Notes About and Text Excerpts From (for Summarizing) the Zel'dovich and Novikov Textbook, *The Structure and Evolution of the Universe*

Les Fishbone

(Transcribed by T. G. Throwe, please report any transcription errors)

(Dated: June 2022 - February 2023)

INTRODUCTION

A. Well-Established

- Observed isotropy of Universe to $\lesssim 0.1 - 1.0\%$ via relic radiation
- Homogeneity deviations $\lesssim 0.1 - 1.0\%$ on a scale of 10^{10} ly
- GTR with cosmo constant is the best basis
- Steady state and changing G are not valid
- Hubble parameter is 50 km/sec/Mpc to within 50%
- Uniform density and pressure
- With $\Lambda \sim 0$, critical density is $0.510^{-29} g/cm^3$
- If density ¡ critical, Universe will expand unbounded and is infinite
- Density is important to know
- The use of celestial objects of a given type to determine the structure of the Universe is complicated by their intrinsic evolution and the evolution of their number as a function of time
- "But the distances over which galaxies can be observed are small compared to cosmological scales. To this day, therefore, the structure of the Universe has not been established through observations of ordinary galaxies either."
- Important to know average density and particle types
- Luminous matter has an average density of $\sim 10^{-31} g/cm^3$, suggesting average number density of baryons $\sim 6 \times 10^{-8}/cm^3$
- Galaxy motions suggest dark matter
- Antimatter absence suggests charge asymmetry
- RR photons now have an average number density of $\sim 400/cm^3$, $10^8 - 10^{10}$ more than the number density of baryons. Their T of 2.7 K corresponds to an energy of 0.0007eV, yielding an overall photon mass-energy density now of $5 \times 10^{-34} g/cm^3$, much lower than that of baryons now.
- Density of neutrinos and gravitational waves is difficult to determine.
- Thus there is as yet no answer to whether total density now is greater or less than critical density, and consequently whether Universe is finite or infinite.
- Going back in time, T increases and radiation and matter are in thermo equilibrium because matter density $\sim V^{-1}$, while radiation density $\sim V^{-4/3}$
- At $t \sim 1sec$ in Friedmann solution, $T \sim 10^{10}$ or $10^6 eV$ and matter density $\sim 10^6 g/cm^3$. There would have been photons plus electrons and positrons and protons and neutrons.
- Expansion leads to disappearance of positrons, while neutrons decay or combine with protons, forming 70% hydrogen and 30% helium by mass, but almost no heavier elements. Also remaining were neutrinos and antineutrinos.
- Further expansion means matter mass density exceeds photon mass-energy density of photons.
- All RR now seen is from $z \sim 1000$, the time of last scattering. The corresponding distance is about 97% of the distance to the singularity, the horizon distance. The observable volume is then about 90% of the maximum possible volume.
- Existing structure indicates early deviations from homogeneity and isotropy.

B. Not Well-Established

- Use perturbation theory by modes on simple time-dependent solution rather than exact solutions of four dimensional spacetime, especially since initial conditions are unknown.

- But how large were density perturbations?
- Present average density of galaxy clusters is roughly characteristic of the average overall density at their formation time; this leads to an estimation of the formation time. For the plasma state in the RD era, this leads to fractional density oscillations of 10^{-3} for $\delta\rho/\rho$
- With theory, observations of RR fluctuations then permit estimation perturbation magnitude was functions of the scale or mass, i.e., the perturbation spectrum.
- Summary of Important Recent Result: Universe picture represents a weakly perturbed (almost homogeneous) expanding Universe with a definite initial (and large) entropy. Measurements of the spectrum and spatial distribution of the RR support this picture.
- But can this picture explain galaxy rotation, magnetic fields and the origin of quasars?
- Primordial magnetic fields are not necessary; plasma motions can generate observed fields.
- But galaxy rotation given vortex-free initial perturbations? Possible given galaxy interactions.
- Another theory is that galaxies formed from explosions of hyper dense bodies, but this violates known physics.

C. Beginning of Expansion

- Anisotropic expansion before $t \sim 1 sec$?
- Is there infinite density at the beginning or is that a characteristic of the isotropic homogeneous model?
- There is proof a singularity even if expansion was not homogeneous and isotropic?
- Details later, but here consider here aforementioned models plus perturbations. With these bases, do present observations and the laws of physics permit the establishment of the history of the Universe, including after and before (if meaningful) the singularity and the nature of the singularity itself?
- Approach this via thermodynamics: many initial states can lead to a the same final state, which can serve as the new initial state for further evolution; the actual initial state is forgotten.
- Thus find that cosmo model which arises from a wide class of initial early states.
- Many anisotropic lead to isotropic expansion. But are such statistical arguments applicable?
- Why is the entropy of the Universe large? Why hot at the start of expansion? Why are perturbations leading to observed structure of just the correct magnitude?
- Laws of physics seem sufficient to explain all. Intense particle creation can occur from intense gravitational field close to the time right after the singularity, but only given anisotropic expansion.
- Finally, there can be new phenomena given quantization of the metric.
- Historical remark: Friedmann theory 1922-1924; Einstein mention thereof; Lemaitre 1927. Thus Lemaitre did not "independently" establish the laws of the expanding Universe.
- After Hubble discovery in 1929, math solutions became established theory. Einstein remarked in 1931 that Friedmann was the first to follow this way;

I. THE HOMOGENEOUS, ISOTROPIC UNIVERSE: ITS EXPANSION AND GEOMETRICAL STRUCTURE

1. Local Properties of the Homogeneous, Isotropic Cosmological Model

Standard exposition based on Newtonian theory for Hubble expansion, age of the universe, and matter density and pressure

2. Relativistic Theory of the Homogeneous, Isotropic Universe

GTR needed to analyze large regions

See Vol. 1 for a sufficient exposition of GTR; Theory of Fields by Landau and Lifschitz for a complete GTR $\gg$ Friedmann eqns, with results the same as for the Newtonian description

Various models for open, closed, and flat (critical density) geometry

3. The Propagation Of Photons And Neutrinos; Observational Methods For Testing Cosmological Theories

Significant effects of relativistic matter on early expansion; cosmological neutrinos would not be observable today, though photons are

As density becomes infinite as size and age approach zero, visibility to an earlier stage is not possible because the optical depth, dependent on particle density, diverges.

Whereas the theoretical particle horizon is at $t = 0$, practically it is at a later time when the optical depth is of order unity.

Observational quantities: red shift, angular size and luminosity of distant objects, amount of matter as a function of red shift, apparent magnitude

Deceleration parameter and the first approximation

Impossibility of determining the cosmo model if sources evolve in an unknown way

Distance ladder to far-away objects

Redshift vs. apparent magnitude observations rule out steady-state universe

No Olbers paradox in an expanding universe

4. The Cosmological Constant

Cosmo constant would only be manifest on the scale of the universe

History of cosmo constant starting with opinion that universe is static; Einstein desire for corresponding GTR solution and ideas of Mach

Hubble observation of expansion and Friedmann nonstatic GTR solutions Realization that cosmo constant is not needed, especially given new Hubble value of 75 and longer age for universe

Various cosmo models with nonzero cosmo constant.

II. PHYSICAL PROCESSES IN THE HOT UNIVERSE

5. Intro to Part II

Relic radiation (RR = CMB) at T=2.7 K is the most important observational fact, and this RR (nor the equivalent background neutrinos) could not have been produced by astronomical objects

Also, there are about $10^{(9+-1)}$ photons per baryon

These two data allow characterization of the composition of the Universe at earlier time given thermodynamic equilibrium with specific entropy of matter conserved and volume changing smoothly during expansion

In later stages, nuclear reactions cease and nucleosynthesis takes place, with only photons, electrons, nuclei, neutrinos and gravitons surviving, with the last two undetectable

Hot universe proved by observations for the period $10y \leq t \leq 10^{10}y$, and likely only consisting of mattering the large, not antimatter too.

Short historical review of RR prediction and discovery

Complete EM spectrum in the universe, of which a small section is the RR (Fig. 27, p. 126)

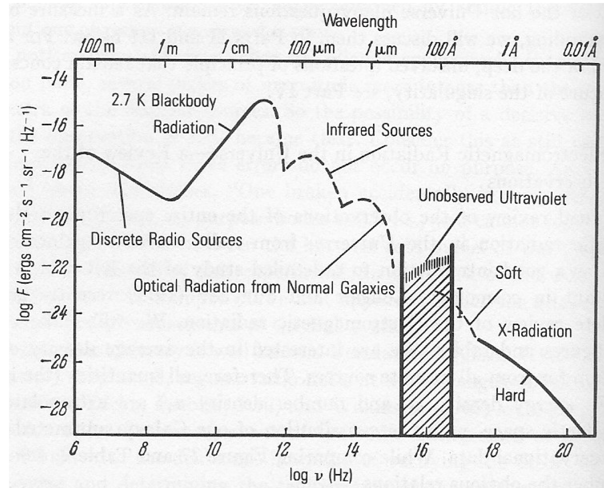


Fig. 27.—The spectrum of electromagnetic radiation in the Universe

6. Thermodynamic Equilibrium...

Early radiation-dominated era with matter and antimatter

Ratio of photons to baryons hardly changes during expansion: thermo equilibrium during early stages and conservation of RR photons later

Given $kT > mc^2$, the number of particles and antiparticles of each kind equals the number of photons. Thus $\sim 10^8$ nucleon-antinucleon pairs in the early universe for each nucleon today. This suggests that the present nucleons result from a small excess (10^{-8}) of nucleons over anti nucleons early.

Expansion eras are therefore:

1. Hadron era: with nucleons and antinucleons and ordinary and anti versions of all other particles; $t \lesssim 10^{-6}s$ and $T > 10^{13}K$

2. Lepton era: with only a small remainder of nucleons, electron and positrons annihilate by the end, leaving a small remainder, and neutrinos decouple; $10^{-6}s < t < 10s$ and $10^{13}K > T > 5 \times 10^9K$

3. Photon-Plasma era: plasma and radiation in equilibrium; $10s < t < 10^{12}s$ and $5 \times 10^9K > T > 10^4K$

4. Post-recombination era: $t > 10^{12}s$ and $T < 10^4K$ when the RR becomes transparent

Gravitons, if they exist, would always be present but would not interact with other particles after Planck time $\sim 10^{-43}s$.

MISSING OR TRANSITION TEXT RESULTING FROM IMAC FAILURE TO BE RESOLVED

At a sufficiently high temperature T such that $kT > Mc^2$, where M is the mass of the most massive particle, photons and other relativistic particles dominate

$P = e/3 = \rho \times c^{2/3}$ and $e = \kappa \times \sigma \times T^4$ to take account of all kinds of relativistic particles.

And $n \sim e/(3kT)$ for the particle density

Consider $T \sim 1MeV$, $t \sim 1sec$, and $n(electron) \sim n(positron) \sim 10^{-31}cm^{-3}$ and annihilation cross section $\sigma(Annihilation) \sim 10^{-24}$ and particles move at c, then time to establish equilibrium is

$\tau \sim 10^{-17}$ small compared to 1 sec.

Similarly for higher mass particles at higher temperatures and correspondingly earlier times.

When $kT > m_{nuc}c^2$, $n_{nuc\,then} - n_{antinuc\,then} = n_{nuc\,now} \sim 10^{-8} \times n_{photons\,now}$

$n_{nuc\,then} \sim n_{photons\,then}$

One could apply the same considerations to quarks and so on and so forth if $kT > m_q c^2$

Hot vs. cold matter as n ≫ infinity ? Different models of Hagedorn and Omnes

Quark theory for nucleons

Conservation of energy and baryon charge and entropy for slow, adiabatic processes ⇒ evolution can be described.

Particle-antiparticle annihilation requires binary collisions, increasingly unlikely as the Universe expands.

1. Residual n(antiparticle) in charge symmetric theory is very small at the end of hadron era, $T \sim 1MeV$ because annihilation σ is large and nucleon excess ⇒ exponentially small n(antiparticle) when their creation ceases.

2. Residual n(q) is large . With respect to photons, it is $\sim [Gm^2/hc]^{1/2} \sim 10^{-18}$; with respect to nucleons, it is about 10^{-9}.

3. In spatially homogeneous, charge-sym universe, nucleon problem similar to quarks and leads to 10^{-18} nucleons/photon, disagreeing with observations by 10^{10}. So we should consider charge-asymmetric universe.

This leads to Omnes theory. Charge symmetry of primordial homogeneity is spontaneously broken on the microscopic scale. Strong interaction leads to separation of matter and antimatter drops of size $\sim 10^{-3}cm$ at $10^{-6}sec$

This separation tendency stops as T decreases and annihilation occurs as usual. But spatial separation means annihilation occurs mainly at the boundary of regions.

There exist regions with 10^{-9} nucleons/photon and regions with 10^{-9} antinuc/photon ⇒ galaxies and antigalaxies. Omnes calculations lead to two characteristic quantities: average n(nuc or antinuc) / n(photons). And characteristic size of matter or antimatter region. But a consistent calc of separation and following annihilation leads to a much smaller concentration of nucleons and antinucleons, disagreeing with present density of nucleons

Annihilation continues during radiation-dominated stage, but expected consequences of prolonged annihilation are not observed. Thus, even with account of phase separation, charge-symmetric theory does not agree with observations.

Consider therefore charge-asymmetric Universe with excess baryons always. Early, excess of baryons is small given number of pairs, so Omnes phase separation is plausible then

For $T = 300MeV$, charge asymmetry manifests itself only in at $T > 1MeV$, when there is an abundance of electrons and positrons and RR spectrum takes equilibrium form.

Finally, charge asymmetry leads to n(baryon today) = n(baryon charge density initially).

It would nevertheless be very interesting to find evidence now of hadron era phase separation.

Hagedorn theory that the number of charged particles is infinite is contradicted by experimental results of QED.

7. Kinetics of Elementary Particle Processes

In earliest stages of the hot Universe, neutrinos (+anti) are in thermo equilib with other particles. Creation of neutrinos mainly by e- + e+ ≫ neutrino (+anti) with relativistic cross section

$$\sigma \sim \frac{g^2 \times E^2}{h^4 \times c^4},$$

where $g \sim 10^{-49} ergscm^3$ is the weak interaction constant

Given particle energy of kT, time to reach equilibrium $\tau = 1/(\sigma \times n \times c)$, and previous relation between universe time t and temperature T, we obtain dependent of τ and t:

$$\tau \sim \frac{[G^{5/4} \times h^{13/4}]}{[g^2 \times c^{1/4}]} \times t^{5/2}$$

(Landau & Lifschitz for statistical factors). When τ is greater than t, neutrinos no longer interact either with other particles or with one another. Equating them leads to $t \sim 0.1s$ without consideration of numerical factors but showing the dependencies of G and g . More accurate calculations follow, including consideration of μ neutrinos.

Present temperatures neutrinos compared to photons is then (Peebles)

$$T_{neutrino} = (4/11)^{(1/3)} \times T_{photon} \sim 0.7 \times 2.7K \sim 2K$$

But mass of neutrinos could be $\leq 100eV$, while background cosmic neutrinos would have energy $5 \times 10^{-4}eV$, so observation of the background would require measurement improvement by 10^6.

RR spectrum tells us about $z \sim 10^6$ at $t \sim 1yr$; neutrino spectrum could tell us abut $z \sim 10^{10}$ at $t \sim 0.1s$!

Particles that decay spontaneously disappear exponentially as Universe expands. Stable particles would remain if annihilation reactions do not occur.

Y FIGURE 30 FOR TEMPERATURE-DEPENDENCE OF PARTICLE RELATIVE ENERGY DENSITY

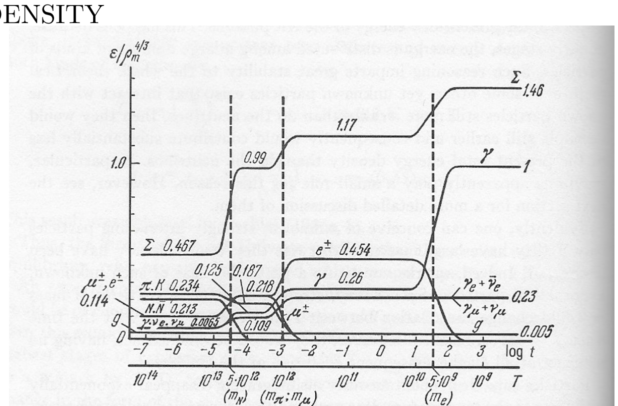


FIG. 30.—Evolution of $\epsilon/\rho_m{}^{4/3}$ for various kinds of particles in the hot, expanding model of the Universe. The quantity ϵ is the energy density, while $\rho_m = m_{\text{nuc}}\, n$, where m_{nuc} is the nucleon rest mass and n is the baryon number density (density of baryons minus density of antibaryons). Four forms of baryons and four forms of antibaryons respectively are encompassed by the symbols N and $\bar{N}$; $\epsilon_\gamma/\rho_m{}^{4/3}$ is taken equal to unity at the present time; and Σ signifies a summation over all particles.

?? MISSING TEXT FOR APRIL AND MAY RESULTING FROM IMAC FAILURE

MAY 12 SUMMARY

Introductory Remarks2

a. Concerning Gravitation by MTW, a new edition (printing) was published in 2017 by Princeton University Press after the original in 1973 by Freeman. This new edition contains by David Kaiser describing the style of the original, the publishing history, and reactions to it. The new edition also contains an additional preface by MT that focuses on the status of the material in the text in light of developments subsequent to the original—chapter by chapter. Perturbations in the early Universe that could lead to structure formation are not addressed in MTW, though the papers by Lifschitz and Khalatnikov are cited (see point b).s

b. The 1961 textbook by Landau and Lifschitz entitled the *The Classical Theory of Fields* does not contain material about GTR cosmology that goes beyond the simple Friedmann models, even though Lifschitz himself in 1946 and with Khalatnikov in 1964 addressed GTR perturbations.

c. The 1972 monograph by Weinberg, Gravitation and Cosmology: Principles and Applications of the GTR, does contain material about GTR cosmology that goes beyond the simple Friedmann models, in particular descriptions of the early hot Universe and perturbations that could lead to galaxy formation. There are citations to work by Z&N and to Lifschitz, among many, many others.

11: Instability in the Hot Model

Approach of last chapter here applied to RD period, when matter completely ionized, the radiation density dominates, and the matter is coupled to the radiation. For $\omega_0 = 1$, there is a short period, when the matter is still completely ionized, and the sound speed varies as $b = \frac{(c/\sqrt{3})}{\sqrt{[1+3\rho_m/4\rho_r]}}$. Matter and radiation densities become equal when $z = 10^4\omega_0$ or $t \sim 2 \times 10^{11} \times (\omega_0)^{-2}$ s.

??????During the period the Universe is filled with a medium whose equation of state is, $P = e/3 \sim a^{-4} \sim t^{-2}$, and sound speed is $b = c/\sqrt{3}$. A definite value for the temperature follows, while the matter density still requires specification of several parameters.

With the help of the Jeans criterion, let us find the conditions dividing regions of stability and instability.

