galilei-Huygens symmetrie, dus dan zitten we weer met het geval $\vec{v} = 0$. Het enige alternatief is

$$\mathcal{Q}_{ij} = \text{constant} \tag{4.4}$$

Daarmee ligt de vorm van het snelheidsveld meteen vast, want als $\mathcal{Q}$ geen voorkeursrichtingen mag hebben dan moet $\mathcal{Q}$ een scalair veelvoud zijn van de eenheidsmatrix, en zodoende moet

$$\frac{\partial v_i}{\partial x_j} = H\,\delta_{ij} \tag{4.5}$$

waarin δ_{ij} de Kronecker delta. De scalar H mag niet van de ruimte-coördinaten afhangen, maar desnoods wel van de tijd. Het geestige is nu dat Eq.(4.5) direct geïntegreerd kan worden, en zo vinden we dat

$$\vec{v} = H\,\vec{r} \tag{4.6}$$ ♠

Dit is de `Hubble relatie`, een verband tussen de afstand $\vec{r}$ tussen twee sterrenstelsels en hun snelheidsverschil $\vec{v}$.

De kinematische relatie Eq.(4.6) betekent ten eerste dat de snelheden van alle sterrenstelsels ten opzichte van elkaar zo zijn dat, gezien vanuit een gegeven stelsel, alle andere *radieel* bewegen. Er is dus sprake van een *uitdijing* als het Hubble-getal H positief is, en van een *inkrimping* als $H < 0$. Uit waarnemingen blijkt dat aan Eq.(4.6) zeer goed voldaan is, en $H > 0$: ons Heelal dijt uit. Ten tweede zien we bij nadere beschouwing dat Eq.(4.6) betekent dat, in een gegeven tijdsinterval, alle afstanden in het Heelal met een vaste factor toenemen. *Het enige spoor van structuur dat in een homogeen isotroop heelal overblijft is de* `schaalfactor` *van het heelal.* Anders geformuleerd: pak de blauwdruk van zo'n heelal. Op die tekening staat in het geheel niets, want een homogeen isotroop heelal heeft geen structuur. Het enige wat er op staat is: "Heelal, stuks 1, schaal 1:a". Hierin is de schaalfactor a een getal dat nog van de tijd kan afhangen (omdat ook H in Eq.(4.6) een functie van t kan zijn).

Deze verrassende vondst is de basis van het `Oerknal-model`. Uitgaande van de *waarneming* van isotropie, en in de *veronderstelling* dat wij geen bijzondere plaats in tijd of ruimte innemen, hebben we *bewezen* dat het Heelal alleen maar kan bewegen volgens de Hubble-regel Eq.(4.6) . Wie dus aan het Oerknal-model wil tornen mag dat doen, maar het heeft weinig zin om het aan te pakken bij H, zoals af en toe met veel tamtam gebeurt. Als er een afwijking is, dan moet die te vinden zijn in het gedrag van de isotropie: pas zodra zou blijken dat ons Heelal anisotroop is, zou het Oerknal-model in ernstige problemen komen.

Hierboven hebben we alleen gekeken naar de `kinematica`, dat wil zeggen het type beweging dat mogelijk is, maar we hebben geen bewegingsvergelijkingen opgesteld of opgelost. Die `kosmische dynamica` gaan we nu aanpakken, en dat moet ook want we zagen boven dat H nog wel van de tijd mag afhangen.

Kies een willekeurig punt in het Heelal. Dat mag, want in een homogeen heelal doet het er niet toe waar je gaat staan. Trek om dat centrum een bol met straal R. Dat is zinvol, want in een isotroop heelal blijft een bol een bol wegens Eq.(4.6) , en de grootte van R doet er niet toe want in een homogeen heelal is elk bolletje, hoe klein ook, representatief voor het geheel. Laat de massa binnen de bol M zijn; dan is

$$M = \frac{4}{3}\pi\rho R^3 \tag{4.7}$$ ♠

In een homogeen heelal is de druk overal hetzelfde, dus versnellingen tengevolge van drukverschillen zijn afwezig. De bewegingsvergelijking voor de bol is dus

$$\frac{dV}{dt} = \frac{d^2R}{dt^2} = -\frac{GM}{R^2} = -\frac{4}{3}\pi G\rho R \qquad (4.8) \spadesuit$$

De energie-integraal hiervan is direct op te schrijven:

$$\frac{1}{2}V^2 = \mathcal{E} + \frac{GM}{R} \qquad (4.9) \spadesuit$$

Hieruit leiden we de eerste-orde bewegingsvergelijking voor R af als

$$V = \frac{dR}{dt} = \sqrt{2\mathcal{E} + \frac{2GM}{R}} \qquad (4.10) \spadesuit$$

Voordat we proberen dit op te lossen, merken we op dat, ten eerste, in de limiet voor zeer kleine R de beweging niet meer afhangt van de constante $\mathcal{E}$. Ten tweede, in de limiet $R \to \infty$ is

$$V(R \to \infty) = \sqrt{2\mathcal{E}} \qquad (4.11)$$

Ten derde, als $\mathcal{E} < 0$ kan R niet oneindig groot worden, maar heeft een maximale waarde

$$R_{\max} = -\frac{GM}{\mathcal{E}} \qquad (4.12)$$

Ten vierde, als $\mathcal{E} = 0$ is de uitdijingssnelheid V exact nul op het moment dat de gemiddelde afstand R tussen de sterrenstelsels oneindig is. Ten vijfde, als $\mathcal{E} > 0$ is de uitdijingssnelheid V positief, zelfs als de sterrenstelsels oneindig ver uit elkaar staan. Ten zesde, omdat de kinematische analyse oplevert $V = HR$, volgt uit Eq.(4.10) dat

$$H = \frac{1}{R}\frac{dR}{dt} = \sqrt{\frac{2\mathcal{E}}{R^2} + \frac{2GM}{R^3}} = \sqrt{\frac{2\mathcal{E}}{R^2} + \frac{8}{3}\pi G\rho} \qquad (4.13)$$

Het interessante hiervan is dat dit een methode oplevert waarmee we (in principe!) de massadichtheid van het Heelal kunnen bepalen uit

$$\frac{8}{3}\pi G\rho = H^2 - \frac{2\mathcal{E}}{R^2} \qquad (4.14)$$

Tenslotte merken we nog op dat de massadichtheid ρ direct samenhangt met de versnelling dV/dt, want Eq.(4.8) zegt dat

$$\frac{1}{R}\frac{dV}{dt} = -\frac{4}{3}\pi G\rho \qquad (4.15)$$

Dat was ook wel te verwachten: de uitdijing wordt immers vertraagd door de onderlinge aantrekking van de materie in het Heelal, die weliswaar geen structuur heeft in het Oerknalmodel maar toch zwaartekracht veroorzaakt. In practische toepassingen wordt vaak gekozen voor een iets andere vorm van Eq.(4.15) waarin de versnelling dimensieloos wordt gemaakt door passende schaling met H. De zo geconstrueerde dimensieloze `vertragingsparameter` is

$$q \equiv -\frac{1}{H^2}\frac{1}{R}\frac{d^2R}{dt^2} \qquad (4.16) \spadesuit$$

Stoppen we de bewegingsvergelijking hierin, dan wordt

$$q = \frac{4}{3}\pi G\rho\left(\frac{2\mathcal{E}}{R^2} + \frac{8}{3}\pi G\rho\right)^{-1} = \left(2 + \frac{6\mathcal{E}}{\pi G\rho R^2}\right)^{-1} \qquad (4.17)$$

Laten we nu Eq.(4.10) proberen te integreren in het unieke geval dat $\mathcal{E} = 0$. Dan is

$$\sqrt{R}\frac{dR}{dt} = \sqrt{2GM} \qquad (4.18)$$

hetgeen onmiddellijk integreert tot

$$\frac{2}{3}R^{3/2} = (t - t')\sqrt{2GM} \tag{4.19}$$

De integratieconstante t' is niet belangrijk, want die geeft alleen maar aan wanneer we het nulpunt van de tijd kiezen. We stellen $t' = 0$ en vinden

$$R = \left(\frac{9}{2}GM\right)^{1/3} t^{2/3} \tag{4.20}$$ ♠

Schrijven we dit in de vorm $R^3 \propto t^2$ dan herkennen we de vorm: alwéér een variant van de Derde Wet van Kepler! Het heelalmodel $\mathcal{E} = 0$ met de oplossing Eq.(4.20) heet het `Einstein-De Sitter model`. Uit het bovenstaande zagen we al waarom dit model zo belangrijk is: *in de limiet $R \to 0$ convergeren alle Oerknal-modellen naar het Einstein-De Sitter model.*

Gebruik makend van Eq.(4.6) en Eq.(4.10) vinden we dat, in het E-DeS model, het getal van Hubble evolueert als

$$H = \frac{1}{R}\frac{dR}{dt} = \frac{2}{3t} \tag{4.21}$$ ♠

Dus H is bepaald niet constant, en het is misleidend om over de "Hubble constante" te praten. Door het waarnemen van snelheden en afstanden van zeer vele sterrenstelsels kunnen we in principe de waarde H_0 van het Hubble-getal op dit moment bepalen, en daarmee hebben we een middel om de leeftijd van het Heelal te berekenen uit

$$t_0 = \frac{2}{3H_0} \approx 1.0 \times 10^{10} \quad \text{jaar} \tag{4.22}$$

Dit geldt voor het Einstein-De Sitter model, dus $q = 1/2$. In werkelijkheid meten wij $q \approx 0.1$, zodat t_0 iets groter is, namelijk

$$t_0 = 1.3 \times 10^{10} \quad \text{jaar} \tag{4.23}$$ ♠

met een statistische onzekerheid van 8% en een onbekende, maar waarschijnlijk vergelijkbare, systematische fout. Merk op dat het vinden van R en V voor slechts een paar sterrenstelsels, hoe nauwkeurig ook, maar bitter weinig zegt over H, omdat H een *gemiddelde* eigenschap is van de beweging van het Heelal. Dus is middeling vereist over een zeer groot aantal sterrenstelsels tot op zeer grote afstand; immers, we vonden dat het Heelal *pas op grote afstanden* isotroop is! Dat betekent dat we direct al tegen problemen oplopen. Ten eerste is het een forse inspanning om van sterrenstelsels de afstand te bepalen (de snelheid gaat vrij gemakkelijk met de roodverschuiving). Ten tweede moeten we liefst zeer grote waarden van R bereiken, maar dat betekent dat de sterrenstelsels erg lichtzwak zijn. Het laatste woord over H, en vooral q, is voorlopig nog niet gezegd.

5. Basic Cosmological Relations

Solutions of the cosmic equation of motion, which will be used extensively in what follows.

To begin with, let us assume that the internal energy density of the gas is dynamically unimportant ("dust"). Take a sphere with radius R and mass density ρ, expanding uniformly with speed v at the surface. It contains a mass $M = \frac{4}{3}\pi\rho R^3$, so that the acceleration on its surface is

$$\frac{dv}{dt} = -\frac{GM}{R^2} \tag{5.1}$$ ♠

Since $v = dR/dt$, one deduces

$$\frac{dv}{dt} = \frac{dR}{dt}\frac{dv}{dR} = \frac{1}{2}\frac{dv^2}{dR} \tag{5.2}$$

$$v^2 = \frac{2GM}{R} + 2\mathcal{E} \tag{5.3}$$

$$\frac{dR}{dt} = \sqrt{\frac{2GM}{R} + 2\mathcal{E}} \tag{5.4} \spadesuit$$

For future reference, we note that the energy constant $2\mathcal{E}$ is often written as $-k$. For the moment we consider the `Einstein-De Sitter model` $\mathcal{E} = 0$, which integrates to

$$\frac{2}{3}R^{3/2} = t\sqrt{2GM} \tag{5.5}$$

(the integration constant can be set to zero by a suitable choice of zero point for the cosmic time). Note the analogy with Kepler's Third Law:

$$R^3 = \frac{9GM}{2}t^2 \tag{5.6}$$

Because M is arbitrary, one usually relates it to local observables by

$$M = \frac{4}{3}\pi\rho_0 R_0^3 \tag{5.7}$$

The function $R(t)$ (sometimes written as a or another symbol in the literature) is called the *scale factor*. In the Newtonian view used here, R is identified with the mean separation between constituents of the Universe (e.g. galaxies). In the correct relativistic form, $R(t)$ says by what fraction the Universe has changed its spatial dimensions by the time t. Obviously

$$\frac{R}{R_0} = \left(\frac{t}{t_0}\right)^{2/3} \tag{5.8} \spadesuit$$

$$\rho = \rho_0\left(\frac{R_0}{R}\right)^3 \tag{5.9} \spadesuit$$

$$t_0^2 \equiv \frac{1}{6\pi G\rho_0} \tag{5.10} \spadesuit$$

Now consider the case $\mathcal{E} \neq 0$:

$$\frac{dR}{dt} = \sqrt{\frac{2GM}{R} + 2\mathcal{E}} \tag{5.11}$$

First, we simplify the expression by scaling all lengths with a factor x' and all times with a factor t',

$$x' \equiv \frac{GM}{\mathcal{E}} \tag{5.12}$$

$$t' \equiv \frac{GM}{\mathcal{E}\sqrt{2\mathcal{E}}} \tag{5.13}$$

so that we can use R/x' instead of R and t/t' instead of t, to obtain

$$\frac{dx}{dt} = \sqrt{\frac{1}{x} - k} \tag{5.14}$$

in which k is $+1$ or -1, depending on the sign of $-\mathcal{E}$. First, we consider the case $k = 1$. Rewriting Eq.(5.14) as

$$dt = \frac{\sqrt{x}}{\sqrt{1-x}}\,dx \tag{5.15}$$

one immediately guesses the substitution

$$x = \sin^2 \phi \tag{5.16}$$

which leads to the solution

$$t = \phi - \sin \phi \cos \phi \tag{5.17}$$

Using this, one finds the parametric solution of the equation of motion,

$$R = \frac{GM}{\mathcal{E}} \sin^2 \phi \tag{5.18}$$

$$t = \frac{GM}{\mathcal{E}\sqrt{2\mathcal{E}}} (\phi - \sin \phi \cos \phi) \tag{5.19}$$

This is the representation of a *cycloid*.

Exercise.

Solve the equation of motion for $k = -1$.

---

6. Some Expressions for Observables

From the geometric behaviour of the cosmic expansion, we can derive certain observable quantities, such as the instantaneous expansion rate and the deceleration of the expansion. These are often given in terms of the redshift, which is not a Doppler shift but a change in wavelength due to the change of the structure of space-time.

Kinematic observables may be derived from the function $a(t)$ by defining

$$H \equiv \frac{1}{a}\frac{da}{dt} = \frac{d \log a}{dt} \tag{6.1}$$ ♠

$$q \equiv -a \left(\frac{da}{dt}\right)^{-2} \frac{d^2 a}{dt^2} = -\frac{1}{H^2}\frac{1}{a}\frac{d^2 a}{dt^2} \tag{6.2}$$ ♠

The first is called the `Hubble parameter`, describing the instantaneous expansion rate; the second is the `deceleration parameter`. Using the equation of motion, we see straight away that

$$H = \sqrt{\frac{8}{3}\pi G \rho} \tag{6.3}$$ ♠

$$q = \frac{4\pi G \rho}{3H^2} \tag{6.4}$$ ♠

The current value of H, indicated by H_0, is observed to lie somewhere between 50 and 90 km s^{-1} Mpc^{-1}, with a best estimate in the neighbourhood of

$$H_0 \approx 65 \text{ km s}^{-1} \text{ Mpc}^{-1} = 2.107 \times 10^{-18} \quad \text{s}^{-1} \tag{6.5}$$

in which case, if our Universe is Einstein-De Sitter,

$$\rho_0 = \rho_c \equiv \frac{3H_0^2}{8\pi G} = 7.94 \times 10^{-27} \text{ kg m}^{-3} = 117.2 \ \text{M}_\odot \text{ kpc}^{-3} \tag{6.6}$$

Instead of q, one often uses the ratio of the present cosmic density and the critical density:

$$\Omega \equiv \frac{\rho_0}{\rho_c} = 2q_0 \qquad (6.7)\ ♠$$

The consequences of the model universe for the observation of distant objects can be found by calculating observables such as brightness, colour, and angular extent. Observations indicate that our Universe is isotropic (unevenness of the order of a few per cent at a distance above 100 Mpc). Assuming that we occupy no special position in space or in time, we conclude that the Universe is isotropic everywhere, so that it must be *homogeneous* too.

Hence we conclude that the most plausible metric for our Universe is *maximally symmetric*. Such spaces obey

$$ds^2 = c^2 dt^2 - (\text{scalar multiple of spherical volume element}) \qquad (6.8)$$

In spherical spatial coordinates about an arbitrary point, this can be written as

$$ds^2 = c^2 dt^2 - a^2(t)\{f(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\} \qquad (6.9)\ ♠$$

The function $a(t)$, the `scale factor` which describes the evolution of all length scales in this metric, can be determined in several ways. In a maximally symmetric space, an infinitesimal patch is representative of all the rest, so we can take the Newtonian limit for the motion (unless the cosmic gas itself is so hot that it is relativistic; we will see more about that later).

In the Einstein-De Sitter case, it can be shown that the function $f(r) = 1$, which makes things easier. In the general case, we can deduce its form by exploiting the requirement of maximal symmetry, in which case the spatial term in the line element must remain unchanged if we pick a different origin; or simply by noting that a maximally symmetric space must be globally symmetric, i.e.

$$a^2 = x^\mu x_\mu = x_0^2 - x_1^2 - x_2^2 - x_3^2 \qquad (6.10)$$

with some constant a^2. This a is a fictitious direction, an "embedding coordinate", which is used for mathematical convenience to express the spherical symmetry of the model (it does certainly not mean that we can travel in a physical direction a.) By eliminating $x^0 x_0$ from this and inserting the result in the line element

$$ds^2 = dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 \qquad (6.11)$$

one finds that

$$ds^2 = \frac{(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)^2}{a^2 + x_1^2 + x_2^2 + x_3^2} - \left(dx_1^2 + dx_2^2 + dx_3^2\right) \qquad (6.12)$$

We can now go to spherical spatial coordinates by using

$$x_1 \equiv r' \sin\theta\cos\phi \qquad (6.13)$$

$$x_2 \equiv r' \sin\theta\sin\phi \qquad (6.14)$$

$$x_3 \equiv r' \cos\theta \qquad (6.15)$$

$$r \equiv r'/a \qquad (6.16)$$

so that we get

$$f(r) = \frac{1}{1 - kr^2} \qquad (6.17)\ ♠$$

The `curvature constant` k is zero in the Einstein-De Sitter case.

It is important to realize that this $a(t)$ is fundamentally different from the function $R(t)$ in Newtonian cosmology. Whereas R is the mean distance between galaxies, a is the `scale factor` of the Universe, that is, the scale of space itself. In the general relativistic case space-time is considered to be real stuff, the cement between the bricks of the particles, while in the Newtionian picture space is merely invisible graph paper.

Furthermore, one notes that a photon has interval zero, so that it obeys

$$0 = c^2 d^2 t - a^2 (1 - kr^2)^{-1} d^2 r \tag{6.18}$$

assuming that it is not scattered on its way, so that we may use $d\phi = d\theta = 0$. In the E-DeS case, this becomes simply

$$c^2\ d^2 t = a^2(t)\ d^2 r \tag{6.19}$$

If the light ray is emitted at (t_1, r_1), it arrives here at $(t_0, 0)$, and integration of the path equation gives a `look-back distance`

$$r_1 = c \int_{t_1}^{t_0} \frac{1}{a} dt = 3ct_0 \left(1 - \tau_1^{1/3}\right) = \frac{2c}{H_0} \left(1 - \tau_1^{1/3}\right) \tag{6.20}$$

in which I have introduced the dimensionless time

$$\tau \equiv \frac{t}{t_0} \tag{6.21}$$ ♠

that is to say, the time expressed in terms of the current age of the Universe. The index 1 simply refers to t_1, namely $\tau_1 = t_1/t_0$. The quantity $2c/H_0$ is usually called the `horizon distance`

$$r_H \equiv \frac{2c}{H_0} \tag{6.22}$$ ♠

From this we can calculate the shape of the past light cone. We look back to a certain time in the past. At a fixed time, we see a hypersurface on which the geometry is described by $dt = 0$ and $dr = 0$:

$$ds^2 = -a^2 r^2 (d\theta^2 + \sin^2\theta\, d\phi^2) \tag{6.23}$$

This is a sphere on which the arc element is ra. We obtain the shape of the past light cone by multiplying the lookback distance r_1 with a_1/a_0, because this gives us the radius over we observe a light wave to have expanded at that time in the past:

$$r_1^* = \frac{a_1}{a_0} r_H \left(1 - \tau_1^{1/3}\right) = r_H \left(\tau_1^{2/3} - \tau_1\right) \tag{6.24}$$

The geometry of the Universe could be observed if we could persuade some galaxies to emit specific signals at a sequence of times t_1. Of course that is impossible. But we can use spectral lines as intrinsic atomic clocks with period P. Then we get a distance between wave crests given by the difference between the r_1 above and

$$r_1' = c \int_{t_1+P_1}^{t_0+P_0} \frac{1}{a} dt \tag{6.25}$$

The period is always much less than $1/H$, so that

$$\frac{P_0}{P_1} = \frac{a_0}{a_1} = \frac{\nu_1}{\nu_0} = \frac{\lambda_0}{\lambda_1} \tag{6.26}$$

That is to say, the wavelength scales with a, entirely as expected ("rubber sheet" analogy). Usually, this ratio is expressed by means of the `redshift` z:

$$1 + z \equiv \frac{\lambda_0}{\lambda_1} = \frac{a_0}{a_1} \tag{6.27}$$ ♠

A very important aspect of this equation is, that it shows that *cosmic redshift has nothing to do with the classical Doppler shift.* The change of wavelength is not due to a velocity, but is due to the change of the structure of space itself, as is evident from the factor a_0/a_1. Accordingly, the derivative da/dt is not a velocity in the usual sense; and the fact that $da/dt \to \infty$ as $t \to 0$ by no means implies that things move faster than light in the early Universe. But the expression $v = Hr$, where $H \equiv a^{-1}da/dt$, *is* seen as a velocity between neighbouring points in the Universe.

Using this expression, we get

$$r_1 = \frac{2c}{H_0}\left(1 - \sqrt{\frac{a_1}{a_0}}\right) \tag{6.28}$$

$$= \frac{2c}{H_0}\left[1 - (1+z)^{-1/2}\right] \tag{6.29}$$

$$\frac{a_1}{a_0} = \left(1 - \frac{H_0 r_1}{2c}\right)^2 \tag{6.30}$$

In a patch near our Galaxy, we expect z to be small. Expansion in powers of z gives

$$r_1 \simeq \frac{c}{H_0}(z - \frac{3}{4}z^2) \tag{6.31}$$

confirming that local distances scale with z, as they should.

Looking out in space is looking back in time. In a static space, we would expect that the circumference of a circle increases linearly with the circle's radius. But along any radius, the circle we see effectively spans less space, because we look back to a time when the scale factor a was less than unity. Thus, the circumference of the circle is proportional to ra, while the lookback radius is given by Eq.(6.20) . Accordingly, the shape of a lookback angle is no longer a wedge $y = r$, but a "petal" with the shape

$$y = ra = r\left(1 - \frac{rH_0}{2c}\right)^2 \tag{6.32}$$

If we have a source with a finite size D at distance r_1, then its `angular size` δ as seen by us is D/r_1, corrected for the geometric change due to the cosmic scale factor:

$$\delta = \frac{D}{a_1 r_1} \tag{6.33}$$

Using the expressions for the E-DeS model,

$$\delta = \frac{DH_0}{2c}(1+z)\left[1 - (1+z)^{-1/2}\right]^{-1} \tag{6.34}$$

This expression has a minimum in the vicinity of $z \approx 1$:

$$\frac{d\delta}{dz} = \frac{DH_0}{2c}\left[1 - (1+z)^{-1/2}\right]^{-2}\left[1 - \frac{3}{2}(1+z)^{-1/2}\right] \tag{6.35}$$

which is zero when $1 + z = 9/4$. Therefore, galaxies appear smallest at a redshift

$$z_{\min} = \frac{5}{4} \tag{6.36}$$

at which point an object with intrinsic size D subtends an angle

$$\delta_{\min} = \frac{27DH_0}{8c} \tag{6.37}$$

Using the value of $H_0 = 65\ \mathrm{km\,s^{-1}Mpc^{-1}}$ and $D = 50$ kpc, one finds

$$\delta_{\min} = 3.66 \times 10^{-5}\ \ \mathrm{rad} = 7.55\ \ \mathrm{arcsec} \tag{6.38}$$

Along the same lines, the power P one receives from a source with intrinsic luminosity L is affected by the increase in distance (r_1^{-2}) and the cosmic scale factor (a_1^{-2}). We look at a source at a fixed distance, that is to say, at a fixed time in the past. The line element at fixed time t_1 has $dt = 0$, so that

$$ds^2 = -r^2 a^2 (d\theta^2 + \sin^2\theta d\phi^2) \tag{6.39}$$

is the arc length on a sphere with radius ra surrounding the observer. Thus, the light of a source that emits a power $S(t_1)$ at time t_1 is spread out over a sphere that is seen to have surface $4\pi r^2 a^2$ by an observer at time t_0. The energy received per unit time is then inversely proportional to this factor. Furthermore, this received power is inversely proportional to $1+z$ due to time dilatation. There is another factor $1+z$ because of the photon redshift (these two factors are the same as in the usual relativistic Doppler shift equation). The net result is that the `received power` P is related to the intrinsic power L by

$$P = \frac{L}{4\pi a_1^2 r_1^2 (1+z)^2} \tag{6.40}$$

Because $1+z$ is proportional to a_1 this is a steep dependence on $1+z$, and explains in part why it is so very difficult to see extended objects at cosmological distances.

The corresponding `surface brightness` is

$$\Sigma \equiv \frac{P}{\delta^2} = \frac{L}{4\pi a_1^2 r_1^2 (1+z)^2} \left(\frac{a_1 r_1}{D}\right)^2 = \frac{L}{4\pi D^2} \left(\frac{a_1}{a_0}\right)^2 \tag{6.41}$$

Remember that in the 'usual' propagation of light, the surface brightness is *independent* of distance!

7. Cosmic Nucleosynthesis

When tracing the cosmic expansion back in time, we find that the energy per particle (the 'temperature') was much higher in the past than it is today. Indeed, there once was a time when energies were high enough to permit nuclear fusion. However, the fusion processes were impeded by photodissociation. the reason for this is, that our Universe is exceedingly rich in photons: about 10^{10} photons per nucleon.

The physical properties and processes that are relevant for cosmic nucleosynthesis are microscopic and macroscopic. The microscopic ones are the fundamental particles and interactions between them. The macroscopic ones are the parameters of the cosmological model, especially the expansion rate, the radiation temperature, the particle densities (especially the photon/baryon ratio) and the species of particles available.

The microscopic physics we must take almost for granted, although some uncertainties in the Standard Model (such as the number of lepton families, a possible non-zero rest mass for neutrinos, the possible presence of exotic particles such as magnetic monopoles or axions) can be narrowed down by means of cosmological tests. The physical parameters of the macroscopic model (almost always a variant of the standard FRW Big Bang) have a profound influence on the microscopic processes, because of the way in which they determine the thermal history of the Universe.

It is remarkable that the interplay between particles impresses itself on the Universe as a whole, thus forging a link between the properties of our Universe on a very small scale and those on the largest possible scales. Whether these two extremes are somehow fundamentally related is a completely open question.

First, consider the microscopic entities involved in nucleosynthesis. All fundamental interactions are important: the electromagnetic force because electric repulsion impedes the merging of protons; the weak force because it allows transformation from proton to neutron and *vice versa*; and the colour force, which makes nucleons out of quarks and binds them together by the residual colour force, called the 'strong force'.

The basic interactions*2 are all examples of `gauge forces` (but the strong force – being in essence derived from the colour force – is not). In a gauge force, a basic multiplet of fermions ('matter') is interconnected by the exchange of vector bosons ('force'). The dimension of the multiplet determines the complexity of the interaction; the detailed behaviour of the interaction is governed by the `symmetry group` corresponding to that dimension, in particular to the structure constants of the group. The group is introduced as a `local symmetry`, that is to say the symmetry operation depends on the position in spacetime.

For example, the electromagnetic interaction is due to local $U(1)$ symmetry. The $U(1)$ group corresponds to phase rotation. The way in which local symmetry generates an interaction is briefly explained in the introduction of the section on the gravitational Lagrangian. The weak interaction is associated with the group $SU(2)$; in the unified electroweak theory, one uses the product group $U(1) \otimes SU(2)$. The colour force is associated with $SU(3)$. The number of field bosons in an N-dimensional SU group is $N^2 - 1$; thus, electromagnetism is carried by a single boson (the photon), weak forces by three (the W^+, Z^0 and W^-), and colour forces by eight (the eight gluons).

Now consider the macroscopic entities, in particular the overall thermodynamic variables in the Universe. The temperature at which nuclear fusion is expected to occur must be less than the rest mass temperature of the lightest baryons, which are the nucleons p and n:

$$kT_p = m_p c^2 = 1.50 \times 10^{-10} \text{ J} = 938 \text{ keV}$$
$$T_p = m_p \frac{c^2}{k} = 1.09 \times 10^{13} \text{ K} \tag{7.1}$$

The time at which this temperature prevails is estimated as follows. For simplicity, and because it probably isn't far wrong, we will use the Einstein-De Sitter case for the scale factor a:

$$a = a_0 \times \begin{cases} \tau^{2/3} & \text{nonrelativistic dust} \\ \tau_e^{1/6}\,\tau^{1/2} & \text{relativistic gas} \end{cases} \tag{7.2}$$ ♠

Here and in all the following cases, the top line after the brace refers to the solution for nonrelativistic dust, and the bottom one to the solution for relativistic matter. In this expression we have used the abbreviations

$$\tau \equiv \frac{t}{t_0}\ ; \qquad \tau_e \equiv \frac{t_e}{t_0} \tag{7.3}$$

The time t_e indicates the time at which the density of matter and that of photons are equal. The density of baryons obeys

$$\rho_b \propto a^{-3} \tag{7.4}$$ ♠

Because the energy of a photon is inversely proportional to its wavelength, and because this wavelength scales with a, we get

$$\rho_\gamma \propto a^{-4} \tag{7.5}$$ ♠

Inserting the E-DeS values, we obtain

$$\rho_b = \rho_{be} \times \begin{cases} (t/t_e)^{-2} \\ (t/t_e)^{-3/2} \end{cases} \tag{7.6}$$

and

$$\rho_\gamma = \rho_{\gamma e} \times \begin{cases} (t/t_e)^{-8/3} \\ (t/t_e)^{-2} \end{cases} \tag{7.7}$$

*2 See e.g. my book *The force of symmetry*, Cambridge Univ. Press, 1997.

Here the index e indicates the value at equality, $t = t_e$. Relating the baryon density to its present value by $\rho_b = \rho_0 (t_0/t)^2$, we get

$$\rho_b = \rho_0 \times \begin{cases} \tau^{-2} \\ \tau_e^{-1/2}\,\tau^{-3/2} \end{cases} \tag{7.8}$$

and for the photons

$$\rho_\gamma = \rho_0 \times \begin{cases} \tau_e^{2/3}\tau^{-8/3} \\ \tau^{-2} \end{cases} \tag{7.9}$$

These expressions can only be equated if we have a good observed value of the mean mass-to-light ratio in the Universe today. The mean particle mass density follows from observation of Ω; we will use the E-DeS-value and assume critical density (which can be calculated directly from the Hubble parameter). The mean photon mass density follows from observation of the temperature of the microwave background and the `Stefan-Boltzmann Law` for the energy density of thermal radiation:

$$\mathcal{E} = \frac{\pi^2 k^4}{15\hbar^3 c^3} T^4 \tag{7.10}$$

Division by c^2 yields the corresponding mass density:

$$\rho_\gamma = \frac{\pi^2 k^4}{15\hbar^3 c^5} T^4 = 8.40 \times 10^{-33}\ T^4 \tag{7.11}$$

Inserting, as before, $H_0 = 65$ km s^{-1} Mpc^{-1} $= 2.107 \times 10^{-18}$ s^{-1} and the observed value $T_{\rm CMBR} = 2.737$ K, we get

$$\frac{\rho_\gamma}{\rho_0} = \tau_e^{2/3} \tag{7.12}$$

because today, by definition, $\tau = 1$; therefore

$$\tau_e = 4.57 \times 10^{-7} \tag{7.13}$$

The temperature is now estimated as follows. We will neglect possible endo- or exothermic effects, for example those associated with the recombination energy of hydrogen. The temperature evolves according to $T \propto 1/a$ so that in the present (nonrelativistic) universe we have

$$T_\gamma = T_{\rm CMBR}\ a^{-1} = T_{\rm CMBR}\ \tau^{-2/3} \tag{7.14}$$

Matter and radiation were last in equilibrium at the cosmic photosphere, where hydrogen recombined. The corresponding temperature follows from the recombination reaction

$$\mathrm{e}^- + \mathrm{p}^+ \quad \leftrightarrows \quad \mathrm{H} + \gamma \tag{7.15}$$

In or close to equilibrium, the number densities n_e, n_p, and $n_{\rm H}$ of the particle species involved is found from `Saha's equation`. This formula relates the number densities on the left and on the right of the reaction equation by the expression

$$\frac{n_e n_p}{n_{\rm H}} = \frac{[\text{occupied velocity volume}]}{[\text{number of phase space cells}]} \times [\text{Boltzmann factor}] \tag{7.16}$$

Using the definition of the ionization factor x, and assuming a pure hydrogen plasma with ionization potential χ, the particle densities are related by

$$n_e = n_p = x\,n \tag{7.17}$$

so that the Saha equation reads

$$\frac{x^2}{1-x} = \frac{(2\pi m_e k T)^{3/2}}{n(2\pi\hbar)^3} \mathrm{e}^{-\chi/kT} = 2.415 \times 10^{21}\ n^{-1} T^{3/2}\, \mathrm{e}^{-1.578\times 10^5/T} \tag{7.18}$$

Now solve τ from the expression Eq.(7.14) for T, put it into the equation

$$n = n_c = \frac{3H_0^2}{8\pi G m_{\rm H}}\,\tau^{-2} = 4.747\;\tau^{-2} \quad {\rm m}^{-3} \tag{7.19}$$

for the Einstein-De Sitter density, substitute this n into Eq.(7.18) , and take the logarithm. This gives an equation for T of the form

$$T = \frac{1}{p - q\,\log\,T} \tag{7.20}$$

which can be solved by iteration: guess T, put it into the right hand side, compute a new T from Eq.(7.20) , and so forth. Taking the half-ionization point $x = 0.5$ for the time of recombination, we conclude that the ionization has dropped to 0.5 when

$$T_{\rm rec} = 4053 \quad {\rm K} \tag{7.21}$$

which corresponds to a dimensionless time

$$\tau_{\rm rec} = 1.76 \times 10^{-5} \tag{7.22}$$

This is a very interesting equation in the sense that it relates a global property of the universe to purely local, small-scale, atomic properties, namely the particle masses and the ionization energy of hydrogen.

The thermal properties of this stage of the Universe are quite remarkable. To begin with, the Universe was very nearly in thermal equilibrium, due to the close coupling between photons and charged particles, in particular electrons. Thus, the name “Big Bang” turns out to be very inappropriate (small wonder: it was meant as ridicule by its inventor). After all, the closeness to equilibrium nowhere nearly corresponds to a violent, non-equilibrium event like an explosion. Besides, it suggests some sort of bomb going off in an already-existing space-time continuum, an endless source of confusion in the minds of the general public. This near-equilibrium is broken only a very little by the fact that the Universe expands.

The second remarkable property is that the number η of photons per baryon is so stupendously large: something like $\eta \approx 10^9 - 10^{10}$. It is immediately obvious that annihilation processes of the type

$$X + Y \quad \rightarrow \quad 2\gamma$$

make practically no difference at all: if we got one extra photon per baryon in this way, we’d have $\eta = 10^9 + 1$. The improbable value of η immediately implies that, even though the Universe as a whole is close to thermal equilibrium at the epoch we are considering, nucleochemistry cannot be anywhere near equilibrium. At the energy-per-particle that corresponds to nuclear binding energies, photodissociation will undo all the gains made by fusion processes. Thus we must wait until the Universe has cooled sufficiently for these photons to have lost their destructive power. However, at such a low temperature fusion reactions between charged particles are much more difficult. Nuclear barriers can be penetrated by neutrons, but these become rare; besides, for the atomic number A to increase we must wait for weak processes to convert n to p, a slow process, 100 sec compared with a local Hubble parameter of about 0.1 $\rm s^{-1}$.

The nuclear equilibrium can be calculated from equations of the type Eq.(7.15) which is, in fact, a model for all cosmic reactions of the type

$$A + B \quad \leftrightarrows \quad C + D \tag{7.23}$$

Note, however, that this type of equation can be used only if the particle collision time is much shorter than the current cosmic expansion time scale; if not, we have a regime of particles `decoupling`. The collision time t_c is found by dividing the mean free path λ (see Sec.9) by the particle speed v:

$$t_c = \frac{\lambda}{v} = \frac{1}{n\sigma v} \tag{7.24}$$

We get decoupling when this is of the order of the instantaneous Hubble time. In the E-DeS case we have

$$\frac{1}{n\sigma v} = t = \frac{2}{3H} \tag{7.25}$$

In the nonrelativistic case,

$$\frac{1}{2}mv^2 = \frac{3}{2}kT \tag{7.26}$$

The evolution of the matter temperature is found from the ideal gas law and the adiabatic equation of state:

$$PV \propto Pa^3 \propto T \tag{7.27}$$

$$P \propto \rho^\gamma \propto a^{-3\gamma} \tag{7.28}$$

from which it follows that

$$T \propto a^{3(1-\gamma)} \tag{7.29}$$

or, in the case of a monatomic gas (for which $\gamma = 5/3$),

$$T = T_0\, a^{-2} = T_0\, \tau^{-4/3} \tag{7.30}$$

The current matter temperature T_0 evolved by adiabatic expansion from the time when matter and radiation were last in equilibrium. We will see later (Sec.10) that $T = T_d \approx 4053$ at $\tau = \tau_d$, so that

$$T = T_d\, \tau_d^{4/3}\, \tau^{-4/3} \tag{7.31}$$

With Eq.(7.26) one then finds

$$v = \sqrt{\frac{3k}{m}T} \tag{7.32}$$

This can be used in Eq.(7.25) to obtain

$$\frac{1}{n\sigma v} = \frac{m}{\rho\sigma v} = \frac{ma}{\rho\sigma}\sqrt{\frac{m}{3kT_0}} = \frac{ma^4}{\rho_c\sigma}\sqrt{\frac{m}{3kT_0}} = t = t_0\, a^{3/2} \tag{7.33}$$

In the E-DeS model, the time is related to the density by Eq.(5.10), so that

$$a^5 = \left(\frac{H_0\sigma}{4\pi Gm}\right)^2 \frac{3kT_0}{m} \tag{7.34}$$

At this value of the scale factor a, collisions occur on a time scale equal to the cosmic evolution time scale. The equation takes on a slightly more intuitive form when we note that $3kT/m$ is equal to the square of the sound speed s, and so

$$a = \left(\frac{H_0\sigma s_0}{4\pi Gm}\right)^{2/5} \tag{7.35}$$

A similar exercise, albeit more involved, can be done for the collision rate in the relativistic epoch of the Universe.

Having found Eq.(7.13) , Eq.(7.14) and Eq.(7.22) , we can compute the temperature at any time in the past. If reactions like Eq.(7.23) occur, any endo- or exothermic effects can be taken into account via a Saha-type equation, in which χ plays the role of the reaction energy (or the `chemical potential`). The species A, B, C and D could be anything. In the above, we used the formalism to compute the recombination of hydrogen, but we might as well have used the nucleochemical reaction

$$H_1^1 + H_1^1 \quad \leftrightarrows \quad D_1^2 + e^+ + \nu \tag{7.36}$$

which is one step in the synthesis of helium in the early Universe. Most of the details of this stuff are given as cookbook recipes, because it is fairly messy; compare the way in which nucleosynthesis is computed in stellar evolution.

The potential barrier can be overcome, as we saw in the discussion of the Gamow factor, by protons coming in at high energy. There is a trade-off between the the Boltzmann distribution, which says that particles with high energy are rare, and the tunnel effect, which says that particles with high energy can pass more easily. At the Gamow energy, there is a compromise between these two. There is a fairly broad range of energies around this special value for which appreciable tunneling occurs. In nucleosynthesis we have to start somewhere, and for the moment we will take the raw materials (protons and neutrons) as given, provided by processes (baryosynthesis) in the earlier Universe. These, in turn, can be described by expressions of the form Eq.(7.23) but the behaviour of quarks is much more complicated and less understood.

In the phase before that, we would have to compute why there is matter at all: why hasn't everything annihilated against its antiparticles? The reason is probably the breaking of CP symmetry (charge-parity). For more details, see the book by Kolb & Turner. Apparently, the asymmetry was extremely snall, because there are now so many photons left over. Since there are about 10^{10} photons per baryon, we tentatively conclude that, of the 10 billion pairs of particle/antiparticle pairs, only one did not manage to annihilate due to the CP violation.

This immense ocean of photons interferes with nucleosynthesis. Eventually, the expansion of the Universe freezes all reaction rates. This heavily non-equilibrium form of nucleosynthesis differs markedly from the more sedate processes that occur in stars, which can go to completion because there is so much time (remember that the characteristic time scale for the expansion is of the order of 10 seconds, whereas the time scale for stars is more like 10 billion years). In the Universe, nucleosynthesis occurs on the exponential part of the tunneling curve.

Because these synthesis processes take place so far from equilibrium and are frozen out, the reaction products are very dependent on certain global properties of the Universe, most notably the photon/baryon ratio η and the kinematic parameters H and Ω. Conversely, observations of the cosmic abundances can give us information about these quantities. The present data imply that the relative baryon density Ω_b is smaller than the global value of Ω. Thus, there is a possibility that some or most of the dark matter is not baryonic. In any case, the value of Ω found from nucleosynthesis is much lower than the mythical value 1. It is quite remarkable that the value of Ω from nucleosynthesis calculations and from observations of the dynamics of galaxy clusters gives about the same result: $\Omega = 0.3$. For the moment, it seems that the champions of $\Omega = 1$ will have to resort to some rather strange contortions to get what they want.

8. De Gamow factor

Uit het verband tussen massa, straal en temperatuur van bollen zien wij dat in het binnenste van een voorwerp als de Zon de temperatuur van de orde van 10 miljoen kelvin is. Onder die omstandigheden zijn de electronen geheel los van de atoomkernen. Die kernen botsen tegen elkaar en zo kan er kernfusie optreden, wardoor de ster kan blijven stralen. Om te berekenen hoe dat gaat moeten we weten hoe twee protonen zo dicht bij elkaar kunnen komen dat de 'sterke kernkracht' ze bijeen bindt. Dit is een toepassing van de quantummechanica. We volgen de methode van Gamow om te schatting bij welke energie kernreacties in een ster het best verlopen.

Het nucleaire kookboek dat bepaalt hoe in een ster de verschillende reacties verlopen, is uitermate ingewikkeld en niet geschikt voor een korte presentatie. Maar het is wel mogelijk om de essentie van het fusie-gedrag te doorzien, met behulp van een combinatie van twee stukken fysica: de quantummechanica, die bepaalt hoe deeltjes met een gegeven energie wisselwerken, en de thermodynamica, die de energie-verdeling van deeltjes beschrijft. Het resultaat van beide is een uitdrukking voor de waarschijnlijkheid dat bepaalde reacties optreden.

Eerst kijken we naar de verdeling van de energieën E van een verzameling deeltjes met massa m. Bij een botsing tussen twee deeltjes is het gemiddeld even waarschijnlijk dat een deeltje aan energie wint, als dat het verliest; anders zou er geen evenwicht zijn. Laat de kans op een energieverlies ΔE gegeven zijn door

$$P = \beta\,\Delta E \tag{8.1}$$

De kans dat een deeltje bij de botsing energie wint is dan $1-P$. De waarschijnlijkheid dat een deeltje na N botsingen steeds energie heeft gewonnen is, wegens de vermenigvuldigingsregel voor samengestelde kansen, gelijk aan

$$w = (1-P)^N = (1-\beta\,\Delta E)^N \tag{8.2}$$

Na N botsingen is van dat deeltje de energie opgelopen tot

$$E = N\,\Delta E \tag{8.3}$$

en zodoende is

$$w = \left(1 - \beta\frac{E}{N}\right)^N \tag{8.4}$$

Wie snel is met analyse, weet uit het hoofd dat dit, in de limiet voor zeer veel stapjes ($N \to \infty$), geschreven kan worden als

$$w = \mathrm{e}^{-\beta E} \tag{8.5}$$

De kans dat we in een thermisch gas een deeltje met energie E aantreffen neemt dus exponentieel af met die energie. De bijbehorende e-waarde is de `thermische energie`

$$\frac{1}{\beta} \equiv kT \tag{8.6}$$

De waarschijnlijkheid w moet nog worden genormaliseerd, zodanig dat de integraal over de hele verdeling gelijk is aan 1, maar dat zullen we hier niet doen. Wij zien wel dat de `verdelingsfunctie` f van de energie in een thermisch gas wordt gegeven door de `Boltzmann-verdeling`

$$f \propto \mathrm{e}^{-E/kT} \tag{8.7}$$ ♠

dat wil zeggen: de kans dat een deeltje in een gas met temperatuur T een energie E heeft, neemt exponentieel af met toenemende E.

Nu vragen wij ons af, wat de kans is dat een deeltje met zekere energie E een ander deeltje 'raakt'. Wat wij daar hier mee bedoelen is dit: stel dat het deeltje afgestoten wordt door een centrale potentiaal $V(r)$. Wat is dan de kans dat het deeltje het punt $r = 0$ bereikt? Klassiek is het antwoord duidelijk: die kans is nul, want de minimale waarde van r is de oplossing van de vergelijking voor het omkeerpunt,

$$V(r) - E = 0 \tag{8.8}$$

Maar in de quantummechanica ligt dat subtieler. De gevolgen van het quantumgedrag kunnen we schatten uit de tijdsonafhankelijke `Schrödingervergelijking`

$$\frac{d^2\psi}{dx^2} = \frac{2m}{\hbar^2}(V-E)\psi \tag{8.9}$$ ♠

In het geval dat $V - E$ constant is, kennen we de oplossing, namelijk de vlakke golf

$$\psi = e^{ir\sqrt{2m(V-E)/\hbar^2}} \tag{8.10}$$

In het algemeen is V een functie van r, maar als dat een redelijk brave functie is, kunnen we hopen dat de vorm

$$\psi = e^{iu(r)} \tag{8.11}$$

een goede benadering geeft. Substitutie van deze ψ in Eq.(8.9) levert

$$i\frac{d^2u}{dr^2} - \left(\frac{du}{dr}\right)^2 = \frac{2m}{\hbar^2}(V - E) \tag{8.12}$$

Nu weten we al, dat voor een deeltje bij constante V geldt $u \propto r$ (zie Eq.(8.10)), zodat de tweede afgeleide nul is. Wij maken nu de *veronderstelling* dat die tweede afgeleide verwaarloosbaar is (dat komt neer op 'braaf' gedrag van V). Dit noemt men de `WKBJ-benadering`. In de klassiek-verboden zone, $E < V$, vinden wij dan

$$u = i\int_a^0 \sqrt{\frac{2m}{\hbar^2}(V(r) - E)}\, dr \tag{8.13}$$

Nu kunnen we uit Eq.(8.11) en uit het product $\psi^*\psi$ de kans uitrekenen dat het deeltje inderdaad het punt $r = 0$ bereikt.

In het vervolg zijn wij aleen geïnteresseerd in de manier waarop die kans afhangt van de energie E van het deeltje. Numerieke factoren zijn van later zorg. Het gaat dus om de integraal

$$I = \int_a^0 \sqrt{V(r) - E}\, dr \tag{8.14}$$

Neem nu aan dat V een afstotende Coulomb-potentiaal is:

$$V = \frac{C}{r} \tag{8.15}$$

Het klassieke omkeerpunt in Eq.(8.8) is dan

$$a = \frac{C}{E} \tag{8.16}$$

Gebruiken we nu de variabele

$$y \equiv \frac{r}{a} \tag{8.17}$$

dan wordt de integraal in Eq.(8.14)

$$I = a\sqrt{E}\int_1^0 \sqrt{\frac{1}{y} - 1}\, dy = \frac{Q}{\sqrt{E}} \tag{8.18}$$

(merk op dat wij Eq.(8.16) hebben gebruikt) waarin Q een constante die onder andere afhangt van de massa van het deeltje, de sterkte van de afstoting, en wiskundige constanten. De waarschijnlijkheid dat een deeltje het punt $r = 0$ bereikt, is dan

$$\psi^*\psi \propto e^{-2Q/\sqrt{E}} \tag{8.19}$$ ♠

Nu zijn we tenslotte in staat om te zien hoe groot de kans is dat een deeltje in een thermisch gas de Coulomb-barrière doordringt. Enerzijds is er Eq.(8.19) die zegt dat die kans exponentieel *toeneemt* naarmate E groter is. Immers, bij hogere energie hoeft het deeltje maar een kleine weg af te leggen op klassiek verboden terrein. Anderzijds zegt Eq.(8.7) dat die kans exponentieel *afneemt* naarmate E groter is. Immers, de kans dat je in een thermisch gas bij N botsingen steeds maar energie wint, en dus een energie $N\,\Delta E$

bereikt, wordt steeds kleiner naarmate N toeneemt. Het ligt voor de hand te kijken of die twee tegengestelde tendensen ergens een gulden middenweg vinden. Daartoe is het nodig dat de exponent in de gezamenlijke waarschijnlijkheid

$$w \propto \mathrm{e}^{-E/kT-2Q/\sqrt{E}} \tag{8.20}$$

(ook wel de `Gamow-vergelijking` genoemd) een minimum bereikt. Differentiatie leert dat het minimum bestaat, en bereikt wordt bij

$$E = (QkT)^{2/3} \tag{8.21}$$

Door ook de tweede afgeleide uit te rekenen kan men eenvoudig nagaan dat dit maximum zeer scherp is. Daarom noemt men de waarde in Eq.(8.21) wel de `Gamow-piek` voor thermische tunnel-reacties. Een voorbeeld (Arnett, p.68) is de reactie waarbij een ^{12}C-kern een proton vangt en onder uitzending van een foton overgaat in een ^{13}N-kern:

$$E = 3.93\, T_6^{3/2} \quad \mathrm{keV} \tag{8.22}$$

waarin T_6 de temperatuur in miljoenen kelvin.

Met behulp van formules van het type Eq.(8.20,22) kunnen wij het hele thermonucleaire kookboek samenstellen. In de praktijk komen daar experimentele gegevens bij uit laboratoriumproeven. Bovendien zijn zeer veel belangrijke reacties niet te schrijven als Eq.(8.20) omdat de atoomkern ook *resonanties* kan vertonen, zodat het energiespectrum van de reacties niet langer een continu spectrum is maar ook discrete toestanden heeft, overeenkomend met spectraallijnen in optische spectra. Het berekenen van deze toestanden en hun overgangswaarschijnlijkheden is buitengewoon ingewikkeld.

9. Vrije weglengte

In de astrofysica komt het vaak voor dat we willen uitrekenen hoe een deeltje zich gedraagt dat zeer veel botsingen ondergaat. Eerst moeten wij daarvoor uitrekenen wat de kans is dat het deeltje met een ander botst. De meest gebruikte maat hiervoor is de gemiddelde vrije weglengte. Hiervoor leiden we een eenvoudige formule af, die zeer veel toepassingen heeft. Voorlopig passen wij de formule toe op fotonen in een ster. Later zullen wij ook zien wat de vrije weglengte is voor sterren in sterrenstelsels en voor sterrenstelsels in het Heelal.

Het is alles goed en wel om energie te maken in het binnenste van een ster, maar de straling moet zich ook nog een weg naar buiten banen. Bij het berekenen van de structuur van een ster betekent dat een forse extra complicatie: niet alleen moeten we het stralingsverlies meeberekenen (met de Stefan-Boltzmann regel), en moeten we de energie-opbrengst van kernfusie weten (uit een 'kookboek' voor kernreacties), we moeten ook nog rekening houden met het feit dat straling er niet zomaar uit kan (behalve de neutrino's die bij kernfusie vrijkomen, maar daar zien we hier even van af).

Om te weten hoe erg dit probleem is, gaan we na hoe ver een lichtdeeltje (foton) kan vliegen voordat het in botsing komt met een electrisch geladen deeltje (zoals een electron) in de ster. We bekijken daartoe een stukje Δx van het pad van een foton. In een cylindertje met lengte Δx en dwarsdoorsnede (oppervlak) S, waarin zich n deeltjes per kubieke meter bevinden, vindt het foton een aantal deeltjes op zijn weg gelijk aan

$$\Delta N = nS\,\Delta x \tag{9.1}$$

omdat het aantal simpelweg gelijk is aan de dichtheid maal het volume. Het vermogen van een deeltje om een foton te onderscheppen wordt uitgedrukt met Σ, de `werkzame doorsnede`, dat aangeeft hoeveel vierkante meter trefvlak het deeltje aan het foton presenteert. De kans op een treffer bij het doorlopen van Δx is dus

$$\frac{\text{totaal trefvlak}}{\text{doorsnede}} = \frac{\text{trefvlak}}{\text{doorsnede}} \times \Delta N \equiv \Delta p = \frac{\Sigma}{S} \Delta N = n\Sigma\,\Delta x \tag{9.2}$$

Hoe ver komt het foton door k maal een stukje Δx spitsroeden te lopen? Als het foton niet gevangen wordt, legt het een weg

$$\text{padlengte} = x = k\,\Delta x \tag{9.3}$$

af, maar zover zal het niet altijd komen. Het is het eenvoudigst om eerst te kijken wat de kans is dat het foton *niet* wordt onderschept. We hebben

$$\text{kans op vrije doortocht} = 1 - \text{kans op botsing} = 1 - n\Sigma\,\Delta x \tag{9.4}$$

Doen we dit k maal, dan is de kans p op vrije doortocht gelijk aan het product van alle afzonderlijke kansen, en dus

$$p = (1 - n\Sigma\,\Delta x)^k = \left(1 - \frac{1}{k} n\Sigma x\right)^k \tag{9.5}$$

In de limiet voor zeer veel stapjes ($k \to \infty$) wordt dit

$$p = \mathrm{e}^{-n\Sigma x} \tag{9.6}$$ ♠

De kans dat een foton vrijelijk kan passeren neemt dus exponentieel af met de afgelegde weg. De bijbehorende e-waarde is de `vrije weglengte`

$$\lambda_{\mathrm{mfp}} \equiv \frac{1}{n\Sigma} \tag{9.7}$$ ♠

('mfp' staat voor *mean free path*, gemiddelde vrije weglengte).

Vaak gebruikt men niet alleen de vrije weglengte, maar ook de `botsingstijd` t_c. Die vinden we als we nagaan hoe lang het duurt voordat een deeltje met snelheid v botst:

$$t_c = \frac{\lambda_{\mathrm{mfp}}}{v} \tag{9.8}$$ ♠

In het geval van fotonen is de snelheid gelijk aan de lichtsnelheid c, en dus

$$t_c = \frac{1}{n\Sigma c} \tag{9.9}$$ ♠

Laten we nu de waarden voor de Zon eens invullen. De gemiddelde dichtheid van de modale ster is ruwweg gelijk aan de dichtheid van oceaanwater op Aarde, een ton per kuub. Het zijn vooral de electronen die met de fotonen botsen. Per proton is er één electron in de ster; de massa van het proton is 1.67×10^{-27} kg, dus de electrondichtheid is om en nabij 10^{30} deeltjes per kuub. De botsingsdoorsnede van het electron is de Thomson-doorsnede, $\Sigma_T = 6.67 \times 10^{-29}$ m^2. Dus vinden we de schatting

$$\lambda_{\mathrm{mfp}} \approx \frac{1}{66.7} = 1.5 \quad \text{cm} \tag{9.10}$$

De vrije weglengte van een foton in een ster met een straal van zowat een miljoen kilometer is dus op z'n hoogst een paar centimeter! Blijkbaar is het voor een foton ontzettend moeilijk om de weg naar buiten te vinden.

De bijzonder kleine vrije weglengte van fotonen heeft nog een ander gevolg wanneer we opmerken dat de lichtkracht L zeer steil toeneemt met de massa M. Als M erg groot is, is L zo enorm dat de botsingen tussen fotonen en deeltjes een bijdrage gaan leveren aan de druk. In dat geval is P niet meer de gewone gasdruk die we uit de thermodynamica kennen, maar de `stralingskracht`

$$F_{\text{rad}} = \frac{L}{c} \tag{9.11}$$ ♠

waarin c de lichtsnelheid. In de vergelijking van de hydrostatica moeten we dan een extra drukgradiënt invoeren. De `stralingsdruk` P_r is de kracht per oppervlakte, dus

$$P_r = \frac{F_{\text{rad}}}{4\pi r^2} \tag{9.12}$$

Als de straling zich over een afstandje Δr naar buiten wurmt, is de verandering ΔP_r evenredig met het aantal vrije weglengtes λ_{mfp} in Δr, dus

$$\Delta P_r = \frac{L/c}{4\pi r^2} \frac{\Delta r}{\lambda_{\text{mfp}}} \tag{9.13}$$

ofwel, gebruik makend van Eq.(9.7) voor de vrije weglengte,

$$\Delta P_r = \frac{L}{c} \frac{\Sigma n}{4\pi r^2} \Delta r \tag{9.14}$$

en in de limiet voor infinitesimale stapjes vinden we de stralingsdrukgradiënt

$$\frac{dP_r}{dr} = \frac{L}{c} \frac{\Sigma n}{4\pi r^2} = \frac{L}{c} \frac{\Sigma \rho}{4\pi \mu r^2} \tag{9.15}$$

Omdat L zo sterk toeneemt met M, besluiten we dat in zeer massieve sterren de stralingsdruk de ster uiteen kan blazen. Daardoor is er een tamelijk *lage* bovengrens aan de massa van sterren: bij uitrekenen blijkt dat sterren met een massa van meer dan zo'n honderd zonsmassa's niet stabiel kunnen zijn.

Laten we eens schatten wat die bovengrens zou kunnen zijn. Als de stralingsdruk de structuur van de ster domineert, kunnen we Eq.(9.15) invullen in de vergelijking voor het hydrostatisch evenwicht en vinden

$$\frac{\Sigma}{4\pi\mu} \frac{L}{c} = GM(r) \tag{9.16}$$

Inplaats van deze lastige vergelijking op te lossen, schatten we brutaalweg dat $M(r) \approx M$ (de totale massa van de ster) en vinden: *de lichtkracht van een door stralingsdruk gedomineerde ster is evenredig met de massa van de ster*,

$$L = 4\pi \frac{c}{\Sigma} G\mu M = 6.30\ M \quad \text{W} \tag{9.17}$$

Exercise.

Reken met Eq.(9.17) uit wat de maximale lichtkracht is van de Zon. Wat is de conclusie uit het feit dat de waargenomen lichtkracht zoveel kleiner is dan de berekende waarde?

Vervolgens roepen we de massa-lichtkracht relatie te hulp. Als we even niet op de constanten letten, hebben we dan

$$L = \alpha M \quad \text{en} \quad L = \beta M^{10/3} \tag{9.18}$$

waaruit we besluiten dat de maximale massa van een ster gevonden kan worden uit

$$M_{\max} = \left(\frac{\alpha}{\beta}\right)^{3/7} \tag{9.19}$$

Vullen we de constanten in, dan komt er

$$M_{\max} = 3.36 \times 10^{32} \quad \text{kg} \ = 169 \quad M_\odot \tag{9.20}$$

Dit is een primitieve versie van de `Eddington limiet`. De bijbehorende lichtkracht is

$$L_{\max} = 2.1 \times 10^{33} \quad \text{W} \ = 5.5 \times 10^{6} \quad L_\odot \tag{9.21}$$

Een ster met 5 à 6 miljoen maal de lichtkracht van de Zon zou uiteraard bijzonder opvallend zijn! De meest massieve sterren ooit gevonden, hebben een massa ergens tussen de 100 en 150 zonsmassa's. Dat klopt dus wel met de hier gegeven schatting.

Wanneer je met behulp van de vrije weglengte wilt uitrekenen hoever je komt bij herhaalde verstrooiing, dan moet je er rekening mee houden dat bijna altijd de richting van het onderschepte deeltje tussen twee opeenvolgende botsingen drastisch verandert. Het deeltje voert een `dronkemanswandeling` uit (in het Engels: `random walk`). Dat is nogal wat anders dan bewegen in een rechte lijn. Als de vrije weglengte λ_{mfp} is, en de af te leggen weg is R, dan zou je verwachten dat het aantal stappen N waarin R wordt overbrugd gelijk is aan

$$N = \frac{R}{\lambda_{\text{mfp}}} \tag{9.22}$$

Maar zo is het niet, want in een meerdimensionale ruimte ligt het $i+1$-ste stapje $\vec{r}_{i+1}$ niet precies in het verlengde van $\vec{r}_i$. In het algemeen zijn twee opeenvolgende stappen niet gecorreleerd, en dus is de gemiddelde afstand (aangegeven met $<\ >$) die na N stappen is bereikt, nul:

$$< \vec{r} >= \sum_{i=1}^{N} \vec{r}_i = 0 \tag{9.23}$$

Met andere woorden, als je niet weet waar je heen wilt, kom je nergens. Maar dat wil niet zeggen dat een deeltje dat een dronkemanswandeling uitvoert, precies op $\vec{r} = 0$ blijft staan, want de verwachtingswaarde van de absolute waarde van $\vec{r}$ is *niet* gelijk aan nul. Immers,

$$< r^2 >= (\vec{r}_1 + \vec{r}_2 + \vec{r}_3 + \ldots)^2 = \sum_{i=1}^{N} r_i^2 + \sum_{i,j} \vec{r}_i \cdot \vec{r}_j \tag{9.24}$$

Omdat de stappen niet met elkaar gecorreleerd zijn, is de gemiddelde som over de producten $\vec{r}_i \cdot \vec{r}_j$ nul, en daarom is

$$< r^2 >= \sum_{i=1}^{N} r_i^2 \tag{9.25}$$

De gemiddelde waarde van $|\vec{r}_i|$ is de vrije weglengte λ_{mfp}, zodat

$$< r^2 >= N\,\lambda_{\text{mfp}}^2 \tag{9.26}$$

Hieruit concluderen wij, dat de gemiddelde afstand R vanaf het vertrekpunt die het deeltje in N stappen heeft bereikt gegeven wordt door

$$R = \lambda_{\text{mfp}} \sqrt{N} \tag{9.27}$$

Dat is nogal wat anders dan Eq.(9.22) ! Laten wij eens uitrekenen hoeveel tijd een foton nodig heeft om van het centrum van een ster tot aan het oppervlak te komen. Uit Eq.(9.7,9,27) zien we dat

$$t = N\,\frac{\lambda_{\text{mfp}}}{c} = \frac{R^2}{\lambda_{\text{mfp}}^2}\frac{\lambda_{\text{mfp}}}{c} = \frac{R^2}{c\lambda_{\text{mfp}}} = \frac{1}{c} R^2 n \Sigma \tag{9.28}$$

Vullen wij de gegevens in die we ook gebruikten in Eq.(9.10) dan blijkt dat voor een ster als de Zon, waar $R = 7 \times 10^8$ m,

$$t = 1.1 \times 10^{11} \quad \text{s} = 3500 \quad \text{yr} \tag{9.29}$$

Het duurt dus op zijn minst vele duizenden jaren voordat een foton zich naar buiten heeft geworsteld. Preciezere berekeningen laten zien dat de vrije weglengte in de kern van een ster nog veel kleiner is dan het gemiddelde Eq.(9.10) zodat de tijd in Eq.(9.29) uitkomt op miljoenen jaren. *Fotonen zitten gevangen* in een ster, en worden slechts sporadisch vrijgelaten. Dit heeft belangrijke gevolgen voor het gedrag van een ster. Zo zien wij bijvoorbeeld dat de tijdschaal in Eq.(9.29) veel korter is dan de dynamische tijdschaal $1/\sqrt{G\rho}$, waaruit wij concluderen dat bij dynamische veranderingen van een ster (zoals oscillaties) het thermische gedrag van de fotonen altijd meeberekend moet worden. Bij trillingen van de ster zitten de fotonen gevangen en bewegen met de materie van de ster mee.

Exercise.

Beredeneer dat dit inhoudt dat een ster maar langzaam van toestand kan veranderen. Maak hiertoe een schatting van de karakteristieke dynamische tijdschaal van een ster.

10. The Microwave Background

In the past, there was less space for a fixed amount of mass-energy. Thus, in the past the mass- and energy densities were higher. At some point in the past, the energy per particle exceeded the ionization energy of hydrogen. Before that time, the Universe was not transparent, due to Thomson scattering. Thus, we expect to see a cosmic photosphere when we look back far enough. The radiation from this photosphere is the CMBR.