Perturbations are of the form $\delta = \delta\kappa(t) \times exp(k \cdot xi)$, where $k = k_0 a(t_0)^{-1}(1+z)$ and $\lambda = \lambda_0/(1+z)$. k_0 and λ_0 refer to the present. With the Jeans criterion taking the form $(b \times k)^2 = 4\pi G\rho$, and substituting values of b and ρ for the RD era, we obtain

$$k_{Jeans} = \frac{3}{(\sqrt{8} \times ct)}$$

and

$$\lambda_{Jeans} = 2\pi/k_{Jeans} = ct \times 4\pi \times \sqrt{2}/3$$

The Jeans length is therefore of the order of the distance over which pressure gradients (sound waves) can equalize density.

The regions of stability and instability are conveniently seen in Figures 43 &44 Also see Weinberg, Figure 15.6,

See slides PPT 36 & 38

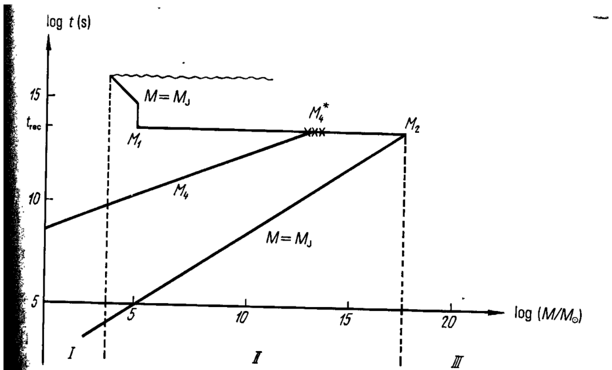


FIG. 43.—The regions of stability and instability on the (log t, log M)-plane for $\Omega =$ 0.025. The line $M = M_J$ separates the stability region from the instability region. The line $M = M_4$ is the boundary of the region in which the perturbations effectively dissipate (cf. § 2 of this chapter). Masses $M > M_J$ are located in the instability region, masses $M < M_J$ are located in the stability region, and masses $M < M_4$ exponentially decay with time. The time of recombination $t = t_{rec}$ coincides, for $\Omega = 0.025$, with the time when the matter density and the radiation density are equal. M_1 is the Jeans mass after recombination; it decreases somewhat after $t \approx 10^{15}$ s (cf. § 8 of chap. 14). The wavy line denotes an arbitrarily chosen time when secondary heating of the matter begins (cf. chap. 15). The remaining details are discussed in the text of this section.

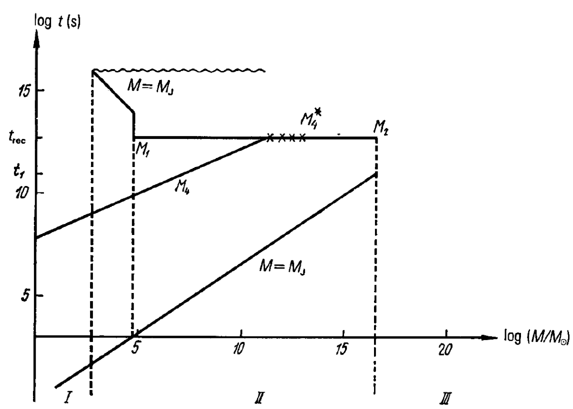


FIG. 44.—The regions of stability and instability in the (log t, log M)-plane for $\Omega = 1$. The notation is the same as in Fig. 43. The matter density and radiation density are equal at the time $t = t_1$; and recombination takes place at the time $t_{rec} > t_1$. During the interval $t_1 \leq t \leq t_{rec}$, the Jeans mass is constant.

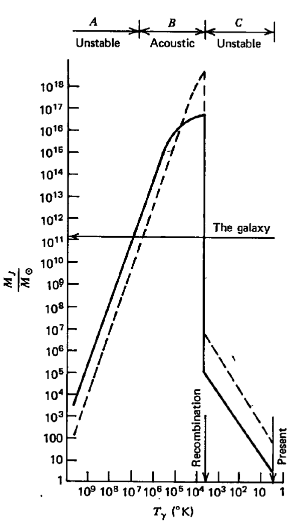


Figure 15.6 Jeans mass as a function of radiation temperature. Solid line is for $\sigma = 0.8 \times 10^8$, corresponding to $T_{\gamma 0} =$ 2.7°K, $\rho_0 = 3 \times 10^{-29}$ g/cm^3. Dashed line is for $\sigma = 2.4 \times 10^9$, corresponding to $T_{\gamma 0} =$ 2.7°K, $\rho_0 = 10^{-30}$ g/cm^3. The drop in Jeans mass at recombination is somewhat more gradual than shown here.

Other points in this chapter:

a. as a result of dissipative processes, decay of a wave is determined by conditions during the last part of every period considered because the increase in photon MFP overpowers the increase in the wavelength;

b. conclusion that difficult-to-observe particles give rise to a rather small decrease in the amplitude of oscillations compared to neglecting them;

c. hypotheses that supermassive stars or globular clusters result from entropy perturbs;

d. conservation of vortex velocity upon early stages of expansion, matching perturbs when the eqn of state change; and

e. Sakharov oscillations

12: Gravitational Instability in the GTR

GTR approach necessary for perturbations with $\lambda \gtrsim ct$ in a fluid with eqn of state $P = e/3$.

Method is take homogeneous, isotropic Friedmann model but then to replace metric g by $g_0 + h$, where h represents perturbations; analogously, stress-energy tensor changes from e to $e_0 + \delta e$, with δP determined by eqn of state, and finally perturbed velocities u are assumed small, with u_0 fixed via the identity $u \cdot u = 1$. These expressions are then substituted into GTR eqns, relating h, δe, and u , and yielding their time evolution once initial perturbations are specified.

See slides PPT 41 & 42 for the metric and the resulting eqns for the perturbs from the GTR eqns. All computations are done in the linear approx in which the quantities h are first-order. Then perturbed values of e, P and U occur only in second order.

Two caveats:

1. GTR eqns put some restrictions on these initial values.

2. Coordinate system choice allows apparent unphysical results; distinguishing between them is important.

Follow approach of Lifschitz and many others: synchronous reference system is used to study the perturbations, and their consequences are studied in other coord systems. Nonlinear approximations likely lead to important changes to be examined in later chapters.

So consider perturbations in spatial regions which may be large with respect to ct but small with respect to radius of curvature during the period studied; $a \gg ct$ during the early stages of evolution, conditions equivalent to perturbs in a flat model with $\rho_{critical}$. This will yield possible galactic evolution paths; Lifschitz studied more general cases.

Solutions of the eqns will be as plane waves of the form $Q = e^{[i\kappa \cdot x]}$ on the background of a spatially homogeneous and isotropic evolving Universe with the same invariances as the unperturbed solution, with κ a certain vector and Q a scalar.

Then tensors can be constructed from κ and Q. There will also be a vector $P = \kappa \times Q$, and still another vector $\Sigma = Es \times Q$, with Es perpendicular to κ.

These are scalars, vectors and tensors only in 3-d space.

The scalar Q will describe density perturbations and Es will describe velocity perturbs. Another tensor with a plane wave dependence will describe gravitational waves.

Scalar : The main term for very early times is that the fractional energy density varies linearly with time. For late times, the fractional energy density would vary as $\cos[(\kappa \times \eta)/\sqrt{3}]$, corresponding to acoustic oscillations with a speed of $c/\sqrt{3}$ and to constant-amplitude density perturbations. This exact result supports the results of the intuitive analysis. The exact theory also works with metric perturbs, which would be finite as t tends to zero. There was a line missing in the translation on page 292..

The independence of the metric perturbations with respect to time during early stages of the expansion accords completely with the idea underscored by the intuitive analysis of long-wave perturbs. This independence is also in accord with the independent evolution of different regions of the Universe with different initial conditions.

For $\lambda \ll ct$, the metric perturbs tend to zero and we are left with a description of sound waves. But this result applies to the eqn of state of a RD plasma, where $b = c/\sqrt{3}$, with the pressure gradient smoothing perturbs, so the "sound" horizon is important. For dust, $P = 0$, there is no sound propagation to smooth geometry and density perturbs. Density perturbs continue to grow for $\lambda < ct$, but metric perturbs remain constant.

One very important conclusion of Lifschitz's work that remains valid is that to explain finite perturbs today (galaxies!), it is sufficient for δ to tend to 0 as t tends to 0, whereas the metric perturb must remain nonzero as t tends to 0. Novikov found $h \sim 10^{-2}$ to 10^{-3} as t tends to 0, corresponding to galaxy clusters.

Vector (Rotational) perturbs case exhibits differences from Hubble expansion. The solutions show that metric perturbs grow as time approaches zero, leading to the conclusion that initial rotational perturbs are incompatible with a Friedmann model. This conclusion is important in any discussion of galaxy formation in the vortex theory.

Tensor perturbs have two independent polarizations for a given wave vector. If the wavelength is less that ct , the solution describes a wavelike gravitational field. So when the wavelength becomes less than the horizon size, usual energy density computations apply:

Amplitude $h \sim (1 + z) \sim a^{-1}$, and energy density $e \sim (1+z)^4 \sim a^{-4}$. The wave velocity is c . The density and velocity perturbs are not connected with gravitational waves.

Matter (coming?) velocity relative to perturbed coordinates is zero in the field of a gravitational wave, but particle velocities do change, so a sphere becomes a time-varying ellipsoid in the directions perpendicular to the velocity of the wave. If the wave passes through an ideal fluid, energy is not dissipated, so entropy does not grow and new waves are not created. Viscosity would change this

Entropy perturbs could arise as inhomogeneities in the eqn of state, with one approximate form being

$$P = e/3[1 - B(x)e^{-1/4}],$$

with consequent effects on the metric and motion. Initial entropy perturbs would give rise to adiabatic density perturbs, and, in particular, to growing-mode adiabatic perturbs if the wavelength is sufficiently large, though this is not a relativistic effect. Entropy perturbs corresponding to masses between one solar mass and 10^4 solar masses would only cause decaying RD plasma oscillations before recombination. For masses greater than 10^4 solar masses, the entropy irregularities are preserved. Such perturbs could be related to the formation and evaporation of n PBHs, leading to entropy production, possibly all entropy!y!

One interesting idea is a quasi-isotropic solution with a uniform distribution of perturbs described by this metric:

$$ds^2 = (cdt)^2 - [t \times a(x) + t^2 b(x) + \dots]dxdx$$

The functions a and b are of the spatial-metric type, with 3x3 indices, and x is a spatial vector.

A general result is that Friedmann behavior near the singularity is compatible with density perturbs and gravitational waves, which may not be small, but not with vortex perturbs. Thus the quasi-isotropic solution, or entropy perturbs, represent cosmo solutions not in conflict with the present state of the Universe, deviating least from the strictly homogeneous solution,

Perturbs whose wavelength is comparable with the size of the Universe requires analysis with math beyond plane-wave theory. Assume ω unequal to unity to exclude the flat model. The ratio of the wavelength of the perturb to the radius of the model is constant during the expansion. Solutions are constructed similar to the methods for constructing spherical harmonics. Scalar functions are considered.

Some remarks about the possible periodic distribution of quasars as a function of redshift.

The concluding remark in this chapter is as follows:

ZN: "One indeed ought to expect that the spherical-wave method will find wide application in the theory of perturbations of the homogeneous, isotropic Universe in the very near future." N.B.:This is a prediction from 1975!.

LF: See slide 43 interpreting the CMB Planck observations as a function of the spherical harmonic parameter l, a result from 2009 and later, as well as from WMAP from 2001 and later.

JUNE 15 SUMMARY

13: Statistical Theory

Any small perturbation can be represented as a linear combo of independent plane waves. Further steps are to examine wave interactions and to solve nonlinear problems.

Galaxy forms and locations are random, suggesting random initial perturbations and a statistical description of the Universe subject to the fundamental laws of physics.

An initial assumption of density perturbations in boxes fails because of interactions among neighboring boxes.

Consider instead a plane wave expansion of the density perturbations:

$$\Delta\rho = \sum_k (A_k \Psi_k(x)$$

where $\Psi_k(x) = V^{-0.5} e^{(ik\cdot x)}$

The Ψ functions satisfy orthonormal conditions. A convenient dimensionless quantity is

$$\Delta k = [k^3/(2\pi^2 \times n)] \times \sum_k (|A_k|)^2.$$

If for all k Δ is small, then the density perturb is small.

For a bounded volume $V = L^3$, then $kx = 2\pi nx/L$, etc., where the n values are whole numbers.

A reasonable definition of a random function is one whose Fourier coeffs are random, and the randomness is not resolved by the physical interaction while the perturb is small. The randomness hypothesis is connected with the idea that we can choose many different volumes in the Universe. Each has a definite density function and a single set of amplitudes A_k. How often is a given A_k value encountered? l

Let $A_k = B_k + iC_k$ and consider over N volumes and respective k values the probability $P(B, C, \ldots)$ for the appearance of given values of the Fourier coeffs. A natural form for this probability is proportional to

$$e^{(-B_k^2/[2\beta_k^2])} \times e^{(-C_k^2/[2\gamma_k^2])}$$

At an early stage near the singularity, the integral which determines the Fourier coeffs reduces to a sum over causally disconnected regions (if there is no period "before the singularity"). Hence the assumption of a normal distribution for the A_k is natural.

Even for small inhomogeneities of 10-20%, the astronomer wants to know their form and amplitude, not Fourier coeffs. While the average value of δ vanishes at each point of space, the average of its square does not. The properties of the Fourier series leads to

$$\langle\delta^2\rangle = \sum_k \langle A_k^2\rangle/V$$

With normal distributions for B_k and C_k, we then find for the prob that a given value of δ is obtained at any given point, with Δ the same for all points and independent of time,f

$$P(\delta) = [1/(2\pi\Delta)^{0.5}] \times e^{(-\delta^2/[2\Delta])}$$

Note well: δ and Δ represent different quantities; δ is the dimensionless density amplitude.

This function describes the amplitude of the inhomogeneities, but says nothing about their spatial structure

The correlation function $f(r)$ characterizes their spatial structure, with

$$f(r) = \frac{\langle[\delta_x \times \delta_y]\rangle}{\langle\delta^2\rangle}$$

and where $r = x - y$

This correlation function will be positive for small values of r , but its sign will vary for larger values. The natural conclusion is that the first zero of this function will define regions with the same sign of δ. For example, if $f(r)$ is given in terms of the spectral function β_k characterizing the amplitude of waves of various length and is concentrated in a narrow interval around a wave number k_0, then the first zero is at $r_0 = \pi/k_0$, half a wavelength

When has a significant fraction of the mass passed into gravitationally bound objects? Suppose that the β_k with $k < k_{Jeans}$ grow with time as a consequence of gravitational instability. Fragmentation will have occurred when the growing Δ is of order unity.

A good and simple description of the matter inhomogeneity is given by average mass and its deviation through

$$\mu = \frac{\langle\delta M\rangle}{\langle M\rangle} = \frac{\sqrt{\langle M^2\rangle - \langle M\rangle^2}}{\langle M\rangle}$$

Assume that all matter is distributed in the form of isolated bodies of mass M_1 with average density ρ_1, so the number density of bodies is $\langle\rho\rangle/M_1$ and the volume of each body is $v = M_1/\rho_1$. Then μ has the following behavior:

1. For $\langle M\rangle \gg M_1$, many bodies are found in the volume under consideration, so μ is small and tends to zero as $\langle M\rangle$ tends to infinity

2. If $(\langle\rho\rangle/\rho_1) * \times M_1 < \langle M\rangle < M_1$, then sometimes there is only one or not even one body in the volume under consideration and $\mu \sim \sqrt{M_1/\langle M\rangle} > 1$ For $\langle M\rangle < \langle\rho\rangle \times M_1/\rho_1$, i.e., for $\langle M\rangle < \langle\rho\rangle \times v$, the volume under consideration is less than that of a single body and

3. $\mu \sim \sqrt{\rho_1/\langle\rho\rangle} > 1$

How does μ behave for $\langle M\rangle \gg M_1$, for objects containing many of the smaller bodies? Naively, much as does a statistically independent distribution distribution of particles. The average number of little bodies in such an object is $\langle N\rangle = \langle M\rangle/M_1$, while

$$\mu = \langle \delta N\rangle/\langle N\rangle = 1/\sqrt{N} = \sqrt{M_1/\langle M\rangle} < 1$$

for $\langle M\rangle > M_1$

But this is not always correct. There is no universal description of the behavior of the function μ since it depends on the isolation process. Indeed, a study of the processes in large volumes containing many objects and leading to small μ values can give valuable info about the universe and its large-scale structure.

The law $\delta N = \sqrt{N}$ is only obtained given a random—not correlated—arrangement of discrete objects in space, corresponding to the hypothesis of a God who, from outside, sows space with galaxies and that they fall into regions independent of how the preceding ones are distributed. But this hypothesis is evidently unacceptable, since gravity from existing objects affects the growth of small perturbs. An evolutionary formulation is necessary from a uniform distribution. Seemingly perturbs grow from the inflow of matter from neighboring regions. But this reasoning is not valid in the case of gravitational instability, taking account of the long-range nature of gravity.

Considerations of Jeans theory allows one to say that the increase of matter at the center (in a spherical configuration) is not from neighboring regions, but from infinity! Thus there is not an anticorrelation among neighboring galaxies.

The final conclusion is that the fluctuation law $\delta N = f(N)$ for the distribution of discrete objects depends on the law for the original small perturbs. In principle, $\delta N \sim 1/N^{1/6}$ or $\delta N \sim N^{2/3}$ are possible, depending on the spectrum of small perturbs. Observational studies can give insight into the initial state.

Concerning limitations of the linear theory, the first is that it is practically difficult to calculate the properties of a surface of given δ for a distribution function whose Fourier expansion is specified. Second, there are conceptual problems in matching the topology of regions with particular δ values, as small islands say, with the topology of known astronomical objects.

The root of these difficulties is that astronomical objects result from strong nonlinearities, which are addressed in the next chapter.

14: Nonlinear Theory and Thermal Instability

Three ways to approach the nonlinear problem, possibly in combination:

1. Exact solution with special initial conditions;
2. Approximation method for extrapolating the linear solution to the general case; and
3. Qualitative explanation of the properties of the general exact solution.

Spherically symmetric perturbs can be analyzed exactly because the effect of neighboring perturbs vanishes. Second method takes account of the tidal action of neighboring perturbs, but this is an approx.

Analyses are only for dust, for which pressure vanishes. The simplest spherically symmetric case is of a sphere with perturbed density Ω' on a Friedmann background with Ω, where $\Omega' > \Omega$. With no extra mass, there is a hole in the shell outside the higher density region.

If $\Omega \leq 1$, then the two cases $\Omega' \leq 1$ and $\Omega' > 1$ are possible. In the first case, the perturbed sphere expands indefinitely and a bound object does not form. In the second the perturb behaves as a closed Friedmann model, going from expansion to contraction and eventual recollapse. In this "Swiss-cheese model", the many such perturbs do not affect the average expansion. Differences occur if there is not initial symmetry or inhomogeneity. If there is pressure, then infinite density does not occur over the whole volume simultaneously, leading to shock waves and nonzero pressure and entropy.