If temperature is meaningful (i.e. near thermodynamic equilibrium), the early Universe should radiate like a black body. Uniform expansion maintains the spectrum (until T drops below the hydrogen recombination temperature). The Wien shift law says that $\lambda_{\rm max} \propto 1/T$, and since the wavelengths scale with a, we get

$$\frac{T_{\rm obs}}{T_{\rm em}} = \frac{a_0}{a_1} = \left(\frac{t_0}{t}\right)^{2/3} = \tau^{-2/3} = 1+z \tag{10.1}$$

The temperature of the matter in an ideal adiabatic gas is determined by $T \propto P/\rho$ and $P \propto \rho^\gamma$, so that $T \propto \rho^{\gamma-1}$. Mass conservation gives $\rho \propto a^{-3}$, so that for a monatomic gas ($\gamma = 5/3$)

$$T_b \propto a^{-2} = T_{b0}\left(\frac{t_0}{t}\right)^{4/3} = T_{b0}\,\tau^{-4/3} \tag{10.2}$$

Clearly, there must have been a time in the past when radiation dominated over matter. In that case, we can no longer use the above solution $a(t)$. But this equality occurs before the period of interest, namely when T has dropped to the value at which hydrogen recombines, so we won't derive the proper a yet.

The redshift z_1 at which recombination occurs can be estimated as follows. The distance $a_1 r_1$ to that epoch is roughly equal to the mean free path λ_{mfp} of photons from $t = t_0$ until the time t_1 of recombination. Let the density of scatterers with cross section σ be n. Then

$$\lambda_{mfp} = \frac{1}{\sigma n_0} \tag{10.3}$$

This value can be calculated as follows. The probability that an interception occurs is proportional to the density n and the product of the cross section σ and the path increment δx:

$$p = n\sigma\,\delta x \tag{10.4}$$

The probability that a photon is *not* intercepted (and consequently gets through) is then $1-p$. Thus, the compound probability that a photon travels a distance λ, chopped up into k segments, is

$$P = \lim_{\delta x \to 0} (1 - n\sigma\,\delta x)^k \tag{10.5}$$

Because the path length is cut into k pieces, we have $x = k\,\delta x$, and therefore

$$P = \lim_{k \to \infty} (1 - \frac{n\sigma x}{k})^k \tag{10.6}$$

This expression is an exponential, as can be seen from the series expansion

$$\begin{aligned}\left(1+\frac{x}{k}\right)^k &= 1 + k\frac{x}{k} + \frac{k(k-1)}{2!}\frac{x^2}{k^2} + \frac{k(k-1)(k-2)}{3!}\frac{x^3}{k^3} + \ldots \\ &= 1 + x + \frac{x^2}{2!}(1-1/k) + \frac{x^3}{3!}(1+\mathcal{O}(1/k)) + \ldots \end{aligned}\tag{10.7}$$

which, in the limit $k \to \infty$, converges to e^x, so that

$$P = \mathrm{e}^{-n\sigma x} \tag{10.8}$$

which implies that the e-folding mean free path is $1/n\sigma$.

When we put $\lambda_{mfp} = a_1 r_1$ and use the E-DeS expressions, the result is

$$\frac{1}{\sigma n_0} = \frac{2ca_1}{H_0}\left[1-(1+z)^{-1/2}\right] \tag{10.9}$$

Because in an E-DeS universe the density is equal to the critical value, we can immediately substitute

$$n_0 = \frac{3H_0^2}{8\pi G m} \tag{10.10}$$

In most of the history of the Universe, m is very nearly equal to the hydrogen mass m_H. Assembly of the above expressions produces

$$\frac{4\pi G m_H}{3\sigma c H_0} = (1+z)^{-1}\left[1-(1+z)^{-1/2}\right] \tag{10.11}$$

It will be found that $z \gg 1$, so that to good approximation

$$z_r = \frac{3\sigma c H_0}{4\pi G m_H} \approx 1350 \tag{10.12}$$

for the value $\sigma = 10^{-24}$ m^2. Thus we expect to see, along every line of sight, a cosmic photosphere with a temperature

$$T \approx T_{\mathrm{rec}}/(1+z_r) = 2.96 \quad \mathrm{K} \tag{10.13}$$

The current best COBE fit gives 2.737(4) K, one of the most convincing fossils of the early Universe. If $H_0 = 65$, then $t_0 = 9.36 \times 10^{17}$ s which corresponds to about 520,000 years.

11. Background Inhomogeneities

The microwave background is expected to show ripples, due to various effects: the Sachs-Wolfe effect, due to a combination of gravitational redshift and time dilatation; the Doppler effect, due to possible motions of the cosmic gas; and the Rees-Sciama effect, due to the propagation of light through a contracting potential well.

The CMBR is not homogeneous if there are ripples in density or velocity. The gravitational redshift causes

$$\left(\frac{\delta T}{T}\right)_{\rm redsh} = \frac{1}{c^2}\,\delta\Phi \tag{11.1}$$

while the gravitational time dilatation has the opposite effect: we see the cosmic photosphere at a time further back, namely

$$\frac{\delta t}{t} = \frac{1}{c^2}\,\delta\Phi \tag{11.2}$$

Since the cosmic scale factor is proportional to $t^{2/3}$ and the temperature is inversely proportional to the scale factor (Wien law: the wavelength scales with the scale factor a), one gets

$$\left(\frac{\delta T}{T}\right)_{\rm dilat} = -\frac{2}{3c^2}\,\delta\Phi \tag{11.3}$$

and overall we have a net positive effect of

$$\frac{\delta T}{T} = \frac{1}{3c^2}\,\delta\Phi \tag{11.4}$$

This is called the `Sachs-Wolfe effect`. The relation to density perturbations is found from Poisson's Equation (at such late times, and small amplitudes, relativistic treatment is not necessary):

$$\Delta\delta\Phi = 4\pi G\,\delta\rho \tag{11.5}$$

Fourier transformation allows us to write this equation as a function of the fluctuation wave number k:

$$-\left(\frac{k^2}{R^2}\right)\delta\Phi_k = 3H^2 q\,\delta\rho_k \tag{11.6}$$

which works out to

$$\delta T_k = -\frac{\Omega H_0^2(1+z_1)}{2c^2}\frac{\delta\rho_k}{k^2} \tag{11.7}$$

using Eq.(11.4) for $\delta\Phi$ and Eq.(4.16) for qH^2. The Doppler amplitude is simply

$$\frac{\delta T}{T} = \frac{1}{c}\vec{r}\cdot\delta\vec{v} \tag{11.8}$$

which can be related to the density spectral amplitude if one assumes that all velocities are initially zero and have grown due to the density amplitude only. In that case, one integration of Poisson's Equation produces a linear equation of motion which relates velocity to density, and we get

$$\delta T_k = -i\frac{H_0}{c}\frac{\delta\rho_k}{k}\vec{k}\sqrt{\Omega(1+z_1)} \tag{11.9}$$

The `Rees-Sciama effect` can be roughly estimated by noting that the photon temperature scales as the photon energy, which, in turn, scales as the gravitational potential of the potential well through which the photon drops:

$$\frac{\delta T}{T} \propto \frac{\delta E}{E} \propto \frac{\Phi}{c^2} \propto \frac{GM}{r}\frac{1}{c^2} \propto \frac{GM}{c^2}\frac{1}{r} \propto \frac{R_s}{r} \tag{11.10}$$

in which R_S is the Schwarzschild radius of the potential well. A more detailed calculation would start from the general relativistic equation of motion for a photon in a weak potential Φ. The metric is

$$ds^2 = (c^2 + 2\Phi)\,dt^2 - \left(1 + \frac{2\Phi}{c^2}\right)^{-1} dr^2 \tag{11.11}$$

so that for photons, which have $ds = 0$, we get

$$\left(1 + \frac{2\Phi}{c^2}\right) c\,dt = dr \tag{11.12}$$

Putting $r = ct + \lambda$, in which λ is a small deviation from the unimpeded propagation, we have

$$\frac{2}{c^2}\Phi(r,t)\,c\,dt = d\lambda \tag{11.13}$$

which must be integrated to estimate the effective propagation delay, and from this the net remaining redshift.

12. Motion Derived from Minimum Principles

All dynamics can be derived from minimum principles coupled to certain forms of symmetry. As a simple introduction, I show how fermat's Principle (path of least time) leads to Snell's Law, and this, in turn, gives the equation of motion through a medium with variable propagation speed. In mechanics, the corresponding formalism leads to the Euler-Lagrange Equation. In relativity, this equation is the basis for the Einslein Equation.

The basic theories we will use in our exposition of cosmology are all related to `minimum principles`. That is to say, the paths of particles in spacetime are found by considering some sort of stationarity property.

How a path can be related to a minimal-value problem is seen for example in `Fermat's Principle` and the consequent law of motion known as `Snell's Law`. Consider two media through which a signal propagates: with velocity v_1 in medium 1, and v_2 in medium 2. The times t_1 and t_2 along the two segments d_1 and d_2 of the path between two fixed points obey

$$t_1 = \frac{d_1}{v_1} = \frac{1}{v_1}\sqrt{x_1^2 + y_1^2} \tag{12.1}$$

$$t_2 = \frac{d_2}{v_2} = \frac{1}{v_2}\sqrt{(x_2 - x_1)^2 + y_2^2} \tag{12.2}$$

The total travel time is $t = t_1 + t_2$, which is minimal if its derivative under displacement of the midpoint coordinate $(x_1, 0)$ is zero (the end points $(0, y_1)$ and (x_2, y_2) are kept fixed):

$$\frac{\partial t}{\partial x_1} = \frac{2x_1}{2v_1\sqrt{x_1^2 + y_1^2}} - \frac{2(x_2 - x_1)}{2v_2\sqrt{(x_2 - x_1)^2 + y_2^2}} = 0 \tag{12.3}$$

from which it follows that

$$\frac{1}{v_1}\frac{x_1}{d_1} - \frac{1}{v_2}\frac{x_2 - x_1}{d_2} = 0 \tag{12.4}$$

or, in the more familiar formulation involving the incoming and outgoing angles ('incidence and refraction'),

$$\frac{1}{v_1}\sin\alpha_1 = \frac{1}{v_2}\sin\alpha_2 \tag{12.5}$$

$$\sin\alpha_2 = n\,\sin\alpha_1\ ; \qquad n \equiv \frac{v_2}{v_1} \tag{12.6}$$

Thus we see that a minimum requirement (move along the fastest path) prescribes the way the particle propagates (Snell's law).

In the above we have assumed that there are two media with constant propagation velocities, separated by a plane boundary. However, we may immediately generalize our equation of motion by noting that Eq.(12.5) amounts to a conservation law along the path:

$$\frac{1}{v}\sin\alpha = \text{constant} \tag{12.7}$$

Here α is the angle of the particle path with respect to the y-axis, and v the absolute value of the velocity. Along the path we may decompose the velocity as

$$\sin\alpha = av \tag{12.8}$$

$$\cos\alpha = \sqrt{1 - a^2v^2} \tag{12.9}$$

We use the fact that the tangent of the angle is the derivative of the path:

$$\tan\alpha = \frac{av}{\sqrt{1 - a^2v^2}} = \frac{dx}{dy} \tag{12.10}$$

or, rewriting in a slightly more familiar way,

$$\frac{dy}{dx} = \sqrt{\frac{1}{a^2v^2} - 1} \tag{12.11}$$

with arbitrary propagation speed $v = v(x, y)$. The constant a is determined by the initial conditions.

Exercise.

Find a relationship between this a and the initial direction of the path $y(x)$.

Exercise.

Consider the simple case

$$v = \beta y \tag{12.12}$$

and prove that the path is a sector of an ellipse. Figure out why the ellipse is such that the maximum radius of curvature is found in the region where the propagation speed is highest.

The above is a simple illustration of something quite general. In fact, all modern theories of motion relate the path to some sort of minimum principle. The general idea is a `principle of least action`. Let us see how we can retrieve classical mechanics from such a principle.

First the simplest case: a particle without any force acting on it, in one dimension. Suppose one wants to get from point A to point B and cover the distance $x_B - x_A$ in a fixed time $t_B - t_A$. The mean velocity of the trip is of course $v = (x_B - x_A)/(t_B - t_A)$. The actual velocity $v(t)$ between the end points at any time can be anything you like. We know from experience that a path without force is a straight line traveled with constant speed. So v cannot give us a good minimum principle, because it can be anything.

However, we can say: The path actually taken is the one where the mean of the *square* of the velocity is minimal. The trick here is that we're using the fact that a mean deviation from a square is always positive! So we get that the `action` S^*, defined by

$$S^* \equiv \int_{t_A}^{t_B} v^2 \, dt \tag{12.13}$$

must be minimal along the path.

We can see straight away that a deviation from the average speed $(x_B - x_A)/(t_B - t_A) =$ constant always makes the time integral over v^2 bigger, but in preparation for what follows let us prove it formally. Suppose that the fastest path through space-time between A and B is described by the function

$$x = x_0(t) \tag{12.14}$$

The corresponding speed is

$$v_0 = \frac{dx_0}{dt} \tag{12.15}$$

Now allow the particle to take a slightly different path

$$x = x_0(t) + \epsilon(t) \tag{12.16}$$

The function ϵ is free, as long as its absolute value is small, and as long as it is zero at the end points A and B, which are supposed to remain fixed in space and time for all paths we consider. The velocity along this modified path will be different too, and so

$$v^2 = v_0^2 + 2v_0\frac{d\epsilon}{dt} + \left(\frac{d\epsilon}{dt}\right)^2 \tag{12.17}$$

To a first approximation we may neglect the term that is quadratic in ϵ, so that the deviation of the action is

$$\delta S^* = \int_{t_A}^{t_B} 2v_0 \frac{d\epsilon}{dt}\, dt \tag{12.18}$$

Now beware – we cannot jump to the conclusion that v_0 is the constant average speed and take it outside the integral! For all we know, v_0 is some hideous function of time, so we need one more step, and write

$$\delta S^* = 2 \int_{t_A}^{t_B} \frac{dx_0}{dt} \frac{d\epsilon}{dt}\, dt \tag{12.19}$$

which we can integrate by parts:

$$\delta S^* = 2 \frac{dx_0}{dt}\, \epsilon \, \Big|_{t_A}^{t_B} - 2 \int_{t_A}^{t_B} \frac{d^2 x_0}{dt^2}\, \epsilon \, dt \tag{12.20}$$

Because we are evaluating the deviation from minimum S^* between fixed points in space-time, ϵ must vanish at the end points, so that

$$\delta S^* = -2 \int_{t_A}^{t_B} \frac{d^2 x_0}{dt^2}\, \epsilon \, dt \tag{12.21}$$

Finally, we use the fact that we had kept ϵ totally free between its end points. If S^* is to be minimal, so that $\delta S^* = 0$, we must then have

$$\frac{d^2 x_0}{dt^2} = 0 \tag{12.22}$$

In other words, if there is no force acting on the particle, *its velocity* v_0 *must be constant.*

Accordingly we see that our prescription of the action S^* reproduces the `law of inertia`: the velocity of a free particle is constant.

The above may seem like a lot of machinery, but its value becomes apparent when we consider the motion of a particle that is not free. The deviation from the force-free action is written as a function Φ, which we will suppose to depend on position only (dependence on time is possible but we'll wait with that). Moreover, due to the algebraic expansion of the square one gets a factor 2, which we absorb by means of the factor 1/2 before the v^2; with hindsight we will not use S^* and v^2 but S and $\frac{1}{2}mv^2$. Subtracting the 'deviant function' Φ we get

$$S \equiv \int_{t_A}^{t_B} \left(\frac{1}{2} m v^2 - \Phi \right) dt \tag{12.23}$$

We proceed exactly as above as far as the deviation from the fastest path is concerned; that is, we use Eq.(12.14) and Eq.(12.16) . This leads to an expansion like Eq.(12.17,18) for the velocity term. However, because $\Phi = \Phi(x)$, this function must be Taylor-expanded too:

$$\Phi(x_0 + \epsilon) = \Phi(x_0) + \epsilon \frac{d\Phi}{dx} \tag{12.24}$$

The first-order expansion, and the subsequent partial integration, then lead to

$$\delta S = - \int_{t_A}^{t_B} \left(m \frac{d^2 x_0}{dt^2} + \frac{d\Phi}{dx} \right) \epsilon \, dt \tag{12.25}$$

instead of Eq.(12.21) .

Exercise.

Verify this – it is important!

Because ϵ is arbitrary between its zero-point ends, we conclude that a non-free particle obeys the `equation of motion`

$$m\frac{d^2x_0}{dt^2} + \frac{d\Phi}{dx} = 0 \tag{12.26}$$

When we identify Φ as the potential energy, this is our old friend $F = ma$.

Our minimum principle can reproduce all of classical mechanics. But the fun is that it can do much more than that. In fact, *all* known dynamics can be formulated in such a way. The integrand we used in the expression for the action has the form "kinetic energy minus potential energy"; this beast is called a `Lagrangian` or, more properly – because it is to be integrated over time – a `Lagrangian density`, called $\mathcal{L}$. Its usefulness in formulating minimum-action principles makes it a formidable instrument.

For later use, let us formulate the above equations of motion by using the minimum principle for general $\mathcal{L}(\vec{x}, \vec{v}, t)$. The action is then

$$S = \int \mathcal{L}\, dt = \int \mathcal{L}(x_i, v_i, t)\, dt \tag{12.27}$$

in which I have written the position and velocity in component notation, which is easier in the derivations. Following the trail of Eq.(12.16-26) we find the variation along the path

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial x_i}\,\epsilon + \frac{\partial\mathcal{L}}{\partial v_i}\frac{d\epsilon}{dt} \tag{12.28}$$

For purposes of partial integration the second term is extracted from the total derivative

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial v_i}\,\epsilon\right) = \epsilon\,\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial v_i} + \frac{\partial\mathcal{L}}{\partial v_i}\frac{d\epsilon}{dt} \tag{12.29}$$

so that we obtain

$$\delta S = \frac{\partial\mathcal{L}}{\partial v_i}\,\epsilon\,\Big|_A^B - \int\left(\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial v_i} - \frac{\partial\mathcal{L}}{\partial x_i}\right)\epsilon\, dt = 0 \tag{12.30}$$

Because, as before, our minimum principle requires $\delta S = 0$, and ϵ is arbitrary between its zero end points, the equation of motion is the `Euler-Lagrange Equation`

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial v_i} - \frac{\partial\mathcal{L}}{\partial x_i} = 0 \tag{12.31}$$

How's this in relativity, then? Amazingly enough, it is much simpler than in classical mechanics, because it suffices to consider a free particle only! That is why people find general relativity so beautiful and simple, even though its machinery can be rather heavy. It all begins with the bizarre fact (which Nature has invented for us, but nobody yet understands) that there is in our world a velocity c that is invariant (it turns out to be equal to the speed of light). This invariance means that the interval

$$s^2 = c^2t^2 - x^2 - y^2 - z^2 \tag{12.32}$$

is constant. Note that in most texts on relativity one takes $c \equiv 1$, which of course is wise; but these are astronomy lectures so we'll let it stand. A system in which c is invariant is said to be `Lorentz symmetric`: it is invariant under Lorentz transformations.

The invariance of c also implies that this must be a maximum speed (from which, incidentally, it also follows that there can be *only one* such speed). But if everything is bound to propagate at most with speed c, we conclude that not all parts of the Universe can simultaneously react to something that happens somewhere. If we apply a Lorentz transformation L somewhere, how would the rest of the Universe know? Apparently, global symmetry such as Eq.(12.32) is not self-consistent. That is to say, we must restrict ourselves to the infinitesimal patch around the origin of our arbitrarily chosen standpoint:

$$ds^2 = c^2 dt^2 - dr^2 \tag{12.33}$$

Local Lorentz symmetry means that, wherever you are, you can always find coordinates such that the above holds ("freely falling coordinates"). Thus, your neighbours in spacetime will also be able to do this. However, it isn't guaranteed that you will agree with the neighbours that their coordinate system $\{x^\mu\}$ is the same as yours. The best you can hope for is a patch-up between the two of you, by means of a bilinear form in the infinitesimal coordinates of some common coordinate system $\{\xi^\mu\}$ which you've both agreed to use, and for which local Lorentz symmetry holds too:

$$ds^2 = g_{\mu\nu} d\xi^\mu d\xi^\nu \tag{12.34}$$

Because this expression is invariant, we can measure distances along space-time paths using the interval ds. This leads us to the idea that *the Lagrangian density for motion in locally Lorentz symmetric spacetime is simply a scalar multiple of* ds, or, more precisely, $\sqrt{ds^2}$. The corresponding action is then

$$S = \int \sqrt{ds^2} = \int \sqrt{\left(\frac{ds}{d\tau}\right)^2} d\tau = \int \left(g_{\mu\nu} \frac{d\xi^\mu}{d\tau} \frac{d\xi^\nu}{d\tau} \right)^{1/2} d\tau \tag{12.35}$$

with suitable parametrization variable τ along the space-time path (world-line). And indeed, we will see (in the section on geodesics) that the equation of motion gives precisely the correct prescription for motion in a space-time described by the metric tensor $g_{\mu\nu}$.

This idea can be extended to make $\mathcal{L}$ a function of $g_{\mu\nu}$ and its 4-coordinate derivative. The resulting action is called the `Einstein-Hilbert action`. Instead of using integration over a scalar, namely $\sqrt{ds^2}$, we use a four-volume in space-time. Such a volume is, of course, not Lorentz invariant, because of the Lorentz-FitzGerald contraction of the spatial lengths and the time dilatation of time intervals. Thus, dx^μ is not the object to use; we must multiply it with the Jacobian of the metric, to be prepared for all possible coordinate transformations.

Symbolically, then, the progression in the construction of a Lagrangian-based theory is

$$\mathcal{L}(x, v) \quad \Longrightarrow \quad \mathcal{L}(x_\mu, u_\mu) \quad \Longrightarrow \quad \mathcal{L}(g_{\mu\nu}, g_{\mu\nu,\kappa}) \tag{12.36}$$

in which the comma notation means partial derivation:

$$g_{\mu\nu,\kappa} \equiv \frac{\partial g_{\mu\nu}}{\partial x^\kappa} \tag{12.37}$$

The Euler-Lagrange equations corresponding to this sequence are

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial v} - \frac{\partial \mathcal{L}}{\partial x} = 0 \tag{12.38}$$

$$\frac{d}{ds}\frac{\partial \mathcal{L}}{\partial u^\mu} - \frac{\partial \mathcal{L}}{\partial x^\mu} = 0 \tag{12.39}$$

$$\frac{d}{dx^\kappa}\frac{\partial \mathcal{L}}{\partial g_{\mu\nu,\kappa}} - \frac{\partial \mathcal{L}}{\partial g_{\mu\nu}} = 0 \tag{12.40}$$

Notice that the choice of making $(g_{\mu\nu}, g_{\mu\nu,\kappa})$ dynamical variables introduces a new level of complexity because these variables have different tensorial orders! Accordingly we must integrate over all of space-time, and integration over the interval no longer suffices. This forces us to include changes of the integration volume element. We have

$$d^4x = dx^0 dx^1 dx^2 dx^3 \tag{12.41}$$

and transformation to another set of coordinates y_μ gives

$$dx^0 dx^1 dx^2 dx^3 = J\, dy^0 dy^1 dy^2 dy^3 \tag{12.42}$$

in which J is the Jacobian of the transformation $x_\mu \to y_\mu$. On the other hand we have

$$g_{\alpha\beta} = \frac{\partial y^\mu}{\partial x^\alpha} \frac{\partial y^\nu}{\partial x^\beta} g^*_{\mu\nu} \tag{12.43}$$

in which the asterisk indicates that this g is to be read as a function of the y-coordinates. This expression can be seen as the product of three tensors; contraction yields an expression for the Jacobian,

$$g = J^2 g^* \tag{12.44}$$

Because of the signature of the space-time interval one has $g < 0$, so that

$$J\sqrt{-g^*} = \sqrt{-g} \tag{12.45}$$

This teaches us that the integral of a scalar over a four-volume has the following transformation property (using that for scalar $\mathcal{L}$ one has $\mathcal{L} = \mathcal{L}^*$):

$$\int \mathcal{L}\sqrt{-g} d^4x = \int \mathcal{L} J \sqrt{-g^*}\, d^4x = \int \mathcal{L}\sqrt{-g^*}\, d^4y \tag{12.46}$$

whence we conclude that the action S, in the case that $\mathcal{L}$ has the dynamical variables $(g_{\mu\nu}, g_{\mu\nu,\kappa})$, requires the factor $\sqrt{-g}$:

$$S = \int \mathcal{L}\sqrt{-g}\, d^4x = \text{invariant} \tag{12.47}$$

Thus, instead of using a Lagrangian, as we did in the simpler cases, we use a `Lagrangian density` $\mathcal{L}\sqrt{-g}$.

In a local field theory, the field itself is a dynamical variable. the field is 'tangible', it is building material. The choice of dynamical variables in GRT means that space-time is explicitly regarded as real stuff, building material. According to the recipe sketched above, we pick the Ricci curvature R as Lagrangian and get the `Einstein-Hilbert action` for matter-free space-time

$$S = \frac{-1}{16\pi G} \int (R + 2\Lambda)\sqrt{-g}\, d^4x \tag{12.48}$$

including a possible cosmological constant Λ.

13. Symmetries and the Form of the Lagrangian

In the formulation of the minimum principle that leads to the equations of motion, the Lagrange density could, in principle, be anything. Its actual form is dictated by a symmetry.

One might wonder whether one must simply guess at the form of the Lagrangian, such as the choice v^2, or if there is a more basic and compelling way for finding a useful expression.

There may be many ways, but it seems for the moment that the most sweeping and general way for adapting our minimum principle to the facts of Nature is by way of a *symmetry*. That is to say, we note that "certain things remain the same when we change certain other things," and derive a Lagrangian which embodies that regularity with which Nature presents us.

The remarkable thing is, that certain symmetries are so powerful and so restrictive that they entirely determine the functional form of the Lagrangian, and therefore the equations of motion. The beginnings of this concept go back a long way. In one of Huygens' publications on the motion of bodies he shows a ball dropping from the top of the mast of a boat, straight down the mast. He then asks: how does someone standing on shore, seeing the boat glide by, see the path of that ball?

The very least we can say about our world is: **the Universe is made of particles, space and time**. Note that this implies that space and time must be seen on an equal footing with particles; that is to say, they are not some sort of invisible graph paper, but real stuff, the cement between the particles. As such, space and time may have a structure, just as matter does. Matter is an arrangement of particles; spacetime is an arrangement of points $(\vec{x}, t)$.

Having noted that Nature is built of particles, space and time, we must of course ask about their interrelationship: *where is what when?* We are thus interested in the positions of particles, and the change of position in the course of time. In other words, we try to describe the Lagrangian as a function of position $\vec{x}$, time t, and velocity $\vec{v} \equiv d\vec{x}/dt$ (we will see that the observed symmetries of Nature make it unnecessary to consider higher derivatives such as $d^2\vec{x}/dt^2$). The action is then

$$S = \int \mathcal{L}\, dt = \int \mathcal{L}(x_i, v_i, t)\, dt \tag{13.1}$$

in which I have written the position and velocity in component notation. The variation along the path is

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial x_i}\epsilon + \frac{\partial\mathcal{L}}{\partial v_i}\frac{d\epsilon}{dt} \tag{13.2}$$

Partial integration of the second term produces

$$\delta S = \frac{\partial\mathcal{L}}{\partial v_i}\epsilon\Big|_A^B - \int\left(\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial v_i} - \frac{\partial\mathcal{L}}{\partial x_i}\right)\epsilon\, dt = 0 \tag{13.3}$$

Because, as before, our minimum principle requires $\delta S = 0$, and ϵ is arbitrary between its zero end points, the equation of motion is

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial v_i} - \frac{\partial\mathcal{L}}{\partial x_i} = 0 \tag{13.4}$$ ♠

The choice of a minimum principle as the basis of a theory of dynamics immediately implies that the Lagrangian has certain properties. First, minimising a function means that we get the same equations of motion if we multiply that function with an arbitrary constant. Second, because it is an integral which we minimise, we get the same motion if we add to the integrand the total time derivative of an arbitrary function of space-time:

$$\mathcal{L}^* \equiv \mathcal{L} + \frac{df(\vec{x}, t)}{dt} \tag{13.5}$$

because then

$$S^* = \int \mathcal{L}^* dt = \int \mathcal{L}\, dt + \int \frac{df(\vec{x}, t)}{dt}\, dt = S + f_B - f_A \tag{13.6}$$

The additional terms are fixed, so that they have no influence on the equations of motion.

The above two properties of $\mathcal{L}$ are mathematical consequences of our decision to use a minimum principle. But further restrictions on $\mathcal{L}$ must be built in by using clues from Nature. These may be of any kind, but in practice the most powerful and general have proven to be *symmetries*.

To begin with, consider the Lagrangian $\mathcal{L}_0$ of a free particle. The first symmetry is that of the `isotropy of space`. That is to say, the motion of a free particle does not depend on the orientation of our coordinate system. We conclude immediately that this means that $\mathcal{L}_0$ cannot depend on the direction of $\vec{v}$, but must be a function

$$\mathcal{L}_0 = \mathcal{L}_0(v^2, \vec{x}, t) \tag{13.7}$$

The second symmetry is the `homogeneity of space and time`. That is to say, the motion of a free particle does not depend on the place from which we measure its progress, nor does it depend on the time at which we start our measurement. Accordingly, $\mathcal{L}_0$ cannot explicitly depend on $\vec{x}$ or t, so that

$$\mathcal{L}_0 = \mathcal{L}_0(v^2) \tag{13.8}$$

The forceful influence of these symmetries becomes apparent when we derive the equation of motion from this $\mathcal{L}_0$:

$$\frac{d}{dt}\frac{\partial \mathcal{L}_0}{\partial v_i} - \frac{\partial \mathcal{L}_0}{\partial x_i} = \frac{d}{dt}\frac{\partial \mathcal{L}_0}{\partial v_i} = 0 \tag{13.9}$$

from which we immediately conclude that *a free particle moves with constant speed.* Thus, we find that the Law of Inertia follows from our minimum principle if we take the two basic symmetries of spacetime into account.

In parentheses, I note that the discovery of symmetry involves a lot of thought. In our everyday environment particles do not behave like Kipling's cat, namely as if all places and all times are alike to them. The deduction that space-time is homogeneous and isotropic involves considerable idealization and perspicacity (which is of course why it took so long to get to that point!)

The above symmetries pertain to rotation and translation of our coordinates. But it so happens that there is yet another symmetry to be observed in Nature: the motion of a free particle does not depend on a steady translation of our coordinate system. That is to say, if we observe a free particle, at rest in a given coordinate system, from a vantage point that moves with velocity $\vec{w}$ with respect to this system, we observe that the particle is still free, i.e. it moves with constant velocity, but with velocity $-\vec{w}$.

This remarkable fact of Nature is called `Galilei-Huygens relativity`, or `Galilei-Huygens symmetry`. The homogeneity and isotropy of space-time have already served to derive the form in Eq.(13.8) for the Lagrangian of the free particle; adding G-H symmetry determines $\mathcal{L}_0$ completely. To demonstrate this, suppose that a particle moves with velocity $\vec{v}$, and that it is observed from a coordinate system moving with a velocity $\vec{w}$ that is small compared with $\vec{v}$. In that second system the Lagrangian is

$$\mathcal{L}_0^* = \mathcal{L}_0((\vec{v}+\vec{w})^2) \tag{13.10}$$

To first order in w this can be written as

$$\mathcal{L}_0^* \simeq \mathcal{L}_0(v^2) + 2\vec{v}\cdot\vec{w}\frac{\partial \mathcal{L}_0}{\partial v^2} \tag{13.11}$$

The additional term must lead to the same particle motion. We saw above that only the addition of a term that is a total time derivative leaves the motion unchanged. The extra term is proportional to $\vec{v}$, which is a total time derivative already; so we conclude that we must have

$$\frac{\partial \mathcal{L}_0}{\partial v^2} = \text{constant} \tag{13.12}$$

In other words, *the Lagrangian of a free particle is a scalar multiple of the square of the velocity.* This conclusion allows us to write down the Lagrangian after a Galilei-Huygens transformation with finite velocity $\vec{w}$:

$$\mathcal{L}_0^* = (\vec{v}+\vec{w})^2 = v^2 + 2\vec{v}\cdot\vec{w} + w^2 = v^2 + \frac{d}{dt}\left(2\vec{x}\cdot\vec{w} + t\,w^2\right) \tag{13.13}$$

The additional terms have been written explicitly as a total time derivative which, as we saw above, has no influence on the equations of motion.

This completes the derivation of $\mathcal{L}_0$. Apparently we cannot tolerate any more restrictions in the form of additional symmetries. In particular, Nature tells us that steady rotation does *not* leave the motion invariant. This has always been a source of wonder, from Newton with his rotating bucket of water to Mach with his 'motion with respect to the distant stars.'

Recalling the arguments given above, one may argue that there is one more freedom of the Lagrangian which we have not yet disposed of, namely the possibility of multiplying it with a scalar factor. Leaving aside trivialities, such as a global change of physical units, we see that the freedom is no longer absolute when one considers the motion of *two* free particles, each with its own Lagrangian $\mathcal{L}_1$ and $\mathcal{L}_2$. The simultaneous motion of the particles is found from

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \tag{13.14}$$

which implies that only the *ratio* of the pre-factors is free. We may use one particle in the whole Universe as a standard and scale the rest accordingly, but this amounts to a trivial choice of physical units. Therefore we must conclude that each particle has its own Lagrangian pre-factor. For reasons that become apparent when we consider interacting particles, this factor is written as $m/2$, so that the free Lagrangian finally takes the form

$$\mathcal{L}_0 = \frac{1}{2} m v^2 \tag{13.15}$$

It is clear that this pre-factor must not be negative, or else the Lagrange formalism wouldn't produce the required minimum in the action.