In the early moments of the solution, when the unperturbed density is large and perturbations small, what critical perturb amplitude leads to the formation of gravitationally bound objects. The answer is

$$\Delta_{crit} = \frac{3}{5}\frac{1-\Omega_0}{\Omega_0(1+z)}$$

14.2

From observations in the expanding Universe, the amplitude of density perturbs become of order unity when the linear size of inhomogeneities is much less than the horizon ct and the radius of curvature a . Thus a linear perturb theory loses its validity when the use of Newtonian physics remains valid, i.e. when relativistic effects are insubstantial. Additionally, there are cases where gas pressure negligibly affects perturb growth, especially adiabatic perturbs during the prerecombination era RD era. Below we consider post recombination effects of such perturbs.

Eulerian coords r of particles are written as functions of their Lagrangian coords s as

$$r = a(t)s + b(t)x(s),$$

where the first term corresponds to unperturbed motion. Neglecting the second term, we find

$$u = \frac{dr}{dt} = s\frac{da}{dt} = r\frac{1}{a}\frac{da}{dt},$$

which is the Hubble expansion law. Thus the Lagrangian coords s are defined as the comoving cords of the unperturbed motion. The second term describes perturbs, exact for small, growing perturbs; we shall use it even for large density contrasts.

Earlier it was shown that in the linear approx and when $P = 0$, a perturb of any form grows, but its form remains unchanged:

$$\delta = \delta_0 \frac{r}{a} \phi_1(t)$$

and

$$w = w_0 \frac{r}{a} \phi_2(t)$$

But we also stipulate that, while the density distribution is arbitrary, the peculiar velocity w is vortex-free. A reformulation incorporating this condition is that w_0 is derivable from an arbitrary potential, $w_0 = \nabla\phi$; then $\nabla \times w_0 = 0$. We assume too that the $\delta_0 = -\nabla \cdot w_0$.

In the construction of the approx nonlinear theory, we select as an extrapol the linear formula $w = w_0(s) \times \phi_2(t)$. The peculiar velocity is then

$$W = \frac{dr}{dt} - Hr = \frac{1}{a}(a\frac{db}{dt} - b\frac{da}{dt})x$$

The Hubble parameter is $\frac{1}{a}\frac{da}{dt}$, $x(s)$ is vortex-free, and ϕ_2 and b re related. This variant of the linear approx is useful for the following qualitative reason. In the absence of other forces, the exact solution takes the form

$$r = a_0 ts + tv(s) + s$$

Then particle trajectories intersect and infinite density is achieved. Clearly the perturbs are large near this singularity. But in the general case, the sing. Takes the form of a 2D surface. Only for degenerate cases does the intersection occur along a line or at a point. Now take into account grav. forces: near the 2D sing these forces are finite, so they do not exert a drastic influence on the perturb growth and do not seriously affect the general picture. Thus it is reasonable to to seek a solution for large perturbs in a form valid for small perturbs and small gravitational forces. And it becomes easy to calculate other quantities such as the density and the velocity. For a given $r(s)$, the density equation in Lagrangian coords is exactly soluble.

In the linear approx, $b(t)$ is well known and density perturbs are proportional to the ratio $b(t)/a(t)$, also well known.

The density can be rewritten in this form:

$$\rho = \frac{\langle\rho\rangle}{[1-(b/a)\alpha][1-(b/a)\beta][1-(b/a)\gamma]}$$

Here, α, β and γ are functions of s only, while b/a is a universal function of t (depending on ω_0 also), conveniently expressed in terms of z . The generally unequal functions α, β and γ depend on the specific form $x(s)$ of the initial perturb and thus characterize the deformation along the three orthogonal axes of the deformation tensor. For definiteness, choose $\alpha > \beta > \gamma$. V v

While the case $\alpha < 0$ is possible, if $\alpha > 0$, then $[\alpha(b/a)]$ grows and can reach unity in the course of the evolution. Then it follows from the density equation that the density becomes infinite there. This arises as a result of the 1D contraction along the axis related to α. The picture that results is that when the perturbs become sufficiently large, flat pancakes of collapsed dust form in various places.

This general picture is supported more generally. It is necessary to note that contraction along one coord can be accompanied by contraction or expansion in the plane of the pancake. Based on a complicated probability distribution function (pdf) for α, β and γ obtained by Doroshkevich [LF interpretation of the text; note too that this pdf is given in the text], only 8% of the matter contracts along all three axes, while 84% contracts in one direction but expands in one or two.

The value 8% is close to Oort's estimate of ∼6% that matter is compressed in all three directions. It is however not evident that only this 8% eventually becomes gravitationally bound. [see Oort, J.H. 1958, in La structure et l'evolution de l'univers, 11 Conseil de Physique Solvay (Brussels: Stoops) and 1970, Astronomy. Ap., 7, 384].

Physical applications of this pancake picture will be discussed in the next chapter.

14.3

In nonlinear spectral theory, the approx solution for dust can be written as functions of spatial coords multiplied by functions of time.

Consider now the case where pressure is large, important for processes before recombination or small-scale entropy perturbs after recomb.

In the linear approx, the eqns for Fourier amplitudes satisfy second order, time dependent dif eqns so the amplitudes A_k are independent of one another, and the A_k can vanish.

In the nonlinear case, the eqns become

$$a\frac{d^2 A_k}{dt^2} + b\frac{dA_k}{dt} + cA_k = F(A_{k'}, A_{k''})$$

thus harmonics of the original spectrum can arise and $A_k = 0$ initially can change to $A_k > 0$ or $A_k < 0$ subsequently given a nonzero function F. Higher order approxs are possible, but we shall not consider them.

The function F will describe an interaction yielding a new wave vector $k = k^{'} + k''$. Basic physical results of the nonlinear interactions are as follows:

1. Neglecting viscosity, interacting longitudinal waves will not lead to the emergence of transverse waves.

Consequently, in the quadratic (second-order) approx, vortex-free motion only gives rise to vortex-free longitudinal waves.

2. Since the interaction of transverse waves gives rise to longitudinal waves, density perturbs can arise in the second order—as well as for higher orders. Therefore the amplitude of short-wave acoustic density perturbs generated is of order $\delta \sim (u/b)^2$, where u is the amplitude of the transverse motion. Given vortex (turbulent) motions in the prerecombination RD plasma, this turbulence leads to large density perturbs after recombination, when radiation pressure ceases to act on the matter.

3. The interaction of short-wavelength ($\lambda < \lambda_0$) longitudinal perturbs leads to long-wavelength longitudinal perturbs (including growing density perturbs). If the initial spectrum falls sufficiently steeply on the long-wave side, then this process is the most important source of density fluctuations on large scales. Under reasonable assumptions the newly created (by the nonlinearity) long waves satisfy $A_k \sim k^2$, leading to

$$\frac{\delta_M}{M} \sim \sqrt{(Ak^2 \times k^3)} \sim k^{7/2} \sim M^{-7/6}$$

This is explained below, but is evidently the extreme case of a huge number of short-wave perturbs combined with the initial absence of long wavelength perturbs.

4. Longitudinal waves with a wavelength $\lambda > \lambda_0$ for some λ_0 generate short waves by the nonlinear interaction, important for cosmology. In accord with the linear theory, viscosity suppresses long-wave perturbs corresponding to masses $M < M_J \sim 10^{12} M_\odot$. However, continuous generation maintains the perturbs in this wavelength interval. Shock waves may arise from these short waves. The emergence of a shock wave leads to entropy growth, whose fluctuations remain even after the longitudinal acoustic waves decay.

Nonlinear wave interaction theory is currently incomplete. But here are some properties of the general theory.

In an equilibrium situation, statistical mechanics says that each degree of freedom has the same energy θ; then $A_k \sim \theta^{1/2}$, must be a solution to the kinetic eqn corresponding to $dA_k/dt = 0$ on the left-hand side (LF—the kinetic eqn?).

The equilibrium laws do not depend on the specific interaction mechanism. Creation of waves $k = k' + k''$ must be accompanied by the reverse, keeping the A_k at their equilibrium values. The theories of turbulence and fluctuations are always concerned with nonequilib situations with amplitudes in a small part of phase volume (values of k much smaller than the inverse interatomic distance) much higher that thermal motion amplitudes. Therefore we are usually concerned with a stationary but nonequilib situation.

In diffusion eg, with turbulence or fluctuations, there is ordinarily a flow of energy to the short-wave side, where the energy transforms into heat. In situations where $A_k \to 0$ as $k \to 0$ in part of phase space, a flow of energy in the reverse direction is possible. Keep in mind that equilibrium is characterized by $\langle A_k \rangle$, which is quite appropriate because the phase of the A_k is random and the distribution of the $|A_k|$ is Gaussian.

But these assumptions do not always apply to the spectral distributions under consideration here. For "pancakes" there is a density distribution with a large amplitude for short waves, which are not random. They are correlated with the long waves, all of which together form the flat pancake! Consequently, in problems concerning the emergence of pancakes and problems concerning shock waves, the spectral approach is worse that the direct approach of analysis in spatial coords.

14.4

Consider the question of the formation of large-scale clusters from discrete objects. There are two possible causes. The first is the presence of small perturbs of large linear dimensions in the initial perturb spectrum, which grow via gravitational instability. When their nondimensional amplitude reaches unity, the matter has become clustered on the large scale. The second is due to nonlinear effects, which give birth to long-wave perturbs from short-wave ones.

Let the part of the spectrum of interest have an amplitude δ_M that is a "not too fast" decreasing function of M. In this case small units, discrete gravitationally bound bodies—are formed first but distributed in space nonuniformly due to the presence of weaker perturbs on larger scales.

Formation of larger units depends on the initial spectrum, independent in the first approx of the smaller scale. The growth of perturbs of $10^{15} M_\odot$ is the same for a gas of H atoms as for a gas whose elementary units are star clusters with mass of $10^5 M_\odot$. Assume the density perturbs at recombination satisfy

$$\delta_{rec} = \delta_\rho/\rho = b_0 \times M^{-\nu}$$

for $M > M_1$; and zero for smaller M.

For the growth law with $\Omega_0 = 1$,

$$\delta = \delta_{rec} \times [(1 + z_{rec})/(1 + z)] = \delta_{rec} \times (t/t_{rec})^{2/3}.$$

Formation of discrete gravitational bound bodies will be when $\delta(M, z) = 1$, so the first appearance of bodies with mass M_1 occurs at time t_1 or z_1 given by

$$\Delta = b_0 \times M_1^{-\nu} \times [z_{rec}/(1+z_1)] = 1,$$

yielding

$$z_1 + 1 = z_{rec} \times b_0 \times M_1^{-\nu}$$

The growth of mass occurs via the unification of smaller masses M_1 into a discrete bound aggregate with mass M, but this does not change the law

$$M(z) = M_1 \times [(1+z_1)/(1+z)]^{1/\nu}$$

Consider the case of an upper spectrum cutoff $\delta = 0$ for $M > M_2$. The short wave perturbs with $k > k_2$ (corresponding to M_2) in the process of growth are producing long-wave perturbs by nonlinear interaction . The Fourier amplitude of these new modes is $A_k \sim k^2$. The corresponding density perturbs in a volume containing a large average mass M, much greater than M_1, is given by

$$\delta(M) = (A_k^2 \times k^3)^{1/2} \sim M^{-7/6}$$

for $k = M^{-1/3}$

By dimensional considerations, we obtain from $\delta(M_1) = 1$ at $t = t_1$ the coefficient in this formula, leading to the final conclusion for the growth with time of bound masses in the expanding Universe:

$$M = M_1 \times (t/t_1)^{4/7} = M_1[(1+z_1)/(1+z)]^{6/7}$$

This law must be called a "self-similar clustering law" for every sufficiently steep initial spectrum (with $\nu > 7/6$) for the case of absent or very weak long-wave perturbs. The $\nu = 7/6$, $M \sim t^{4/7}$ law corresponds to the clustering of great masses, causally dependent by nonlinear effects on the early formation of small masses M_1. These considerations are especially important in the analysis of entropy perturbs, where the natural assumption is that the smallest units M_1 are of the order of the Jeans mass after recomb, of order 10^5 or $10^6 M_\odot$ if the H temperature is equal to the RR temp.

Another case depends on a random ("white-noise") initial fluctuation spectrum with $\delta \sim M^{-.5}$ and $\nu = .5$; though the white noise spectrum has no foundation in cosmology, it leads to a growth law $M \sim t^{4/3} \sim (1+z)^{-2}$ and $M_{max} \sim 10^{12} M_\odot$.

The simple clustering theory operates with the average bound mass only.

Thermal instability could lead to a separation of a homogeneous gas into phases, namely dense cold clouds and low-density hot gas between them. Though unrelated to gravity, its effectiveness decreases for long waves. Therefore it is characterized by an optimal scale size. This is not the mechanism for separation of the cosmological plasma into discrete bodies, but could be important for galactic situations such as the formation of stars.

15: Theories of Galaxy Formation

15.1 Introduction

Several theories for galaxy formation in the hot model are in competition. But common are the assumptions that in the RD period (before recombination) the plasma is almost homogeneous and that the present-day structure arises after recombination ($z < 1400$). First, we shall discuss briefly four explanations for the deviations from homogeneity. (I shall not repeat historical notes given by Z&N—LF),

First, adiabatic perturbs are defined by the property that the ratio of numbers of baryons to photons is constant. In the period when radiation is predominant, adiabatic density perturbs represent acoustic oscillations for $M \lesssim 10^{17} M_\odot$. After recomb, the perturbs grow in accord with the law governing gravitational instability. Taking account of photon viscosity and damping of acoustic oscillations leads to a characteristic perturb mass of $10^{13} M_\odot$, close to the observed mass of galaxy clusters. During the nonlinear stage of growth, disks (pancakes) arise and the gas is heated by shock waves. The great diversity of gas states allows explanations for such various objects as quasars and elliptical and spiral galaxies. There is a possible explanation for galactic rotation and magnetic field. Adiabatic perturbation theory uses rather small initial metric perturbs — 10^{-3} to 10^{-4} in dimensionless units — of the ideal cosmological model to obtain presently observed structure.

Second, perturbs of fluctuations in the ratio of baryons to photons are termed entropy fluctuations. In this model, the metric and the expansion dynamics do not "feel" any perturbs during universe evolution when radiation is predominant. Immediately after recomb, the neutral gas with $T \sim 4000K$ is not homogeneous. This temperature and the average density determine the Jeans mass $M_J \sim 5 \times 10^4 \times \Omega_0^{-.5} \times M_\odot$. Large scale perturbs grow scale independently if $M \gg M_J$, while smaller scale perturbs do not grow at all.

A falling initial mass spectrum for perturbs is most probable. Then objects with the least mass consistent with the Jeans threshold form first, this minimum mass being $\gtrsim 5 \times 10^4 \Omega_0^{-.5} \times M_\odot$; in other words, gas pressure limits this mass from below. Assuming instead that pressure is substantial leads to the formation of spherical bodies, not pancakes. We have tried to obtain the entire Universe structure from such entropy perturbs. Another tack suggested that objects with a mass of order $10^5 - 10^6$ transform into globular clusters, their unusual similarity previously unexplained. To explain the entire Universe structure, one could set the initial spectrum of entropy perturbs "by hand" so that the amplitude at about $10^{13} \times M_\odot$ is just that required to obtain at present $\delta_M/M \sim 1$. Then the results are roughly the same as obtained for adiabatic perturbs.

Third is the hypothesis of a charge-symmetric Universe, with annihilation where particles and antiparticles intermingle. However, the resulting gamma rays — from $p + \bar{p} \to \pi_0 +$ other particles, and then $\pi_0 \to 2\gamma$ — have not been observed in the interval between 50 and 200 MeV. There are other arguments against this hypothesis, which has several variants.

Purely in theoretical grounds based on charge symmetry, a symmetry that CP violation shows is not exact, there is no reason to choose a charge-symmetric initial state of the Universe or a charge asymmetric initial state.

One aspect of charge symmetry is connected with entropy fluctuations. Until we examine a plasma with photons and baryons, it is natural to assume that the baryon density varies while the photon density is uniform. When we consider times $t < 10^{-6}s$, when $kT \geq massp \times c^2$, and there are many baryons and anti baryons and as many photons, $n_B/n_{\bar{B}} = 1.00000001$ on average and appearances change. It is necessary to adopt this initial ratio ($n_B - n_{\bar{B}} \sim 10^{-8} n_B$) because those baryons present today are the survivors of the annihilation when the temperature decreased below $\sim massp \times c^2$. Undoubtedly the early stage of the hot Universe is quasi-=charge-symetric! But there must be a baryon excess and moreover, in the theory of entropy fluctuations this excess is not spatially uniform!

These fluctuations are very large compared with "statistical" fluctuations of order $N^{-.5}$ Indeed, $M = 10^6 \times M_\odot$ corresponds to 10^{63} baryons, so $N^{-.5} \sim 3 \times 10^{-32}$. But in the hot Universe with entropy we use $\delta_{n_B}/n_B \sim 10^{-10}$ before annihilation and $\sim 10^{-2}$ after annihilation. Although these numbers are very different, both are much greater than 10^{-32}. Although the fluctuations are not "statistical", one can suppose that the smaller the scale, the larger the fluctuations. Then the assumption is not excluded that on some scale the fluctuations lead to an average baryon surplus of 10^{-8}. This leads to $0.99999997 < n_B/n_{\bar{B}} < 1.00000005$. This means that at the conclusion of the annihilation process, there are regions of matter and antimatter in which $n_B/n_\gamma \sim 5 \times 10^{-8}$ and $n_{\bar{B}}/\gamma \sim 3 \times 10^{-8}$ respectively. Thus the entropy-fluctuation hypothesis is naturally linked with the hypothesis of charge symmetry.

Fourth is the vortex theory, whereby the plasma of protons, electrons and photons is in turbulent motion superposed on the general expansion. The velocity of this motion is or order $(0.05-0.1) \times c$ and the maximum scale of the motion is much less than the horizon at recombination. The turbulent theory predicts a definite spectrum for the motion and a condition connects the scale, velocity and time to establish the motion. One parameter, eg the maximum scale L for the correlation, fixes the entire picture. It is also presumed that turbulence is subsonic during the RD state, so radiation pressure prevents density deviating from the average.

An estimate of the pressure and density fluctuations follows from oppositely moving fluid volumes creating a pressure difference of order $\delta_P = \rho \times u^2$, small compared to $P = \rho \times b^2$, with b the sound speed and u the fluid speed. Correspondingly,

$$\delta_P/p \sim \delta_\rho/\rho \sim u^2/b^2 \ll 1.$$

A spreading out and breaking down of the flow thus occurs, so the motion hardly differs from that of an incompressible fluid. When recombination ensues, the pressure cuts off and the sound velocity goes to zero. The motion then continues in accord with its inertia. But the incompressibility is not preserved and collisions between gas clouds disturb the inertial motion. The clouds first heat because of compression and shock waves, but later they cool. Gravity then begins to act and regions of increase density transform into bound objects when $z \sim 130$ (in contrast to $z \sim 5$ in the adiabatic theory).

The most important feature of the vortex theory, the rotation of the galaxies and clusters is thus inherited from the initial vortex motion of the plasma before recombination. One difficulty is that it is incompatible with small perturbs to the Friedmann model. Another is the difficulty of reconciling the turbulence theory with RR observations with respect to this parameter (which I understand to mean the aforementioned red shift events—LF).