The motion of a non-free particle may be described in many ways, depending on the nature of the influence acting on it. An easily recognizable case is obtained when we try the simplest generalization of the Lagrangian, namely subtracting a function Φ from it:

$$\mathcal{L} = \frac{1}{2} m v^2 - \Phi(\vec{x}) \tag{13.16}$$ ♠

Substitution into the Lagrangian equation gives

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial v_i} - \frac{\partial \mathcal{L}}{\partial x_i} = \frac{d\, m\vec{v}}{dt} + \vec{\nabla}\Phi = 0 \tag{13.17}$$

which is the classical equation of motion if we identify Φ as an external potential (and which, incidentally, justifies *a posteriori* our pre-factor $\frac{1}{2}m$).

As a concluding remark I note that the above mechanism of using symmetries to nail down the functional form of $\mathcal{L}$ works wonderfully in other cases, too. In fact, all known forces of Nature can be derived from a symmetry principle which, remarkably enough, not only prescribes the Lagrangian of a free particle (like the way in which we obtained v^2 above), but *the interactions as well.*

14. Special Relativity

Bij nader inzien blijkt de Galilei-Huygens symmetrie niet exact te zijn. Daarvoor in de plaats komt Lorentz symmetrie, die de lichtsnelheid invariant laat. De wiskundige vorm van deze symmetrie wordt afgeleid naar analogie van de klassieke bewegingsvergelijkingen. Een paar eenvoudige toepassingen zijn: de tijddilatatie, het optellen van snelheden, en beweging met een constante versnelling.

Hierboven werden de bewegingsvergelijkingen van de klassieke mechanica afgeleid op grond van de Galilei-Huygens symmetrie. Maar bij nader inzien blijkt, dat onze natuur niet exact aan die symmetrie voldoet. Uit de experimenten van Michelson en Morley (1887-1888) volgt namelijk dat het licht *niet* gehoorzaamt aan die symmetrie. Als licht beweegt met een snelheid c, zouden we volgens Huygens verwachten dat die snelheid slechts relatief is; meebewegen met diezelfde snelheid (en dat kan, door een Galilei-Huygens-transformatie) zou dan een effectieve snelheid $c - c = 0$ opleveren. Maar het `Michelson-Morley experiment` (dat de afgelopen eeuw in vele varianten, met steeds grotere precisie is herhaald) toont aan dat *de lichtsnelheid niet afhangt van de bewegingstoestand van de bron of de ontvanger.* De lichtsnelheid is niet relatief, maar absoluut: de lichtsnelheid is `invariant`.

Dat is totaal strijdig met Galilei-Huygens symmetrie, en daarmee komt de klassieke mechanica op losse schroeven te staan. Het gaat er nu om, een mechanica te bedenken waarin c invariant is. Je zou zo denken dat een dergelijke constructie 'absoluutheidstheorie' of zoiets zou worden genoemd, maar de naam is 'relativiteitstheorie' geworden (Einstein, de bedenker van de theorie, vond die naam ook maar niks). Om in vogelvlucht [*3] te zien hoe deze werkt, gaan we eerst na hoe een G-H-transformatie wordt geformuleerd. Wij beperken ons even tot één ruimte-dimensie. Laat x de positie zijn van een deeltje op tijd t. Het deeltje wordt dus beschreven met de coördinaten (x, t) in een stelsel $\mathcal{K}$. Stel dat een waarnemer in een ander stelsel, $\mathcal{K}'$, met een snelheid w beweegt ten opzichte van $\mathcal{K}$. De coördinaten in $\mathcal{K}'$ zijn (x', t') en volgens Huygens geldt dat

$$x' = x + wt \tag{14.1}$$
$$t' = t \tag{14.2}$$

Een snelheid is $v = dx/dt$, en dus volgt hieruit dat

$$v' = \frac{dx'}{dt'} = \frac{dx'}{dt} = \frac{dx}{dt} + w = v + w \tag{14.3}$$

zoals verwacht. Het is duidelijk dat er, onder de transformatie Eq.(14.3) , nooit een invariante snelheid kan bestaan.

De vraag is nu: wat moet er in de plaats komen voor Eq.(14.1,2) zodat er wèl een invariante snelheid is? Dit kunnen wij beantwoorden door ons eerst te realiseren dat het bestaan van zo'n bijzondere snelheid betekent dat *ruimte en tijd iets met elkaar te maken hebben.* Immers, er geldt

$$\text{snelheid} = \frac{\text{ruimte}}{\text{tijd}} \tag{14.4}$$

dus meters per seconde, kilometers per uur, en dergelijke. Als een snelheid invariant is, dan moeten teller en noemer samenspannen om ervoor te zorgen dat het quotiënt steeds hetzelfde blijft. Ruimte en tijd hebben dus iets met elkaar te maken, sterker nog, ze kunnen met een en dezelfde maat worden gemeten, door als maat voor de afstand de reistijd van het licht te nemen (omdat c invariant is, is dat een prima manier). De afstand van de Aarde naar de Maan is 1.3 seconden, naar de Zon 8.3 minuten, naar de Andromeda Nevel 2 miljoen jaar.

Blijkbaar stelt de Natuur ons de eis: "Ontwerp een mechanica waarin c onveranderd blijft." De symmetrie die c invariant maakt heet `Lorentz symmetrie`. Hieronder zullen we zien op welke manier je zoiets aanpakt, als illustratie van het thema *een formule is er niet in de eerste plaats om in je rekenmachine te proppen, maar om naar je hand te zetten en te analyseren.*

Omdat tijd en ruimte in deze zin 'hetzelfde zijn', ligt het voor de hand om de transformatie Eq.(14.1,2) uit te breiden tot een symmetrische vorm:

$$x' = L_{xx}x + L_{xt}ct \tag{14.5}$$
$$ct' = L_{tx}x + L_{tt}ct \tag{14.6}$$

[*3] Uitgebreide uitleg is te vinden in E.F. Taylor & J.A. Wheeler, *Spacetime phycics,* Freeman, New York 1966; D. Bohm, *The special theory of relativity,* Routledge, London, 1996; A.P. French, *Special relativity,* Chapman & Hall, London, 1997.

Dit is eigenlijk de belangrijkste stap van de hele behandeling. Uit het experimentele feit dat c invariant is, concluderen wij dat ruimte en tijd met dezelfde maat te meten zijn, en staan wij toe dat een term evenredig met x verschijnt in de transformatieregel voor de tijd t. Dat betekent iets heel merkwaardigs: namelijk, dat tijd *relatief* kan zijn, omdat het nu niet langer gegarandeerd is dat $t' = t$. Gegeven de mogelijkheid dat x en t in een mengvorm optreden, is het verstandig om ze met dezelfde maat te meten. De beste manier daarvoor is zulke eenheden te kiezen dat $c \equiv 1$, maar dat is onder sterrenkundigen helaas niet gebruikelijk. Daarom schrijven we in Eq.(14.5,6) ct inplaats van t. Eq.(14.5,6) komen in de plaats van Eq.(14.1,2) .

Het gaat er uiteraard om, de matrix L te bepalen. Laten we eerst eens zien hoe zoiets gaat in een wat minder exotisch geval: draaiingen. Bij een rotatie van coördinaten (x, y) naar (x', y') hebben we

$$x' = R_{xx}x + R_{xy}y \tag{14.7}$$
$$y' = R_{yx}x + R_{yy}y \tag{14.8}$$

Omdat bij een draaiing alle lengtes hetzelfde blijven, hebben wij

$$x^2 + y^2 = x'^2 + y'^2 = r^2 \tag{14.9}$$

Merk op dat het opleggen van de invariantie Eq.(14.9) een voorschrift is voor het afstandsrecept in de ruimte (Pythagoras)! In een ander type ruimte zou dat wel eens helemaal anders kunnen zijn.

Exercise.

Een cirkel is de verzameling van alle punten met een vaste afstand tot een gegeven punt. Teken de cirkels die overeenkomen met het afstandsrecept Eq.(14.9) , met $r^2 = |x| + |y|$ en met $r^2 = x^2 - y^2$.

Passen we Eq.(14.9) toe op Eq.(14.7,8) dan krijgen wij de eisen

$$R_{xx}^2 + R_{yx}^2 = 1 \ ; \quad R_{xy}^2 + R_{yy}^2 = 1 \ ; \quad R_{xx}R_{xy} + R_{yx}R_{yy} = 0 \tag{14.10}$$

Hieraan wordt voldaan door Eq.(14.7,8) te schrijven als

$$x' = x \cos\phi - y \sin\phi \tag{14.11}$$
$$y' = x \sin\phi + y \cos\phi \tag{14.12}$$

waarin ϕ de draaiingshoek.

Op soortgelijke manier gaan we nu de coëfficiënten vinden in Eq.(14.5,6) . De eerste eis die we stellen is, dat de beweging van het licht in de coördinatenstelsels $\mathcal{K}$ en $\mathcal{K}'$ hetzelfde is:

$$x = \pm ct \tag{14.13}$$
$$x' = \pm ct' \tag{14.14}$$

Met behulp van Eq.(14.5,6) volgt hieruit, als we eerst het geval $x = +ct$ nemen, dat

$$ct' = (L_{xx} + L_{xt})\, ct \tag{14.15}$$
$$ct' = (L_{tx} + L_{tt})\, ct \tag{14.16}$$

en dus

$$L_{xx} + L_{xt} = L_{tx} + L_{tt} \tag{14.17}$$
$$L_{xx} - L_{xt} = -L_{tx} + L_{tt} \tag{14.18}$$

waarin de tweede regel op dezelfde manier is afgeleid als de eerste, maar dan voor licht dat de andere kant opgaat, $x = -ct$. *Merk op dat er van deze hele truc niets terecht zou komen* *als c niet invariant was*, want dan zouden we hebben $x' = \pm c't'$ en schoten we niets op. Uit Eq.(14.17,18) volgen de eerste twee voorwaarden waaraan de transformatiematrix L moet voldoen:

$$L_{xx} = L_{tt} \tag{14.19}$$

$$L_{xt} = L_{tx} \tag{14.20}$$

Hoe komen we nu aan verdere voorwaarden voor L? Het uiteindelijke doel is immers om alle vier de componenten van L dwingend voor te schrijven uit de eis dat c invariant blijft. Merk op dat het helemaal niet voor de hand ligt dat dat ook echt kan! We gaan net zo te werk als in Eq.(14.3) en schrijven

$$\frac{v'}{c} = \frac{dx'}{c\,dt'} = \frac{L_{xx}dx + L_{tx}c\,dt}{L_{tx}dx + L_{xx}c\,dt} = \frac{L_{xx}v + L_{tx}c}{L_{tx}v + L_{xx}c} \tag{14.21}$$

Vervolgens stellen we vast dat voor snelheden ver beneden die van het licht, de GH-symmetrie wel degelijk goed werkt. Dus eisen wij dat, in de limiet voor $c \to \infty$, Eq.(14.21) het resultaat geeft dat $v' = v$, en dus

$$\frac{v}{c} = \frac{v'}{c} \simeq \frac{L_{tx}c}{L_{xx}c} = \frac{L_{tx}}{L_{xx}} \tag{14.22}$$

Er blijft dus nog maar één grootheid te bepalen over, namelijk L_{xx}, en wij kunnen Eq.(14.5,6) schrijven als

$$x' = L_{xx}(x + vt) \tag{14.23}$$

$$ct' = L_{xx}(\frac{v}{c}x + ct) \tag{14.24}$$

Door Eq.(14.23,24) van elkaar af te trekken komt er

$$x' - ct' = L_{xx}(1 - \frac{v}{c})(x - ct) \tag{14.25}$$

Nu komt het slotstuk: in de transformatie die door L wordt beschreven, beweegt het coördinatenstelsel $\mathcal{K}'$ met een snelheid v ten opzichte van $\mathcal{K}$. Uiteraard moet L zo zijn, dat een terugtransformatie met een snelheid $-v$ weer de oorspronkelijke toestand oplevert. Dus hebben we dat naast Eq.(14.25) moet gelden

$$x - ct = L_{xx}(1 + \frac{v}{c})(x' - ct') \tag{14.26}$$

Door substitutie van Eq.(14.25) in Eq.(14.26) komt er

$$L_{xx}^2 \left(1 - \frac{v^2}{c^2}\right) = 1 \tag{14.27}$$

waarmee tenslotte de hele matrix L is vastgelegd:

$$L = \frac{1}{\sqrt{1 - (v/c)^2}} \begin{pmatrix} 1 & v/c \\ v/c & 1 \end{pmatrix} \tag{14.28} ♠$$

Dit is de matrix van de `Lorentztransformatie`, die in de plaats komt van de Galilei-Huygens transformatie Eq.(14.1,2) .

Deze transformatie kan uiteraard ook worden uitgeschreven in componenten. Het is gebruikelijk om dat wat af te korten, door gebruik te maken van de definities

$$\beta \equiv \frac{v}{c} \tag{14.29} ♠$$

$$\gamma \equiv \frac{1}{\sqrt{1 - \beta^2}} \tag{14.30} ♠$$

waarmee de Lorentztransformatie een plezierig symmetrische vorm aanneemt:

$$x' = \gamma(x + \beta ct) \tag{14.31} ♠$$

$$ct' = \gamma(\beta x + ct) \tag{14.32} ♠$$

Exercise.

Bewijs expliciet dat L in de limiet voor $c \to \infty$ inderdaad de Galilei-Huygens-transformatie oplevert. Doe dit door een Taylor-ontwikkeling toe te passen op de wortelvorm in Eq.(14.28) .

Exercise.

Bewijs met behulp van Eq.(14.31,32) dat de grootheid $s^2 = c^2t^2 - x^2$ invariant is onder Lorentztransformaties. Men noemt s het `interval`. De invariantie van het interval is analoog aan de invariantie van de straal r van de cirkel in Eq.(14.9) .

De Lorentztransformatie is de basis van de relativistische mechanica. Hier laten we daarvan maar een zeer klein stukje zien. Ten eerste een stelling over het optellen van snelheden in één dimensie. Laat een deeltje een snelheid u hebben in het stelsel $\mathcal{K}$ en laat $\mathcal{K}'$ met een snelheid v bewegen ten opzichte van $\mathcal{K}$. Met welke snelheid beweegt het deeltje gezien vanuit $\mathcal{K}'$? Passen wij Eq.(14.21) toe op u dan zien we (met behulp van Eq.(14.28)) dat

$$\frac{u'}{c} = \frac{dx + v\,dt}{(v/c)dx + c\,dt} = \frac{u + v}{uv/c + c} \tag{14.33}$$

en zodoende

$$u' = \frac{u + v}{1 + uv/c^2} \tag{14.34}$$ ♠

Exercise.

Bewijs uit Eq.(14.34) dat de lichtsnelheid de grootst mogelijke snelheid is, door in te vullen $v = c$.

Uit Eq.(14.34) kunnen wij ook nog een vergelijking afleiden voor relativistische beweging onder invloed van een constante kracht. Dit is een nuttige oefening, omdat het laat zien hoe je moet oppassen met uitspraken over snelheden in een relativistische omgeving. Evenals boven beperken wij ons hier tot één ruimte-dimensie.

Stel dat wij een ruimteschip waarnemen dat ten opzichte van ons coördinatenstelsel $\mathcal{K}$ een snelheid v heeft. Aan boord van het ruimteschip gebruikt men stelsel $\mathcal{K}'$, waarin $v' = 0$, want ten opzichte van zichzelf staat het schip stil. De stoker van het ruimteschip gooit er een schepje op, en bereikt daarmee een kleine versnelling $\delta v'$. Met 'klein' wordt uiteraard bedoeld $|\delta v'| \ll c$: in de klassieke mechanica heeft zo'n bewering geen zin, omdat er geen absolute snelheid bestaat en je dus ook niet van 'snel' of 'langzaam' kunt spreken! Gezien door ons verandert de snelheid van het schip van v naar $v + \delta v$. Door de transformatie $\mathcal{K}' \to \mathcal{K}$ in Eq.(14.34) vinden we het verband tussen $\delta v'$ en δv:

$$\delta v' = (v + \delta v)' = \frac{(v + \delta v) - v}{1 - v(v + \delta v)/c^2} \tag{14.35}$$

De veranderingen δ worden als zeer klein beschouwd. Wij maken gebruik van de benadering

$$\frac{1}{1 + \epsilon} \simeq 1 - \epsilon \quad ; \qquad \epsilon \ll 1 \tag{14.36}$$

en vinden dan voor Eq.(14.35)

$$\delta v' \simeq \frac{1}{1 - v^2/c^2}\,\delta v \tag{14.37}$$

Stel nu dat de stoker er per tijdseenheid $\delta t'$ een schepje op doet. Vanwege de Lorentztransformatie weten wij dat, gezien vanaf een vast ruimtelijk punt, de tijd transformeert volgens de formule van de `tijddilatatie`

$$t' = t\,\sqrt{1 - (v/c)^2} \tag{14.38}$$

en zodoende is

$$\frac{dv'}{dt'} = \left(1 - \frac{v^2}{c^2}\right)^{-3/2} \frac{dv}{dt} \tag{14.39}$$

Wanneer wij dv'/dt' invullen als een constante a, krijgen we de bewegingsvergelijking

$$\frac{dv}{dt} = a\left(1 - \frac{v^2}{c^2}\right)^{3/2} \tag{14.40}$$ ♠

Die komt dus in de plaats voor de klassieke vorm $F = ma$.

Het treft dat Eq.(14.40) exact opgelost kan worden. Om de berekeningen gemakkelijker te maken, drukken wij alle snelheden uit in eenheden van c door de transformatie

$$\beta \equiv \frac{v}{c} \tag{14.41}$$

De oplossing van Eq.(14.40) is dan

$$\frac{at}{c} \equiv \tau = \int \left(1 - \beta^2\right)^{-3/2} d\beta = \frac{\beta}{\sqrt{1 - \beta^2}} \tag{14.42}$$

Hierin is ook ook de tijd dimensieloos gemaakt. Door inversie van Eq.(14.42) wordt de oplossing voor β

$$\beta = \frac{\tau}{\sqrt{1 + \tau^2}} \tag{14.43}$$

We zien nu direct dat in het begin $\tau \approx 0$, als het ruimteschip nog maar net begint met versnellen,

$$\beta = \tau \; ; \qquad \text{en dus} \qquad v = at \tag{14.44}$$

hetgeen precies overeenkomt met het klassieke geval. Als $\tau \to \infty$ daarentegen, hebben we

$$\beta = 1 \; ; \qquad \text{en dus} \qquad v = c \tag{14.45}$$

Ook een oneindig lang volgehouden versnelling brengt ons niet boven de lichtsnelheid uit!

Tenslote berekenen wij nog welke afstand het ruimteschip aflegt. Omdat $v = dx/dt$ hebben we uit Eq.(14.43)

$$\frac{a}{c^2}\frac{dx}{d\tau} = \frac{\tau}{\sqrt{1 + \tau^2}} \tag{14.46}$$

met als oplossing

$$\frac{a}{c^2}x = \int \frac{\tau}{\sqrt{1 + \tau^2}}\, d\tau = \int \frac{1}{2\sqrt{1 + \tau^2}}\, d\tau^2 = \sqrt{1 + \tau^2} - 1 \tag{14.47}$$

waarin de integratieconstante zo gekozen is dat het ruimteschip vertrekt van $x = 0$ op $\tau = 0$.

Exercise.

Bereken twee limietgevallen voor Eq.(14.47) , namelijk $\tau \ll 1$ en $\tau \gg 1$. Laat zien dat het eerste geval overeenkomt met het klassiek-mechanische $x = \frac{1}{2}at^2$ en het tweede geval met het extreem-relativistische $x = ct$.

Het leven aan boord gaat gewoon door, en we zien pas merkwaardigheden wanneer wij bezien hoe de tijd bij ons verloopt vergeleken met die welke wij (niet zij!) aflezen op de klok van het schip. Wegens Eq.(14.38) is

$$\tau' = \tau\sqrt{1 - \beta^2} \tag{14.48}$$

en dus, gebruik makend van Eq.(14.42) ,

$$d\tau' = \frac{1}{\sqrt{1+\tau^2}}\, d\tau \tag{14.49}$$

De oplossing hiervan is

$$\tau' = \int \frac{1}{\sqrt{1+\tau^2}}\, d\tau = \log\left(\tau + \sqrt{1+\tau^2}\right) \tag{14.50}$$

Opnieuw bekijken we de limietgevallen. Als de versnelling begint, $\tau \simeq 0$, is

$$\tau' \simeq \log(\tau + 1) \simeq \tau \tag{14.51}$$

hetgeen weer precies is wat we klassiek verwachten. Als $\tau \to \infty$ daarentegen, komt er

$$\tau' \simeq \log(\tau + \sqrt{\tau^2}) = \log(2\tau) \tag{14.52}$$

Dus: ten gevolge van de tijddilatatie *zien wij de tijd aan boord van het ruimteschip veel langzamer lopen dan bij ons!*

Exercise.

Reken de versnelling a uit, in de veronderstelling dat de tijdseenheid overeenkomend met $\tau = 1$ precies 1 jaar is. Hoe groot is deze a in vergelijking met de versnelling g van de zwaartekracht aan het oppervlak van de Aarde? Zou het comfortabel zijn aan boord van zo'n ruimteschip?

Exercise.

De afstand van de Zon tot het centrum van de Melkweg is ongeveer 28000 jaar. Als een ruimteschip die afstand overbrugt met de versnelling a uit de vorige som, hoeveel ouder zijn de bemanningsleden dan volgens de boordklok wanneer zij, naar onze tijdrekening, 28000 jaar gereisd hebben? Doe dezelfde berekening voor een sterrenstelsel op een afstand van 10 miljard jaar. Wat zegt dit over de bereikbaarheid (in principe!) van verre plaatsen in ons Heelal?

Tot besluit nog een stuk over de energie van relativistische bewegingen. De arbeid die een kracht F verricht is gelijk aan de kracht vermenigvuldigd met de weglengte. Een klein stukje dx van de weg komt dus overeen met een verandering van de energie E:

$$dE = F\, dx \tag{14.53}$$

In bovenstaande vergelijkingen Eq.(14.43,46) zagen wij, dat

$$dx = \frac{c^2}{a}\frac{\tau\, d\tau}{\sqrt{1+\tau^2}} = \frac{c^2}{a}\beta\, d\tau \tag{14.54}$$

Merk op dat dit niet alleen voor het raket-voorbeeld geldt, maar *algemeen* is, omdat we slechts infinitesimale veranderingen bezien. Gebruiken we Eq.(14.53) in Eq.(14.54) dan komt er

$$dE = mc^2 \beta\, d\tau \tag{14.55}$$

Met behulp van Eq.(14.42) is dit te schrijven als

$$dE = mc^2 \frac{\beta\, d\beta}{(1-\beta^2)^{3/2}} \tag{14.56}$$

Enig geploeter met integratie laat zien dat dan

$$dE = mc^2\, d\frac{1}{\sqrt{1-\beta^2}} \tag{14.57}$$

Omdat dit voor elk infinitesimaal stukje van de baan zo is, hoeven we niets te veronderstellen over F, en dus geldt algemeen dat

$$E = \frac{mc^2}{\sqrt{1-\beta^2}} \tag{14.58}$$ ♠

Dit is Einsteins beroemde massa-energieformule. Dus niet $E = mc^2$, zoals amateurs zeggen! De formule Eq.(14.58) is veel algemener dan dat. In het bijzondere geval van de beweging onder invloed van een constante kracht F (constant zoals gezien door de raket) hebben we uit Eq.(14.43,47,58)

$$E = mc^2\sqrt{1+\tau^2} = mc^2 + Fx \tag{14.59}$$

Hoe realistisch is het idee van zo'n voertuig, waarmee wij in principe overal in het Heelal kunnen komen binnen een redelijke eigen tijd? Hoe je zoiets moet bouwen is onbekend, maar in principe is er geen bezwaar tegen. Een mogelijke tegenwerping is, dat de 'brandstof' voor zo'n raket van buiten moet komen. Maar wij weten dat de ruimte tussen de sterren niet leeg is, en de vraag is dus: kan een ruimtevoertuig voldoende materie verzamelen om zichzelf daarmee te versnellen? We bekijken eerst het geval dat de raket reeds relativistisch beweegt, dus $v \approx c$. Dan hebben we wegens Eq.(14.42,59)

$$E = mc^2\tau = amct = amx \tag{14.60}$$

Als het ruimteschip een doorsnede D heeft, kan het over een afstand x een hoeveelheid massa M opscheppen die wordt gegeven door

$$M = \rho D x \tag{14.61}$$

waarin ρ de massadichtheid van het medium waar de raket doorheen beweegt. Hieruit volgt een schatting voor D. Als $\tau = 1$ overeen komt met een jaar, dan is wegens Eq.(14.42)

$$\frac{c}{a}\tau = \frac{c}{a} = 1 \text{ jaar} = 3.156 \times 10^7 \text{ s} \tag{14.62}$$

en dus is de versnelling van zo'n raket

$$a = 9.51 \text{ m s}^{-2} \tag{14.63}$$

hetgeen prettig dicht bij de aardse waarde 9.8 ligt, zodat wij ons aan boord zeer goed zouden voelen (wegens het equivalentie-principe is er geen verschil tussen een versnelling en de zwaartekracht). De energie die we uit de opgeveegde materie kunnen halen is Mc^2, en dus vinden we uit Eq.(14.60,61) dat

$$amx = \rho D x c^2 \quad \to \quad D = \frac{am}{\rho c^2} \tag{14.64}$$

Nu vullen we een paar schattingen in. Laat de massa m van de raket een miljoen ton zijn (wie weet?) en laat de gewenste versnelling gegeven zijn door Eq.(14.63) . De minimale dichtheid in het Heelal is de gemiddelde massadichtheid tussen de sterrenstelsels. Deze is ongeveer een waterstofatoom per kuub, ofwel

$$\rho = 10^{-26} \text{ kg m}^{-3} \tag{14.65}$$

Zodoende wordt

$$D = \frac{9.51 \times 10^9}{10^{-26} c^2} = 1.06 \times 10^{19} \text{ m}^2 \tag{14.66}$$

De doorsnede van de raket moet dus ongeveer de wortel hieruit zijn, en dat is ruim 3 miljoen kilometer. De baan van de Maan heeft een straal van 384,000 km, de baanstraal van Mercurius is 57.9 miljoen kilometer. Dus de waarde in Eq.(14.66) is niet absurd groot, hoewel duidelijk niet binnen ons technisch bereik. De toestand wordt iets beter als we in plaats van Eq.(14.65) de gemiddelde dichtheid van de interstellaire materie nemen. Deze is ongeveer een waterstofatoom per kubieke centimeter, dus een miljoen maal groter dan Eq.(14.65) zodat D overeenkomstig kleiner wordt:

$$D = 1.06 \times 10^{13} \text{ m}^2 \tag{14.67}$$

hetgeen betekent dat een raket met een 'schep' van ruim 3000 km doorsnede volstaat. Dat is kleiner dan de straal van Aarde (6371 km). Zoals Ya.B. Zel'dovich bij zulke gelegenheden zei: *It is possible, but it is difficult.*

Een iets nauwkeuriger beschouwing leert, dat de schatting Eq.(14.64) een ondergrens is. Immers, wanneer wij Eq.(14.59) in het algemeen gebruiken (dus niet alleen in het geval $v \approx c$) vinden we in plaats van Eq.(14.64)

$$D = \frac{ma}{\rho c^2} + \frac{m}{\rho x} \tag{14.68}$$

Aan het begin van de reis is $x = 0$, dus om te starten hebben we een veel groter schep-oppervlak nodig. De raket moet dus op de een of andere manier gelanceerd worden. Ook dat hoeft geen probleem te zijn. In de buurt van een ster is ρ vele malen groter dan in de interstellaire ruimte; voor een sterrewind met constante snelheid is $\rho \propto r^{-2}$, waarin r de afstand tot de ster. Samen met Eq.(14.68) zien we daaruit dat het lanceren van zo'n denkbeeldige raket vanuit een planetenstelsel in principe haalbaar is.

Exercise.

Laat met behulp van Eq.(14.62-64) zien dat een raket van het beschreven type zijn eigen massa opveegt in precies 1 jaar, dus over een afstand van 1 lichtjaar.

De afstand tot het centrum van de Melkweg is 28000 jaar, dus om die reis te maken moet de raket evenzoveel maal zijn eigen massa opvegen. Ter vergelijking: een auto van 500 kg die 'een-op-tien' loopt, kan 5000 km afleggen bij het verstoken van zijn eigen massa aan brandstof. Rijdt zo'n auto 28000 maal zijn eigen massa op, dan is de afgelegde afstand 140 miljoen kilometer. Dat is bijna precies de afstand tussen de Aarde en de Zon. Met een snelheid van 200 km/h zou de auto hier 80 jaar over doen. De relativistische raket bereikt het centrum van de Melkweg in iets meer dan bijna 11 jaar (zie Eq.(14.50) voor $\tau = 28000$). Het is aardig dat autorijden binnen ons Zonnestelsel in deze opzichten vergelijkbaar is met relativistisch door de Melkweg rossen.

15. Algemene relativiteit

Omdat de lichtsnelheid eindig en maximaal is, kan een globale symmetrie niet bestaan. Elke symmetrie moet lokaal zijn. Eerst berekenen we het interval, een grootheid die onveranderd blijft bij een globale Lorentztransformatie. We gaan over naar lokale symmetrie door te werken met differenties, analoog aan de werkwijze bij de klassieke mechanica. Lokale Lorentz-symmetrie leidt tot een structuur van tijd-ruimte die wordt beschreven met een soort afstandsrecept. De numerieke factoren in dat recept zijn samengevat in de metrische tensor.

In het vorige hoofdstuk zagen wij, dat de ruimte-tijd (x, t) in een coördinatenstelsel $\mathcal{K}$ wordt omgevormd in de (x', t') in een coördinatenstelsel $\mathcal{K}'$ dat met snelheid v ten opzichte van $\mathcal{K}$ beweegt, volgens de Lorentztransformatie. Deze wordt beschreven door de matrix

$$L = \gamma \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \tag{15.1}$$ ♠

Er is echter een beperking die de toepassing van L wat kunstmatig maakt: de transformatie $\mathcal{K} \to (L) \to \mathcal{K}'$ geldt alleen voor *constante* snelheid $v = \beta c$.

Hoe moeten we nu te werk gaan als v verandert in plaats en in tijd? Op soortgelijke manier als hierboven: wij stellen eerst vast dat L zo gek nog niet is. Wij veronderstellen dat de natuur weliswaar *globaal* niet Lorentz-invariant is, maar dat *locaal* wel de vorm L mag worden gebruikt, zij het dan dat β van plaats tot plaats en van tijd tot tijd verschilt. We kunnen dat in ons formalisme inbouwen door vast te stellen dat een baan $\vec{x}$ die in ruimte-tijd gekromd is, kan worden opgebouwd uit infinitesimale stukjes die elk voor zich een rechte lijn willekeurig dicht benaderen. Door niet globaal te kijken maar locaal, beschouwen wij niet (x, t) maar de differenties (dx, dt). Dus inplaats van

$$x' = \gamma(x + \beta ct) \tag{15.2}$$

$$ct' = \gamma(\beta x + ct) \tag{15.3}$$

krijgen we

$$dx' = \gamma(dx + \beta c\, dt) \tag{15.4}$$

$$c\, dt' = \gamma(\beta dx + c\, dt) \tag{15.5}$$

Om te zien welke, gebruiken wij de fundamentele symmetrie-eigenschap van L. Overal zagen we, dat een symmetrie leidt tot een invariantie, een behouden grootheid. In het geval van de speciale relativiteitstheorie zien we uit Eq.(15.2,3) dat

$$\begin{aligned} x'^2 - c^2t'^2 &= \gamma^2(x^2 + 2\beta xct + \beta^2c^2t^2 - \beta^2x^2 - 2\beta xct - c^2t^2) \\ &= \gamma^2(1 - \beta^2)(x^2 - c^2t^2) \\ &= x^2 - c^2t^2 \end{aligned} \tag{15.6}$$

Dus de uitdrukking $x^2 - c^2t^2$ heeft een zeer bijzondere eigenschap: onder Lorentztransformaties is

$$x^2 - c^2t^2 = \text{invariant} \tag{15.7}$$

Dit is de grootheid die behouden is dankzij de Lorentz-symmetrie. Uiteraard verdient deze een eigen naam; we noemen s het `interval`, gedefinieerd door

$$x^2 - c^2t^2 \equiv -c^2s^2 \tag{15.8}$$

De invariantie van het interval onder Lorentztransformaties doet zeer sterk denken aan invariantie onder draaiingen. Een draaiing over een hoek ϕ in een (x, y)-vlak wordt beschreven door de rotatiematrix

$$R = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{15.9}$$

zodat de coördinaten transformeren volgens

$$x' = x\cos\phi - y\sin\phi \tag{15.10}$$

$$y' = x\sin\phi + y\cos\phi \tag{15.11}$$

Hieruit zien we meteen dat

$$x'^2 + y'^2 = x^2 + y^2 \equiv r^2 \tag{15.12}$$

Zoals de voerstraal r invariant is onder draaiingen, zo is het interval s invariant onder Lorentztransformaties. Vergelijken we Eq.(15.8) en Eq.(15.12) , dan zien we deze gelijkenis direct. Alleen dat minteken in Eq.(15.8) is vreemd! Bij verder onderzoek blijkt dat deze malle '−' verantwoordelijk is voor bijna alle tegen-intuïtieve eigenschappen van de relativiteit. Merk nog op dat *een lichtstraal heeft interval nul:* voor het licht geldt $s = 0$. Volgens de meetkunde van ruimte en tijd is de 'afstand' tussen het punt waarop het licht wordt uitgezonden en het punt waarop het aankomt, nul. Kijken we naar Eq.(15.3) dan zien wij dat langs een lichtstraal ook geldt $t' = 0$. Ruwweg kunnen wij dus zeggen: *voor een lichtstraal vallen het moment van uitzending en absorptie samen.*

Wij kunnen de invariantie van s gebruiken door te stellen: waar bij constante v een globale Lorentz-symmetrie geldt, daar geldt een *locale* Lorentz-symmetrie in het algemene geval. Dus schrijven we Eq.(15.8) , in navolging van Eq.(15.4,5) , als

$$(ds)^2 = c^2(dt)^2 - (dx)^2 \tag{15.13}$$

Dit wordt meestal geschreven in de vorm

$$ds^2 = c^2\,dt^2 - dx^2 \tag{15.14}$$

Merk op dat dit een wat onzuivere notatie is, want met ds^2 bedoelen we $(ds)^2$ en *niet* $d(s^2)$. Het probleem van een locale Lorentz-symmetrie is nu, dat hierdoor L een functie van de ruimte-tijdcoördinaten is geworden:

$$L = L(x, t) \tag{15.15}$$

Zodoende kan Eq.(15.14) in die vorm maar op één plaats en tijd in het Heelal worden gebruikt, en dat is een beetje weinig. Wij moeten dus een algemenere versie van deze vergelijking vinden. Net als bij de afleiding van de Lorentztransformatie doen we dat, door een algemene bilineaire vorm te proberen:

$$ds^2 = g_{tt}\,dt^2 + g_{tx}\,dt\,dx + g_{xt}\,dx\,dt + g_{xx}\,dx^2 \tag{15.16}$$

Als (x, t) gewone reële getallen zijn, is $dx\,dt = dt\,dx$. Dus moet ook gelden

$$g_{tx} = g_{xt} \tag{15.17}$$

De matrix g is symmetrisch.

Hoe moeten wij nu Eq.(15.16) interpreteren? We kunnen een analogie maken met de vergelijkingen voor krommen in Eq.(15.12,14) : evenals (x, y) de coördinaten zijn van een cirkel met straal r, en evenals (x, ct) de coördinaten zijn van een hyperbool met pericentrum s, zo zijn (dx, dt) *de infinitesimale lijnstukjes die gezamenlijk een kromme bepalen die wordt beschreven door de matrix* g. Dat is nogal dramatisch: blijkbaar is het zo, dat het toepassen van een *locale* Lorentz-transformatie leidt tot een tijd-ruimtestructuur die wordt beschreven met g. De meetkundige eigenschappen van die algemene (dx, dt)-ruimte worden bepaald door g. Deze grootheid heeft men dan ook de `metrische tensor` genoemd. De uitdrukking Eq.(15.16) is een *afstandsrecept*, dat aangeeft hoever het is 'van A naar B' in een ruimte met een algemene structuur. De afstand wordt gemeten met behulp van het interval s.

De theorie welke g en zijn dynamica beschrijft heet de `algemene relativiteitstheorie`. In werkelijkheid bestaat g uit 16 getallen, in een blok van 4×4, maar omdat g symmetrisch is zijn er slechts 10 onafhankelijke componenten.

De structuur van tijd-ruimte die door g wordt beschreven, kunnen we samenvatten door een *kromming*, evenals Eq.(15.12) de kromming van een cirkelboog beschrijft en Eq.(15.14) de kromming van een hyperbooltak. Enigszins kort door de bocht kunnen we stellen: *een gekromde ruimte geeft gekromde banen.* Op deze manier is de theorie van g bruikbaar voor de beschrijving van de zwaartekracht, waarbij de kromming van banen niet wordt toegeschreven aan de werking van een 'kracht-op-afstand', maar aan de locale structuur van tijd en ruimte. Daarvoor is nog wel nodig dat we de rol van de materie erin betrekken, maar voor dit college gaat dat veel te ver. We moeten ons beperken tot de opmerking dat er talloze `metrieken` zijn van het type Eq.(15.16) , waaronder zeer belangrijke en interessante, zoals de `Friedmann-Robertson-Walker metriek` die het Heelal beschrijft, en de `Schwarzschild metriek` van zwarte gaten.

16. From Global to Local Lorentz Symmetry

Global Lorentz symmetry is inconsistent with relativity, because nothing can propagate faster than the speed of light. Thus, all symmetries must be local, but that implies that they can depend on space-time coordinates. A Lagrangian that is locally symmetric doesn't lead to proper equations of motion, unless it is patched up somehow. Local relativity is patched up by the introduction of space-time curvature.

Special relativity starts from the invariance of the speed of light. The equation of motion for light is the equation for a sphere in 3-space, where the radius is given by the travel distance ct:

$$s^2 = c^2t^2 - r^2 \tag{16.1}$$

where $s = 0$ for a light ray. The general case where the interval obeys $s \neq 0$ must hold for consistency. Now we note that the above leads to a maximum value $v \leq c$ for all speeds. That means that the presumed global Lorentz symmetry cannot be maintained. To stay consistent, we must require *local* symmetry only. That is to say, we must restrict ourselves to the infinitesimal patch around the origin of our arbitrarily chosen standpoint:

$$ds^2 = c^2dt^2 - dr^2 \tag{16.2}$$

Local Lorentz symmetry means that, wherever you are, you can always find coordinates such that the above holds ("freely falling coordinates"). Thus, your neighbours in space-time will also be able to do this. However, it isn't guaranteed that you will agree with them that their coordinate system $\{x^\mu\}$ is the same as yours. The best you can hope for is a patch-up between the two of you, by defining a kind of "euro-coordinates" which are arbitrary but serve as a common coordinate currency. That is to say, you express the *local* infinitesimal increase of the interval by a bilinear form in the coordinates of some common coordinate system $\{\xi^\mu\}$ which all participants have agreed to use, and for which local Lorentz symmetry holds too:

$$ds^2 = g_{\mu\nu}d\xi^\mu d\xi^\nu \tag{16.3}$$

This sequence of arguments is like classical mechanics. First, consider free motion only, the kinematics of motion without force (homogeneity and isotropy of space and time, and Galilei-Huygens symmetry). Then attribute deviations from the results (constant energy and momentum: rest is equivalent to steady motion) to an external force. This was basically the method that Stevin and Huygens used in their quest for the equations of motion. In relativity: from Lorentz symmetry to local Lorentz symmetry (=structure of space-time given by $g_{\mu\nu}$). Thus the equivalent of motion under the influence of a force is force-free motion in curved space-time.

Formally, this goes as follows. The space-time coordinates x_μ are functions of the coordinates $\{\xi^\alpha\}$. The infinitesimal interval ds^2 can be written as

$$ds^2 = d\xi^\alpha d\xi_\alpha = \frac{\partial \xi^\alpha}{\partial x^\mu}dx^\mu \frac{\partial \xi_\alpha}{\partial x^\nu}dx^\nu \tag{16.4}$$

Now if we define the `four-velocity` u^μ as

$$u^\mu \equiv \frac{dx^\mu}{ds} \tag{16.5}$$

we find that

$$u^\mu u^\nu g_{\mu\nu} = 1 \tag{16.6}$$

Let us proceed to the algebraical equivalent of the statement "curved space-time gives curved paths". A classical free particle moves according to the law of inertia: $dv^i/dt =$

$d^2x^i/dt^2 = 0$. If an external force is present, say due to a gravitational potential Φ, we have $d^2x^i/dt^2 = -\partial\Phi/\partial x_i$. In GRT the analogue of classical free motion must be written as $du^\mu/ds = 0$. Note that we have changed from using the time derivative to a derivation with respect to the interval s. In classical mechanics, time is absolute; in relativity, it is the interval we must use. If we arrange with 'the neighbours' elsewhere in the Universe to refer all our descriptions to a global but otherwise arbitrary coordinate system $\{\xi^\mu\}$ (much as at a conference one usually agrees to speak English), we find

$$u^\mu = \frac{dx^\mu}{ds} = \frac{\partial x^\mu}{\partial \xi^\alpha}\frac{d\xi^\alpha}{d\tau} = v^\alpha \frac{\partial x^\mu}{\partial \xi^\alpha} \tag{16.7}$$

Here v^α is the *local* four-velocity, the "free-fall motion".

Accordingly, the differential operator d/ds can be written as

$$\frac{d}{ds} = u^\alpha \frac{\partial}{\partial x^\alpha} \tag{16.8}$$

This produces an expression for the way in which the four-velocity changes with respect to interval:

$$0 = \frac{du^\alpha}{ds} = \frac{d^2\xi^\alpha}{ds^2} = \frac{d}{ds}\left(\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{dx^\mu}{ds}\right) \tag{16.9}$$

and, after working out the differentiations,

$$\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{d^2x^\mu}{ds^2} + \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0 \tag{16.10}$$

From this last expression one may obtain a 'clean' form for d^2x^μ/ds^2 by using the metric tensor:

$$g_{\mu\nu} \equiv \frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial \xi^\beta}{\partial x^\nu}\eta_{\alpha\beta} \tag{16.11}$$

The point is that we can multiply the equation of motion with $\partial\xi^\beta/\partial x^\nu$ and then contract the resulting expression with $g_{\alpha\beta}$. Using the fact that g is unimodular, we get

$$\frac{d^2x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0 \tag{16.12}$$

$$\Gamma^\lambda_{\mu\nu} \equiv \frac{\partial x^\lambda}{\partial \xi^\alpha}\frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \tag{16.13}$$

The above equation of motion can be interpreted as "curved space-time gives curved orbits": the four-acceleration is no longer zero but proportional to Γ. Note also the proportionality to two factors of the four-velocity dx^ν/ds. This makes it very hard to understand the effects of the 'curvature-generated force' intuitively; it is even worse than the Lorentz force in electrodynamics, which also depends on $\vec{v}$ in a peculiar way.

In the factor Γ we have dumped all the garbage due to the 'mismatch' in space-time caused by introduction of local symmetry; we made the mess invisible by choosing a new clean symbol for it. Note that it is not a proper tensor, but transforms in a complicated fashion. Thus, although we can always *locally* transform away the curvature of space-time ("freely falling coordinates"), we cannot do so globally. The above equation of motion is sometimes written yet more compactly as

$$\frac{Du^\lambda}{Ds} = 0 \tag{16.14}$$

where the differential operator D/Ds is called `covariant derivative`. Note the similarity with the classical free particle, $dv^i/dt = d^2x^i/dt^2 = 0$. The particle is still free, but we use curvilinear coordinates. This behaviour is typical for local gauge theories: we can write the equation of motion in covariant form, such that the force disappears as an external entity. The coupling is completely prescribed by the gauge symmetry.

The final thing that needs settling is: what connects matter and space-time curvature? Or, what connects matter and the distance recipe in space-time?

When we compare the Newtonian and the Einsteinian equations for the trajectory of a particle, we notice immediately that Φ and $\Gamma^{\lambda}_{\mu\nu}$ are apparently related. Of course it cannot be that $\Gamma = \Phi$ or anything as simple as that, because then Newtonian theory would be relativistic already! We must incorporate a physical ingredient in our theory. The one that Einstein used is: in the limit for small velocities and small curvatures in a static space-time we must recover the classical equations.

In that case, the only components that remain are those related to the time-indices: everything is zero except Γ^{μ}_{00}. Using the above calculations one may see that this is closely related to the fact that the limit of the four-velocity u^{μ} for small three-velocities is $u^{\mu} = (1, 0, 0, 0)$. The zero-component does not vanish, just like the 0-component of the energy-momentum does not vanish for $v \to 0$ (this is the famous mc^2-term).

Furthermore, from the definition of the metric we know that

$$ds^2 = g_{00} dt^2 \tag{16.15}$$

(the other components of $g_{\mu\nu}$ vanish in the limit for small curvature), and therefore

$$\frac{dx^0}{ds} = \frac{1}{\sqrt{g_{00}}} \tag{16.16}$$

When we write $x^0 \equiv t$, as is customary, we get

$$\frac{d^2 x^{\mu}}{ds^2} + \Gamma^{\mu}_{00} \left(\frac{dt}{ds} \right)^2 = \frac{d^2 x^{\mu}}{ds^2} + \frac{1}{g_{00}} \Gamma^{\mu}_{00} = 0 \tag{16.17}$$

The expression for derivation with respect to the interval then becomes

$$\frac{d}{ds} = \frac{dt}{ds} \frac{d}{dt} = \frac{1}{\sqrt{g_{00}}} \frac{d}{dt} \tag{16.18}$$

and so

$$\frac{d^2 x^{\mu}}{dt^2} + \Gamma^{\mu}_{00} = 0 \tag{16.19}$$

from which we finally conclude that the desired correspondence with classical gravitation can be obtained if we relate Γ to the gradient of the gravitational potential:

$$\Gamma^{\mu}_{00} \Longleftrightarrow \frac{\partial \Phi}{\partial x_{\mu}} \tag{16.20}$$

The `Christoffel symbol` Γ is related to the metric tensor g. We can calculate how by starting from the expression

$$g_{\mu\nu} \equiv \frac{\partial \xi^{\alpha}}{\partial x^{\mu}} \frac{\partial \xi^{\beta}}{\partial x^{\nu}} \eta_{\alpha\beta} \tag{16.21}$$

Looking at the definition of Γ suggests that it should be related to derivatives of g. Differentiation with respect to the four-coordinate x^{λ} produces

$$\frac{\partial g_{\mu\nu}}{\partial x^{\lambda}} = \frac{\partial^2 \xi^{\alpha}}{\partial x^{\lambda} \partial x^{\mu}} \frac{\partial \xi^{\beta}}{\partial x^{\nu}} \eta_{\alpha\beta} + \frac{\partial \xi^{\alpha}}{\partial x^{\mu}} \frac{\partial^2 \xi^{\beta}}{\partial x^{\lambda} \partial x^{\nu}} \eta_{\alpha\beta} \tag{16.22}$$

The definition of Γ can be rewritten as

$$\Gamma^{\lambda}_{\mu\nu} \frac{\partial \xi^{\alpha}}{\partial x^{\lambda}} = \frac{\partial^2 \xi^{\alpha}}{\partial x^{\mu} \partial x^{\nu}} \tag{16.23}$$

so that the derivative of g becomes

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \Gamma^\kappa_{\lambda\mu}\frac{\partial\xi^\alpha}{\partial x^\kappa}\frac{\partial\xi^\beta}{\partial x^\nu}\,\eta_{\alpha\beta} + \Gamma^\kappa_{\lambda\nu}\frac{\partial\xi^\alpha}{\partial x^\mu}\frac{\partial\xi^\beta}{\partial x^\kappa}\,\eta_{\alpha\beta} \tag{16.24}$$

Using the definition of g then yields

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \Gamma^\kappa_{\lambda\mu} g_{\kappa\nu} + \Gamma^\kappa_{\lambda\nu} g_{\kappa\mu} \tag{16.25}$$

Clearly, the expression for Γ can be found from a linear combination of first coordinate derivatives of g. It is straightforward to show that

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\nu} = 2 g_{\kappa\nu}\Gamma^\kappa_{\lambda\mu} \tag{16.26}$$

(notice the minus sign, which rather spoils the transformation properties of Γ). Therefore, using the fact that

$$g^{\alpha\beta} g_{\mu\alpha} = \delta^\beta_\mu \tag{16.27}$$

we finally get

$$\Gamma^\kappa_{\lambda\mu} = \frac{1}{2} g^{\nu\kappa}\left\{\frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\nu}\right\} \tag{16.28}$$

Now we had seen that, in order to connect with the classical Newtonian case, we felt obliged to equate Γ with a coordinate derivative of the classical potential. Because we have now seen how Γ is, in its turn, related to a linear combination of first derivatives of g, we finally conclude that *the role which the potential Φ plays in classical Newtonian gravity is taken over by the metric tensor $g_{\mu\nu}$ in general relativity.*

The classical potential is related to the presence of matter by means of the `Poisson equation`

$$\Delta\Phi = 4\pi G\rho \tag{16.29}$$

which should give us a hint about how the presence of matter can be connected with the structure of space-time. The point here is that the density field ρ could not possibly be used, because it is in no way Lorentz invariant. In fact, we can see immediately that a Lorentz transformation of ρ should go as

$$\rho' = \gamma^2\rho \tag{16.30}$$

One Lorentz factor γ comes from the change of the effective mass via $E = \gamma mc^2$; the other one comes from the fact that a volume in motion decreases by one factor γ because of its Lorentz-FitzGerald contraction. Accordingly we suspect that ρ should be part of a tensor because a Lorentz scalar transforms with γ^0, a vector with γ^1 and a tensor with γ^2. Since ρ is classically a scalar field we similarly suspect that it is the 00-component of a tensor.

In the limit of small velocities without external forces we may take that tensor, which we'll call $T_{\mu\nu}$, to be diagonal. Then ρ is placed in the top-left corner. What will we have on the remaining three places of the diagonal? Since ρ is a mass density, and since in relativity we have to take mass and energy as equivalent, it seems natural to use an energy density. The mass density is the mass of a collection of particles per unit volume. The corresponding energy per unit volume we know as the *pressure* P of the collection of particles (or P/c^2 if we do not choose $c \equiv 1$). Thus we put

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -P/c^2 & 0 & 0 \\ 0 & 0 & -P/c^2 & 0 \\ 0 & 0 & 0 & -P/c^2 \end{pmatrix} \tag{16.31}$$

where we have taken account of the $+\;-\;-\;-$ signature.

The entries elsewhere in the tensor can be found by Lorentz transformation of T. If the Lorentz matrix in a single coordinate direction is

$$L = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16.32}$$

then we get

$$T' = LTL^{-1} = \begin{pmatrix} \gamma^2(\rho + \beta^2 P/c^2) & -\beta\gamma^2(\rho + P/c^2) & 0 & 0 \\ -\beta\gamma^2(\rho + P/c^2) & \gamma^2(\beta^2\rho + P/c^2) & 0 & 0 \\ 0 & 0 & -P/c^2 & 0 \\ 0 & 0 & 0 & -P/c^2 \end{pmatrix} \tag{16.33}$$

Thus we see that, apart from the usual factor γ^2 which we always get when transforming a tensor, the first row and the first column contain a vector that corresponds to $(\rho, \rho\vec{v})$ in the classical case. The entity $\rho\vec{v}$ is the momentum density in the flow, the systematic gas motion of the particles. Accordingly we suspect that the complete form of T is the *energy-momentum tensor*

$$T_{\mu\nu} = \frac{P}{c^2}\,\eta_{\mu\nu} + \left(\rho + \frac{P}{c^2}\right) u_\mu u_\nu \tag{16.34}$$

(or use P if $c = 1$).

The final order of business now is to construct a tensor $G_{\mu\nu}$ from $g_{\mu\nu}$ and its derivatives that has the same transformation properties as $T_{\mu\nu}$. The most obvious choice, taking T simply proportional to g, is not sufficient because it would not include Newtonian gravity; in order to obtain that, as we had seen above, we must include the derivatives of g. In particular, in the Newtonian limit we retain the 00-components only, namely

$$\frac{\partial^2 g_{00}}{\partial x^\alpha \partial x_\alpha} \propto T_{00} \tag{16.35}$$

Apparently, the desired tensor must contain at least second derivatives of g. This implies that *a fourth-rank tensor must be involved!* Einstein showed that the correct expression is related to the monstrous `Riemann-Christoffel curvature tensor`

$$R^\lambda_{\mu\nu\kappa} \equiv \frac{\partial \Gamma^\lambda_{\mu\nu}}{\partial x^\kappa} - \frac{\partial \Gamma^\lambda_{\mu\kappa}}{\partial x^\nu} + \Gamma^\alpha_{\mu\nu}\Gamma^\lambda_{\kappa\alpha} - \Gamma^\beta_{\mu\kappa}\Gamma^\lambda_{\nu\beta} \tag{16.36}$$

This animal must be reduced to second rank before we can equate it to $T_{\mu\nu}$. This is done by contracting it over one index. Then we obtain the `Ricci tensor`

$$R_{\mu\nu} \equiv R^\lambda_{\mu\lambda\nu} \tag{16.37}$$

and its contracted form, the `curvature scalar`

$$R \equiv g^{\alpha\beta} R_{\alpha\beta} \tag{16.38}$$

The most general expression for the required tensor is then

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} \tag{16.39}$$

and in units where the gravitational constant G is retained explicitly, correspondence with the Poisson equation shows that

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{16.40}$$

This is a physical choice; it does not have the mathematical necessity of a gauge theory. It is Einstein's guess, based on correspondence with Newtonian mechanics, and it works very well on large scales: black holes, relativistic stars, the Universe. But it works not at all on small, atomic, quantum-mechanical scales. The hassle is that we are obliged to include the vacuum zero-point energy in $T_{\mu\nu}$. This is infinite or at least very large. The zero-point energy shows up in the form of a finite value of the cosmological constant Λ, which one really should like to be zero in order to conform to cosmological observations. Current field theories require $\Lambda \approx 10^{120}$ or thereabouts, which is totally excluded by cosmological observations of the expansion of the Universe.

Possibly there is another guess one could make by searching for an expression that corresponds with (say) Schrödinger's Equation instead of Poisson's Equation, but nobody has yet succeeded in doing so.

Note that the rather peculiar and counter-intuitive behaviour of Λ, and the related possibility of inflation ("you get something out of nothing") is due to our initial assumption for the connection between $T_{\mu\nu}$ and the (thermo)dynamic quantities. As I emphasized there, the connection between matter and space-time curvature is still a conjecture, since we do not have a quantum gravity theory.

So, maybe all this will have to be drastically revised in the future.

If ρ_P is the `Planck density`

$$\rho_P = M_P L_P^{-3} = \sqrt{\frac{\hbar c}{2G}} \left(\frac{2\hbar G}{c^3}\right)^{-3/2} = \frac{c^5}{4\hbar G^2} \tag{16.41}$$

we expect

$$\Lambda = 8\pi G \rho_P \tag{16.42}$$

Numerically,

$$\rho_P = 1.29 \times 10^{96} \quad \text{kg m}^{-3} \tag{16.43}$$

while the critical density is

$$\rho_0 = \frac{3H_0^2}{8\pi G} = 1.06 \times 10^{-26} \quad \text{kg m}^{-3} \tag{16.44}$$

If Λ were to be so large that it would be easily observable, we ought to have $\rho_P \approx \rho_0$. Accordingly, the implied value of Λ is some 10^{132} times too large.

One immediate cause for worry is that it seems like 'double dipping' to introduce space-time in $g_{\mu\nu}$ as well as in Λ. Ought we not to include the Planck-scale fluctuations in some (possibly extended) form of $g_{\mu\nu}$ rather than in the potential term that produces Λ? After all, a potential is a classical beast that would be wiped out by second quantization.

But if gravity does not couple to vacuum fluctuations, how come it couples to anything at all?

17. Geodesics

Curved space gives curved orbits. These are found as the paths of least distance (more correctly stationary distance) in curved space-time. These paths are called geodesics.

The equation of motion through a curved spacetime we have derived above is

$$\frac{d^2x^\lambda}{ds^2} + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0 \tag{17.1}$$

or, in terms of the four-velocities,

$$\frac{du^\lambda}{ds} + \Gamma^\lambda_{\mu\nu} u^\mu u^\nu = 0 \tag{17.2}$$

It turns out that this is a very special equation: namely, it describes the shortest path in a given metric, called a `geodesic`.

In order to see this, consider a path P^α in spacetime stretching between the events a and b. We may map this curve to the real line by a parametrization $P^\alpha(\tau)$. An infinitesimal piece dP^α always points along the path; thus, the tangent w^α to the path is given by

$$w^\alpha \equiv \frac{dP^\alpha}{d\tau} \tag{17.3}$$

We will construct the path in such a way that its length is minimal. In spacetime, invariant lengths are measured through the interval, and we know that an infinitesimal piece of the path has a length ds which obeys

$$ds^2 = g_{\mu\nu} dP^\mu dP^\nu \tag{17.4}$$

The length S of the path between a and b is therefore

$$S = \int_a^b \sqrt{g_{\mu\nu} dP^\mu dP^\nu} = \int_a^b \left(g_{\mu\nu} \frac{dP^\mu}{d\tau} \frac{dP^\nu}{d\tau} \right)^{1/2} d\tau \tag{17.5}$$

Introducing the `Lagrangian` $\mathcal{L}$ by

$$\mathcal{L}(x^\lambda, u^\lambda) \equiv g_{\mu\nu} \frac{dP^\mu}{d\tau} \frac{dP^\nu}{d\tau} \tag{17.6}$$

this length can be written as

$$S = \int_a^b \sqrt{\mathcal{L}}\, d\tau \tag{17.7}$$

Suppose, now, that we pick a path slightly different from P^α, but with the same beginning a and end b. This new path may be parametrized as

$$P^\alpha(\tau) + \epsilon p^\alpha(\tau) \tag{17.8}$$

with small ϵ and $p^\alpha(a) = p^\alpha(b) = 0$. Inserting this new path in Eq.(17.7) for the path length, and requiring the derivative of S with respect to ϵ to be zero, we get in the limit $\epsilon \to 0$

$$0 = \frac{dS}{d\epsilon} = \int_a^b \frac{1}{2\sqrt{\mathcal{L}}} \left(\frac{\partial \mathcal{L}}{\partial \dot{P}^\mu} \dot{p}^\mu + \frac{\partial \mathcal{L}}{\partial P^\mu} p^\mu \right) d\tau \tag{17.9}$$

in which we have used the shorthand

$$\dot{p} \equiv \frac{dp}{d\tau} \tag{17.10}$$

The first term in brackets can be reduced to a multiple of p^μ through integration by parts, using the fact that p vanishes at the end points:

$$\int_a^b \frac{1}{2\sqrt{\mathcal{L}}} \frac{\partial \mathcal{L}}{\partial \dot{P}^\mu} \dot{p}^\mu d\tau = - \int_a^b \frac{d}{d\tau} \left(\frac{1}{2\sqrt{\mathcal{L}}} \frac{\partial \mathcal{L}}{\partial \dot{P}^\mu} \right) p^\mu d\tau \tag{17.11}$$

The integrand in Eq.(17.9) can therefore be written as

$$\left[\frac{d}{d\tau} \left(\frac{1}{2\sqrt{\mathcal{L}}} \frac{\partial \mathcal{L}}{\partial \dot{P}^\mu} \right) - \frac{1}{2\sqrt{\mathcal{L}}} \frac{\partial \mathcal{L}}{\partial P^\mu} \right] p^\mu \tag{17.12}$$

Because this must hold for all deviations p^μ, we conclude that the path is minimal if

$$\frac{d}{d\tau} \left(\frac{1}{\sqrt{\mathcal{L}}} \frac{\partial \mathcal{L}}{\partial \dot{P}^\mu} \right) - \frac{1}{\sqrt{\mathcal{L}}} \frac{\partial \mathcal{L}}{\partial P^\mu} = 0 \tag{17.13}$$

Such a minimal path is called a `geodesic`.

The most useful form of this equation is obtained if we choose the parametrization with τ of the path in a special way, namely by taking the interval s as parameter. In that case the Lagrangian becomes

$$\mathcal{L}(x^\lambda, u^\lambda) = g_{\mu\nu}\frac{dP^\mu}{ds}\frac{dP^\nu}{ds} = g_{\mu\nu}u^\mu u^\nu = 1 \tag{17.14}$$

along this particular path. Consequently,

$$\frac{d}{ds}\left(\frac{\partial\mathcal{L}}{\partial u^\mu}\right) - \frac{\partial\mathcal{L}}{\partial x^\mu} = 0 \tag{17.15}$$

This general form allows us to easily find conserved quantities if $\mathcal{L}$ is symmetric in some way. For example, if $\mathcal{L}$ is independent of some coordinate x^ν, it is obvious that there is a corresponding conservation law for a 'generalized momentum'

$$\frac{\partial\mathcal{L}}{\partial u^\nu} = \text{constant} \tag{17.16}$$

The Lagrangian is equal to 1 only for one particular path in all of spacetime; a geodesic elsewhere also has a constant $\mathcal{L}$ but with a different value. In general we have

$$\mathcal{L}(x^\lambda, u^\lambda) = g_{\mu\nu}u^\mu u^\nu \tag{17.17}$$

From this we can calculate the derivatives in Eq.(17.15) as

$$\frac{\partial\mathcal{L}}{\partial u^\mu} = 2g_{\mu\nu}u^\nu \tag{17.18}$$

$$\frac{\partial\mathcal{L}}{\partial x^\mu} = \frac{\partial g_{\kappa\nu}}{\partial x^\mu}u^\kappa u^\nu \tag{17.19}$$

Differentiation of Eq.(17.18) with respect to s produces

$$\frac{d}{ds}\left(g_{\mu\nu}u^\nu\right) = g_{\mu\nu}\frac{du^\nu}{ds} + \frac{dg_{\mu\nu}}{ds}u^\nu = g_{\mu\nu}\frac{du^\nu}{ds} + \frac{dx^\kappa}{ds}\frac{\partial g_{\mu\nu}}{\partial x^\kappa}u^\nu = g_{\mu\nu}\frac{du^\nu}{ds} + \frac{\partial g_{\mu\nu}}{\partial x^\kappa}u^\kappa u^\nu \tag{17.20}$$

Thus, the geodesic equation becomes

$$g_{\mu\lambda}\frac{du^\lambda}{ds} + \frac{\partial g_{\mu\nu}}{\partial x^\kappa}u^\kappa u^\nu - \frac{1}{2}\frac{\partial g_{\kappa\nu}}{\partial x^\mu}u^\kappa u^\nu = 0 \tag{17.21}$$

If we write down the same equation with κ and ν transposed, we must have the same result; adding the two together gives

$$2g_{\mu\lambda}\frac{du^\lambda}{ds} + \left(\frac{\partial g_{\mu\nu}}{\partial x^\kappa} + \frac{\partial g_{\mu\kappa}}{\partial x^\nu} - \frac{\partial g_{\kappa\nu}}{\partial x^\mu}\right)u^\kappa u^\nu = 0 \tag{17.22}$$

Contracting this expression with $\frac{1}{2}g^{\mu\alpha}$ and using the unimodularity of g we finally obtain

$$\frac{du^\alpha}{ds} + \Gamma^\alpha_{\kappa\nu}u^\kappa u^\nu = 0 \tag{17.23}$$

$$\Gamma^\alpha_{\kappa\nu} \equiv \frac{1}{2}g^{\mu\alpha}\left(\frac{\partial g_{\mu\nu}}{\partial x^\kappa} + \frac{\partial g_{\mu\kappa}}{\partial x^\nu} - \frac{\partial g_{\kappa\nu}}{\partial x^\mu}\right) \tag{17.24}$$

which is exactly the equation of motion we found before. Thus we conclude that *test particles in spacetime move along geodesics.*

18. Zwarte gaten

Als voorbeeld van de structuur van tijd en ruimte bekijken wij de Schwarzschild-metriek, die het gedrag van bolsymmetrische zwarte gaten beschrijft. Door strikte toepassing van symmetrieregels leiden wij de bewegingsvergelijking af voor licht in de buurt van een zwart gat.

Wij beschrijven de tijd-ruimte nu met bolcoördinaten (t, r, θ, ϕ), waarin t de tijd, r de radiële afstand, θ de elevatie (de hoek langs een meridiaan, gemeten vanaf de noordpool), en ϕ het azimut (de hoek langs een parallel, gemeten vanaf de nulmeridiaan). Een algemene vorm van een bolsymmetrische metriek is te schrijven als

$$ds^2 = \left(1 + 2\Phi/c^2\right) c^2\, dt^2 - \left(1 + 2\Phi/c^2\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{18.1}$$

In het deel waarin $d\theta, d\phi$ voorkomen herkennen wij het oppervlakte-element van een bol. Als wij nu voor de potentiaal Φ de klassieke Newton-vorm nemen, dan is

$$\frac{2\Phi}{c^2} = -\frac{2GM}{c^2 r} \tag{18.2}$$

Wij definiëren de `Schwarzschild-straal` R_S als

$$R_S \equiv \frac{2GM}{c^2} \tag{18.3}$$

en de bijbehorende `Schwarzschild-metriek` is

$$ds^2 = \left(1 - \frac{R_S}{r}\right) c^2\, dt^2 - \left(1 - \frac{R_S}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{18.4}$$

Het is niet gemakkelijk te bewijzen dat dit inderdaad een mogelijke stabiele tijd-ruimtestructuur is. Dat komt pas veel later, bij het onderwerp Algemene Relativiteitstheorie.