15.2 Begin here after July 20, presumably on August 24.2

Now considered is the evolution of perturbs after recombination ($z \lesssim 1400$), carried to the stage of formation of gravitationally bound isolated masses. Limit the analysis to perturbs arising from initially adiabatic perturbs. Five properties are important:

1. The motion is derivable from a potential, so the vorticity is and remains zero, until a shock wave is formed;

2. The initial perturbs are small at recomb, so that afterward the perturbs remain small and linear theory is valid for a prolonged period;

3. Linear theory is applicable before recombination and at an early stage the perturb amplitude is an approximately smooth function of wavelength; The wavelength interval for galaxy formation is not wide, with a power law applying for the amplitude:

   $$A_k \sim k^r. \text{ and } \sqrt{\delta_M} \sim M^{-\nu}$$

   $\sqrt{\delta_M}$ was introduced earlier so that the exponents r and ν are related by $\nu = 1/2 + 1/3 \times r$. Damping is also important and depends on a parameter $M_0 \sim 10^{13} * M_\odot$. Its meaning is that perturbs decay for scales corresponding to masses $M < M_0$ and they do not for masses $M > M_0$.

4. After recombination, a restructuring of the motion takes place whereby growing perturbs develop that have an amplitude

$$\sqrt{\delta_M} = M^{-n} \times e^{-(M_0/M)^{1/3}}$$

with $n = \nu + 1/3$

The formulae presented here are valid for scales such that the perturbs become acoustic waves before recombination, ie for $M < 10^{17} \times M_\odot$. The distinction between n and ν relates to the proportionality between δ_ρ and $u \times t/l$ for a growing mode in the postrecomb era.

5. After recombination one can neglect the pressure of the neutral gas when studying perturb evolution. The basis for this is that $M_0 \gg M_J$, where M_J is the Jeans mass for the neutral gas. As a result, the approx theory for the the growth of perturbs in dust is applicable.

By virtue of the first and fifth points, according to the approx theory, regions of infinite density arise. The nearest gas particles encounter the dense region and are stopped. We will therefore examine the motion of the gas, its heating by shock waves, and the subsequent fate of the compressed gas.

15.3 Shock Waves

{Math for shock wave description mostly too confusing for now—LF}

From a previous consideration of the approximate solution, infinite density occurs when this condition is met: $\alpha(q) \times (b/a) = 1$. This leads to a contraction of an ellipsoid (q is a vector) in one dimension to a flat figure — an ellipse aka pancake with infinite density. The mass of the ellipsoid grows in proportion to $(t - t_c)^{1.5}$, where t_c is the time when the density becomes infinite in the plane $qx = 0$ and $x = 0$. For a one-dim model with a single sinusoidal wave and with $\omega_0 = 1$, the density will be

$$\rho = \langle\rho\rangle \times [1 - \kappa \times (b_0/a_0) \times t^{2/3} \times cos(\kappa \times qx)]^{-1},$$

with $\langle\rho\rangle = 1/(6\pi G t^2)$.

Now define $\mu = \kappa \times qx/\pi$, a type of Lagrangian coord, which is the ratio of the mass included between the origin $qx = 0$ and a given qx to the total mass in half of a period cell.

In the general case, the amount of matter at each point of the pancakes surface grows as $\sqrt{t - t_{c'}}$, with different values of $t_{c'}$ at different points in the plane of the pancake, whose surface grows as $(t - t_{c'})$.

It is particularly necessary to notice what matter is subjected to shock compression. It is the matter which does not manage to achieve infinite density as the nearby layers approach and also the matter which is expanding. For $t \ll t_c$ (during the linear stage) half of the matter contracts and half expands. See Table I.

TABLE I. Characteristics of Shock-Wave Compression

t/t_c	z/z_c	μ
1.0	1.0	0.0
1.025	0.983	0.1
1.33	0.83	0.333
1.96	0.64	0.5
6.1	0.30	0.76
28	0.1	0.9
Infinity	0.0	1.0

A computation gives, conveniently in terms of μ as the independent variable, the speed u at which the shock wave strikes the matter and the matter density in front of the shock wave ρ. Characteristic is the behavior for small values of μ, equivalent to small values of $t - t_c$ or $z_c - z$:

$$u \sim \mu \sim (t - t_c)^{.5} \sim (z_c - z)^{.5}$$

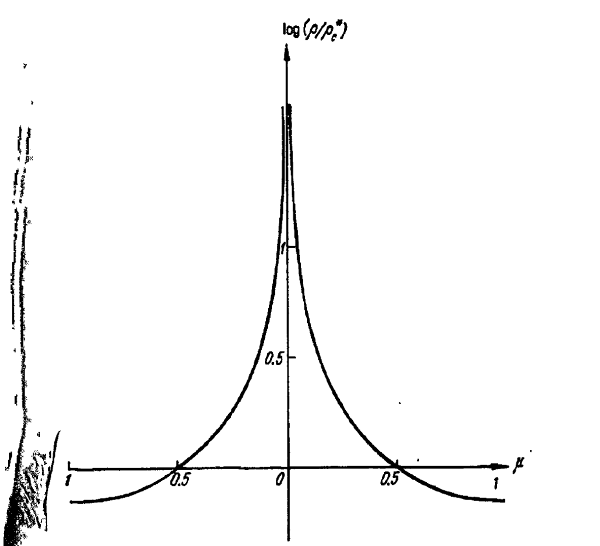
and

$$\rho \sim \mu^{-2} \sim 1/(t - t_c) \sim 1/(z - z_c)$$

and P approximately constant.

Consider the opposite limit, the complete absence of heat losses and transfer. Complete computations require numerical methods, but the asymptotic behavior at the beginning of the contraction, when $\mu \ll 1$, can be found simply. The pressure has a definite limiting value, larger than the first case by 11/9. = 1.22, in the same limit, the amount of matter compressed by a hot shock is $(11/9)^{.5} = 1.1$ times larger that the amount of matter compressed by the shock with instantaneous cooling of the gas. This is because the in the hot case the shock propagates from the plane $x = 0$ toward the infalling matter.

The average density of the compressed matter is 12 times larger that the matter density in front of the shock before compression, ie, three times greater that immediately after the shock. Figure 46 shows the density distribution at t_c, when $\rho = infinity$ is achieved in the plane $x = 0$. Figure 47 gives the distribution at time $t = (7/6) \times t_c$, when about 1/4 of the matter is compressed. Figure 47 applies to rapid cooling, and Figure 48 to the adiabatic case. 1 h



o. 46.—The density distribution (density ρ as a function of Lagrangian coordinate μ) developing cluster at the time $t = t_c$. At the location $\mu = 0$, the density is infinite; q. (15.3.6).

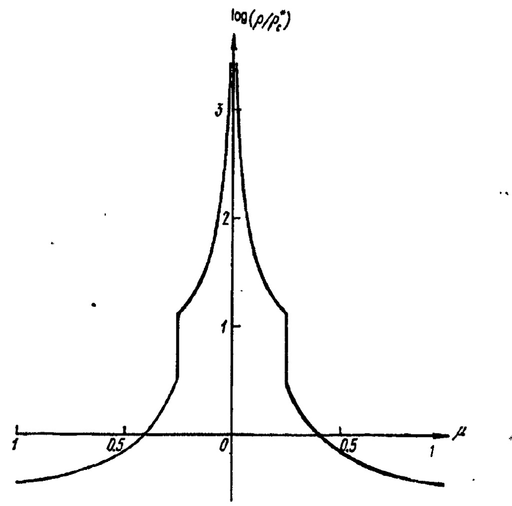


FIG. 48.—The same as Fig. 46, but for adiabatic compression. In this case, with cooling absent, the density does not become infinite after the compression.

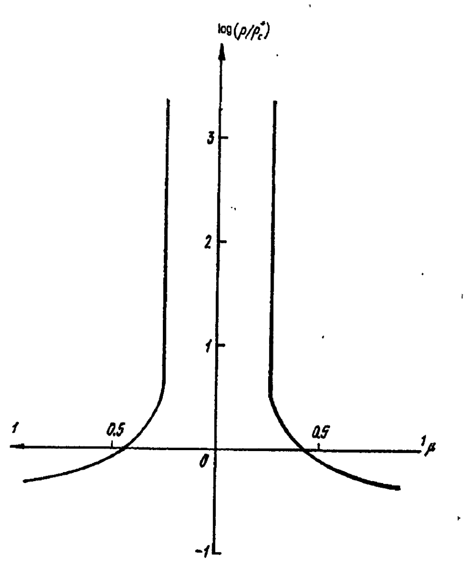


FIG. 47.—The same as Fig. 46, but after rapid cooling—at $t = 7t_c/6$, when the shock wave has just reached the Lagrangian coordinate $\mu = 0.25$. In Eulerian coordinates the region $\rho = \infty$ is compressed into a single point.

NOTES BEGINNING AUGUST 30

From these formulae, when 10% of the matter is compressed by the shock ($\mu = 0.1$), this matter occupies a fraction $\alpha = 3 \times 10^{-4}$ of the volume, so its density is 300 times greater than the average. If instead 30% 0f the matter is compressed, $\alpha = 10^{-2}$ and the density is 30 times the average. The calculation has assumed strict adiabatic behavior, hence without heat loss. The true value of α for a given perturb type can only be smaller and the true density larger that the values presented.

In the future of the process, a mutual attraction among the compressed matter layers switches on, leading to a gravitational pressure. Numerical calcs show that when $t \sim 4t_c$, the process dies away. The compressed matter is held in place by gravity, the falling of matter is suspended, and the shock weakens. An unrealistically large time would be needed for the shock to compress even 90% of the matter. Meanwhile, the new processes of radiation, star birth, etc are beginning, and this schematic picture of a one-dimensional adiabatic compression does not account for them.

15.4 Thermal Processes in the Compressed Gas

In the adiabatic approximation, the temperature of the gas compressed by the shock front is

$$T = 2.5 \times 10^6 \times z_c \times \mu^2 \times [M/(10^{13} \times M_\odot)]^{2/3}$$

for asymptotically small μ, for a sinusoidal wave, for zero initial temperature, and in the absence of heat losses and transfer. Now renounce the simplifying assumptions describing the temperature distribution and the thermal processes. But start from the pressure in the adiabatic picture. In the unperturbed gas, the compression leads to a finite density and a finite temperature.

Examining the processes in a homogeneous universe, we found that thermal exchange between matter and radiation ceases when $z \sim 200$, $T = 540K$, and $\rho \sim 0.8 \times 10^{-22} g/cm^3$ for $\Omega_0 = 1$. Subsequently, the pressure, temperature and density are connected by the adiabats (isentropic curves) of a monatomic gas, so $\rho \sim P^{3/5}$ and $T \sim P^{2/5}$. Thus, when $z + 1 = 5$, $T = 0.34K$ for the unperturbed gas (when the radiation temperature is 14 K) . Upon compression to $P = 1.44 \times 10^{12} dyn/cm^2$, corresponding to the pressure in a pancake if mass $M = 10^{13} M_{\odot}$, we obtain T = 400 K and $\rho = 0.5 \times 10^{-22} g/cm^3$, compared to the average matter density then of $\rho \sim 10^{-27} g/cm^3$. We thus obtain the thickness of an adiabatically compressed layer.

Heat losses are computed next, but I shall only summarize the results. Figure 50 shows the density distribution. The temperature distribution exhibits the opposite behavior in the sense of a reflection through the mu axis.

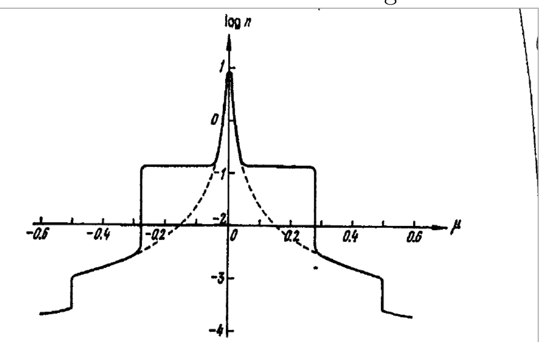


FIG. 50.—The density distribution in a developing cluster at the same moment as in Fig. 49. The number density of particles n is shown as a function of the Lagrangian coordinate μ. The solid curve takes account of radiative cooling, while the dashed curve is for the case of no cooling.

1. About 1% of the mater is subject only to adiabatic compression and has a very low temperature.

2. About 2-3% of the matter is shock heated to $10^2K - 10^4K$. It is not ionized and cools slowly thorough adiabatic expansion.

3. About 20% of the matter is shock compressed to $T > 10^4K$. Then it cools to 10^4K by radiating, and to a large degree, recombines.

4. About 25% of the matter is shock heated to $T > 5 \times 10^5K$ up to $\sim 1.5 \times 10^6K$ and remains hot.

5. Half of the matter is not subject to the shock action.

Indeed, it is observationally known that neutral H is practically absent between clusters. Convincing proof follows from quasar spectra with z ¿ 2. Neutral H would totally absorb shortward of the Lyman-alpha line, but this radiation falls in the spectrum accessible to ground telescopes. Independent of the detailed pancake theory, it is inconceivable that there is no matter between galaxies, so it is must be completely ionized. A suitable explanation for this ionization is on e of the main problems facing cosmological theory.

Possibly the ionization is accomplished by the earlier quasars with $z >\sim 4$, not yet observable directly {SUPERSEDED?—LF}. Perhaps young bright galaxies are important. In both cases the radiation comes from the first pancakes. In any event, the calculational results depend significantly on the fundamental cosmological parameters, the Hubble constant and the matter density.

Furthermore, the hot compressed gas clouds — "pancakes" — cannot be observed directly as protoclusters in optical and ultraviolet against the background of other nearby sources. X-rays might be observable from the statistically rare largest clusters.

Finally, it is apparently possible to observe protoclusters via 21 cm radiation from the neutral H in the pancake region where the temp is too small for ionization to occur. The redshift of the line and the radiation's association with a "pancake" would suggest the pancake's epoch. The observational difficulty is related to the lack of knowledge of the absolute wavelength, unknown now because of the unknown redshift value at which the compressed gas was formed.

15.5 Cluster Masses and Protocluster Fragmentation

One ought to expect a distribution of masses for protoclusters and dependence on the form of the of the perturbation spectrum, i.e., the amplitude change versus wavelength. The mass distribution will be considered here, as well as the fate of the isolated gas bodies (protoclusters) already described. The gravitational instability of the protoclusters leads to further fragmentation into galaxies and quasars.

To determine the time when nonlinear effects become substantial in a significant fraction of the matter, i.e. when a significant part is compressed into pancakes, it is enough in the first approx to fined the fraction of the matter satisfying $\alpha \times b/a > 1$. The one-dimensional hydro calcs showed that the amount of matter shock compressed exceeds the amount of matter that, by itself satisfying the same condition in the absence of the shock, approaches infinite density. By integrating over the distribution function, one can find the fraction of the matter compressed as a function of time, which appears in Figure 51. In the figure, t_c is the time when $(\delta_\rho/\rho)^2 = 1$ as calced via linear theory. The dashed curve for cold clouds is distinguished from the solid curve primarily by the statistical nature, not by the one-dimensional nature.

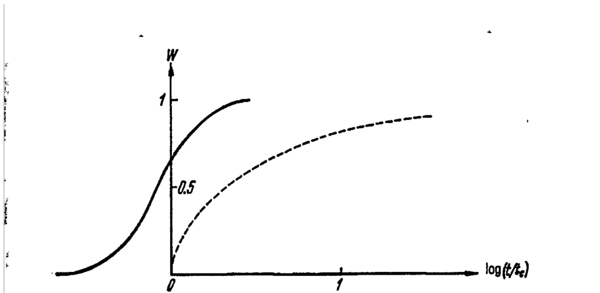


FIG. 51.—The time evolution of the fraction of the matter compressed by the shock wave. The solid curve is the result of the statistical theory, while the dashed curve is for a one-dimensional problem with a sinusoidal perturbation.

More detailed calls show that the thickness of the typical pancake (in Lagrangian coords) is of order R_c, while the area is $\sim 30R_c^2$.

Consider the fragmentation of a protocluster. The peculiarity of the problem is in examining the gravitational instability of a thin layer of matter with a thickness many times less than its longitudinal size. Also, the dense, cold matter is subject to the external pressure of the surrounding hot matter, with much lower density. During the initial pancake growth, this external pressure is greater than the gravitational forces within the dense layer. Though the general problem of a thin disk has been much studied, the problem here is simpler because there is no rotation and no peculiar velocities in the disk plane.

For a homogeneous disk with given surface density sigma (in g/cm^2), assume

$$\delta_\sigma/\sigma \sim e^{(w\times t+ik\times x)},$$

with k and x two-dimensional, while the gravitational potential outside the disk is three-dimensional and satisfies the Poisson equation in empty space. Thus the solution requires a suitable boundary condition at the disk.

Accounting for the nonzero thickness of the disk and the pressure, we obtain

$$w = [2\pi \times G \times \sigma \times abs(k) - P \times h \times k^2/\sigma]^{1/2}$$

Thus, for a disk of nonzero thickness, there is a critical wavelength analogous to the Jeans wavelength for the 3-dimensional problem:

w = 0 at $k = k_j = 2\pi \times G \times \sigma^2/(P \times h)$; and $\lambda = 2\pi/k_j$.

There is also an important quantity not analogous to anything in the 3-d problem. This is the wave vector k_m and the wavelength λ of the most "dangerous" (most rapidly growing) perturb, for which w is a maximum. It satisfies $k_m = k_j/2$ and $\lambda_m = 2 \times \lambda$. If the pressure in the layer is determined only by the weight through gravity of the matter in the layer, then $P_c = \pi \times G \times \sigma^2/2$ in the middle of the layer; at the edges, $P_c = 0$. The average pressure is of order $G \times \sigma^2$. With $\sigma = h \times avg(\rho)$, this leads to

$\lambda_J = h$, $\lambda_m = 2 \times h$, and $\omega = 0.885 \times [4\pi \times G \times avg(\rho)]^{1/2}$

An application of this theory is to the disintegration of a planar protocluster — a pancake — as just described. The inner, most dense (adiabatically compressed) layer must disintegrate into masses of $\sim 10^7 - 10^8 M_\odot$. A layer that cools to $10^4 K$ disintegrates into masses $\sim 10^{11} - 10^{12} M_\odot$. These latter masses are naturally identified with galaxies.

A hot gas with $T > 4 \times 10^5 K$ remains gravitationally unbound in part and in part forms a hot halo around masses of $10^{11} - 10^{12} M_\odot$, leading to a hypothesis that the most dense but small masses represent quasars and galactic nuclei.

With an initial density of 30 atoms / cm^3, the average density can grow with adiabatic compression. After the formation of the first pancakes, the part of the gas enveloped by the shock wave is heated to a high temperature. The hard radiation from theses layers then heats the gas to at least $\sim 10^3 K$, but this does not affect the general picture of the pancakes' evolution. For pancakes formed later, the middle cold layer is then not so dense and cold as was the case for the first pancakes. One might then assume that the formation of quasars (and galactic nuclei?) ceases earlier that does the formation of galaxies. Possibly the sharp fall in the number of quasars and radio sources for $z > 3$ or 4 is related to this circumstance.

In the first pancakes, about 1% of the matter passes, by assumption, into quasars and galactic nuclei. This agrees with rough estimates derived from the assumption that the active life of quasars and quasi-stellar radio-quiet objects extends over $10^5 - 10^6$ years.