Exercise.

Bereken R_S voor de Zon, de Aarde, en een appel. Beredeneer op basis van fysische groothe-den waarom de structuur van tijd en ruimte bij de beschrijving van deze voorwerpen geen rol van betekenis speelt.

Het gaat er nu om, uit Eq.(18.4) bewegingsvergelijkingen af te leiden. Opnieuw, dat is nogal ingewikkeld, en we kunnen het algemene geval hier niet helemaal behandelen. Maar een eenvoudige aanzet kan worden gegeven door te kijken naar fotonen op radiële banen, die dus regelrecht naar het zwarte gat toe bewegen. Voor deze zijn elevatie en azimut constant, dus $d\theta = d\phi = 0$. Voor fotonen is bovendien het interval $s = 0$, en zodoende is

$$0 = \left(1 - \frac{R_S}{r}\right) c^2\, dt^2 - \left(1 - \frac{R_S}{r}\right)^{-1} dr^2 \tag{18.5}$$

Dit levert meteen een vergelijking voor dr/dt op:

$$\frac{dr}{dt} = \pm\, c\left(1 - \frac{R_S}{r}\right) \tag{18.6}$$

Dit kan meteen geïntegreerd worden tot

$$c(t - t_0) = \pm \int \left(\frac{r}{r - R_S)}\right) dr = \pm\,(r + R_S\, \log(r - R_S)) \tag{18.7}$$

Wanneer wij de tijd en de radiële afstand dimensieloos maken door als variabelen te nemen

$$\tau \equiv \frac{ct}{R_S}\ ; \qquad x \equiv \frac{r}{R_S} \tag{18.8}$$

dan komt dit er iets eenvoudiger uit te zien:

$$\tau = \tau_0 \pm (x + \log(x - 1)) \tag{18.9}$$

Exercise.

Maak een tijd-ruimte diagram van deze oplossingen. Zet verticaal de tijd τ en horizontaal de radiële afstand x uit. Teken grafieken van de ingaande lichtstralen (minteken in Eq.(18.9)) en de uitgaande (plusteken). Merk op dat de uitgaande lichtstralen allemaal tegen de afstand $x = 1$ aanliggen. Daaruit blijkt dat er geen licht ontsnappen kan vanuit het gebied $x < 1$. Daarom noemt men de bol met straal $r = R_s$ de `horizon`.

Het geval van niet-radiële banen is een stuk lastiger, maar kan worden opgelost door de bolsymmetrie van het probleem uit te buiten. Eerst voeren wij de coördinaten van Eq.(18.8) in in de metriek Eq.(18.4) en krijgen

$$ds^2 = \left(1 - \frac{1}{x}\right) d\tau^2 - \left(1 - \frac{1}{x}\right)^{-1} dx^2 - x^2(d\theta^2 + \sin^2\theta\, d\phi^2) \tag{18.10}$$

Dan zien wij, dat de factoren in de metriek niet van t, ϕ of θ afhangen. Daarvan maken we nu gebruik. Om te beginnen mogen wij $\theta = \text{const}$ nemen, zodat we – net als in het klassieke Kepler-geval – concluderen: *de banen liggen in een vlak.* Voor de inclinatie van dat vlak mogen we nemen $\theta = \pi/2$, zodat

$$ds^2 = \left(1 - \frac{1}{x}\right) d\tau^2 - \left(1 - \frac{1}{x}\right)^{-1} dx^2 - x^2\, d\phi^2 \tag{18.11}$$

In dat vlak gebruiken we poolcoördinaten (x, ϕ), omdat we ook hier met een centrale kracht te maken hebben. Nu nemen wij uiteraard *niet* $\phi = \text{const}$; dat mag wel, maar dan zijn we weer terug bij de boven behandelde radiële banen. De onafhankelijkheid van τ en ϕ geeft twee behouden grootheden: respectievelijk de energie en het impulsmoment, net als in de klassieke mechanica.

Stel dat wij het tijdsinterval $d\tau$ met een klein beetje ϵ laten toenemen tot $d\tau + \epsilon$. Dan veranderen de diverse grootheden in Eq.(18.11) niet, omdat zij niet van de tijd afhangen. De enige grootheid die kan veranderen is het interval ds, en wel van ds naar $ds + \delta$. Stoppen wij dit in Eq.(18.11) , en verwaarlozen wij hogere-orde termen van het type ϵ^2 en $\epsilon\delta$, dan komt er

$$ds^2 + 2\delta\, ds = \left(1 - \frac{1}{x}\right) d\tau^2 + 2\left(1 - \frac{1}{x}\right) \epsilon\, d\tau - \left(1 - \frac{1}{x}\right)^{-1} dx^2 - x^2\, d\phi^2 \tag{18.12}$$

Met behulp van de oorspronkelijke Eq.(18.11) zien we dat hieruit volgt

$$\delta\, ds = \left(1 - \frac{1}{x}\right) \epsilon\, d\tau \tag{18.13}$$

De verhouding δ/ϵ kan willekeurig worden ingesteld, en is dus een vrije constante. Dit wordt de behouden grootheid die is geassocieerd met de tijdsonafhankelijkheid. We noemen deze grootheid $\mathcal{E}$ en vinden

$$\frac{\delta}{\epsilon} \equiv \mathcal{E} = \left(1 - \frac{1}{x}\right) \frac{d\tau}{ds} \tag{18.14}$$

Evenals in het klassieke geval, is $\mathcal{E}$ de energie per massa-eenheid:

$$\mathcal{E} = \frac{E}{m} = \left(1 - \frac{1}{x}\right) \frac{d\tau}{ds} \tag{18.15}$$

Vervolgens herhalen wij deze truc met het azimut ϕ. Stel dat wij het hoekinterval $d\phi$ met een klein beetje μ laten toenemen tot $d\phi + \mu$. Dan veranderen de diverse grootheden in Eq.(18.11) niet, omdat zij niet van het azimut afhangen: we gebruiken de bolsymmetrie. De enige grootheid die kan veranderen is het interval ds, en wel van ds naar $ds + \delta$. Stoppen wij dit in Eq.(18.11) , en verwaarlozen wij hogere-orde termen van het type μ^2 en $\mu\delta$, dan vinden wij

$$ds^2 + 2\delta\, ds = \left(1 - \frac{1}{x}\right) d\tau^2 - \left(1 - \frac{1}{x}\right)^{-1} dx^2 - x^2\, d\phi^2 - 2x^2\mu\, d\phi \tag{18.16}$$

Met behulp van de oorspronkelijke Eq.(18.11) zien we dat hieruit volgt

$$-\frac{\delta}{\mu} \equiv \mathcal{J} = x^2 \frac{d\phi}{ds} \tag{18.17}$$

Vergelijken wij dit met de klassieke uitdrukking voor het impulsmoment, dan zien we dat $\mathcal{J}$ het impulsmoment per massaeenheid is:

$$\mathcal{J} = \frac{J}{m} = x^2 \frac{d\phi}{ds} \tag{18.18}$$

De vergelijkingen Eq.(18.15,18) stellen ons nu in staat om van Eq.(18.11) een bewegingsvergelijking te maken. Delen we de hele vorm door ds^2, en substitueren we Eq.(18.15,18) dan blijkt dat

$$1 = \left(1 - \frac{1}{x}\right)^{-1} \mathcal{E}^2 - \left(1 - \frac{1}{x}\right)^{-1} \left(\frac{dx}{ds}\right)^2 - \frac{\mathcal{J}^2}{x^3} \tag{18.19}$$

Deze vergelijking geldt voor deeltjes met eindige rustmassa. Voor fotonen is $ds = 0$, en vinden wij inplaats van Eq.(18.19)

$$\left(1 - \frac{1}{x}\right)^{-1} \mathcal{E}^2 - \left(1 - \frac{1}{x}\right)^{-1} \left(\frac{dx}{ds}\right)^2 - \frac{\mathcal{J}^2}{x^3} \tag{18.20}$$

Op deze manier hebben wij, door steevast van de bol- en tijdsymmetrie gebruik te maken, de volgende drie vergelijkingen gevonden:

$$\frac{d\tau}{ds} = \mathcal{E}\left(1 - \frac{1}{x}\right)^{-1} \qquad \text{(tijddilatatie; grav. roodverschuiving)} \tag{18.21}$$

$$\frac{d\phi}{ds} = \frac{\mathcal{J}}{x^2} \qquad \text{(impulsmoment)} \tag{18.22}$$

$$\frac{dx}{ds} = \pm\left\{\mathcal{E}^2 - \left(1 - \frac{1}{x}\right)\frac{\mathcal{J}^2}{x^3}\right\}^{1/2} \qquad \text{(impuls)} \tag{18.23}$$

Bovendien weten we dat de baan in een vlak ligt. De vorm van de baan kan worden gevonden door Eq.(18.22,23) op elkaar te delen, net als in het Kepler-geval; dan komt er

$$\frac{dx}{d\phi} = \pm\frac{x^2}{\mathcal{J}}\left\{\mathcal{E}^2 - \left(1 - \frac{1}{x}\right)\frac{\mathcal{J}^2}{x^3}\right\}^{1/2} \tag{18.24}$$

Tenslotte kunnen wij de verhouding $\mathcal{J}/\mathcal{E}$ nog wegwerken door opnieuw te kijken naar het analoge Kepler-geval. De verhouding van impulsmoment en energie geeft een lengteschaal b:

$$\frac{\text{energie}}{\text{impulsmoment}} = \frac{\mathcal{E}}{\mathcal{J}} = \frac{1}{b} \tag{18.25}$$

Hierin is b de afstand waarop het foton het punt $x = 0$ zou passeren als er geen afbuiging was (in het Engels: *impact parameter*), in de eenheden van Eq.(18.8) . Met behulp van b wordt Eq.(18.24)

$$\left(\frac{1}{x^2}\frac{dx}{d\phi}\right)^2 + \frac{1}{x^3}\left(1 - \frac{1}{x}\right) = \frac{1}{b^2} \tag{18.26}$$

Merk op dat ongebonden banen (overeenkomend met Kepler-hyperboolbanen) hebben $b^2 > 0$, terwijl bij gebonden banen (ellipsbanen in het Kepler-geval) $b^2 < 0$. Dit is uiteindelijk de bewegingsvergelijking voor licht dat in de buurt van een zwart gat beweegt.

19. Mass Scales

The Big Bang model assumes that the Universe has no structure at all, but that is of course a simplification. Actual structures have rather specific masses. The biggest mass is the horizon mass. The masses of stars are determined by microscopic properties, such as the smallest mass for which nuclear fusion occurs, and the mass at which a star is blown apart by its own radiation. Larger (galactic) mass scales may be related to global properties of the cosmic plasma, such as the Jeans mass and the photon damping mass.

What mass scales can we expect to play a special role in a Friedmann model? We will look at four possibilities: the horizon mass, the Jeans mass, the optical horizon mass, and the photon damping mass. Other mass scales can be calculated similarly. Note that we do not make quite the same fine distinction in epochs as Padmanabhan does.

The `horizon mass` is the amount of mass-energy encompassed by a sphere with a radius equal to the distane over which an unscattered light ray could have traveled since $t = 0$. The equation of motion of light is $ds = 0$; in a Robertson-Walker metric, this means that

$$c\, dt = a\, dr \left(1 - kr^2\right)^{-1/2} \tag{19.1}$$

For simplicity, and because it probably isn't far wrong, we will use the Einstein-De Sitter case

$$a = a_0 \times \begin{cases} \tau^{2/3} & \text{nonrelativistic dust} \\ \tau_e^{1/6}\, \tau^{1/2} & \text{relativistic gas} \end{cases} \tag{19.2}$$

Here and in all the following cases, the top line after the brace refers to the solution for nonrelativistic dust, and the bottom one to the solution for relativistic matter. In this expression we have used the abbreviations

$$\tau \equiv \frac{t}{t_0}\ ; \qquad \tau_e \equiv \frac{t_e}{t_0} \tag{19.3}$$

The time t_e indicates the time at which the density of matter and that of photons are equal. This time can be calculated as follows. The density of baryons must obey

$$\rho_b \propto a^{-3} \tag{19.4}$$

Because the energy of a photon is inversely proportional to its wavelength, and because this wavelength scales with a, we get

$$\rho_\gamma \propto a^{-4} \tag{19.5}$$

Inserting the E-DeS values, we obtain

$$\rho_b = \rho_{be} \times \begin{cases} (t/t_e)^{-2} \\ (t/t_e)^{-3/2} \end{cases} \tag{19.6}$$

and

$$\rho_\gamma = \rho_{\gamma e} \times \begin{cases} (t/t_e)^{-8/3} \\ (t/t_e)^{-2} \end{cases} \tag{19.7}$$

Here the index e indicates the value at equality, $t = t_e$. Relating the baryon density to its present value by $\rho_b = \rho_0 (t_0/t)^2$, we get

$$\rho_b = \rho_0 \times \begin{cases} \tau^{-2} \\ \tau_e^{-1/2}\, \tau^{-3/2} \end{cases} \tag{19.8}$$

and for the photons

$$\rho_\gamma = \rho_0 \times \begin{cases} \tau_e^{2/3}\tau^{-8/3} \\ \tau^{-2} \end{cases} \tag{19.9}$$

These expressions can only be equated if we have a proper observation of the mean mass-to-light ratio in the Universe today. The mean particle mass density follows from observation of Ω; we will use the E-DeS-value and assume critical density (which can be calculated directly from the Hubble parameter). The mean photon mass density follows from observation of the temperature of the microwave background and Stefan-Boltzmann's expression for the energy density of thermal radiation:

$$\mathcal{E} = \frac{\pi^2 k^4}{15\hbar^3 c^3} T^4 \tag{19.10}$$

Division by c^2 yields the corresponding mass density:

$$\rho_\gamma = \frac{\pi^2 k^4}{15\hbar^3 c^5} T^4 = 8.40 \times 10^{-33} T^4 \tag{19.11}$$

Inserting, as before, $H_0 = 65$ km s^{-1} Mpc^{-1} and the observed value $T = 2.737$ K, we get

$$\frac{\rho_\gamma}{\rho_0} = \tau_e^{2/3} \tag{19.12}$$

because today, by definition, $\tau = 1$; therefore (see Eq.(8.13))

$$\tau_e = 4.57 \times 10^{-7} \tag{19.13}$$

In the E-DeS case, the lookback distance becomes

$$r = c \int_0^t \frac{1}{a}\, dt = \frac{2c}{a_0 H_0} \times \begin{cases} \tau^{1/3} \\ \tau_e^{-1/6}\,\tau^{1/2} \end{cases} \tag{19.14}$$

The mass within the horizon is then $(4\pi\rho r^3)/3$. In what follows, we will consider the formation of baryonic structures; therefore, we will use ρ_b instead of the total mass-energy density which may also include zero rest mass particles. The baryon density scales as a^{-3}, and in the E-DeS universe $\rho_0 = \rho_c$, which allows us to express ρ_0 in terms of the present Hubble parameter and the gravitational constant. The baryonic horizon mass becomes

$$M_b = \frac{4c^3}{GH_0} \times \begin{cases} \tau \\ \tau_e^{-1/2}\,\tau^{3/2} \end{cases} \tag{19.15}$$

Using $H_0 = 65$ km s^{-1} Mpc^{-1} $= 2.107 \times 10^{-18}$ s^{-1}, the prefactor can be written as

$$\frac{4c^3}{GH_0} = 7.67 \times 10^{53} = 3.85 \times 10^{23}\ \mathrm{M}_\odot \tag{19.16}$$

The next mass scale to be considered is the `Jeans mass`, that is the amount of matter above which a body is gravitationally unstable against gas pressure. The value of this critical mass M_J can be estimated as that of a sphere in which the crossing time for sound waves is equal to the free fall time. For, if we were to perturb a piece of the surface of the sphere by pushing it inward, a pressure wave would result, propagating at the speed s; if that wave could transmit the news that the sphere is being perturbed before the perturbed piece has fallen very far, the sphere would restore itself to equilibrium. Therefore

$$s^2 = v_{ff}^2 = \frac{GM_J}{r_J} \tag{19.17}$$

where r_J is the Jeans length, in this case the radius of the sphere. Because the sphere is supposed to consist of baryons, we get

$$M_J = \left(\frac{s^2}{G}\right)^{3/2} \left(\frac{3}{4\pi\rho_b}\right)^{1/2} \tag{19.18}$$

After matter becomes dominant but before decoupling, $s^2 = \gamma kT/m$, but the matter is coupled to the radiation so that $T \propto 1/a$; for a monatomic ideal gas $\gamma = 5/3$. In the photon-dominated regime, $s^2 = c^2/3$ (because pressure signals are then transmitted by relativistic particles random-walking through three dimensions), so that

$$s^2 = \begin{cases} \frac{5kT}{3m} \\ \frac{1}{3}c^2 \end{cases} \tag{19.19}$$

Inserting this into the equation for the Jeans mass, we obtain

$$M_J = \frac{c^3\sqrt{2/27}}{GH_0} \times \begin{cases} \tau_e \\ \tau_e^{1/4}\,\tau^{3/4} \end{cases} \tag{19.20}$$

Some algebra shows that M_J drops below M_b close to the equalization time $t = t_e$,

$$4\tau_e^{1/2}\,\tau^{3/2} = \sqrt{\frac{2}{27}}\,\tau_e^{1/4}\,\tau^{3/4} \quad \rightarrow \quad \tau_e^{3/4} = \sqrt{8}\sqrt{27}\,\tau^{3/4} = 6^{3/2}\,\tau^{3/4} \tag{19.21}$$

so that $M_b = M_J$ at $t = t_e/36$. The fun is, that this number does not depend on any details like H_0 and so forth. Thus we can conclude quite robustly that Jeans instability at large masses becomes possible just before the time when baryons start to dominate. This is an *a posteriori* justification for the use of ρ_b in our calculations. From t_e onwards, the Jeans mass is constant at the value

$$M_J = 5.22 \times 10^{52}\tau_e \quad \text{kg} = 2.38 \times 10^{46} \quad \text{kg} = 1.2 \times 10^{16} \quad \text{M}_\odot \tag{19.22}$$

At the critical density $\rho_0 = 7.94 \times 10^{-27}$, this mass corresponds to a present-day length scale of about 46 Mpc. This is quite encouraging, seeing that this is rather close to the size of voids and filaments today.

Below M_J, perturbations cannot grow but must oscillate like a sound wave. Their amplitude is changed adiabatically by the cosmic expansion. However, small perturbations can be damped by photon viscosity. Adjacent dense and tenuous regions in a wave can exchange energy by means of radiation and thereby damp out. To get the corresponding mass scales, we must see how far a photon can travel. The mean free path for Thomson scattering is estimated as before by

$$\lambda \approx \frac{1}{\sigma n_e} \tag{19.23}$$

If m is the mean atomic weight, we have

$$n_e = \frac{\rho_0}{m} \times \begin{cases} \tau^{-2} \\ \tau_e^{-1/2}\,\tau^{-3/2} \end{cases} \tag{19.24}$$

If we call M_τ the mass within the `optical horizon` with radius λ, then

$$M_\tau = \left(\frac{2G}{H_0^2}\right)^2 \left(\frac{4\pi m}{3\sigma}\right)^3 \times \begin{cases} \tau^4 \\ \tau_e\,\tau^3 \end{cases} \tag{19.25}$$

If we use the Thomson cross section $\sigma = 6.65 \times 10^{-29}$ m^2 and for m the hydrogen mass 1.67×10^{-27} kg, the prefactor becomes 1.06×10^{57} kg $= 5.32 \times 10^{26}$ M$_\odot$. At t_e, the optical horizon mass then equals 23 M$_\odot$.

The mass M_τ is not the mass that is being damped, however. A photon random-walks around; in a cosmic time t the photon is scattered N times and it travels a mean distance $\lambda\sqrt{N}$. The number of scatterings is evidently

$$N = \frac{ct}{\lambda} \tag{19.26}$$

wherefore the length covered by N scatterings is

$$\lambda_N = \sqrt{\lambda ct} = \sqrt{\frac{ct}{\sigma n_e}} \tag{19.27}$$

In the damping process, the photons *do* contribute to the inertia, so we must use the total density in calculating the damping mass M_D:

$$\rho = \frac{3H_0^2}{8\pi G}\tau^{-2} \tag{19.28}$$

The final result then follows from $M_D = \frac{4}{3}\pi\rho\lambda_N^3$ as

$$M_D = \frac{32}{H_0}\sqrt{G}\left(\frac{\pi mc}{3\sigma H_0}\right)^{3/2} \times \begin{cases} \tau^{5/2} \\ \tau_e^{3/4}\,\tau^{7/4} \end{cases} \tag{19.29}$$

Inserting the usual numbers, the prefactor becomes 2.85×10^{55} kg, so that

$$M_D = 1.43 \times 10^{25}\ \mathrm{M}_\odot\ \times \begin{cases} \tau^{5/2} \\ \tau_e^{3/4}\,\tau^{7/4} \end{cases} \tag{19.30}$$

At the epoch of recombination, the importance of photon damping evidently disappears; at that point, $\tau \approx 10^{-5}$, so that $M_D \approx 4 \times 10^{12}\ \mathrm{M}_\odot$. It is tempting to associate this mass with the scale of galaxies: below it, all perturbations from the early Universe will have been wiped out by photon viscosity.

Finally, we consider what happens after recombination. Because the Universe becomes transparent to photons, M_τ and M_D become effectively infinite. Of course M_b keeps increasing proportional to t. The only scale that changes is M_J. Since the temperature of matter after recombination scales as a^{-2}, we have $T = T_{\rm rec}(t_{\rm rec}/t)^{4/3}$, where $T_{\rm rec} \approx 4000$ K is the temperature at which recombination has reduced the ionization to about one half. Accordingly,

$$M_J = \left[\frac{5kT_{\rm rec}}{3mG}\left(\frac{t_{\rm rec}}{t}\right)^{4/3}\right]^{3/2}\left(\frac{2G}{H_0^2}\tau^2\right)^{1/2} \tag{19.31}$$

from which follows

$$M_J = \frac{\sqrt{2}}{H_0 G}\left(\frac{5kT_{\rm rec}}{3m}\right)^{3/2}\left(\frac{t_{\rm rec}}{t_0}\right)^2\tau^{-1} \tag{19.32}$$

The prefactor turns out to be $1.9 \times 10^9\ \mathrm{M}_\odot$, so that the Jeans mass starts at about $10^4\ \mathrm{M}_\odot$ right after recombination. It has been suggested that this mass scale corresponds to that of globular clusters.

Exercise.

Calculate the angles subtended by various interesting length scales, as seen by us now. For example, what is the value of λ_D, and what angle does it subtend as seen by us?

20. Motion on a Smooth Background

To see whether the Friedmann solutions are stable, we must consider the motion of the cosmic gas as a perturbation on a smooth mean expansion. This produces certain expressions for the growth rates.

The proper separation of two comoving particles is given by a fixed distance $\vec{x}$ multiplied with the cosmic scale factor:

$$\vec{r} = a(t)\vec{x} \tag{20.1}$$

Because the relative importance of the curvature term (proportional to k) gets smaller for earlier times, and because structure formation can only occur after decoupling of radiation and matter anyway, it is not too bad to assume the Einstein-De Sitter dust model with $k = 0, P \ll \rho c^2, \Lambda = 0$, so that the cosmic potential Φ_b due to the background mass density ρ_b at $\vec{r} = 0$ is (see earlier discussion about Newtonian models)

$$\Phi_b = \frac{2\pi}{3} G\rho_b r^2 \tag{20.2}$$

The usual Newtonian equation of motion $d\vec{v}/dt = -\vec{\nabla}\Phi$ then becomes

$$\frac{d^2a}{dt^2} = -\frac{4\pi}{3} G\rho a \tag{20.3}$$

The proper velocity is

$$\vec{w} = a\frac{d\vec{x}}{dt} + \vec{x}\frac{da}{dt} \tag{20.4}$$

and the energy with respect to the comoving background of a particle with mass m is

$$E_{\text{pec}} = \frac{1}{2} m a^2 \left(\frac{d\vec{x}}{dt}\right)^2 \tag{20.5}$$

This quantity can be seen as the derivative of a potential term; using this, we can introduce the potential ϕ of perturbations on the smooth background as

$$\phi \equiv \Phi - \frac{1}{2} x^2 a \frac{d^2a}{dt^2} \tag{20.6}$$

If we stipulate that $\vec{\nabla}$ and δ denote derivatives with respect to the comoving coordinates, then Poisson's Equation becomes

$$\Delta\phi = 4\pi G\rho a^2 + 3a\frac{d^2a}{dt^2} \tag{20.7}$$

Using the background solution for a, this yields

$$\Delta\phi = 4\pi G a^2 \left(\rho(\vec{x}, t) - \rho_b(t)\right) \tag{20.8}$$

Consequently, the equation of motion for the peculiar velocity $\vec{v} \equiv a(d\vec{x}/dt)$ with respect to the comoving particles is

$$\frac{d\vec{v}}{dt} + \frac{1}{a}\frac{da}{dt}\vec{v} = -\frac{1}{a}\vec{\nabla}\phi \tag{20.9}$$

Note that, in the case of vanishing perturbations ($\phi = 0$), this equation has the solution $v \propto 1/a$. This is a reflection of the fact that adiabatic expansion suppresses random motions; it is analogous to the equation for the decrease of the particle temperature as the Universe expands. This decay of random velocities is the reason that the Doppler term in the equation for $\Delta T/T$ of the CMBR is likely to be unimportant.

If we want to calculate the evolution of inhomogeneities in a practical case, we may solve the comoving Poisson Equation by

$$\phi(\vec{x}) = -Ga^2 \int \frac{\rho(\vec{y}) - \rho_b}{|\vec{y} - \vec{x}|} d\vec{y} \tag{20.10}$$

The peculiar acceleration can simply be found by applying the comoving gradient $\vec{\nabla}/a$:

$$\vec{g} = Ga \int \rho(\vec{y}) \frac{\vec{y} - \vec{x}}{|\vec{y} - \vec{x}|^3} d\vec{y} \tag{20.11}$$

where we have used that the integral over ρ_b vanishes due to isotropy of the background universe. Often, one wants to use this expression in an N-body code; in that case we get a sum over all particles,

$$\vec{g} = Ga^{-2} \sum_j m_j \frac{\vec{x}_j - \vec{x}}{|\vec{x}_j - \vec{x}|^3} \tag{20.12}$$

21. The Boltzmann/Vlasov Equation

As a preparation for the equations of motion of a gas, we consider the kinetic properties of equilibrium gases.

The main differences between the Newtonian dynamics we use in N-body calculations and the dynamics of a gas are the following. In the former case, each particle has a definite identity. If a particle has a mass much greater than that of a galaxy, we pretend that it stands for a small group or cluster of galaxies; if its mass is much smaller, several such particles together may be taken to represent one galaxy. In the calculation of the particle orbits we still follow each particle individually, albeit sometimes with gross approximations made necessary by the large number of particles. Furthermore, barring some special applications, we assume that the particles are collisionless.

Contrariwise, in hydrodynamics we renounce following the particles individually and take the average over a great many particles, which are taken to be collision dominated, to the effect that the particle trajectories are so well mixed up on a microscopic scale that we can take quasi-equilibrium thermodynamic averages over them.

Suppose that the particles form a closed `Hamiltonian system`. In that case the motion of each one is fully characterized by its position x and momentum p, and there is no external force. Each orbit is found from the Hamiltonian function $\mathcal{H}$ through the equations of motion

$$\frac{dx}{dt} = \frac{\partial \mathcal{H}}{\partial p} \tag{21.1}$$

$$\frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial x} \tag{21.2}$$

Now assume that we have a giant set $\{x^i\}$ of such particles, each with its own combination (x, p). Then the number dN of particles with positions between x and $x + dx$ and momenta in the range from p to $p + dp$ is

$$dN = f(x, p, t)\, dx\, dp \tag{21.3}$$

Note, first, that the notation becomes a little more complicated if we explicitly take into account that x and p are multi-dimensional; second, that this equation is relativistically invariant if we write x and p as the appropriate four-vectors.

The evolution of the whole system can be described in the average if we know how the `distribution function` f changes in the course of time. Now we note that f is a density function on (x, p)-space. If no particles are created or destroyed, the time rate of change of f is composed of two terms, one that describes how the occupancy of a cell (dx, dp) changes due to streaming along the x-coordinate, and one that does something similar for streaming in the p-direction. In the x-direction, streaming goes with the speed $v \equiv dx/dt$, whereas along the p-axis the streaming 'speed' is the derivative $F \equiv dp/dt$, the force:

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\left(f\frac{dx}{dt}\right) - \frac{\partial}{\partial p}\left(f\frac{dp}{dt}\right) \tag{21.4}$$

Working out the differentiations gives

$$\frac{\partial f}{\partial t} + \frac{dx}{dt}\frac{\partial f}{\partial x} + \frac{dp}{dt}\frac{\partial f}{\partial p} = -f\left[\frac{\partial}{\partial x}\frac{dx}{dt} + \frac{\partial}{\partial p}\frac{dp}{dt}\right] \tag{21.5}$$

Inserting the Hamiltonian equations of motion in the term between brackets finally produces

$$\frac{\partial f}{\partial t} + \frac{dx}{dt}\frac{\partial f}{\partial x} + \frac{dp}{dt}\frac{\partial f}{\partial p} = 0 \tag{21.6}$$

It is important to note that we have implicitly assumed that no particles enter or leave phase cells due to collisions. Taking these into account is possible; then the right-hand side of Eq.(21.6) is not zero but equal to a `collision term`. Under certain general conditions this term can be calculated exactly for strict two-body collisions; in that case we speak of the `Boltzmann Equation`. Deriving this would go to far in the present context.

Instead, we assume that collisions are so frequent and effective that our many-particle system is always *locally* in equilibrium. In that case f has the form of a Maxwellian velocity distribution and the collision term averages to zero. Thus we retain the equation

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} + F\frac{\partial f}{\partial p} = 0 \tag{21.7}$$

To make the following discussion clearer we shall assume that all particles have the same mass. Then we can replace p by mv and the force F by the acceleration A to obtain

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} + A\frac{\partial f}{\partial v} = 0 \tag{21.8}$$

In three dimensions this becomes, in component notation,

$$\frac{\partial f}{\partial t} + v_i\frac{\partial f}{\partial x_i} + A_j\frac{\partial f}{\partial v_j} = 0 \tag{21.9}$$

This equation, which occurs frequently in plasma physics and gravitational N-body work, is called the `Vlasov Equation`. As it stands, it is just a simple bookkeeping equation and doesn't seem to involve much physics. However, we must realize that the acceleration A_j is due to the average interaction between all particles, microscopic collisions excluded. Calculation of A is no small feat. In fact, it is practically impossible. In actual applications of the Vlasov Equation we must specify very carefully how we are going to derive A from f. Note, by the way, that the fact that A depends on f makes the Vlasov Equation strongly nonlinear, which it would not seem to be at first sight.

22. The Maxwell-Boltzmann Distribution

The most probable distribution of particles in phase space is found from a maximum-likelihood permutation argument.