In every pancake with a dense layer, one can expect the formation of about 100 dense objects. In such a case, must the quasars be arranged in groups? Observations do not suggest this {STILL THE CASE?—LF}, and it should not be so. This follows if one keeps in mind that the quasar active period is $10^5 - 10^6$ yr, many times less that the 10^{10} yr cosmological age and the formation time of the pancakes ($3\times10^8 - 10^9$ yr). At each time we observe $\sim 10^{-3}$ of the total dense objects — potential quasars — so that the average number of observed quasars for each protocluster is $\ll 1$. Expected pairs of quasars must constitute ¡ 10% of the total number.

Regarding galaxies, the temperature in the layer from which galaxies presumably form is relatively stable at $\sim 10^4 K$; this is determined by the features of the recombination law and of the radiation from the optically thin gas. The gas radiates most when partially ionized, whence the emission is mostly in lines and in recombination radiation. After recombination, the emission decreases sharply. Here the temperature that the gas achieves depends weakly on the history and mass and formation time of the pancake. This temperature stability might well reflect the same scatter in the density of galaxies from the average density.

15.6 Galaxy Rotation

Is the emergence of rotation possible in a theory if the initial perturbs are vortex-free? The Kelvin-Helmholtz theorems say that gravitational forces and radiative pressure are capable during the RD stage of creating only a potential-type velocity field, i.e., a field with zero vorticity. Before recomb, the density and pressure are uniquely related to each other and the pressure forces do not give rise to vortex motion. Finally, viscosity does not give rise to vorticity in potential-type motion.

In principle, the emergence of vortex motion is not excluded when potential-type motion leads to shock waves. However, the effect is likely to be small and we conclude that adiabatic perturb theory probably leads to vortex-free motion at recombination.n.

The second aspect of the question concerns the relation between the body's angular momentum and its vorticity. The problem of the emergence of angular momentum in vortex-free motion is connected with violation of axial symmetry or to variable density. A general property of rigid-body rotation is that it represents motion with a minimum KE for a given total angular momentum, so that a body isolated from external forces tends to rigid-body rotation as time passes, absent internal sources of energy. But rigid-body rotation is vortex motion. Viscosity evidently must give rise to a transition from potential-type motion with angular momentum to rigid-body rotations; i.e., the viscosity must give rise to vorticity. It does so by transforming into heat the surplus of KE exceeding the energy of rigid-body rotation with the same ang mom. Vorticity arises first at the boundary and gradually encompasses the whole volume.

Further discussion is now necessary of the theorem of the conservation of of vorticity and the relation between vorticity and ang mom. Three situations will be considered to explain how rotation might arise in galaxies and clusters.

A. Early Postrecombination Period of Small Adiabatic Perturbs

If we take account of the deviation of the density from the average value, then in the approximation quadratic with respect to perturb amplitude, the ang mom of a sphere differs from zero. For a density perturb amplitude of order $\delta_\rho/\rho \sim 1$, the speed is $u \sim l/t$, where l is the scale of the motion and t is the cosmological time. Thus

$$AngMom \sim \delta_\rho \times u \times l^4 \sim \rho \times l^5/t.$$

Substituting $t = (6\pi \times G \times \rho)^{1/2}$ yields

$$AngMom = G^{1/2} \times \rho^{3/2} \times l^5 \sim G^{1/2} \times M^{3/2} \times l^{1/2},$$

where $M = \rho \times l^3$. But this is just the ang mom for which the centrifugal force keeps the mass M in a disk of size l.

B. Shock-Wave Induced Compression of the Cooled Gas at $z < z_c$, after Recombination

In the nonlinear picture, perturb evolution leads to strong shocks; in this situation conservation of vorticity no longer holds. Formally, this is connected with the growing entropy of the shock. When, after the shock passes, the gas cools from about $4{*}10^5 K$ to $10^4 K$, its vorticity is not conserved. In one example, a shock wave coinciding with the plane x = 0 compressed the gas by a factor of 4, so that the motion parallel to the x axis is slowed by this factor behind the wave front. Examination of the governing equations shows that the compensation effect expressed by the equation $\partial u_y/\partial x - \partial u_x/\partial y = 0$, which leads to vortex-free motion before the shock compression, is completely absent after the compression. A rough approach then gives the estimate $\nabla \times u = H \times \rho/Avg(\rho)$, where H is the Hubble parameter at the time of compression. This is based on the nonlinear compression when the gradients of the velocity perturbs in one direction locally exceed the Hubble parameter. For a particular set of values, this leads to an angular velocity and rotation period of

$\nabla \times u = 1/[7.5 \times 10^5 yr]$ and $\tau = \pi/[Abs(\nabla \times u)] = 2.3 \times 10^6$ yr.

But this rotation period is 100 times less than what is observed!

C. Gravitational Interactions of Discrete Bodies

Consider now the last stage, when the perturbs lead to spatially separated discrete bodies interacting gravitationally. For point or spherical bodies, momentum could be exchanged, leading to changes in the direction and speed of motion. An irregularly shaped body has a nonzero quadrupole moment. This could lead to a change in the orbital angular momentum of both bodies upon encounters. Because part of the external field of a body with a quadrupole moment decreases as $1/r^3$, closest encounters with the other body dominate the interaction.

The torque acting on a given body with mass M_1 and quadrupole moment Q_1 due to a second body with Mass M_2 is of the order of

$$Abs[d(ang.Mom.)/dt] \sim G \times Q_1 \times M_2/r_{12}^3.$$

For passage at the impact parameter b^*, effectively $r_{12} = b^*$ and $t \sim b^*/u$, where u is the speed, so

$$Abs[\delta(ang.Mom.)] \sim G \times Q_1 \times M_2/[(b^*)^2 \times u]$$

A statistical estimate for chaotic collisions follows, leading to a rough result for the process of separating galaxies of

$$\delta(ang\ mom) = (G \times M^3 \times R)^{1/2},$$

where M is the galactic mass and R is the galactic size. Note that this is just the ang mom necessary that the centrifugal force of mass M and radius R counterbalance the gravitational attraction force. An initial quadrupole moment could arise by the tidal effect of the second body on the first.

Begin here after August 31, presumably on September 15

15.7 Galactic Magnetic Fields

The origin of a galaxy's magnetic field is related to the dynamo effect, the strengthening of a field caused by the plasma motion. Previously believed insurmountable difficulties with the dynamo effect led to the proposal of a specific type of singularity and a comparatively strong primordial magnetic field frozen in the relic plasma. A galactic field of about 10^{-6} gauss is obtained by compression of the magnetic lines of force when the dilute gas condenses into galaxies. Between galaxies, a field of $\sim 10^{-9}$ gauss must remain. This formulation now appears artificial.

Consider therefore theories describing the generation of fields within a hot model of the Universe with no primordial field, but with small density and velocity perturbs superposed on a homogeneous and isotropic Friedmann solution.

Conservation of ang mom for a sphere of radius R yields the condition

$$Ang\ mom = I \times w = (8\pi/15) \times \rho \times R^5 \times w = const$$

During the expansion, R grows, but the total density falls as $\rho \sim R^{-4}$ since the photon density dominates. Therefore the angular velocity decreases as $w \sim 1/R$ while the linear velocity $u = w \times R$ remains constant. The matter density decreases as $\rho \sim R^{-3}$. By virtue of ang mom conservation, therefore, as the expansion proceeds, the matter alone "would like" to rotate more slowly, with $w \sim R^{-2}$ and $U \sim 1/R$.

Under these conditions, the light electrons are dragged along by the radiation, but the heavy protons fall behind the rotation of the RD plasma. Given this difference, there arises an electric current and consequent magnetic field H directed along the rotation axis. But this in turn leads to an electromotive force that tends to equalize the electron and proton velocities. In the first approx, we assume that the two velocities differ little. We shall then find the values of H and dH/dt necessary so that the protons do not lag behind the electrons. Then we shall verify the difference between the velocities required for the creation of such a field is actually small. Analysis of the equations of motion for the protons leads to an asymptotic solution for the magnetic field

$$H = -2 \times m_p \times c \times w/e$$

In the vortex theory, the motion is examined for which the scale of the largest vortex equals the product of the velocity and the cosmological time, which is also the condition for the realization of turbulence. Thus $w \sim 2\pi/t$, where t is the cosmological time, which at recombination (or at the end of the RD period) is $\sim 3 \times 10^{12} s$. With this value, we find that $H = 0.4 \times 10^{-15} gauss$ when $\rho_m \sim 3 \times 10^{-20} g/cm^3$. The density of matter in galaxies if less by $\sim 10^4$. If the expansion continues, then the frozen-in condition — the conservation of the magnetic flux — leads to a later decrease of H by an additional 10^2 to a value of 2×10^{-18} gauss.

A different mechanism leads to a still smaller value of 10^{-23} gauss

Thus the field due to rotation turns out to be quite small in comparison with that observed in the Galaxy, $\sim 10^{-6}$ gauss. A satisfactory solution of the galactic magnetic field problem is thus impossible without making use of the dynamo effect, which can produce exponential growth with time. Exact theorems on the impossibility of a dynamo have been proved for axisymmetric and planar motion.

One simple nonplanar situation illustrating rapid growth is the following: Imagine a torus with cross section S and radius R with a field H_0. While preserving its volume, stretch the torus to a radius $2R$; the cross section decreases $\rho R/2$ but the field grows to $2H_0$ because of flux conservation. Now bend the torus into a figure eight and then fold the two rings into one torus. The linear dimensions return to the original, but the field is now $2H_0$. This operation can be repeated indefinitely, with exponential growth time if each cycle requires the same duration. One possible governing equation that gives rise to an exponential solution is

$$dH/dt = \nabla \times (\alpha H)$$

where "X" denotes the cross product. An absolute upper limit on the growth rate of the field is $e^{(wt)}$, where w is the angular velocity. With $w \sim 3 \times 10^8/yr$ and $t = 10^{10} yr$, such a law would give an increase of $e^{(300)}$, more than sufficient. "Only" $10^{10} - 10^{20}$ is needed.

There is no reliable theory, but neither is there a blind alley. . . .

15.8 The Theory of Entropy Perturbations; & 15.9 The Vortex Theory

While explained in the text, I am skipping these topics in these notes because they are disfavored

by the authors as explicitly stated at the end of Section 15.10: . . . From this account it should be apparent that the present authors prefer the adiabatic theory of galaxy formation. . .

============================

15.10 A Comparison of the Evolutionary Theories of Galaxy Formation

The previously stated theoretical assertions are consequences of the initial conditions posited for an early time, lacking a theory for choosing these initial conditions. Very likely only one set of initial conditions, i.e., types of perturbations are likely, though this is a matter of personal taste.

In principle, this choice could be determined by comparing the evolutionary calculations with observational data about the characteristics of galaxies and clusters, as well as about the RR and nucleosynthesis.

The subsequent discussion about galaxy formation in this section will be based on all of the information developed previously in this book and that to be presented later, however inconvenient this may be.

Though at present there is no generally accepted theory, we will give our preference with the reasons behind that choice.

The first choice is that between processes involving the condensation of a dilute gas versus the explosion of a superdense body. The latter should not be confused with the overall expansion of the Universe, but rather the expansion of a "white hole." However, consider our own Galaxy, with a mass of $\sim 2 \times 10^{11} M_{\odot}$, and an ang. mom. of $\sim 2 \times 10^{74} gcm ** 2/s$; this corresponds to a circular velocity of 300 km / s at a radius of 10 kpc. But if the Galaxy formed from an initially superdense body with a radius of $\sim 1pc$, the ang. mom. would have corresponded to a linear velocity at the equator greater than c, a glaring contradiction. Moreover, the body as a whole would have been inside its gravitational radius.

This objection is invalid with respect to quasars, because their ang. mom. is unknown. On the other hand, the whole idea of exploding white holes has the serious difficulty that the QM process of particle creation near the singularity leads to the early explosion of the white hole [???-LF] or to no explosion at all.

These remarks do not concern ideas about the present activity of quasars or galactic nuclei, though they limit the notions of a possible source of of their energy.

Thus the viewpoint of galaxies arising from a dilute gas is dominant. Then the remaining question is the type of initial perturbs of the homogeneous and isotropic Friedmann model, i.e., between the galaxy formation theories set forth in the preceding sections.

To begin, consider perturbs near the singularity. Adiabatic and entropy perturbs are compatible with small metric perturbs, with a quasi- isotropic solution. Near the singularity therefore, such a solution is not distinguished by its local properties from a strictly isotropic and homogeneous solution, so the nucleosynthesis results are the same. In particular, the mass abundance results for He of $\sim .26$ and for D of less than .0001 are supported by observations,

The vortex theory requires a substantially anisotropic singularity, thereby changing the expansion law then. This leads to quite different nucleosynthesis results, probably disagreeing with observations: either > 0.3 for He or $\sim 2.5\%$ for D and 97.5% for H. To avoid the need for unusual initial conditions, it is much more plausible to assume that all types of perturbations are equally represented since a solution with a non-Friedmann-like beginning approaches a Friedmann form. In that case, vortex perturbs decay rapidly, entropy perturbs remain constant, and adiabatic perturbs grow. Thereafter, the evolution follows the tracks of the theory of perturbs.

Suppose we do not consider the period from the singularity until nucleosynthesis. Consider instead the later period from the end of e-e+ annihilation to the present. This is the period $10^8 > z > 0$, including the epochs of the RD plasma, He and H recombination, perturb growth in the neutral gas, structure formation, and the secondary ionization for the neutral gas not a part of gravitationally bound objects. Begin with a comparison of the theory and RR observations.

A distinctive feature of the vortex theory is the assumption of a large perturb amplitude of a large velocity for the vortex motions superposed on the Hubble expansion. Another feature, unavoidable, is that vortex motion does not give rise to density perturbs immediately after recomb.

In one variant, it was assumed that the velocity of vortex motion at recomb is of the order of $0.1 - 0.4c$. The corresponding RR temperature fluctuations due to the Doppler effect are evidently $\Delta T/T = 0.1 - 0.4$. But what is observed is $\Delta T/T < 3 \times 10^{-4}$, and other effects could not have eliminated the incompatibility.

A slow motion variant with a comparatively low initial vortex velocity of $u/c \sim 0.03$ leads to effects that do not agree with the observed RR isotropy. And what about the attractiveness of the vortex theory? Is there not an element of arbitrariness? Are there other consequences?

To explain the current structure of the Universe via adiabatic perturbs, the metric perturbs must have an amplitude of order 0.001 to 0.0001 on a scale corresponding to a mass $M \sim 10^{13} - 10^{14} M_{\odot}$. One can continue the perturb spectrum smoothly with the same amplitude to higher and lower masses. Perturbs of longer wavelengths, corresponding to much higher masses, grow more slowly and result in $\delta\rho/\rho << 1$ today. They also produce RR perturbs of order 0.0001, which do not contradict observations. Perturbs corresponding to much lower masses decay and do not affect the creation of galaxies. If the amplitude of these perturbs is of order 0.001 to 0.0001, then upon decay no observable effects remain.

Thus in the adiabatic theory, one can choose an initial spectrum without any preferred scale by giving one characteristic quantity—the amplitude—and leaving the scale of galaxy clusters to the the natural laws. The vortex theory lacks this simplicity, requiring a special choice of initial spectrum.

From this account, it should be apparent that the present authors prefer the adiabatic theory of galaxy formation.

Not discussed here is the entropy theory, which in recent years has been made more attractive through new observations and theoretical analysis.

15.11 Observational Data Concerning the Properties of Galaxies and Galaxy Clusters; The Average Density of Matter in the Universe

There are many sources of data, but the parameters of galaxies and clusters have not been well determined and error estimates are very subjective.

One parameter is the luminosity function, with the following cited by Peebles. Given MG as the absolute magnitude, then the average number of galaxies per unit volume brighter than MG is

$$n(< MG) = A \times 10^{al} \times MG \text{ with } al = 0.75 \text{ for } MG < MG^*$$

$$n(< MG) = B \times 10^{bt} \times MG \text{ with } bt = 0.25 \text{ for } MG > MG^*$$

For $MG = MG^*$, the two formulas must give the same answer, so

$$MG^* = -19.5 + 5 log h_0,$$

with $H_0 = 100 \times h_o km/s/Mpc$.

To astronomers, the most reliable path to galaxy mass determination follows measurements of the rotation speed and their linear size via Newtons's law for irregulars and spirals. For ellipticals, the masses can be determined via the dispersion in the line-of-sight star velocities. The mass function appears in Figure 52 and the overall mass contributions in Table 12.

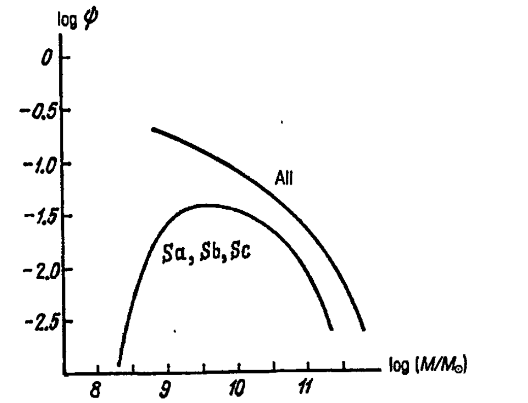


FIG. 52.—The mass function for galaxies of various types

TABLE 12

PERCENTAGE OF THE TOTAL MASS CONTRIBUTED BY GALAXIES WITH VARIOUS MASSES

	$M_\odot/M$			
	$<10^9$	10^9–10^{10}	10^{10}–10^{11}	$>10^{11}$
Percent contribution . . .	3	11	40	46

The mass-to-luminosity ratio M/L in solar units is about 5 for irregulars and about 7-20 for spirals and about 40 for ellipticals.

Radio observations indicate that the rotation curves of most spirals are very flat: V(R) = const to galactocentric distances of 50 kpc. Hence the cumulative (inner) mass satisfies

$$M(R) \sim V(R)^2 \times R \sim R$$

On the other hand, the spatial luminosity density in galaxy disks is well represented by an exponential model

$$I(a) = I_0 e^{(-a/a_0)},$$

So the cumulative luminosity

$$L(R) \sim \int_0^R I(a) a^2 da$$

converges rapidly to the total galaxy luminosity L_{tot}. The cumulative ratio M(R) / L(R) has a value 3 -10 in solar units inside of the optically visible spiral boundaries depending on the stellar population.

The diameters of the brightest galaxies are $\sim 30kpc$, while dwarfs are much smaller. Galaxy angular momentum is known poorly.

The % distribution of galaxies by type is as follows:

TABLE II. Distribution of Galaxies by Type

E and S0	Sa.	Sb	Sc	Irr.	Others
22.9	7.7	27.5	27.3	2.1	12.5

Properties of galaxies are closely related to the properties of their respective systems. In rich clusters, most galaxies are type E or S0, explainable by the ram pressure of the intracluster gas. In poor clusters, ellipticals tend to populate inner regions; spirals and irregulars, the other regions.

The velocity dispersion of stars in the main galaxies of systems is equal to the main velocity dispersion of the galaxies, indicating that systems of galaxies form together with galaxies, since a later adjustment of dynamics takes too long.