The equilibrium solution of the Boltzmann Equation can be found by simple combinatorial arguments. Consider a Hamiltonian system of N particles in a volume V, with total energy E. Because E is finite there is a maximum velocity v beyond which (x, v) phase space is unoccupied: one particle has usurped all the energy. This finite extent implies that we can divide the occupied volume of phase space into a finite number M of cells $dx\,dv$. Give each cell a unique serial number i. The number of particles in the cell is N_i, the energy is ϵ_i. Because of conservation of particles and energy we have

$$\sum_i^M N_i = N \tag{22.1}$$

$$\sum_i^M N_i \epsilon_i = E \tag{22.2}$$

The probability of generating any particular f is proportional to the number of ways in which f can be obtained from the same N particles, under the above constraints. This number is proportional to the number of permutations of N objects, divided by the number of trivial permutations, which are those that permute the particles within one cell. Therefore, the probability W associated with any distribution $\{N_i\}$ is

$$W(\{N_i\}) = \frac{N!}{\prod_i (N_i!)} M^{-N} \tag{22.3}$$

The physical assumption now is that short-range binary collisions are properly modeled by permutations between phase cells; the equilibrium situation is obtained when W is maximal under its constraints of energy and particle conservation. To find this maximum it is plausible to maximise $\log W$ instead, because W contains a product Π. This is permissible since the logarithm is strictly monotonic. Therefore we try to make

$$\log W = \log N! - \sum_i \log N_i! - N \log M \tag{22.4}$$

extremal. We use Lagrange multipliers α and β to obtain

$$\frac{\partial}{\partial N_i} \left(\log W - \alpha \sum_i N_i - \beta \sum_i N_i \epsilon_i \right) = 0 \tag{22.5}$$

For very large values of N we can use `Stirling's Rule`

$$\log N! \simeq N \log N - N \tag{22.6}$$

and therefore we conclude

$$\log N_i + \alpha + \beta \epsilon_i = 0 \tag{22.7}$$

and therefore, summing over i and using the conservation equations,

$$\mathrm{e}^{-\alpha} \sum_i \mathrm{e}^{-\beta \epsilon_i} = N \tag{22.8}$$

$$\mathrm{e}^{-\alpha} \sum_i \epsilon_i \mathrm{e}^{-\beta \epsilon_i} = E \tag{22.9}$$

If the number of particles N_i in each cell is large enough, so that the fluctuations $(1/\sqrt{N_i})$ of cell quantities are small compared with their mean, we can replace the sums by integrals over continuous variables. The discrete distribution $\{N_i\}$ corresponds to the continuous distribution function f, which is effectively interpreted as a function proportional to the probability of finding a particle in a cell. Thus we get

$$\sum_i N_i g_i \to \int f(v) g(v) V \, dv \tag{22.10}$$

The condition Eq.(22.8) we have derived above then becomes

$$\int f(v)V\,dV = N \quad \text{or} \quad \int f(v)\,dV = \frac{N}{V} = n \tag{22.11}$$

where n is the particle density in space.

Similarly we identify ϵ_i with the energy $\frac{1}{2}mv^2$, so that Eq.(22.9) produces

$$f(v) = A\,\mathrm{e}^{-\frac{\beta}{2}mv^2} \tag{22.12}$$

This is the expression for the equilibrium distribution f with an as yet undetermined normalization constant A. Note that the energy ϵ_i could have been identified with other energy expressions, depending on the particular formulation of the problem, for example the difference between kinetic and potential energy, if the particle system sits in an externally applied potential. The normalization constant obeys

$$\int f\,dv = \frac{N}{V} = A\int \mathrm{e}^{-\frac{\beta}{2}mv^2}dv \tag{22.13}$$

and thus, for one degree of freedom,

$$A = \frac{N}{V}\left(\frac{\beta m}{2\pi}\right)^{1/2} \tag{22.14}$$

The constant β is found by defining the temperature as a measure of the energy in a single degree of freedom:

$$\frac{E}{N} \equiv \frac{1}{2}kT \tag{22.15}$$

Using the Lagrange energy constraint

$$E = V\int \frac{1}{2}mv^2 f(v)\,dv \tag{22.16}$$

we conclude that

$$\beta = \frac{1}{kT} \tag{22.17}$$

and finally the equilibrium distribution $f(v)$ is found to be, if D is the number of degrees of freedom of the system ($D = 3$ in three-space),

$$f(v) = n\left(\frac{m}{2\pi kT}\right)^{D/2}\mathrm{e}^{-mv^2/2kT} \tag{22.18}$$

This is the Maxwell-Boltzmann distribution. In a spatially homogeneous system, the equilibrium distribution function has this form.

23. The Euler Equation

The equation of motion of a gas is basically the same as that for Newtonian particles, but with an additional term due to the unseen presence of microscopic particle motions. This term is a pressure gradient.

So far we have still not said anything about averages, except to sweep the effects of microscopic collisions under the rug. But having found the equilibrium form of F we are finally in a position to derive averages of the Boltzmann-Vlasov Equation to produce macroscopically useful equations. First, we use f as a probability density in phase space and calculate its velocity moments, i.e. the integrals over f of v^0, v^1, and v^2 (higher moments produce nothing new in a Hamiltonian system):

$$n = \int f\, dv \qquad \text{density} \tag{23.1}$$

$$u_i = \frac{1}{n}\int v_i f\, dv \qquad \text{velocity} \tag{23.2}$$

$$E = \frac{1}{n}\int \frac{1}{2} m v^2 f\, dv \qquad \text{energy} \tag{23.3}$$

Next, we calculate the same type of moments by integration of the whole B-V Equation over the velocity part of phase space. The zeroth moment produces

$$\int \left(\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} + A_j \frac{\partial f}{\partial v_j}\right) d^3v = \frac{\partial}{\partial t}\int f\, d^3v + \int v_i \frac{\partial f}{\partial x_i} d^3v + A_j \int \frac{\partial f}{\partial v_j} d^3v \tag{23.4}$$

The first term is the time derivative of the density; the third is integrable and is therefore zero, because in a finite system f must be zero at infinity. The spatial differentiation in the second term commutes with the integration ov er velocity, so that

$$\int v_i \frac{\partial f}{\partial x_i} d^3v = \frac{\partial}{\partial x_i} n u_i \tag{23.5}$$

It is customary to use the mass density ρ instead of the particle density n; because

$$\rho \equiv mn \tag{23.6}$$

the final equation obtained from the zeroth moment of the B-V Equation is then

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_i}{\partial x_i} = 0 \tag{23.7}$$

This is the equation of `mass conservation`, or `continuity equation`. It is sometimes writen in vectorial form as

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}\cdot(\rho \vec{u}) = 0 \tag{23.8}$$

It is the first equation of hydrodynamics.

In the same way the first velocity moment of the B-V Equation is found to be

$$\int \left(v_j \frac{\partial f}{\partial t} + v_j v_i \frac{\partial f}{\partial x_i} + v_j A_k \frac{\partial f}{\partial v_k}\right) d^3v = \frac{\partial}{\partial t}\int v_j f\, d^3v + \int v_j v_i \frac{\partial f}{\partial x_i} d^3v + \int v_j A_k \frac{\partial f}{\partial v_k} d^3v \tag{23.9}$$

The first term is straightforward:

$$\frac{\partial}{\partial t}\int v_j f\, d^3v = \frac{\partial n u_j}{\partial t} = n\frac{\partial u_j}{\partial t} + u_j \frac{\partial n}{\partial t} \tag{23.10}$$

The third term is not too difficult either, since it can be dealt with through integration by parts:

$$\int v_j A_k \frac{\partial f}{\partial v_k} d^3v = \int A_k \left(\frac{\partial f v_j}{\partial v_k} - f\frac{\partial v_j}{\partial v_k}\right) d^3v = -\int f A_k \delta_{jk} d^3v = -nA_j \tag{23.11}$$

Only the second term in Eq.(23.9) requires some thoughtful manipulation. We have

$$\int v_j v_i \frac{\partial f}{\partial x_i} d^3v \tag{23.12}$$

Effectively, this is an expression for the gradient of the velocity correlation $v_j v_i f$. In an isotropic gas the off-diagonal elements are zero, and thus the only contribution that remains is of the form

$$\frac{\partial}{\partial x_i} \int v_j v_j f \, d^3 v \tag{23.13}$$

The integral is the expectation value of the energy density of the microscopic random motions in the gas. It is called the `gas pressure` P. In terms of P we can write Eq.(23.9) as

$$\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} = -\frac{1}{\rho}\frac{\partial P}{\partial x_i} + A_i \tag{23.14}$$

or, in vector notation,

$$\frac{\partial \vec{u}}{\partial t} + (\vec{u} \cdot \vec{\nabla})\vec{u} = -\frac{1}{\rho}\vec{\nabla} P + \vec{A} \tag{23.15}$$

This is the equation of motion or `Euler's Equation`. The new function P is the gas pressure, defined as

$$P \equiv m \int (\vec{v} - \vec{u})^2 d^3 v \tag{23.16}$$

Thus, P is a measure of the energy density of the random motions in the system, after we have split off the systematic velocity ("wind speed") $\vec{u}$. It therefore summarizes the microscopic processes which we have averaged out. In order to be able to solve the equation of motion, we need a prescription for P, known as an equation of state. For ideal gases such equations are not difficult but the general case is rather nasty. The most frequently used is the equation of state for an ideal gas:

$$P = \kappa \rho^{\gamma} \tag{23.17}$$

The constant κ is proportional to the logarithm of the entropy. The constant γ, called `Poisson's exponent`, is equal to 5/3 for a classical monatomic gas and 4/3 for an extremely relativistic gas.

The third moment equation, describing the evolution of the energy density, can be calculated directly by integrating the B-V Equation multiplied by v^2. However, it is much easier to obtain the energy equation by multiplication of Euler's Equation with $\rho\vec{u}$ and deriving from that the equation for the energy density $\frac{1}{2}\rho v^2$.

24. Stability of Friedmann Models

The preceding techniques are applied to the Friedmann solutions, and it is shown that certain perturbations of the cosmic gas can grow under the influence of their self-gravity. However, due to the expansion of the Universe this growth is not exponential but follows a power law instead.

In the phase plane of the Friedmann equation, the phase orbits crowd together asymptotically as $a \to 0$. That suggests that small perturbations might cause the Universe to proceed *locally* on a different phase path, and maybe even to collapse (i.e. acquire local characteristics similar to a closed $k = 1$ universe).

To see if and how this may happen, we consider the evolution of small perturbations on a background Friedmann model. We start with the usual solution

$$\frac{\rho}{\rho_0} = \left(\frac{a_0}{a}\right)^3 \tag{24.1}$$

$$v = \frac{1}{a}\frac{da}{dt} x \equiv Hx \tag{24.2}$$

$$\left(\frac{da}{dt}\right)^2 + k = \frac{8\pi}{3} G\rho a^2 \tag{24.3}$$

Let us assume that every variable V^* can be written as the sum of a part V which solves the Friedmann equation and a small perturbation V_1. Then to a first approximation (i.e. neglecting all terms of order $V_1 W_1$ and higher) we get a perturbed equation of motion, Poisson equation, continuity equation and equation of state:

$$\frac{\partial \vec{v}_1}{\partial t} + H\vec{v}_1 + H(\vec{x}\cdot\nabla)\vec{v}_1 = -\frac{1}{\rho}\vec{\nabla}P_1 + \vec{g}_1 \tag{24.4}$$

$$\vec{\nabla}\times\vec{g}_1 = 0 \tag{24.5}$$

$$\vec{\nabla}\cdot\vec{g}_1 = -4\pi G\rho_1 \tag{24.6}$$

$$\frac{\partial \rho_1}{\partial t} + 3H\rho_1 + H(\vec{x}\cdot\vec{\nabla}\rho_1) + \rho\vec{\nabla}\cdot\vec{v}_1 = 0 \tag{24.7}$$

$$P_1 = s^2\rho_1 \tag{24.8}$$

This set of equations is linear (by construction), so it makes sense to analyze them by Fourier decomposition; every variable V_1 is written in the form

$$V_1 \rightarrow V_1(t)\exp(\frac{i}{a}\vec{x}\cdot\vec{q}) \tag{24.9}$$

Since the gas has zero shear strength, the only interesting modes are compressional, that is to say, the velocity perturbation is a scalar multiple of the wave vector:

$$\vec{v}_1 = i\epsilon(t)\vec{q} \tag{24.10}$$

For convenience, we will use the fractional density

$$\rho_1(t) \equiv \rho_0\left(\frac{a_0}{a}\right)^3\delta(t) \tag{24.11}$$

Substitution of the above in the linearized equations produces

$$\frac{d\epsilon}{dt} + H\epsilon = \left(-\frac{s^2}{a} + \frac{4\pi G\rho a}{q^2}\right)\delta \tag{24.12}$$

$$\frac{d\delta}{dt} = q^2\frac{\epsilon}{a} \tag{24.13}$$

The second equation can be substituted into the first, to yield finally

$$\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} + \left(\frac{s^2q^2}{a^2} - 4\pi G\rho\right)\delta = 0 \tag{24.14}$$

Substitution of the value of ρ in an E-dS universe produces the corollary

$$\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} + \left(\frac{s^2q^2}{a^2} - \frac{3}{2}H^2\right)\delta = 0 \tag{24.15}$$

The second term is due to damping by adiabatic expansion (the same term we encountered earlier in the comoving velocity); the third term describes oscillation or growth of the perturbation. Obviously, oscillatory behaviour is determined by the sign of the expression in brackets: if this is negative, we get a runaway solution. The dividing line between oscillation and growth is then found by putting this term to exactly zero. This gives

$$q = \left(\frac{4\pi G\rho a^2}{s^2}\right)^{1/2} \propto \left(\frac{GM/a}{s^2}\right)^{1/2} \propto \frac{v_{\rm free}}{s} \tag{24.16}$$

This is the famous `Jeans condition`: if a perturbed mass can fall freely more rapidly than pressure waves can travel, then counterpressures inside always come too late to stabilize the perturbed mass against gravitational collapse.

Let us suppose that we are well above the Jeans limit given by this condition. Then the pressure term (i.e. the one proportional to s^2) can be neglected, and the perturbation equation becomes

$$\frac{d^2\delta}{dt^2} + \frac{4}{3t}\frac{d\delta}{dt} - \frac{2}{3t^2}\delta = 0 \tag{24.17}$$

in which we have assumed an Einstein-De Sitter background. This equation has two solutions:

$$\delta \propto t^{2/3} \; ; \qquad \delta \propto t^{-1} \tag{24.18}$$

Thus, there is a growing solution. Unfortunately, because the perturbation is taken on an expanding background, the growth is not exponential but merely a power law. We will see that this slow growth means trouble for the reconciliation of the smoothness of the CMBR and the origin of big perturbations like (clusters of) galaxies.

The velocity perturbation corresponding to the growing density mode goes as

$$\epsilon \propto \frac{d\delta}{dt} \propto t^{-1/3} \tag{24.19}$$

or, using the full Fourier superposition of a velocity power spectrum $V(q)$,

$$\vec{v}_1 = \int \frac{2ia\vec{q}}{3q^2 t_0^{2/3}}\, t^{-1/3} \exp\left(\frac{i}{a}\vec{q}\cdot\vec{x}\right) V(q) d\vec{q} \tag{24.20}$$

Because the velocities damp out with respect to the comoving background, we can expect that integration of $\vec{v}(t)$ allows a reasonable approximation of the positions of perturbed particles, an approximation which should hold fairly well for long times even if the density amplitude δ increases into the nonlinear regime:

$$\vec{x} = \vec{x}_0 + F(t) \times (\text{direction of initial perturbation}) \tag{24.21}$$

25. Growth or Oscillation

Because of the expansion of the Universe, the growth, oscillation, or damping of perturbations is not the same as in a static medium. In an Einstein-De Sitter universe, the various rates can be expressed as powers of the cosmic time.

We saw that small density perturbations obey

$$\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} + \left(\frac{s^2q^2}{a^2} - \frac{3}{2}H^2\right)\delta = 0 \tag{25.1}$$

Well above the Jeans limit, we can use $s \approx 0$, so that we get

$$\begin{aligned} &\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} - \frac{3}{2}H^2\delta = 0 \qquad \text{(nonrelativistic)} \\ &\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} - 4H^2\delta = 0 \qquad \text{(relativistic)} \end{aligned} \tag{25.2}$$

(for the relativistic case see Weinberg p.588). Inserting the value of H in the two cases, one obtains

$$\begin{aligned} &\frac{d^2\delta}{dt^2} + \frac{4}{3t}\frac{d\delta}{dt} - \frac{2}{3t^2}\delta = 0 \\ &\frac{d^2\delta}{dt^2} + \frac{1}{t}\frac{d\delta}{dt} - \frac{1}{t^2}\delta = 0 \end{aligned} \tag{25.3}$$

The growing solutions are

$$\delta \propto \begin{cases} t^{2/3} \\ t \end{cases} \tag{25.4}$$

These are the only growing solutions before decoupling, because we have taken masses above the Jeans mass. Below that critical mass we have the other limiting case of the perturbation equation, namely

$$\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} + \frac{s^2q^2}{a^2}\delta = 0 \tag{25.5}$$

This equation has constant amplitude oscillations in the relativistic regime, whereas after t_e we find oscillations with a slowly decreasing amplitude proportional to $t^{-1/6}$.

26. Growth of Structures

Small perturbations locally look like ellipsoidal features. Their equations of motion shows that overdensities become more and more aspherical as they collapse, whereas underdensities become more and more spherical when they expand, a little faster than the Hubble expansion.

On intermediate scales, Newtonian gravitational instability should suffice for describing the formation of structure. Hydrodynamical details such as pressure effects are unlikely to be important in the progenitors of structures in the 10–500 Mpc regime, so we can restrict ourselves to a 'dust' equation of state. The potential Φ near any point (x, y, z) of a self-gravitating medium can be written as

$$\Phi = \sum_{ijk} a_{ijk} x^i y^j z^k \tag{26.1}$$

Near a density maximum, the leading terms are the quadratic ones, which, by a suitable orientation of Cartesian coordinates, can be written as

$$\Phi = Ax^2 + By^2 + Cz^2 + \cdots \tag{26.2}$$

Neglecting terms of higher than second order, this is the potential of a homogeneous ellipsoid. That should be no surprise: the smallest closed contours in any topographical map are ellipses.

The collapse of high-density regions can thus be approximated by considering the homologous motion of ellipsoids. Consider, then, the Newtonian collapse of a homogeneous ellipsoid. Suppose that a particle of such a mass distribution were initially located at (a, b, c), and that at some later time t it had moved to the point $(aX(t), bY(t), cZ(t))$, then the density ρ would evolve according to

$$\rho(t) = \rho_0/XYZ \tag{26.3}$$

The equations of motion for the scaling functions X, Y, and Z are found as follows. The potential Φ obeys

$$\begin{aligned}\Phi &= k(\alpha x^2 + \beta y^2 + \gamma z^2) = \\ &= k(\alpha a^2 X^2 + \beta b^2 Y^2 + \gamma c^2 Z^2)\end{aligned} \tag{26.4}$$

and Poisson's Equation demands that

$$k(\alpha + \beta + \gamma) = 2\pi G\rho \tag{26.5}$$

The components of the gravitational force are $-\partial\Phi/\partial x = -\frac{1}{X}\partial\Phi/\partial a$ *et cycl.*, so that the equations of motion become

$$-\frac{1}{X}\frac{d^2X}{dt^2} = 2\pi G\rho\alpha \ ; \quad -\frac{1}{Y}\frac{d^2Y}{dt^2} = 2\pi G\rho\beta \ ; \quad -\frac{1}{Z}\frac{d^2Z}{dt^2} = 2\pi G\rho\gamma \tag{26.6}$$

where the functions α, β, and γ are defined as

$$\alpha = abc\int_0^\infty \frac{ds}{(a^2+s)\Delta} \ \text{ et cycl.} \tag{26.7}$$

$$\Delta^2 \equiv (a^2+s)(b^2+s)(c^2+s) \tag{26.8}$$

Here a, b, and c are identified with the axes of the ellipsoid.

Now comes a crucial observation: without loss of generality, one can order the axes according to $a > b > c$, in which case $\alpha < \beta < \gamma$, so that

$$-\frac{1}{X}\frac{d^2X}{dt^2} < -\frac{1}{Y}\frac{d^2Y}{dt^2} < -\frac{1}{Z}\frac{d^2Z}{dt^2} \tag{26.9}$$

Consequently, the axial ratios $a : b : c$ always increase with time, and *slight initial asphericities are amplified during the collapse.* This secular increase of aspherical perturbations provides an explanation for the pancake-like, and later filamentary, appearance of megaparsec structures. Note also that, for the contraction described, the velocities inside the ellipsoid are linear functions of position: *the collapse produces a Hubble-type velocity field.*

In order to avoid nonlinearities and other complications in high-density regions, we may view the development of structure in a selfgravitating pressure-free medium by considering the evolution of the *low*-density regions. These are the progenitors of the observed voids. The arguments presented above can still be applied, except that the sense of the final effect is reversed: because a void is effectively a region of negative density in a uniform background, the voids expand as the overdense regions collapse, while *slight asphericities decrease as the voids become larger* ("Bubble Theorem", Icke 1984). The proof holds strictly only on a non-expanding background, though this should be no objection for structures which are much smaller than the particle horizon. Because $|\delta\rho/\rho|$ does not exceed unity in a void, the approximation will remain good for a longer period, except, of course, near the outer parts of the voids, where the matter gets swept up.

According to the above, taking voids as the dominant dynamical component of the Universe, one may think of the megaparsec structure as a close packing of spheres of different sizes, out of which matter flows in a slightly super-Hubble expansion towards the interstices of the spheres. Thus, *the importance of the Bubble Theorem is that it provides a specific physical mechanism for producing the non-Poissonian matter distribution in the large scale Universe.*

27. Voronoi Foam

Underdensities become more and more spherical when they expand, which they do a little faster than the Hubble expansion. Consequently, they form convex voids, from which the mass flows towards the zones separating the low-density regions. The asymptotic shape of such a mass distribution is a Voronoi foam.

Continuing the above argument, I can construct the "skeleton" of the mass distribution by considering the locus of points towards which the matter streams out of the voids. Suppose that some cosmic process produces a collection of regions where the density is slightly less than average (the origin and statistical properties of the requisite fluctuations is a very important unsolved problem). As we have seen, these regions are the seeds of the voids, because underdense patches become expansion centres, from which matter flows away until it encounters similar material flowing out of an adjacent void. Making the approximation that the excess Hubble parameter is the same in all voids, the matter must collect on planes that perpendicularly bisect the axes connecting the expansion centres.

For any given set of expansion centres, or `nuclei`, the arrangement of these planes define a unique process for the partitioning of space, a `Voronoi tessellation`. A particular realisation of this process (i.e. a specific subdivision of N-space according to the Voronoi tessellation) may be called a `Voronoi foam`. In three dimensions a Voronoi foam consists of a packing of Voronoi cells, each cell being a convex polyhedron enclosed by the bisecting planes between the nuclei and their neighbours. A Voronoi foam consists of four geometrically distinct elements: the polyhedral cells (`voids`), their walls (`pancakes`), edges (`filaments`) where three walls intersect, and nodes (`clusters`) where four filaments come together.

In the cosmological context, each Voronoi cell is a void. The planes are identified with the "walls" in the galaxy distribution, the filaments are identified with the elongated "super"clusters, and the vertices correspond to the virialised `Abell clusters`.

We have constructed two– and three-dimensional Voronoi foams geometrically. The example shows the characteristic appearance of two-dimensional Voronoi foams; the similarity with numerical simulations of gravitational clustering of collisionless particles is indeed quite striking. In three dimensions, the similarity is even more remarkable, in projection as well as in a slice.

The advantage of using these geometrically constructed models is that one is not restricted by the resolution or number of particles. A cellular structure can be generated over a part of space beyond the reach of any N-body experiment. This makes the Voronoi model particularly suited for studying the properties of galaxy clustering in cellular structures on very large scales, for example in very deep pencil beam surveys, and for studying the clustering of clusters in these models.

The Voronoi foams are expected to give a good asymptotic description of the matter distribution at late times in several physical models. Models to which the Voronoi foams apply can be either gravitational instability models in which the structure formation is dominated by the negative density fluctuations, as in models with fluctuation spectra having a high-frequency cutoff or considerable power at low wavelengths; or models in which the driving force is due to explosions. The Voronoi foams outline the "skeleton" of the mass distribution, around which matter assembles during the evolution of the Universe.

28. Growth of Voronoi Features

It is found that the mass in the Voronoi features evolves according to a simple exact law.

As with other models, the key problem is: given an initial fluctuation small enough to keep the CMBR as smooth as 10^{-5}, does the amplitude of the consequent structure become large enough by today?

In the Voronoi model, which focuses on the key role of the evolution of the voids, it is easy to calculate the relative growth of the other features (walls, filaments, and nodes). Let m_v, m_w, m_f, m_n be the mass in voids, walls, filaments, and nodes, respectively. The first three features lose mass in a Voronoi "cascade":

$$\text{void} \Longrightarrow \text{wall} \Longrightarrow \text{filament} \Longrightarrow \text{node}$$

Each feature gains mass from the one immediately above it in the hierarchy. The mass loss term in N dimensions has the form

$$\text{mass change} = -NmH^* \, dt \tag{28.1}$$

where H^* is the excess Hubble expansion and dt is the time increment.

In the quasi-linear case, $H^* \propto t^{-1/3}$, because the excess velocity increases as $H^*ax = v \propto t^{1/3}$, and in the Einstein-De Sitter case $a \propto t^{2/3}$. The power law dependence means that one may absorb H^* into the time by defining $d\theta = H^*dt$. Thus,

$$\theta \propto t^{2/3} \tag{28.2}$$

The mass loss in N dimensions is then

$$dm = -Nm\,d\theta \tag{28.3}$$

The effective excess Hubble parameter H^* is *the same* in all topological features. This is one of the cute properties of the Voronoi model. Let N be a node, let P be the point where a `Delaunay line` (that is a line connecting two nuclei, perpendicular to a wall) intersects a wall, let W be a point in the wall, and let the angle WNP be called θ. If we indicate the distance NP by a, the distance PW by b, and NW by c, then the velocity at W is cH^*. The components perpendicular and perpendicular to the wall are

$$cH^* \cos\theta = aH^* \tag{28.4}$$

$$cH^* \sin\theta = bH^* \tag{28.5}$$

Thus, *the excess velocity in any Voronoi feature is simply found by multiplying H^* with the length along the feature.* This allows us to use the above formula for $N = 3$, 2, 1, and 0.

The mass gain is found simply by reversing the sign of the loss term of the feature higher in the hierarchy. This gives the following equations:

$$\frac{dm_v}{d\theta} = \qquad - 3m_v \tag{28.6}$$

$$\frac{dm_w}{d\theta} = 3m_v - 2m_w \tag{28.7}$$

$$\frac{dm_f}{d\theta} = 2m_w - m_f \tag{28.8}$$

$$\frac{dm_n}{d\theta} = m_f \tag{28.9}$$

These equations are simply solved by noting that the N-dimensional mass loss equation has a solution of the type $\psi\exp(-N\theta)$:

$$m_v = \mathrm{e}^{-3\theta} \tag{28.10}$$

$$m_w = 3\mathrm{e}^{-2\theta}(1 - \mathrm{e}^{-\theta}) \tag{28.11}$$

$$m_f = 3\mathrm{e}^{-\theta}(1 - \mathrm{e}^{-\theta})^2 \tag{28.12}$$

$$m_n = (1 - \mathrm{e}^{-\theta})^3 \tag{28.13}$$

If we call $m_h = m_w + m_f + m_n$ the total mass of the high density regions, while $m_l = m_v + m_w + m_f$ is the mass in low density features, I find that

$$m_h/m_v = \exp(3\theta^{2/3}) - 1 \tag{28.14}$$

$$m_l/m_n = \Big(1 - \exp(-\theta^{2/3})\Big)^{-3} - 1 \tag{28.15}$$

The scaling of the time is given by the initial amplitude. Both mass ratios can be compared with published galaxy surveys. This gives us a way to determine what time it is in the simulations and in the Universe.

29. What Time is it?

The time scale of the Voronoi growth can be related to the initial amplitude of the perturbations.

In the quasi-linear case, $H^* \propto t^{-1/3}$, because the excess velocity increases as $H^* ax = v \propto t^{1/3}$, and in the Einstein-De Sitter case $a \propto t^{2/3}$. The power law dependence means that one may absorb H^* into the time by defining $d\theta = H^* dt$. Consequently,

$$\theta = \theta_r (t/t_r)^{2/3} \tag{29.1}$$

where t_r is the time at which the Universe becomes transparent to radiation. The constant θ_r is related to the amplitude δ_r of the voids at decoupling; one readily finds that

$$3\theta_r = \delta_r \tag{29.2}$$

The amount dm of mass lost in a dimensionless time interval $d\theta$, in N dimensions, is then

$$dm = -Nm\, d\theta \tag{29.3}$$

These equations allow one to relate the dimensionless time parameter θ to the cosmic time and to the redshift z at decoupling:

$$\theta = \frac{1}{3}\delta_r (t_0/t_r)^{2/3} (t/t_0)^{2/3} = \frac{1+z}{3}\, \delta_r (t/t_0)^{2/3} = \frac{1+z}{3}\, \delta_r \tau^{2/3} \tag{29.4}$$

The scaling of the time is given by the initial amplitude. The mass ratios derived in the previous section can be compared with published galaxy surveys. This gives us a way to determine what time it is in the simulations and in the Universe. When one determines the best fitting values, one finds $\theta \approx 1$ (note that this is in reasonable agreement with age determinations from the two-point correlation of galaxies in the kinematic Voronoi model). That means that the initial amplitude must be of the order of 2×10^{-3}. The Sachs-Wolfe effect (i.e. the gravitational redshift incurred by photons which must climb out of a potential well in which they were formed) is then about 70 times larger than that which is allowed by CMBR observations. But COBE averages over a field of 7 degrees, do maybe the above isn't that bad: such a field contains many Voronoi cells, the fluctuations of which will average out.

Numerical calculations of the fate of CMBR photons on their way through an evolving Voronoi universe show, that the Rees-Sciama effect (i.e. the growth of $\delta T/T$ due to the fact that photons falling into an evolving cosmic potential well see a deeper well when they climb out) is on the order of a few times 10^{-6}, which is allowed by the CMBR observations.

Apparently, there is not really enough time to make the contrast between voids and the rest, or between nodes (=Abell clusters) and the rest, as large as seems necessary, unless some special pleading is invoked; e.g. observational selection effects, or a particular dependence of the M/L ratio on the local mass density (called `biasing`).

30. The Gravitational Lagrangian

We revisit the derivation of the Einstein Equation, by looking at it from the perspective of local Lorentz symmetry.

The Yang-Mills method creates fields from local symmetry. Start with a guessed global symmetry, then note that using this symmetry globally is contrary to the spirit of relativity.

But it turns out that using a *local* symmetry is not possible unless one inserts a new field to counteract the "mismatch" caused by locality. The quanta of this mismatch field transmit a force. In this way, a local symmetry of a basic multiplet of fermions produces bosons that couple to the fermions in a way that is prescribed by the symmetry.

Suppose that our mechanical system is described by a Lagrangian $\mathcal{L}$, which is a function over spacetime $\{x_\mu\}$ of a generalized coordinate vector q and its corresponding momentum $q_{,\mu}$ (we use the abbreviation $q_{,\mu} \equiv \partial q/\partial x^\mu$). The action corresponding to this is found by integrating the Lagrangian density over an arbitrary four-volume Ω:

$$\mathcal{S} = \int_\Omega \mathcal{L}(q, q_{,\mu}) dx_\mu \tag{30.1}$$

where the dynamical variables q and $q_{,\mu}$ are to be seen as functions of x_μ. The requirement $\delta\mathcal{S} = 0$ implies that

$$0 = \int_\Omega \left(\frac{\partial \mathcal{L}}{\partial q}\delta q + \frac{\partial \mathcal{L}}{\partial q_{,\mu}}\delta q_{,\mu} \right) dx_\nu = \frac{\partial \mathcal{L}}{\partial q_{,\mu}}\delta q \bigg|_\Omega + \int_\Omega \left(\frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dx_\mu}\frac{\partial \mathcal{L}}{\partial q_{,\mu}} \right) \delta q \, dx_\nu \tag{30.2}$$

from which the Lagrangian equations of motion follow directly because Ω is arbitrary. Note that we are allowed to subject $\mathcal{L}$ to the same symmetry under which q is supposed to be symmetric, because $\mathcal{L} = f(q, q_{,\mu})$.

Now let us change q infinitesimally by some symmetry $\mathbf{L}$. A global symmetry does not change the equations of motion, because $\mathbf{L}$ commutes with δ. However, if $\mathbf{L}$ is a *local* symmetry, then $\mathbf{L} = \mathbf{L}(x_\mu)$, and therefore

$$q \to q + \delta q \quad \text{and} \quad \delta q = \epsilon(x_\mu) q \tag{30.3}$$

It follows immediately that if $q_{,\mu} \to q_{,\mu} + \delta q_{,\mu}$, we get

$$\delta q_{,\mu} = \epsilon_{,\mu} q + \epsilon(x_\mu) q_{,\mu} \tag{30.4}$$

so that the integrand of $\delta\mathcal{S}$ becomes

$$\delta\mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial q} q + \frac{\partial \mathcal{L}}{\partial q_{,\mu}} q_{,\mu} \right) \epsilon + \frac{\partial \mathcal{L}}{\partial q_{,\mu}} \epsilon_{,\mu} q \tag{30.5}$$

The term in brackets drops out because of the equations of motion, and we conclude that $\delta\mathcal{L} \neq 0$ because of the derivative $\epsilon_{,\mu}$: the *local* character of the transformation $\mathbf{L}$ spoils the proper extremum behaviour of $\mathcal{L}$, and no good equations of motion result!

In other words, the fact that $\mathbf{L}$ changes from event to event in spacetime produces a *mismatch* between $\mathbf{L}\mathcal{L}$ at one event and the $\mathbf{L}\mathcal{L}$ elsewhere. The key idea now is, to patch this up by adding extra terms to the Lagrangian to correct the mismatch. It is by no means obvious that this can be done successfully!

Because the culprit is a vector $\epsilon_{,\mu}$, we try to patch up $\mathcal{L}$ by adding a vector field to it. For the moment, let us call this field A', and the corresponding Lagrangian is

$$\mathcal{L}' = \mathcal{L}'(q, q_{,\mu}, A') \tag{30.6}$$

of which we will now rigorously require that $\delta\mathcal{L}' = 0$. This requirement prescribes a functional dependence of $\mathcal{L}'$ on its arguments, as follows. The δq and $\delta q_{,\mu}$ are found as before; the variation $\delta A'$ is, of course, a linear combination of ϵ and $\epsilon_{,\mu}$ (for infinitesimal transformations). The most general form for $\delta A'$ is then

$$\delta A' = U A' \epsilon + C^\mu \epsilon_{,\mu} \tag{30.7}$$

with constant scalar U and vector C^μ, to be determined afterwards. To get a proper equation of motion from the patched-up Lagrangian, we require

$$\delta\mathcal{L}' = \frac{\partial \mathcal{L}'}{\partial q}\delta q + \frac{\partial \mathcal{L}'}{\partial q_{,\mu}}\delta q_{,\mu} + \frac{\partial \mathcal{L}'}{\partial A'}\delta A' = 0 \tag{30.8}$$

Inserting the expressions for δq and so forth yields a linear equation in ϵ and $\epsilon_{,\mu}$. Because the magnitude of ϵ is arbitrary (provided it is infinitesimal), each coefficient of ϵ and $\epsilon_{,\mu}$ must vanish independently. This gives

$$\frac{\partial \mathcal{L}'}{\partial q} q + \frac{\partial \mathcal{L}'}{\partial q_{,\mu}} q_{,\mu} + \frac{\partial \mathcal{L}'}{\partial A'} U A' = 0 \tag{30.9}$$

$$\frac{\partial \mathcal{L}'}{\partial q_{,\mu}} q + \frac{\partial \mathcal{L}'}{\partial A'} C^{\mu} = 0 \tag{30.10}$$

with the consistency requirement that

$$C^{\mu} C_{\mu} = 1 \tag{30.11}$$

The latter means that C_μ has an inverse; if it did not, then some of the above equations would be linearly dependent and the system could not be solved. Now define the vector field A_μ as

$$A_\mu \equiv C_\mu A' \tag{30.12}$$

to find that

$$\frac{\partial \mathcal{L}'}{\partial q_{,\mu}} q + \frac{\partial \mathcal{L}'}{\partial A_\mu} = 0 \tag{30.13}$$

The remarkable thing is, that *this equation is in fact a prescription for the way in which the Lagrangian must depend on its arguments.* We see directly that it requires that the vector field A_μ, which was introduced to patch up the mismatch created by the locality of $\mathbf{L}$ (i.e. the dependency $\mathbf{L} = f(x_\mu)$) occurs in $\mathcal{L}'$ *only* through the combination

$$q_{,\mu} - q A_\mu \tag{30.14}$$

If we now define the shorthand notation

$$q_{;\mu} \equiv q_{,\mu} - q A_\mu \tag{30.15}$$

the `covariant derivative` of q, we see that the form $\mathcal{L}' = \mathcal{L}'(q, q_{,\mu}, A')$ allows us only one way to insert $q_{;\mu}$ into $\mathcal{L}'$, namely in exactly the same way as $\mathcal{L}$ depends on $q_{,\mu}$. This must be so because $q_{;\mu}$ contains a term that is linear in $q_{,\mu}$, and another term that can be made zero by letting $\mathbf{L}$ equal the identity. Thus we get

$$\mathcal{L}' = \mathcal{L}(q, q_{;\mu}) \tag{30.16}$$

and from now on we use this form.

Note that in the covariant derivative the local symmetry prescribes that q and A_μ couple by means of the product qA_μ; in quantum electrodynamics this appears in the form where the dynamical variables q and $q_{,\mu}$ are replaced by the derivative ∂ and a constant factor ie:

$$q_{,\mu} - q A_\mu \quad \Longrightarrow \quad \partial - ieA \quad \Longrightarrow \quad i\psi^*(\gamma \cdot \partial - ie\gamma \cdot A)\psi \tag{30.17}$$

which is the famous `minimal coupling term` in the Dirac equation (the γ's are Dirac matrices).

Having now found that there is only one functional form of the Lagrangian which allows us to patch up the mismatch due to the local symmetry, it remains to determine the constants U and C^μ. First, we note that

$$\delta A_\mu = C_\mu \delta A' = C_\mu U \epsilon(x_\alpha) A' + \epsilon_{,\mu} = C_\mu C^\nu U \epsilon A_\nu + \epsilon_{,\mu} \tag{30.18}$$

Second, we recall the expressions for the variations $\delta\mathcal{L}$ and $\delta\mathcal{L}'$, which lead directly to

$$\frac{\partial \mathcal{L}'}{\partial q} = \frac{\partial \mathcal{L}}{\partial q}\bigg|_{q_{;\mu}} - \frac{\partial \mathcal{L}}{\partial q_{;\mu}}\bigg|_{q} A_\mu \tag{30.19}$$

$$\frac{\partial \mathcal{L}'}{\partial q_{,\mu}} = \frac{\partial \mathcal{L}}{\partial q_{;\mu}}\bigg|_{q} \tag{30.20}$$

$$\frac{\partial \mathcal{L}'}{\partial A'} = -\frac{\partial \mathcal{L}}{\partial q_{;\nu}}\bigg|_{q} C_\nu q \tag{30.21}$$

Inserting these into the equation resulting from $\delta\mathcal{L}' = 0$, we find

$$\left(\frac{\partial\mathcal{L}}{\partial q}q + \frac{\partial\mathcal{L}}{\partial q_{;\mu}}q_{;\mu}\right) - \frac{\partial\mathcal{L}}{\partial q_{;\nu}}qUA_\nu = 0 \tag{30.22}$$

The term in brackets vanishes because of the equation of motion for $\mathcal{L}$, and because we had $\mathcal{L}' = \mathcal{L}(q, q_{,\mu})$. It follows immediately that

$$U = 0 \tag{30.23}$$

The definition of A_μ then gives, by means of the expression for $\delta A'$, that

$$\delta A_\mu = \epsilon_{,\mu} \tag{30.24}$$

This demonstrates quite clearly how the vector field A_μ comes in because of the *local* character of the symmetry: if **L** were independent of x_α, we would have $\epsilon_{,\mu} = 0$!

One further point remains to be settled. By patching up the Lagrangian, we have let a genie out of a bottle, namely the field A_μ. We are now obliged to take this field seriously, and to identify it with an actual particle. In that case, we must of course allow A_μ to occur in the Lagrangian as a free field (i.e. as more than just an entity which couples to the q-field by qA_μ). It may be a trifle much to ask, but can the locality of **L** prescribe the form of the occurrence of this free field too?

The patch-up vector field A_μ may occur itself in the Lagrangian as a dynamical variable, together with its spacetime derivative $A_{\mu,\nu}$, in the same way that we had a dependence on the dynamical variables of the q-field. Because $\mathcal{L}'$ is linear, we can insert extra terms by simple addition, so we can restrict ourselves to finding the sub-part $\mathcal{L}'$ that depends on the A's only, and then add it to what we had found already (note that the coupling term has already been disposed of!) We use the same variational form:

$$\delta\mathcal{L}' = \frac{\partial\mathcal{L}'}{\partial A_\mu}\delta A_\mu + \frac{\partial\mathcal{L}'}{\partial A_{\mu,\nu}}\delta A_{\mu,\nu} \tag{30.25}$$

As usual, we insert δA_μ, and require that each coefficient of ϵ and $\epsilon_{,\mu}$ vanish independently. This yields the equations

$$\frac{\partial\mathcal{L}'}{\partial A_\mu} = 0 \tag{30.26}$$

$$\frac{\partial\mathcal{L}'}{\partial A_{\mu,\nu}} + \frac{\partial\mathcal{L}'}{\partial A_{\nu,\mu}} = 0 \tag{30.27}$$

Accordingly, we find that the patch-up field *cannot itself occur in the Lagrangian.* Consequently, the field A_μ *is not an observable;* but it *can* couple to the dynamical variable q by means of the term qA_μ. The above shows that the new field can occur in $\mathcal{L}$ only through the combination

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu} \tag{30.28}$$

That is to say, *the curl of the field is an observable!* This should of course look very familiar to aficionados of Maxwell's Equations.

This completes the demonstration that the requirement of local symmetry of the Lagrangian is so severe, that not only the way in which the patch-up field A_μ couples to the q-field, but also the way in which it must occur in the Lagrangian is prescribed entirely!

It can be shown that the same kind of construction works for vector fields q^β (in fact, this is what the original Yang-Mills paper was all about). In that case, it can be shown that the strictness and cleanness with which the form of the Lagrangian is prescribed is due to the group structure of the symmetry.

We have four such cases in Nature:

(1) the case of a `phase-rotation symmetry` $U(1)$, (i.e. the multiplication with a complex scalar function as treated above), which produces electromagnetism;
(2) the case of the `isospin symmetry` $SU(2)$ (i.e. multiplication with a factor derived from a 2×2 symmetry via $\exp(\frac{1}{2} i g \tau \cdot \omega)$, where τ are Pauli matrices and ω is an arbitrary smooth function over spacetime), which produces the weak interaction;
(3) the group $SU(3)$, leading to the `colour interaction`;
(4) Lorentz symmetry, which gives rise to the gravitational interaction (General Relativity).

In the Yang-Mills case, the group has nonzero structure constants f^a_{bc}, and following exactly the same line of resoning one may show that the "wrinkle" or "patch-up" field A can occur in the Lagrangian only through the combination

$$F^a_{\mu\nu} = A^a_{\nu,\mu} - A^a_{\mu,\nu} - \frac{1}{2} f^a_{bc} \left(A^b_\mu A^c_\nu - A^b_\nu A^c_\mu \right) \tag{30.29}$$

which clearly shows the occurrence of nonlinear terms due to the non-Abelian character of the group.

If the local symmetry is Lorentz symmetry, one may show in precisely the same way – though with much more effort – that the patch-up fields (which in gravity theory are traditionally called Γ instead if A) can occur in the Lagrangian only through the combination

$$R^\lambda_{\mu\nu\kappa} \equiv \Gamma^\lambda_{\mu\nu,\kappa} - \Gamma^\lambda_{\mu\kappa,\nu} + \Gamma^\eta_{\mu\nu} \Gamma^\lambda_{\kappa\eta} - \Gamma^\eta_{\mu\kappa} \Gamma^\lambda_{\nu\eta} \tag{30.30}$$

which is the Riemann-Christoffel tensor.

31. Spontaneous Symmetry Breaking

As a preparation for the section on inflation, I consider the possible influence of the zero-point energy of a scalar field.

A second key concept for particles is `spontaneous symmetry breaking`, which comes about when the solution of a scalar field equation does not exhibit the manifest symmetry of its Lagrangian. We will only sketch the Goldstone model here. Let ϕ be a complex scalar field with components (ϕ_1, ϕ_2):

$$\phi = \frac{1}{\sqrt{2}} (\phi_1 + i \phi_2) \tag{31.1}$$

The Lagrangian density is then

$$\mathcal{L} = \partial_\alpha \phi^\dagger \partial^\alpha \phi - V(\phi) \tag{31.2}$$

where ∂_α is an abbreviation for $\partial / \partial x^\alpha$. Suppose now that the potential has the wine-bottle form

$$V = \frac{1}{2} \lambda^2 |\phi|^4 + \frac{1}{2} \mu^2 |\phi|^2 \tag{31.3}$$

Because the potential depends only on the absolute value of the field ϕ, it is invariant under global phase transformations (i.e. rotations of the axes in the complex plane over a constant angle ω):

$$\phi \to \exp^{i\omega} \phi \tag{31.4}$$

The term with a fourth power is supposed to come about by a self-interaction of the field. If μ^2 were positive, it would be an ordinary mass term. But is it is negative something interesting happens. Then V has two local minima: one at the point $\phi = 0$ (an unstable maximum) and a stable one on the ring

$$|\phi|^2 = -\frac{\mu^2}{2\lambda^2} \tag{31.5}$$

On that ring, the vacuum expectation value of the field is obviously not zero, but has the value

$$|\langle 0|\phi|0\rangle|^2 = \frac{\mu^2}{2\lambda^2} \tag{31.6}$$

The phase is arbitrary, provided that we choose it to be the same in all of the vacuum (this is called "fixing a gauge"). Let us choose the phases such that one component of the field vanishes:

$$\langle 0|\phi_1|0\rangle = \frac{\mu}{\lambda} \tag{31.7}$$

$$\langle 0|\phi_2|0\rangle = 0 \tag{31.8}$$

This means that the vacuum, represented by the state $|0\rangle$, is not invariant under the phase rotation. The vacuum solution does not exhibit the same symmetry as its potential V. Analogy: the Lagrangian of a point mass is spherically symmetric, but a planetary orbit lies in a plane. In the above, we see that the vacuum potential can give mass to some particles ϕ_1 (but it creates massless ϕ_2 too, the `Goldstone boson`).

32. Inflation

If the Universe has a non-zero cosmological constant, its motion in the limit for long times will tend towards an exponential expansion, first discovered by De Sitter. Nowadays, such a possible phase in the evolution of the Universe is called inflation.

The equations of motion in the general case are

$$3\frac{d^2a}{dt^2} = -4\pi G\left(\rho + \frac{3P}{c^2}\right)a + \Lambda a \tag{32.1}$$

$$a\frac{d^2a}{dt^2} + 2\left(\frac{da}{dt}\right)^2 + 2kc^2 = 4\pi G\left(\rho - \frac{P}{c^2}\right)a^2 + \Lambda a^2 \tag{32.2}$$

which can be integrated once to give

$$\left(\frac{da}{dt}\right)^2 = \frac{8\pi}{3}G\rho a^2 - kc^2 + \frac{\Lambda}{3}a^2 \tag{32.3}$$

$$\frac{d}{dt}\left(\rho a^3\right) + \frac{P}{c^2}\frac{da^3}{dt} = 0 \tag{32.4}$$

It is instructive to compare Eq.(32.3,4) with the equation of motion in the nonrelativistic case. First Eq.(32.3) , which is identical with the exception of the cosmological constant Λ. because Eq.(32.3) is in effect an energy equation, we see that Λ corresponds to a boundary term in the potential energy. In the Newtonian case, the Poisson equation for the potential has a solution Φ up to an additive constant which corresponds to the boundary 'at infinity'. But in a homogeneous universe thios term does not disappear, as it does for the potential of finite mass distributions.

In the relativistic case, Λ comes in as the mass-energy density of the vacuum, as can be seen immediately from the fact that Λ occurs as a multiplier of $g_{\mu\nu}$. Lorentz invariance requires the appearance of a pressure term whenever a population of particles appears in the vacuum. The energy-momentum tensor of such a population is found by averaging over all of phase space. One could use a diagonal $(\rho, 0, 0, 0)$ for any one particle, but not for an ensemble.

Second, consider Eq.(32.4) , which is more easily recognized in its difference form

$$d(\rho c^2 a^3) + P\,da^3 \tag{32.5}$$

We see straight away that this is the difference form of the relativistic energy equation,

$$E = mc^2 + E_{\text{kin}} \tag{32.6}$$

but then in the form of an energy density, that is to say the energy in a volume V:

$$E = \frac{m}{V}c^2\, V + \frac{E_{\text{kin}}}{V}\, V = \rho c^2 a^3 + Pa^3 \tag{32.7}$$

In the guise of the first law of thermodynamics, one gets a term $P\,dV$, which can be observed in Eq.(32.4,5) .

In the case of a dust equation of state, one has $P = 0$, and the equations of motion become

$$\rho a^3 = \text{constant} = \frac{3}{4\pi}M \tag{32.8}$$

$$\left(\frac{da}{dt}\right)^2 = \frac{2GM}{a} - kc^2 + \frac{\Lambda}{3}a^2 \tag{32.9}$$

If the cosmological constant Λ is positive, and if the initial mass density is not so large that the first two terms on the right hand side make the Universe recollapse at early times, then the Λ term may overcome the others, due to the growth of a. This then gives a model in which

$$\left(\frac{da}{dt}\right)^2 = \frac{\Lambda}{3}a^2 \tag{32.10}$$

This `De Sitter model` obviously has an exponentially expanding solution:

$$a = \exp t\sqrt{\Lambda/3} \tag{32.11}$$

At present, the observational evidence is that $\Lambda = 0$ to very high precision (this actually presents a problem, which won't be discussed here). But there may have been early phases in the Universe when Λ was not zero because of a difference in energy between the expectation value for the vacuum energy and its actually realized value. That is to say, gauge theories of elementary interactions start with a quantum Lagrangian which is manifestly symmetric under some group, but the solutions of the equations from this Lagrangian need not be so symmetric. This is called `spontaneous symmetry breaking`. We see that sort of thing all over the place: for example, the Lagrangian of a point mass is clearly spherically symmetric, and yet a planetary orbit is not a sphere at all, but lies in a plane. It isn't even a circle centered on the point mass: it is an ellipse, with the central mass in a focus!

But we expect the symmetric Lagrangian to have a physical meaning. In quantum terms, it means that the expectation value of the energy-momentum tensor $T_{\mu\nu}$ must be zero in a symmetric vacuum. The vacuum which is physically realized has a *nonzero* expectation value, i.e. there are particles present due to the spontaneous symmetry breaking. The energy of the state with particles can be lower than the state without (analogy: a rain cloud without raindrops has a higher energy than a cloud where drops have condensed out).

Because the potential V is presumed to be due to a *scalar* field ϕ, the $T_{\mu\nu}$ is proportional to the metric tensor:

$$< T_{\mu\nu} >= g_{\mu\nu} < V(\phi) > \tag{32.12}$$

On the other hand, the general expression for $T_{\mu\nu}$ at zero four-velocity has ρ and P/c^2 on the main diagonal, so that we must conclude that there is a cosmological constant and

$$\Lambda = V(\phi) \tag{32.13}$$

$$T_{\mu\nu} = \Lambda g_{\mu\nu} \tag{32.14}$$

$$\rho = \Lambda \tag{32.15}$$

$$\frac{P}{c^2} = -\Lambda \tag{32.16}$$

Thus, there is an effective `inflation`, because negative pressure in an expanding system means that the expansion actually generates energy! Apparently, the transition of the Universe through a scalar field freeze-out can generate an exponential runaway, as long as the expectation values of the actual space and the vacuum differ.

Note that this rather peculiar and counter-intuitive behaviour ('you get something out of nothing') is due to our initial assumption for the connection between $T_{\mu\nu}$ and the (thermo)dynamic quantities. As we emphasised there, the connection between matter and spacetime curvature is still a conjecture, since we do not have a quantum gravity theory. So, maybe all this will have to be drastically revised in the future.

In the Goldstone model, the potential has the wine-bottle form

$$V = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{32.17}$$

Here $\mu^2 < 0$, otherwise it would be a plain mass term. Now $v/\sqrt{2}$ is the real vacuum expectation value of ϕ, so that we get a particle at the location

$$V(\phi) = \mu^2 v^2/4 \tag{32.18}$$

with the properties

$$m_\phi^2 = -2\mu^2 \tag{32.19}$$

$$v^2 = 4\left(\frac{M_W}{g}\right)^2 = \frac{1}{G_F\sqrt{2}} \tag{32.20}$$

where M_ϕ is the Higgs mass, M_W is the weak vector boson mass, g is the electroweak coupling constant, and G_F is the Fermi coupling constant. In units where $\hbar = c = 1$, we get

$$\Lambda = -\left(\frac{\pi G}{G_F\sqrt{2}}\right) m_\phi^2 \tag{32.21}$$

The inflation epoch can in principle account for the homogeneity and isotropy of the Universe and for its causal connectedness. It also requires $\Omega = 1$ to a very high order of precision. Some say that is nice, others find it a worry that the observed value of Ω is about ten times less than that.

33. Bias

Cosmic structures made of dark and of luminous (baryonic) matter should be commensurate, because gravity always produces the same acceleration, being a geometric 'force'. If dark matter – as appears to be the case – is less clumped on small scales than baryonic matter, there must be a physical mechanism that causes this effect. The difference in the distributions is called *bias*.

Steepening of smaller features occurs because of two concurrent effects. First, Poisson's Equation

$$\Delta\Phi = 4\pi G\rho \tag{33.1}$$

shows that mass is more sharply peaked than its potential. This is obvious from the Fourier transform of this equation:

$$k^2 \delta\Phi(k) = 4\pi G \delta\rho(k) \tag{33.2}$$

Consequently, in the equation of motion (which has a force proportional to $\vec{\nabla}\Phi$), sharp features grow faster than fuzzy ones. Second, the equation for the energy in the cosmic gas flow reads

$$\frac{\partial e}{\partial t} + \vec{\nabla} \cdot (\vec{v}(e + P)) = -\vec{v} \cdot \vec{\nabla}\Phi + \Gamma_{\rm rad} - \lambda_{\rm rad} \tag{33.3}$$

The radiative gain term Γ is due all photon-genetaring mechanisms. The loss term λ can in most cases be written as

$$\lambda_{\rm rad} = \Lambda(T)\,\rho^2 \qquad (33.4)$$

Thus, the cooling occurs preferentially in the densest regions. That implies that sharp features are most dissipative. These two effects in concert lead us to expect

$$\rho_{\rm dark}(k) \otimes B(t,k,\rho) \to \rho_{\rm bary}(k) \qquad (33.5)$$

where the biasing function B is due to: the cooling properties (e.g. by LiH); the density distribution; and the local thermal and dynamic history. Therefore, B can be quite complicated and is sure to be very nonlinear in all its arguments.

34. References

For more details, you should consult the literature, in particular the books by Achterberg, Coles, Narlikar, Padmanabhan, Peacock, Peebles or Weinberg.

Achterberg, A., *Van Oerknal via niets tot straling en stof*, Epsilon Uitgaven, Utrecht 1994

Barrow, J.D., Bhavsar, S.P., & Sonoda, D.H., *Minimal spanning tree, filaments and galaxy clustering*, 1985 *Monthly Notices Roy. Astron. Soc.* **216**, 17-35

Bertotti, B., Balbinot, R., Bergia, S., & Messina, A. (Eds.), *Modern cosmology in retrospect*, Cambridge University Press, Cambridge 1990

Centrella, J., & Melott, A.L., *Three-dimensional simulation of large scale structure in the universe*, 1983 *Nature* **305**, 196-198

Coles, P., & Ellis, G., *Is the Universe open or closed?*, Cambridge Univ. Press, Cambridge 1997

Coles, P., & Lucchin, F., *Cosmology, the origin and evolution of cosmic structure*, 2nd Ed., Wiley, Chichester 2002

Efstathiou, G ., & Silk, J., *The formation of galaxies*, 1983 *Fundam. Cosmic Phys.* **9**, 1-138

Einstein, A., *The meaning of relativity*, Princeton University Press, Princeton 1956

Feynman, R.P., Leighton, R.B., & Sands, M., *The Feynman lectures on physics*, Addison-Wesley, Redwood City 1977

Frenk, C.S., Ellis, R.S., Shanks, T., Heavens, A., & Peacock, J.A. (Eds.), *The epoch of galaxy formation*, NATO ASI Series C/264, Kluwer, Dordrecht 1989

Geroch, R., *General relativity from A to B*, University Chicago Press, Chicago 1978

Harrison, E.R., *Standard model of the early universe*, 1973 *Ann. Rev. Astron. Astrophys.* **11**, 155-186

Harrison, E.R., *Cosmology*, Cambridge Univ. Press, second edition, Cambridge 2000

Hawking, S.W., & Ellis, G.F.R., *The large scale structure of space-time*, Cambridge University Press, Cambridge 1973

Icke, V., *Voids and filaments*, 1984 *Monthly Notices Roy. Astron. Soc.* **206**, 1P-3P

Icke, V., & Van de Weygaert, R., *The galaxy distribution as a Voronoi foam*, 1991 *Quart. J. Roy. astron. Soc.* **32**, 85-112

Jones, B.J.T., *The origin of galaxies: a review of recent theoretical developments and their confrontation with observation*, 1976 *Revs. Mod. Phys.* **48**, 107-149

Kolb, E.W., & Turner, M.S., *The early universe*, Addison-Wesley, Redwood City 1990

Kolb, E.W., & Turner, M.S., *The early universe: reprints*, Addison-Wesley, Redwood City 1988

Linde, A.D., *Inflation and quantum cosmology*, Academic Press, Boston 1990

Linde, A.D., *Particle physics and inflationary cosmology*, Harwood Academic Publishers, Chur 1990

Martin, J.L., *General Relativity*, Prentice-Hall, London 1996

Misner, C., Thorne, K.S., & Wheeler, J.A., *Gravitation*, Freeman, San Francisco 1973

Narlikar, J.V., *Introduction to cosmology*, Cambridge University Press, Cambridge 1993

Oort, J.H., *Superclusters*, 1983 *Ann. Rev. Astron. Astrophys.* **21**, 373-428

Padmanabhan, T., *Structure formation in the universe*, Cambridge University Press, Cambridge 1993

Padmanabhan, T., *Cosmology and astrophysics through problems*, Cambridge Univ. Press, Cambridge 1996

Peacock, J., *Cosmological physics*, Cambridge University Press, Cambridge 1999

Peebles, P.J.E., *Principles of physical cosmology*, Princeton Univ. Press, Princeton 1993

Rees, M.J., *Perspectives in astrophysical cosmology*, Cambridge Univ. Press, Cambridge 1995

Rowan-Robinson, M., *The cosmological distance ladder*, Freeman, New York 1985

Shanks, T., Banday, A.J., Ellis, R.S., Frenk, C.S., & Wolfendale, A.W., *Observational tests of cosmological inflation*, Kluwer, Dordrecht 1991

Taylor, F. & Wheeler, J.A., *Spacetime physics*, Freeman, San Francisco 1966

Taylor, J.C., *Gauge theories of weak interactions*, Cambridge University Press, Cambridge 1976

Utiyama, R., *Invariant theoretical interpretation of interaction*, 1956 *Phys. Review* **101**, 1597-1607

Weinberg, S., *Gravitation and cosmology*, Wiley, New York 1972

Wilczek, F., *Whence the force F=ma. I. Culture shock*, 2004a *Physics Today*, October, p.11

Wilczek, F., *Whence the force F=ma. II. Rationalizations*, 2004b *Physics Today*, December, p.10

Physical constants for Introductory Astrophysics

Physics Today, August 1996, BG9-BG16

Quantity	**Symbol**	**Value**	**Units**
speed of light	c	299792458	m s^{-1}
vacuum permeability	μ_0	$4\pi \times 10^{-7}$	N A^{-2}
vacuum permittivity	ϵ_0	$1/\mu_0 c^2$	F m^{-1}
Newtonian gravity constant	G	6.6726×10^{-11}	$m^3 kg^{-1} s^{-2}$
Planck constant	h	6.626076×10^{-34}	J s
modified Planck constant	$\hbar$	1.0545727×10^{-34}	J s
electron charge	e	1.6021773×10^{-19}	C
proton mass	m_p	1.672623×10^{-27}	kg
neutron mass	m_n	1.674929×10^{-27}	kg
electron mass	m_e	9.1093897×10^{-31}	kg
Thomson cross section	σ_T	6.6652462×10^{-29}	m^2
Boltzmann constant	k	1.380658×10^{-23}	J K^{-1}
Stefan-Boltzmann constant	σ	5.67051×10^{-8}	W $m^{-2}K^{-4}$
solar mass	$M_\odot$	1.989×10^{30}	kg
solar luminosity	$L_\odot$	3.85×10^{26}	W
solar radius	$R_\odot$	6.96×10^8	m
mean Earth radius	$R_\oplus$	6.371×10^6	m
Earth mass	$M_\oplus$	5.98×10^{24}	kg
semimajor axis of Earth orbit	AU	1.496×10^{11}	m
reference Hubble parameter	h	100	km $s^{-1}Mpc^{-1}$
		3.24149×10^{-18}	s^{-1}
Hubble parameter	H_0	67	km $s^{-1}Mpc^{-1}$
		2.172×10^{-18}	s^{-1}
parsec	pc	3.085×10^{16}	m
year	yr	3.1558×10^7	s
light year	ly	9.4609×10^{15}	m

electron volt	eV	1.6021773×10^{-19} J
kilometre per second	1 km s^{-1}	1.023 pc Myr^{-1}

Glossary

-A-

Abell clusters	75
action	31
algemene relativiteitstheorie	49
angular size	14

-B-

bewegingsvergelijking	5
biasing	77
Boltzmann Equation	66
Boltzmann-verdeling	21
botsingstijd	24

-C-

chemical potential	20
Christoffel symbol	52
clusters	75
clusters van sterrenstelsels	6
collision term	66
colour interaction	81
continuity equation	69
covariant derivative	51,79
curvature constant	13
curvature scalar	53

-D-

deceleration parameter	11
decoupling	19
deformation tensor	2
Delaunay line	76
De Sitter model	83
distribution function	66
divergence	2
dronkemanswandeling	26

-E-

Eddington limiet	26
eenparig versnelde beweging	6
Einstein-De Sitter model	9
Einstein-De Sitter case	10
Einstein-Hilbert action	34,35
equation of motion	33
Euler's Equation	70
Euler-Lagrange Equation	34

-F-

Fermat's Principle	30
filaments	75
four-velocity	50
Friedmann-Robertson-Walker metriek	49

-G-

Galilei-Huygens relativity 37
Galilei-Huygens symmetrie 5,37
Gamow-piek 23
Gamow-vergelijking 23
gas pressure 70
gauge forces 16
geodesic 55,56
Goldstone boson 82

-H-

Hamiltonian system 65
homogeen 6
homogeniteit van de ruimte 4
homogeneity of space and time 37
horizon 58
horizon distance 13
horizon mass 60
Hubble parameter 11
Hubble relatie 7
Hubble's Law 3

-I-

inflation 84
integratieconstanten 6
interval 42,47
invariant 39
isospin symmetry 81
isotroop 6
isotropy of space 37

-J-

Jeans condition 71
Jeans mass 61

-K-

kinematica 8
kosmische dynamica 8

-L-

Lagrangian 33,55
Lagrangian density 33,35
law of inertia 32
local symmetry 16
look-back distance 13
Lorentz symmetric 34,40
Lorentztransformatie 42

-M-

mass conservation 69
metrieken 49
metrische tensor 49
Michelson-Morley experiment 39
minimal coupling term 79

minimal spanning tree 2
minimum principles 30

-N-

nuclei 75

-O-

Oerknal-model 8
optical horizon 62

-P-

pancakes 75
partiële afgeleide 7
phase-rotation symmetry 81
Planck density 54
Poisson equation 52
Poisson's exponent 70
principle of least action 31

-R-

random walk 26
received power 15
redshift 14
Rees-Sciama effect 29
Ricci tensor 53
Riemann-Christoffel curvature tensor 53
rotation 2

-S-

Sachs-Wolfe effect 29
Saha's equation 17
scale factor 12,13
schaalfactor 8
Schrödingervergelijking 22
Schwarzschild metriek 49,57
Schwarzschild-straal 57
shear 2
snelheid 5
Snell's Law 30
spontaneous symmetry breaking 81,83
Stefan-Boltzmann Law 17
Stirling's Rule 67
stralingsdruk 25
stralingskracht 25
summation convention 3
surface brightness 15
symmetry group 16

-T-

thermische energie 21
tijddilatatie 43
two point correlation function 2

-V-

verdelingsfunctie 21
versnelling 5
vertragingsparameter 9
vervormingstensor 7
Vlasov Equation 66
voids 75
Voronoi foam 75
Voronoi tessellation 75
vrije weglengte 24

-W-

wet van de traagheid 5
werkzame doorsnede 24
WKBJ-benadering 22

Contents

1. Observations of Cosmic Structure 1
2. The Oort Constants of the Universe 2
3. Symmetrie en klassieke mechanica 4
4. Newton-kosmologie 6
5. Basic Cosmological Relations 10
6. Some Expressions for Observables 12
7. Cosmic Nucleosynthesis 16
8. De Gamow factor 21
9. Vrije weglengte 24
10. The Microwave Background 28
11. Background Inhomogeneities 29
12. Motion Derived from Minimum Principles 30
13. Symmetries and the Form of the Lagrangian 36
14. Special Relativity 39
15. Algemene relativiteit 47
16. From Global to Local Lorentz Symmetry 49
17. Geodesics 55
18. Zwarte gaten 57
19. Mass Scales 60
20. Motion on a Smooth Background 64
21. The Boltzmann/Vlasov Equation 66
22. The Maxwell-Boltzmann Distribution 67
23. The Euler Equation 69
24. Stability of Friedmann Models 71
25. Growth or Oscillation 73
26. Growth of Structures 74
27. Voronoi Foam 75
28. Growth of Voronoi Features 76
29. What Time is it? 77
30. The Gravitational Lagrangian 78
31. Spontaneous Symmetry Breaking 82
32. Inflation 83
33. Bias 85
34. References 87