The best studied clusters are regular clusters. For example, the size of the Coma is about 4 Mpc, with about an estimated several tens of thousands of galaxies. The velocity dispersion along the line-of-sight is about $\delta V = 1000 km/s$, perhaps falling to half that value from the center to the edge of the cluster. By means of the viral theorem, one can estimate the cluster mass to be the very large $3 \times 10^{15} M_{\odot}/[H/(75km/s/Mpc)]$

By counting the galaxies, one can determine the integrated luminosity and then the ratio M / L. One estimate is $300[H/(75km/s/Mpc)]$, which is many times larger that the M/L for ellipticals, the type with the largest M/L. This signifies either (1) the cluster is nonstationary; (2) there is "invisible mass" between the visible parts of galaxies; or (3) there are systematic errors in the observations and their interpretations.

[The observations of Rubin et al. are neither mentioned nor cited!]

Begin here next, presumably on November 15

Practically all of the groups of galaxies studied turn out to be unstable if the viral theorem is applied to them; their kinetic energy based on the measured line-of-sight velocities greatly exceed the potential energy calculated on the basis of their masses. Thus all such systems turn out to be unstable, as Table 14 illustrates.

TABLE 14

AVERAGE CHARACTERISTICS OF SYSTEMS OF GALAXIES

Galaxy Systems	Luminosity ($10^{10} L_{\odot}$)	Average Number of Members	Radius R (10^4 pc)	Dispersion of the Line-of-Sight Velocities (km s^{-1})	Observable Density (10^{-25} g cm^{-3})	Virial Density (10^{-25} g cm^{-3})	Dispersal Time $R/\Delta V$ (10^8 yr)
Triplets	4.9	3	8.8	121	0.10	1.0	4.5
Groups	10.5	8	39	287	0.002	9×10^{-2}	11
Poor clusters	43	35	114	354	4×10^{-4}	2×10^{-2}	20
Rich clusters	270	220	271	827	2×10^{-4}	2×10^{-2}	18
Superclusters	1500	1200	1640	1100	4×10^{-6}	6×10^{-4}	87

411

At the same time, the average density of matter $<\rho>$ in the Universe turns out to be much less than the critical density. Oort made the best early determinations of $<\rho>$ based on the average luminosity per unit volume in the Universe by means of galaxy counts and the use of the $<M/L>$ for galaxies. If this value is $15 \times [H/(75km/s/Mpc)]$, then $<\rho> \sim 2 \times 10^{-31} \times [H/(75km/s/Mpc)] g/cm^3$, so

$$<\rho> /\rho_{crit} \sim 0.02.$$

The existence of this paradox concerning the instability of galaxy clusters has led to the assumptions that a large mass of invisible matter is present sufficient to gravitationally bind the clusters. And then $<\rho_p>$ would be closer to ρ_{crit}. One candidate for such "invisible mass" is intergalactic ionized gas. But theoretical considerations concerning gas accretion onto rich clusters, along with observational data about the clusters leads to the conclusion that $\omega <= 0.05$ for such gas.

Observations still cannot exclude the possibility that a significant amount of gas with $T \sim 10^5 - 10^7$ K is present in the so-called de Vaucouleurs groups of galaxies, but the presence or absence of this gas has no relation to the problem of the stability of rich clusters.

Recent observation have revealed a considerable amount of neutral gas near galaxies, including ours. Some of this gas forms long streams connected to companion galaxies.

Quite recently has arisen the idea of huge halos around galaxies of faint stars or other dark matter. This matter would not influence the dynamics of the internal parts of galaxies. Recent studies of the motion of companions of giant galaxies, leading to estimates of of $<M/L> \sim 65 - 100$ for systems with spiral main galaxies and about 200 for systems with elliptical main galaxies. These values lead to $\omega \sim 0.2 - 0.6$ for the total mass density of the Universe.

15.12 Astrophysical Consequences of the Existence of a Heavy Neutral Lepton

Gravitational attraction between galaxies in clusters, and the motion of stars in galaxies show that the bulk of the matter in galaxies is outside the region of luminous objects, leading to the hypothesis of dark matter.

One suggested solution to this puzzle is that galactic halos consist of neutrinos. For example, if the matter density in galaxies is $< 0.1 \times \rho_{crit}$, then the sum of neutrino masses is < 0.6 eV. However, considerations based on the Pauli exclusion principle lead to a limit on the maximum neutrino density in phase space of 2 g_{ν}, later refined to g_{ν}, where g_{ν} is the number of helicity states. Combined with the notion that neutrinos in a galactic halo form a Maxwellian thermal gas, the limit on the neutrino density leads to $m_{\nu} > 10eV$. This contradiction led to the conclusion that light neutrinos could not solve the missing mass problem. (Heavy leptons with a mass of several GeV could solve it.) But the accuracy of the aforementioned expression for the matter density in galaxies is not great, so light neutrinos could supply the missing mass.

A recent study claimed that the electron spectrum in tritium decay proves that the rest mass of the electron neutrino is of the order of 30 eV. Another study found oscillations in an antineutron beam inducing the reaction

$$\overline{n} + d \rightarrow 2n + e+$$

The difference $m_1^2 - m_2^2$ was estimated to be $\sim 1ev^2$. Claims of oscillations were also made through comparisons of the charged-current reaction above and the neutral current reaction

$$\overline{n} + d \rightarrow p + n + \overline{n}$$

induced by a beam of reactor antineutrinos.

All of this has led to the attitude in the astrophysics establishment toward neutrinos with nonzero rest mass of the order of some eV or tens of eV. The first success of this approach is the explanation of the deficit of solar neutrinos in the Davis experiment by oscillations $\nu_e \leftrightarrow \nu_\tau \leftrightarrow \nu_\mu$ as proposed by Pontecorvo.

One astrophysical implication is that if the sum of the neutrino masses exceeds 25 eV, the Universe is closed and its expansion must change to contraction in the remote future. The age of the Universe is also shortened. Since

$$U_{age} \sim 21 \times 10^9 yr/[(H/50) + (\Sigma/40)^{.5}],$$

where H is the Hubble constant in $km/s/Mpc$ and Σ is the sum of the neutrino rest masses in eV.

If all neutrinos have the same mass, then when $\Sigma = 90$, $U_{age} = 8.5 \times 10^9 yr$, which is less that the values obtained by Os / Re chronometry and by stellar evolution theory applied to old globular clusters.

One resolution is that U_{age} can be lengthened by introducing a cosmological constant, thereby getting a closed but till expanding Universe. The upper limit on m_ν is then about 200 eV. Another possibility is to assume that an unknown source of background photons existed, so the ratio N_ν/N_{photon} is smaller that in the standard model, and consequently $\Sigma \gg 25eV$ would be required to close the Universe.

"Despite the fact that massive neutrinos could solve some cosmological problems, we would like to end this discussion with a warning: astronomy will use gratefully any neutrino rest mass found in laboratory experiments, but astronomy cannot be used to prove or even to support a definite value of or even the very existence of a nonzero mass for neutrinos or other unknown particles."

16. RESEARCH ON PERTURBATIONS USING THE RELIC RADIATION

16.1 Introduction

The spectrum and angular distribution of the RR can in principle give quite detailed information about deviations of the Universe from the homogeneous and isotropic model. There is no doubt such deviations exist, and indeed are necessary for structure formation in the Universe. But determinations of the deviations based upon the presently observed structure of the Universe is difficult and fraught with ambiguities. Research on RR substantially supplements our knowledge of the perturbs. For example, small-scale perturbs decay even before H recomb occurs and leave no visible trace in the subsequently emerging structures. In decaying, however, these perturbs liberated energy and distort the RR spectrum, giving rise to spectrum deviations away from that described by the equilibrium Planck law.

On the other hand, the largest scale perturbs apparently have small amplitude, so that statistical methods concerned with discrete bodies are used to analyze extragalactic objects such as clusters, quasars and radio sources. [Are quasars and radio sources extrgalactic?-LF] Yet this discreteness itself create an additional source of random errors. Thus, the precision of the Hubble parameter determination is no better no that 20% [Superseded-LF], as is its constancy with respect to objects in the northern hemisphere.

Meanwhile, measurements of the RR temperature in various directions show no differences to 0.1% at centimeter wavelengths. The subsequent discussion in this chapter is about "nonexistent" effects for two reasons. First, today's negative observational results can put limits on the amplitude of the perturbs. Second, one can expect advances in the accuracy of experimental methods, such as permitting a more accurate separation of the RR fluctuations from those caused by nearby sources. This would be observable after an increase in the sensitivity to $\delta T/T \sim 10^{-5}$.

Thus, the following topics are discussed in this chapter:

A. Matter-antimatter annihilation

B. Transformation of the kinetic energy of peculiar motion into heat

C. General scheme for calculating the temperature distribution for a specified peculiar motion and for a specified distribution of density perturbs.

D. Application of this general scheme to adiabatic perturbs

E. Possibility that that RR temperature fluctuations arise from vortex perturbs and long-wavelength gravitational waves.

An overall conclusion is that the observable high degree of RR isotropy is a strong argument that perturbs of the Friedmann homogeneous and isotopic model are small. In particular, this isotropy favors the the theory of adiabatic perturbs and argues against the vortex (turbulence) theory of galaxy formation.

16.2 Antimatter Annihilation

Charge symmetry or asymmetry, quite unchangeable, must be established at the singular instant; we assume symmetry. [The general principles behind this preference will be discussed in Part V of this book.]. Now we shall relate the distortion of the radiation spectrum caused by annihilation.

Consideration of the RD period is important theoretically since one can compute the annihilation process characteristics, in particular the diffusion of matter and antimatter to the boundaries between their regions. Also playing a role are the friction arising from the motion of e+ or e- relative to the radiation and the impossibility of the e-, p separation. The process is in some sense similar to electrolyte diffusion in electrochemistry.

A diffusion constant can be calculated, leading to a distance that a particle diffuses during the RD period up to the instant when the redshift achieves a given value. Its value is

$$D = 0.6 \times 10^{32}/z^3 cm^2/s$$

Based on the diffusion distance and today's density, a mass can be calculated, but the result is only applicable to the RD period before recomb, when $10^8 > z > 1400$. For example, it is $3 \times M_\odot$ when $\omega = 0.1$ and $z = 1400$.

Return now to the general picture: matter and antimatter exist in separate regions, nowhere coexisting appreciably, so annihilation takes place at the boundaries. One can then compute the disappearance time for an "island" of antimatter surrounded by matter, and ultimately the number of instances of annihilation per unit of time per unit of boundary area. This would be a first approx, since there would be overlap regions. Note that the high temperature prevents the formation of proton-antiproton "atoms."

This situation is similar to the flame of a candle, with a region of excess hot fuel and a region of excess oxygen, so the methods of combustion theory can be applied to calculate the characteristics of the overlap region.

For larger red shifts (higher temperatures) two additional circumstances enter.

First, for $T > 1$ MeV, some neutrons will be in thermodynamic equilibrium with protons (and antineutrons with antiprotons). Neutron and antineutron diffusion will take place more rapidly in the absence of restoring electrostatic fields, not requiring e- and e+ displacement to compensate.

Second for $T > 3 \times 10*$ K, there are e+e- pairs along with the electrons in thermo equilibrium in the matter region. The total concentrations will satisfy n(positron) + n(electron) $\gg$ n(proton). In compensating for the electric field that arises due to the proton displacement, all the light charged particles participate, significantly increasing the diffusion coef for the protons.

With due regard for both circumstances, the diffusion and annihilation disposes of the small islands of antimatter by the instant when $z \sim 10^8$. These islands of antimatter are surrounded by matter (the case of opposite signs is also possible) so these processes also serve to smooth out any gradients on scales up to $\sim 5 \times 10^{11}$ cm, or 5×10^{19} cm today, The corresponding mass with respect to the present-day density is $\sim 10^{-3} \times \omega M_\odot$. The liberation of energy when $z > 10^8$ is not expressed in the RR spectrum because the e- and e+ accelerate the establishment of thermo equil at these times.

So what conclusion can one thereby draw for a charge-symmetric Universe? Assume that clusters with masses of the order of $10^{12} M_\odot$ are either matter or antimatter, since 70-100 MeV quanta from

$$p + \overline{p} \to \pi_0 \to 2\gamma$$

are not observed. There are additional possibilities.

16.3 Adiabatic Perturbations, Acoustic Oscillations, and Their Effect on the Relic Radiation Spectrum

...

16.4 Perturbations of the Spatial Homogeneity and Isotropy of the Relic Radiation

The observed independence of the RR intensity, spectrum or temperature with respect to direction confirms that the Universe is to a first approx homogeneous and isotropic. To this day, no deviations have been observed with a sensitivity for which $\delta T/T < 10^{-3}$ or higher. In the next approx, one must consider perturbs that destroy the homogeneity and isotropy, are of various types, range over all spatial scales, and have amplitudes that can vary with respect to scale size.

Given the objective of learning about the initial state, perturbs with an amplitude of $\sim 10^{-4}$ for metric changes on a scale $3 \times 10^{12} M_\odot$ are as interesting as those with an amplitude of $\sim 10^{-7}$ on a scale of $10^{24} M_\odot$ or an amplitude of 0.1 on a scale of $10^{-10} M_\odot$. One can then consider it "accidental" that perturbs with scale sizes from $\sim 10^{-4} M_\odot$ to $\sim 10^{12} M_\odot$ lead to $\delta \sim 1$ at the present time and clearly manifest themselves in the observable structure of the Universe.

Small-scale perturbs, even if their initial amplitude is not small, decay long before recomb and therefore do not manifest themselves directly as temperature fluctuations.

Small-amplitude, large-scale perturbs manifest themselves at the present time quite weakly as a small irregularity in the density and the velocity. We shall consider the possibility of observing them via the RR. To do so requires a theory describing the angular distribution of

RR in a perturbed universe. We shall construct here a simple theory.

Note first that, in analogy with geometrical optics, RR isotropy is in way affected by gravitational lenses at rest with respect to the general Hubble expansion. On the other hand, a brightness change always occurs as a result of a redshift or blueshift of the spectrum. As result, the Planck spectrum remains a Planck spectrum, though with a temperature shift. However, this shift for different bundles of rays can differ depending on their "history," the velocity of emitter and observer, on the gravitational potential at the points of emission and reception, and on the nonstationary gravitational fields along the path.

The gravitational field of an isolated body does not destroy isotropy because it gives rise to both blue and red shifts for a passing photon. And the gravitational potential of the cluster or supercluster to which our galaxy belongs only gives rise to a general but unobservable blueshift.

Thus the possible RR anisotropy is completely characterized by the distribution $T(\theta, \phi)$ on the celestial sphere of the observed temperature characterizing the Planck spectrum. It is sufficient to keep track of the redshift along the path of a ray to determine $T(\theta, \phi)$ and this redshift depends on the Doppler effect and on gravitational fields, including gravitational waves.

In the absence of perturbs, the velocity field near any particle is given by the isotropic Hubble expansion law $u = H(t) \times r$. In the presence of perturbs, this generalizes to the expression $u = H(L, K) \times r(L, K)$, where H is a tensor that depend on the time and the coordinate r. Both the isotropy and homogeneity of the unperturbed model are lost. Thus the redshift can be computed by determining the motion of the test particles based on GTR. Additionally, for small-amplitude perturbs, both the change in the ray trajectory and the redshift (leading to δT) are small. In the unperturbed field of particle motions, a change in the trajectory does not change the redshift. The change in the redshift from the change in the trajectory depends on the second order of a small quantity and is not taken into account.

So consider a Universe filled with zero pressure matter, a good approx after recomb. Though before recomb photons undergo repeated Thomson scattering from free electrons, after recomb the average energy of photons in the Rayleigh-Jeans part of the spectrum is much less that the ionization energy of H, so the neutral gas is transparent to these photons.

Detailed computation leads to a relatively simple result:

$\delta T/T$ = Doppler shift corresponding to the peculiar velocity / c + Pure gravitational shift / c^2.

The result if also presented in this form in by Sachs and Wolfe based on more complicated GTR calculations.

16.5 The Detection of Density Perturbations Using the Relic Radiation

. . .

16.6 The Perturbation Spectrum and the Low-Density, Hyperbolic Model

. . .

16.7 The Angular Distribution of the Relic Radiation Fluctuations

Since RR probably approaches us without undergoing scattering after recomb at $z \sim 1400$, its temperature provides an opportunity to observe directly the prestellar plasma early in evolution, and thereby its homogeneity and how well its velocity corresponds to the that of a Friedmann model.

The local density and velocity deviations must produce small angle temperature fluctuations dependent on the angular variables ϕ and θ. The function $\delta T(\theta, \phi)$ also characterizes the amplitude and spatial scale of the perturbs of the Friedmann model.

Here we shall use the values $75 km/s/Mpc$ for H, $\rho_{crit} = 10^{-29} g/cm^3$, and $P \ll \rho \times c^2$. We will be considering the linear size of an inhomogeneity at recombination lr, the size now l0, the characteristic mass M, and the angle θ it subtends. The analysis yield

$$\theta \sim 10 \times [M/10^{14} \times M_\odot]^{1/3} \times (\omega)^{2/3} arcminutes$$

One study of whether the perturbs leading to galaxy formation simultaneously give RR fluctuation led to this amplitude estimate corresponding to adiabatic contraction of the material and radiation

$$\delta T/T = \delta\rho_m/3\rho_m = \delta/3$$

Given $\delta \sim 1$ when $z \sim 4$, then $\delta \sim 1/300$ when $z = 1400$. This led, on a scale of several minutes of arc, to

$$\delta T/T \sim 10^{-3}.$$

But measurements at the wavelength 2.8 cm gave the initial result $\delta T/T < 3.0 \times 10^{-5}$ on a scale of 3 arc minutes; a refined analysis gave $\delta T/T \sim (1-3) \times 10^{-4}$ on a scale of 3 - 5 are minutes.

Nevertheless this analysis does not signify that a perturb theory description of the formation of structure in the universe is untrue because serious alterations are required for these estimates.

START HERE ON NOV. 22

17. GRAVITATIONAL WAVES IN COSMOLOGY

17.1 Introduction

The distinctive feature of grav waves is their weak interaction with ordinary matter and with one another. Recent sensational reports about discovery of grav waves have not been confirmed by later experiments [superseded!-LF], but they have stimulated great interest in various aspects of grav waves, and in particular to their cosmological role.

Does the Universe consist primarily of grav waves? Can their density now exceed the density of ordinary matter, or, at least, the density of the EM background radiation? Finally, what is the grav wave spectrum?

In the absence of experimental evidence [superseded!-LF], we shall in this chapter collect indirect information that bears on grav waves and assess their possible role in filling the Universe by their various astrophysical effects:

1. Effect of grav radiation on the overall expansion of the Universe and on its age;
2. Effect of the radiation through the expansion rate on the perturb growth law and consequent connection with galaxy cluster formation;
3. Effect on cosmological nucleosynthesis when the temperature of the Universe is of order $10^9 - 10^{10}$ K, at an earlier stage of the expansion than for points 1 & 2.
4. Effect of energy transfer from the waves to galaxies in clusters or to the clusters themselves.
5. Effect of the grav waves on RR propagation. This approach is most informative for long waves [EM or grav??-LF], because the measurement accuracy can only show that the energy density of the grav radiation is cosmologically small in the long-wave region.
6. Possible existence of short-wave, thermal grav radiation, analogous to the EM RR. Can thermo equilibrium be established near the singularity?
7. Connection between long-wave grav radiation and other metric perturb types—e.g., adiabatic perturbs.
8. Generation of grav waves by stars and in galactic nuclei.

9. & 10. Before studying the aforementioned, begin first with fundamental classical results about grav waves, including general expressions and numerical estimates from standard theory and the evolution of grav waves in an expanding Universe.

The general conclusion from these points is only a set of inequalities that limit the density of grav radiation, though limited by knowledge gaps connected with unstudied spectral regions. A program of further research is also presented.

17.2 General Information about Gravitational Waves

The exposition below is based on that of Landau and Lifschitz and in volume 1 of this work.

In empty space, for which the energy momentum tensor $T_{ik} = 0$, a possible solution of the GTR equations is one with the weakly perturbed Minkowski metric form

$$ds^2 = c^2 dt^2 - g_{uv} dx_u dx_v, \text{ where } g_{uv} = \delta_{uv} + h_{uv}(t, x)$$

Then the equation $G_{ik} = R_{ik} - (1/2) g_{ik} R = 0$ leads in the linear approx to

$$\Box h_{uv} = (1/c^2 d^2/dt^2 - \nabla^2) h_{uv} = 0.$$

With the conditions $h_{uv}^{(0)} k^u = h_{uv}^{(0)} k^u = 0$, the elementary solutions are

$$h_{uv} = h_{uv}(\mathbf{k}) = h_{uv}^{(0)} e^{i\mathbf{k} \cdot \mathbf{x} - i\omega t}, \text{ with } \omega = c \times |\mathbf{k}|.$$

Any general solution can be expressed as a superposition of these via an integral over the wave vector $\mathbf{k}$, taking the real part only.

Hence a grav wave propagates transversely at c. Because the metric is synchronous, $h_{00} = 0, g_{00} = 1$ and $h_{oa} = g_{0a} = 0$, the motion of test particles in such a wave field is $x^u(t) = const$, so mutual separations and relative motions of the particles are completely determined by the metric changes.

By considering terms quadratic in the h_{uv}, one can calculate the stress energy pseudotensor of the metric, whose general form is

$$t_{00} = t_{01}/c = t_{11}/c^2$$
$$= \left[\frac{c^2}{32\pi G}\right] [(dh_{23}/dt)^2 + (1/4)(dh_{22}/dt - dh_{33}/dt)^2]$$

The wave vector $\mathbf{k}$ is directed along the x^1 axis, and from transversality, only h_{22}, h_{23}, and h_{33} are nonzero. Although the pseudotensor can be set to zero at any point by a coordinate transformation, its integral over a space that is flat at infinity has an invariant significance as the energy-momentum four-vector $(E, \mathbf{p})$ of the grav field in the volume under consideration.

A packet of unidirectional, almost flat waves satisfies $abs(p) = E/c$, i.e., its energy-momentum four vector is almost null: $E^2 - p^2 c^2 = 0$, corresponding to the speed of grav waves equalling the speed of light or that the rest mass of the graviton is zero. The qualification "almost" is necessitated because there may be small wave vector components in other directions if the packet is limited in space in these directions.

For a distribution of randomly directed waves, the energy flux goes to zero while the spatial components of the pseudotensor satisfy $< \varepsilon_g >= 3 < P_g >$. The quantities ε_g and P_g, the grav wave energy density and pressure, along with the corresponding terms for "ordinary" matter, enter into the right side of the Einstein equations for the slowly varying part of the metric.

In the cosmological problem, for an isotropic (on average) model filled only with grav radiation, we have the order-of-magnitude equation

$$-1/ad^2a/dt^2 = (1/a\ da/dt)^2 = \omega^2 \times \alpha^2 \sim (dh/dt)^2$$

Here h characterizes any of the components h_{uv}, and α is the corresponding Fourier amplitude. Thus, even a small amplitude

$$h \sim 1/(\omega * t) \sim H/\omega$$

for the metric perturbs can produce important cosmological effects. With the Hubble parameter $H = 10^{-18} s^{-1}$ and $\omega = 1\ s^{-1}$, the value for h is $\sim 10^{-18}$. For $\omega = 10^4 s^{-1}$, the Weber detector resonant frequency, $h \sim 10^{-22}$. For $\omega = 10^{10} s^{-1}$ (the average frequency of the RR photons), $h \sim 10^{-28}$.

These small values for h characterize the weak interaction of grav waves with matter even if their cosmological role is important. On the other hand, for the longest possible waves with a wavelength $\lambda \sim ct$ and a frequency $\omega \sim t^{-1} \sim H$, an amplitude of unity is necessary for a cosmo effect. Inasmuch as no dimensionless quantities differing from order unity are needed to characterize the effects of such waves, their effect on radiation from the horizon is also of order unity. Below we shall see that RR isotropy thus excludes grav waves with an order unity amplitude and a wavelength of order the horizon size.

A random (isotropic on average) collection of short-wavelength grav waves behaves in a Friedmann Universe as an ideal ultra-relativistic gas with an equation of state $P = \varepsilon/3$. Consequently the energy density of the waves changes as $\varepsilon \sim [a(t)]^{-4} \sim (1+z)^4$. For each separate mode the wavelength is proportional to the scale length of the Universe: $\lambda \sim a(t) \sim (1+z)^{-1}$ and $\omega \sim (1+z)$. Bearing in mind that $\varepsilon \sim (h\omega)^2$, we thus have that $h \sim a^{-1} \sim (1+z)$; i.e., the wave amplitude expressed in terms of the dimensionless metric coeffs decreases with time.

The ratio of the wave energy in a given coming volume to the wave frequency, an adiabatic invariant and in some sense the number of quanta or gravitons, remains constant during the evolution. Thus the number of gravitons is conserved during the adiabatic period of expansion when the wave frequency is greater than the Hubble parameter and the wavelength is less than the horizon size. This is the case for an individual wave or a gas of waves.

An isotropic random distribution remains isotropic in an isotropically expanding Universe. However, given any global deviations from homogeneity and isotropy, since the gas is collision-less, the isotropy of the wave distribution is rapidly destroyed either by a global anisotropic expansion of the Universe or by any local anisotropic effects.

17.3 Gravitational Waves in the Theory of Small Perturbations of a Cosmological Model

Referring back to the Chapter 12 discussion of small perturbs to the homogeneous and isotropic Friedmann model, we saw that tensor perturbs are so defined that one cannot construct from them and the wave vector any nonzero scalar or vector quantities. Hence tensor perturbs represent grav waves.

During the early stages of expansion of the Universe, when the wavelength of the wave is greater than the horizon size, tensor perturbs represent "waves to be." Consider the relation between the horizon size ct and the grav wavelength. The wavelength is proportional to a(t), so its instantaneous value is $\lambda = \lambda_0 a(t)/a(t_0)$, where $a(t_0)$ and λ_0 are the values at an arbitrary time t_0. Now a(t) grows more slowly than t; in particular $\sim t^{2/3}$ for $P = 0$; $\sim t^{1/2}$ for $P = \varepsilon/3$; and $\sim t^{1/3}$ for $P = \varepsilon$. For each wave, therefore, there is a time t_1 such that for $t < t_1$, $\lambda > ct$, while for $t > t_1$, $\lambda < ct$. For example, $t_1 = \lambda_0^2/(c^2 t_0)$ when $P = \varepsilon/3$.

It is more convenient to express quantities in terms of a different temporal parameter η (eta), where $d(\eta) = dt/a(t)$, so that the metric is

$$\begin{aligned} ds^2 &= dt^2 - a(t)^2[dx^2 + dy^2 + dz^2] \\ &= a(t)^2[d\eta^2 - dx^2 - dy^2 - dz^2]. \end{aligned}$$

In this coord system, the wave vector $\mathbf{k}$ is constant as is the comoving wavelength $2\pi/|\mathbf{k}|$. Since the equation for light propagation is ds = 0, the distance to the horizon is simply η and the periods when $k\eta < 1$ and $k\eta > 1$ are naturally differentiated as the initial period, when the distance to the horizon is less than the wavelength, and the later period, when the distance is greater.

For weak grav waves in a RD Universe, there can be two linearly independent solutions:

$$h_a^{(1)b} = C_1 G_a^b \cos(\mathbf{k}\,.\,\mathbf{x} + \phi_1)[\sin(k\eta)]/(k\eta);$$

and

$$h_a^{(2)b} = C_2 G_a^b \cos(\mathbf{k}\,.\,\mathbf{x} + \phi_2)[\cos(k\eta)]/(k\eta).$$

Of these, the first remains finite near the singularity at $\eta = 0$. Consequently, finite or small metric perturbs at the singular state lead to finite or small-amplitude grav waves at late times when $k\eta > 1$.

For any particular wave, with given values $C_1 < 1$ and $C_2 = 0$, the grav wave energy during a late, but still RD, stage is a fraction of order $\alpha = (C_1)^2$ of the energy of the matter. We emphasize, however, that it is in principle inadmissible to ask about the density of grav waves at such an early stage that $k\eta < 1$.

We have been discussing the evolution of an individual wave. For small-amplitude metric perturbs and correspondingly small waves, superposition holds, so the results above generalize to a random superposition of waves. The condition that the perturbs be finite near the singularity limits the randomness, already accounted for in the previous equations. Indeed, this condition distinguishes standing grav waves.

Taking the most general linear combination of complex solutions with given values of $|k|$ and given directions $\pm\mathbf{k}/|k|$, and requiring a real solution, leads to

$$f = [A/(k\eta)]\cos(\mathbf{k}\,\mathbf{.}\,\mathbf{x} - k\eta + \phi) + [B/(k\eta)]\cos(-\mathbf{k}\,\mathbf{.}\,\mathbf{x} - k\eta + \psi)$$

This expression represents two traveling waves moving oppositely and with arbitrary and unequal amplitudes, though near the singularity its amplitude is (almost everywhere) infinite. Satisfying the amplitude limitation on f as $\eta \gg 0$, we are ultimately left with

$$f = [2A\sin(k\eta)/k\eta]\sin(\mathbf{k}\,\mathbf{.}\,\mathbf{x} + \phi)$$

The amplitudes for the two oppositely moving waves are equal, so they form a standing wave. The same condition was obtained for acoustic waves (scalar waves accompanied by a density change) and led to Sakharov's prediction for the amplitude of oscillations after recomb. For grav waves however, no method has yet been devised for distinguishing between traveling and standing waves.

17.4 The Expected Intensity of Relic, Short-Wavelength Gravitational Radiation

We shall consider scenarios in which background grav radiation is expected, beginning at the Planck time $t_{Pl} = 10^{-43}$ s.

The simplest variant results from assuming that at t_{Pl}, thermo equilibrium holds between the gravitons and the finite number of other elementary particles and antiparticles, which are weakly interacting [??-LF]. We shall consider the situation when the temperature T is higher than the rest energy mc^2 of all of the particles, so we have an ultra-relativistic gas.

Suppose that the energy density of all of the particles including the gravitons is $\varepsilon = \mathcal{N}\sigma T^4$, where σ is the constant for EM radiation and $\mathcal{N}$ is approximately the number of particle kinds. The corresponding graviton entropy is $4\sigma T^3/3$, while the entropy for the remaining particles is $(\mathcal{N}-1)4\sigma T^3/3$.

During the expansion, the graviton entropy per unit coming volume is conserved. But with the exception of a small number of protons, all of the heavy particles and antiparticles eventually annihilate, leaving additionally the gravitons (accounted for separately), and when $T \sim 5-10$ MeV, only the electrons, the two [superseded-LF] kinds of neutrinos, their antiparticles, and photons. These leptons inherit the entropy of the heavy particles, so the lepton temperature T_l, and the graviton temperature T_g are related by [with 2 below now 3-LF]

$$(1 + 7/4 + 2 \times 7/8)(4/3)\sigma T_l^3 = (\mathcal{N}-1)(4/3)\sigma T_g^3.$$

In the subsequent evolution, the photons and the electron-positron pairs decouple from the neutrinos. When the e+e- pairs annihilate, their entropy is inherited by the photons and the radiation temperature is thereafter related to T_l by $T_r^3 = (11/4)T_l^3$, and we thus find

$$T_g = [(4/11)(9/2)/(\mathcal{N}-1)]^{1/3}T_r.$$

Setting $\mathcal{N} \sim 20-40$ and $T_r = 2.7K$, we obtain $T_g = 1K$ for the present-day graviton temperature. In this variant, we can always neglect the graviton contribution to the total density since their energy density is only of order 0.01 times the radiation density.

The grav wave wavelength where the maximum energy is concentrated is of order a fraction of a centimeter, as is also the case for the relic EM radiation.

Now consider short-wave RR in the case where equilibrium conditions are assumed not to exist at the Planck time. Two extreme assumptions are that: (1) there are no grav waves at all and the matter is hot and has normal entropy; and (2) the energy of the short-wavelength grav waves specifies the overall density almost entirely. Today's resultant state depends on the relaxation rate or of subsequent physical processes, or, more precisely, on the relation between the relaxation and expansion rates.

A calculation shows that at the Planck time, the product of the relaxation rate and the characteristic expansion time is of order unity.

The transformation of two gravitons into a e+e- pair corresponds to pair creation by two photon collision in Feynman diagram language, so the gravitational cross section is proportional to G^2. To obtain a true cross section, it is necessary to multiply by the square of the energy but to have no dependence of the rest masses of the involved particles, so the result is $\sigma = G^2E^2c^{-8}$.

In units where $\hbar = c = 1$, at the moment $t_{Pl} = G^{1/2}$, the corresponding energy density is $\varepsilon = G^{-2}$. The equilibrium spectrum of grav waves in accord with assumption (2) above then corresponds to a temperature $T = G^{-1/2}$ since $\varepsilon_g = T^4 = G^{-2}$.

This means that the energy of individual gravitons is $G^{-1/2}$, the energy corresponding to the Planck mass, while their density is $n = G^{-3/2}$. The cross section is then $\sigma = G^2 E^2 = G$. Given motion at the speed of light, the total number of pairs created at the Planck time is then $n' = \sigma n^2 t_{Pl} = G^{-3/2} = n$. This crude calculation leads to the conclusion that the number of created pairs is of order the initial number of gravitons.

In accord with the scenario based on assumption (1) above, the inverse process will produce a number of gravitons comparable to the number of other particles, so the energy density of short-wave grav waves will in both cases be of order the equilibrium value, i.e., in accord with the temperature of the remaining matter and around 1 K today. The difference in the energy densities calculated according to the two assumptions amounts to a factor of perhaps five, but does not involve factors like $Gm^2/(\hbar c)$, which might in principle have entered.

If there is equilibrium at t_{Pl}, then because of the annihilation of particles with nonzero rest mass, the energy density of grav waves today should be $\varepsilon_g \sim 0.01\varepsilon_r$. One can suppose that the the result $\varepsilon_g < \varepsilon_r$ today remains valid even if $\varepsilon_g \gg \varepsilon_r$ at t_{Pl} because the situation would then change through relaxation to one of equil. On the other hand, again because of relaxation, it is doubtful that $\varepsilon_g < 10^{-3}\varepsilon_r$ today even if $\varepsilon_g = 0$ in the initial state.

Finally, keep in mind that the result applies to a particular set of assumptions.

17.5 The Equipartition Theorem and Long-Wavelength Gravitational Radiation

In the theory of small perturbs of the homogeneous, isotropic cosmo model, grav waves are formally independent of other perturb types. And waves of different wave vectors are are independent of one another.

The homogeneous and isotropic Friedmann solution does not precisely describe our Universe, as is apparent from the existence of structure, e.g. galaxies. On the other hand, the theory of nucleosynthesis in homogeneous, isotropically expanding matter accords well with observations. Thus the true solution should possess the property of behaving locally as a Friedmann solution during an early stage.

If one assumes the g_{uv} are close to their Friedmann values near the singularity, but that the differences of the various components are independent, then one is drawn to the "equipartition hypothesis." It says that the metric perturbs related to density (scalar) perturbs are equal in amplitude to metric perturbs related to grav waves (tensor).

Now consider the spectral characteristics of quasi-isotropic solutions. Given the separation of the perturbs into plane waves, spatial isotropy implies that the perturb amplitude depends on the wave vector magnitude, but not on its direction. The equipartition hypothesis speaks about the equality of two functions of $|k|$ - the amplitudes of the scalar and tensor perturbs.

Consider the time when the wavelength equals the horizon size, i.e., when $\lambda = ct$ or $k\eta = 1$. According to the hypothesis, the relative density perturb amplitude has the same order of magnitude as the metric perturb,

$$(\delta\rho/\rho)_k = h_k, \ k\eta = 1$$

Hence one can express the amplitude of the grav wave in terms of the amplitude of the density perturb.

For a scale corresponding to galaxy clusters (mass $M \sim 10^{13} M_\odot$ or a present linear scale of $\lambda \sim 30$ Mpc), we infer a grav wave amplitude of $\sim 10^{-4}$. Thus

$$dh/dt \sim 10^{-4} t^{-1}, \ \rho_g \sim 10^{-8} < \rho > = 10^{-8} \rho_r$$

Here we have passed from separate Fourier amplitudes to quantities integrated over a single section of the logarithmic wavelength scale. The quantity ρ_g is ε_g/c^2, where ε_g is the energy density of the grav waves and ρ_r is the radiation density.

For the scale considered, the time when $k\eta = 1$ occurs during the RD era, when ρ_r is practically the same as the total density. The ratio ρ_g/ρ_r itself remains constant until the present. Thus the estimated present energy density of grav waves whose scale is that of galaxy clusters is

$$\rho_g(M \sim 10^{13} M_\odot \sim 10^{-8} \rho_r = 10^{-42} g \ cm^{-3};$$

$$\varepsilon_g \sim 10^{-21} ergs \ cm^{-3}; dh/dt = 10^{-24} s^{-1} = 10^{-6} H$$

Therefore, on the scale over which density perturbs are known, the grav waves are only weak.

Over still larger scales, the amplitude of grav waves is limited by the EM RR isotropy. Over scales smaller than those characterized by a mass of $10^{13} M_\odot$, however, it is difficult to observe such perturbs.

But this is a scale where the equipartition hypothesis is useful. Density perturbs decay before recomb, and the their decay heat must change the RR spectrum. Such changes are not observed, so this allows one to impose the limit $\varepsilon_{acous} < 0.05 \ \varepsilon_r$ the mass interval $10^{13} M_\odot > M > 10^5 M_\odot$. The equipartition then implies $\varepsilon_g < 0.05 \ \varepsilon_r$ for the same mass interval.

According to the equipartition hypothesis, a flat spectrum of scalar metric perturbs, which could explain the structure and entropy of the Universe, would imply that grav waves only make a small contribution to the total energy density over all wavelengths from millimeters to 10^{10} pc. But the exisitence of such a flat spectrum has not been demonstrated.

17.6 The Generation of Gravitational Waves in the Present Epoch

The generation of grav waves in astrophysical objects concerns cosmology only indirectly. Apparent detection by Weber in the 1960s and 1970s was disproved.

Based on general energy considerations, in all cases except the collision of two black holes, the release of EM radiation exceeds that of grav radiation for ordinary stars, white dwarfs and neutron stars. Thus the energy density of such grav waves is less that the energy density of non relic EM radiation, i.e., $< 10^{-13}$ergs cm^{-3}, corresponding to $\rho_g < 10^{-34}$g cm^{-3}.

However, for a collision of two black holes, the grav radiation reaches several percent of the rest mass if there is relative orbital motion, so grav radiation can play a large role in the evolution of clusters containing a large number black holes. But such a cluster becomes unstable when the grav mass defect becomes several percent.

Only a part of this energy loss goes into grav radiation, though. So one can draw an important conclusion: noncosmological grav radiation cannot close the universe! For if the total mass of ordinary matter and of BHs of all sizes is insufficient to close the universe, then the grav radiation emitted by such bodies can only change the total density slightly: if the BHs constitute 10% of the matter density and if in turn 10% of their mass is converted into grav radiation, then the radiation density can only be 1% of the matter density, i.e., $\sim 10^{-32}$g cm^{-3}.

Consider the related parameters of the age of the universe t, its matter density, and the Hubble parameter. With high probability, the age satisfies $t > (10-12)\times 10^9$ yr, and with absolute confidence, $> 5\times 10^9$ yr, the age of the solar system. From this follows $\rho < 3\times 10^{-29}$g cm^{-3} if $H = 75$km s^{-1}Mpc^{-1} as one set of values if $P = \varepsilon/3$. This density exceeds that of EM radiation by a large factor and in this sense is not interesting for our present purposes.

A second limit is related to the indirect effect of weakly interacting particles, grav waves in particular, on the development of density perturbs. In essence, density perturbs grow in a stationary Universe as $\delta\rho/\rho \sim e^{\omega t}$, where $\rho = (4\pi G\rho_m)^{1/2}$. From an exact calculation, $e^{\omega t}$ generalizes in an expanding Universe to $e^{[(2/3)^{1/2}\int \omega dt]}$. So now consider the dependence of the expansion law on the density components in units of the critical density. The growth factor becomes

$$(2/3)^{1/2}\int \omega dt \equiv \alpha = \int_{z=0}^{z=z_r} \Omega_m^{1/2}\{(1+z)[1+\Omega_m z + (\Omega_r+\Omega_g)z(2+z)]\}^{-1/2}dz$$

For $\Omega_m = 1$ and $\Omega_r = \Omega_g = 0$, this gives the well-known result

$$\alpha = \ln(z_r+1),\ e^{\alpha} = (z_r+1) = 1401.$$

Consider now the effect of nonzero Ω_r and Ω_g terms. Given $\Omega_r = 10^{-4}$, let us set$\Omega_m = 0.2$ as an example. If $\Omega_g = 0$, then the term with Ω_r has almost no effect: $\alpha = \alpha_0 = 5.6$, $e^{\alpha} = 270$. If, however, we set $\Omega_{g,1} = 100\Omega_r$, then we obtain a significant decrease in the perturb growth:

$$\alpha = \alpha_1 = 2,\ e^{\alpha_1} = 10.$$

Clusters of galaxies form during the present epoch or perhaps when $z \sim 4$. This requires that the initial perturbs are larger for smaller values of e^{α}. In the case $\Omega_{g,1} = 100\Omega_r$, therefore, the density perturbs at recomb must reach the value $\delta\rho/\rho \sim 0.1$. For adiabatic perturbs, this corresponds to $\delta T/T \sim 0.03$. It is probable, however, that such fluctuations in the RR are excluded by current observational data.

A third limit follows from the possible effect of grav radiation on nucleosynthesis. Calculations show that if $\Omega_g/\Omega_r = 10$ or 50, then the respective amounts of ^{4}He are 40% and 50% (greater than the 25% observed-LF)

Table III (Table 15 in text) below shows the limits on grav rad by the three methods. The results are characterized by the most likely limits on the ratio of the grav wave energy to the energy of the RR. The wavelengths refer to the present. Keep in mind that Ω_r/Ω_m lies within the limits 5×10^{-5} and 5×10^{-3}, and that because the methods lack complete reliability, one should not just choose the method with the lowest value for Ω_g.

TABLE III. Table 15 in text. Relative Density Ω_g/Ω_r of Grav Radiation in the Universe by Different Arguments

Argument	Result for Ω_g/Ω_r	Spectral Region of Applicability
Nucleosynthesis	< 15	$0 < \lambda < 3\times 10^{17}$ cm
Perturbation evolution	< 200	$0 < \lambda < 5\times 10^{26}$ cm
Age of the Universe	$< 2\times 10^4$	$0 < \lambda < 10^{28}$ cm

17.7 The Effect of Gravitational Waves on the Relic Radiaton

Like any stationary metric perturb, grav radiation influences the distribution of the 3K RR background. This influence causes differences in the observable temperature with direction but not in the form of the spectrum. If the RR is of Planck equilibrium form at recomb, then it remains so in every direction. The Liouville theorem with GTR guarantees a change in the photon flux density proportional to T^3, corresponding to the change in the photon frequency and temperature. So to describe the temperature in the presence of grav waves, one must introduce a function $T(\theta, \phi)$ or $\Delta T(\theta, \phi) = T(\theta, \phi) - < T >$.

The dimensionless frequency change of an EM wave along a given ray can be expressed via the dimensionless grav wave amplitude h. Since h decreases during the expansion, the relevant quantity is its value at recomb. Because the sum total of the fluctuation-induced changes satisfies $\Delta T/T < 10^{-4}$, we obtain the estimate that $h < 10^{-4}$ at recomb. Allowing for the subsequent expansion, this corresponds to the bound at present of $h_0 < 10^{-7}$. But because the grav wave energy density depends on dh/dt, it is necessary to explain the wavelength and frequency applicability of these limit.

In the case of high-frequency waves, one cannot assume that the recombination is instantaneous. Rather, we see simultaneously several layers on which the grav wave effects are not all of the same sign. However, there is an undiminished effect for $\lambda_r > 0.05ct_r$, which corresponds to $\lambda_0 > 2 \times 10^{25}$ cm at present-day scales. On the other hand, to ensure that we are dealing with a grav wave that has formed, we require $\lambda < ct_r$. Together, we need 2×10^{25} cm $< \lambda < 5 \times 10^{26}$ cm today.

For a wavelength in this interval, we estimate the ratio $\varepsilon_g/\varepsilon_r = \Omega_g/\Omega_r$ to be in the interval 10^{-8} through 4×10^{-4}.

To see whether the greater sensitivity is achieved for wavelengths equal to the horizon size at recombination, we must examine other spectral regions. Note that horizon-sized waves give rise to temperature anisotropies ΔT on an angular scale of order $ct_r(1 + z_r)/(ct_0)$, i.e., about 2 degrees or 0.03 rad.

Several curious features arise because the speed of a grav wave equals the speed of light. A radio wave propagating at a small angle with respect to a grav wave experiences the same phase of the grav wave for a long time, intensifying the interaction. However, the grav wave is transverse, weakening its effect on a ray sliding by at a small angle. Thus, close to the propagation direction of a grav wave (say the z axis, or $\theta = 0$ in polar coords) there arises a distinctive form for the background photon distribution: $\Delta T \sim cos(2\phi + \alpha)$. [What is α ?-LF]

The temperature perturb ΔT remains finite, but its gradient tends to infinity near the axis. In a decomposition over the sphere (θ, ϕ) into Legendre polynomials, the high harmonics decrease correspondingly slowly with harmonic number for a given wavelength, decreasing by a power law, not exponentially. For perturbs which become grav waves after recomb, the sensitivity of $\Delta T/T$ to h is less. For a given h, however, the energy density is also smaller. For a wavelength $\lambda = ct_0 = 10^{28}$ cm (horizon radius today), the perturb ΔT is of quadrupole character. The amplitude $\Delta T/T$ is still of order h. An observationally based estimate then gives $h < 3 \times 10^{-4}$ and $\Omega_g < 10^{-7}$, correspondingly, so that $\Omega_g/\Omega_r < 10^{-3}$.

The situation in the intermediate region $5 \times 10^{26} < \lambda < 10^{28}$ cm follows from an interpolation between the limiting cases. Waves with $\lambda > 10^{28}$ cm do not now exist in the sense that the corresponding tensor perturbs are still not waves. The temperature fluctuations are very sensitive to waves with $\lambda \sim (1 - 5) \times 10^{26}$ cm because the metric distortion that occurs during the recomb period, when the interaction between the EM radiation first ceases, is observable. Presumably, in this case the secondary ionization from cluster and quasar formation or even earlier formation of shock waves occurs sufficiently late $(z < 10 - 20)$ that the fluctuations are not smeared out by the last scattering. Finally, short wavelength grav waves with $\lambda < 2 \times 10^{25}$ cm produce only small fluctuations in ΔT since these fluctuations are inevitably smeared out in the recomb process, which cannot be regarded as instantaneous. Any grav waves, including short waves, participate in the creation of quadrupole ΔT perturbs with an amplitude corresponding to the current value of h. We thus obtain for the energy density

$$\Omega_g/\Omega_r = 10^{-3}(10^{28}cm/\lambda)^2 \text{for} \lambda \leq 10^{28}\text{cm},$$

covering the limiting case of 10^{28} cm considered above and supplementing the ranges given in Table III (table 15 in text).

17.8 The Peculiar Motion Induced by Gravitational Waves

17.9 The Interconversion of Gravitational and EM Waves

The processes of wave interconversion are interesting for several reasons: (1) they may occur in the early Universe; (2) they may be a method for laboratory generation of grav waves; and (3) the detection of grav waves requires their conversion into other energy forms, into EM waves in particular.

One important case is the conversion of waves in the presence of a magnetic field. A plane EM wave in empty space does not emit grav waves, because its stress-energy tensor does to contain a variable quadrupole moment. But such a moment does appear for an EM wave propagating through a constant electric or magnetic field H_o.

Analysis shows that the fraction α of EM wave energy converted into grav wave energy is

$$\alpha = GH_o^2 x^2/c^4,$$

where the grav wave metric coefficient satisfies

$$h_{yy} = axcos(kx - \omega t),$$

And where the field begins at $x = 0$. [But why should the conversion fraction α depend on location x ?=LF] The conversion fraction is the same for a reverse process. But it is small for laboratory and pulsar conditions, with $\alpha = 10^{-35}$ and $\alpha = 10^{-11}$ respectively.

For a cosmo field , one could make the extreme assumption that $H_o^2 = \varepsilon_r$ with $H_o = 10^{-6}$ gauss today and 1 gauss at recomb. One would then obtain $\alpha \sim 10^{-3}$ at $z = z_r$. Such a conversion would be noticed in the RR by the weakening for a single polarization of the radio waves that would result. If the conversion occurs in a medium of atoms and electrons with nonzero conductivity, coherence is destroyed and α would remain immeasurably small under cosmo conditions.

The problem of the EM detection of grav waves is similar to the preceding problem. . .

The use of EM detectors is apparently the most promising method for lab experiments for detection of high-frequency grav waves themselves generated in the lab. However, even for cumbersome devices with fields of 10^5 gauss and with sizes of the order of meters, the estimates are still not acceptable.

On there other hand the theoretical advantage of a resonator would be lost in the detection of a wide-band spectrum of cosmo grav radiation. A calculation shows that the conversion coefficient for photon detection from conversion by a field of 10^5 gauss and $x = 100$ cm is of order 10^{-35}. Given a number density of cosmo gravitons as 10^3cm^{-3}, their flow through the area $x^2 = 10^4\text{cm}^2$ amounts to 10^{17}/s. Hence the probability of such photon creation over the volume is $10^{17} \times 10^{-35}/\text{s} = 10^{-18}/\text{s}$, amounting to one photon over the age of the Universe! —obviously not realizable!

IV. ANISOTROPIC COSMOLOGY

18. INTRODUCTION TO PART IV

19 VERY SIMPLE ANISOTROPIC COSMOLOGICAL MODELS

Begin here Jan. 10, 2023

V. THE SINGULARITY AND NEW THEORETICAL DEVELOPMENTS

23. The Cosmological Singularity

23.1 Introduction

Previous conclusions are that the universe expands isotropically and homogeneously at least from the time when $\rho_m \approx \rho_r \approx 10^{-12}$ g cm^{-3}.

Earlier, is there Friedmannian expansion from the singularity, or at least from the Planck time $t_{Pl} \approx 10^{-43}$ s? Does matter pass through a state of infinitely large density, or at least $\rho_{Pl} \approx 10^{93}$ g cm^{-3} ? Or has the Universe changed from contraction in a previous epoch to expansion in our epoch at a finite density, according to a suggestion of an "elementary length" $l \approx 10^{-17}$ cm, a corresponding time $\sim 10^{-27}$ s, and a density $\rho \approx \hbar/[cl^4] \approx 10^{30}$ g cm^{-3}? Might the laws of physics change during this epoch? We think not.

Is the existence of a singularity at the start of expansion a special property of Friedmann and other symmetric models, or will it disappear upon the introduction of peculiar velocities of matter or rotation? In the Newtonian problem, peculiar velocities will make the time reversal of expansion avoid an infinite density singularity. This bears on the stability of a cosmological model with contraction preceding expansion.

If expansion does begin from a singularity, the initial conditions are specified by unknown processes under conditions not describable by present theory.

Two questions have remained: (1) Does there exist a singularity-free general (in the sense of "stable") cosmo solution? (2) Does a singularity occur in the past, given the real conditions in the Universe?

At the end of the 1960s, Penrose, Hawking and Geroch (PHG) proved that the expansion of the universe begins from a singularity if GTR is valid. However, near the singularity quantum effects will be important and GTR may not be valid. In the 1970s, Belinsky, Lifschitz and Khalatnikov constructed a general (stable) solution with a singularity, qualitatively the same near the singularity as the solution for Misner's "mixmaster" model.

We shall below discuss the proof that the Universe expanded from a singularity. Then we shall discuss the physical processes that may occur near the singularity. These analyses and their consequences might lead to an understanding of the true character of the early Universe expansion, when the densities are much greater than nuclear density.

23.2 The Singularity at the Beginning of the Expansion

The PHG theorems make use of geometrical methods for the proof of the existence of a true ST singularity

— not a coordinate singularity — and do not provide analytical solutions [see by comparison the Creutz paper-LF]. They are thus existence theorems, and do not say anything about the structure of the singularity.

A true, invariant, non-removable singularity should be a place where coordinate-system independent invariants of the curvature tensor R_{iklm} achieve infinitely large values. At these places, GTR can be violated. Although one could remove the singularity by excising a piece of the ST around it, the ST should still be termed singular because particle world lines would go right to the edge of the small hole; thus particles would be "created" or "disappear" there. Excluding such holes by requiring a singly connected ST does not solve the problem.

Most natural is to consider a ST singularity-free if one can continue all the timeline and null world lines into both the past and future without bound. This would mean that all particles move along all allowable paths and neither arise nor disappear spontaneously anywhere either in the past or future. There are neither infinite curvatures nor small holes. But such a requirement on the world lines is too strong, for it cannot be fulfilled even in flat Euclidean ST.

We pass to the following definition: A ST is nonsingular if any timeline or null geodesics can be continued into the past and the future without bound, i.e., to infinite proper length if timelike and infinite affine parameter if null. Such a ST is termed "causally, geodesically complete." These requirements are the minimum necessary for the ST not to contain a singularity. However, a ST not satisfying these requirements, one with a singularity, does not necessarily contain points with infinite curvature or a with a small hole.

From a physical viewpoint, we confine ourselves to the definition that a ST is singular if a particle world line cannot be continued without bound with respect to its proper time. For such a singular ST would lead to a violation of conservation laws.[my emphasis-LF]

As applied to the cosmo problem, the Hawking-Penrose theorem is as follows:

THEOREM. A ST M cannot be causally, geodesically complete if the GTR equations hold and the following conditions are fulfilled:

1. The ST M does not contain closed timelike lines.
2. These conditions on the equation of state are fulfilled, where ε is the energy density and P_1, P_2, and P_3 are the three principal values of the pressure tensor:

$$\varepsilon + \Sigma_\alpha P_\alpha \geq 0;$$

$$\varepsilon + P_\alpha \geq 0, \alpha = 1, 2, 3$$

3. A mathematical statement relating the existence of a tangent to timeline and null geodesics and the curvature along those tangents.
4. The ST M contains either (a) a point P such that all diverging rays from this point begin to converge if one traces then back into the past, or (b) a compact spacelike hypersurface.

Requirement 1. is natural, equivalent to not violating causality, though there is a subtle argument that it is not necessary.

Requirement 2. is always fulfilled for all known forms of matter when the ST curvature is far from the Planck value of $(10 - 33\text{ cm})^{-2}$ and quantum effects are not important.

Requirement 4. represents the global physical conditions that lead to a singularity, suggesting that the gravity of all matter is so strong that a singularity is inevitable. Requirement (4a) is in principle subject to direct observation but in reality is tested through a combination of observational data and simple calculations. E.g., if one proves that the condition is fulfilled in a Friedmann model and then shows - in agreement with observations - that this model applies to the past when the rays begin to converge, then one has proved that (4a) is fulfilled. Indeed it is fulfilled in the real Universe for sufficiently large redshift values $> 10 - 20$.

Consequently, the expansion of the Universe begins from a singularity if GTR is valid, and almost all specialists assume that in the general, nondegenerate case, a singularity signifies infinite ST curvature.

Two questions remain: Does all or only a part of matter go through a singular state in the past? And how does the expansion proceed near the singularity itself?

23.2 The General Cosmo Solution with a Singularity

The character of the expansion at its start is determined by the initial conditions at the singularity, conditions that we do not know and conditions that can determine this character according to some special solution, not the the most general. But the solution describing the most general expansion near the singularity is of great interest to understanding what can occur and what in reality did occur.

But it is indeed the general solution that describes collapse, the contraction to a singularity, both for cosmo models (if the expansion changes to contraction, i.e, if $\rho > \rho_{crit}$) and for the collapse of an isolated body.

The general solution near a singularity has been constructed by Belinsky, Lifschitz and Khalatnikov. In the approach to the singularity, the solution describing the deformation has in the neighborhood of every point the character of the asymmetric Bianchi type VIII or IX models. It consists of alternating Kasner epochs and describes an oscillatory regime of approach to the singularity.

The metric of this generalized Kasner solution is written in a synchronous reference system as

$$g_{\alpha\beta} = a^2 l_\alpha l_\beta + b^2 m_\alpha m_\beta + c^2 n_\alpha n_\beta,$$

where

$$a \sim t^{p_l}, b \sim t^{p_m}, c \sim t^{p_n}$$

and

$$p_l + p_m + p_n = p_l^2 + p_m^2 + p_n^2 = 1$$

The Greek indices run over 1, 2, and 3. The exponents p_l, p_m , and p_n and the spatial unit vectors l, m , and n are functions of the spatial coordinates. We can then write the Einstein equations in the synchronous reference system with c = 1 as follows:

$$\frac{\partial \mathcal{D}_\alpha^\alpha}{\partial t} - \mathcal{D}_\alpha^\beta \mathcal{D}_\beta^\alpha = 8\pi G(T_0^0 - \frac{1}{2}T);$$

$$\mathcal{D}_{\alpha;\beta}^\beta - \mathcal{D}_{\beta;\alpha}^\beta = 8\pi G T_\alpha^0;$$

$$-\frac{1}{g^{1/2}} \frac{\partial [g^{1/2} \mathcal{D}_\alpha^\beta]}{\partial t} - R_\alpha^\beta = 8\pi G(T_\alpha^\beta - \frac{1}{2}\delta_\alpha^\beta T)$$

Here $\mathcal{D}_\alpha^\beta = 1/2 \frac{\partial g_\alpha^\beta}{\partial t}$; [all the derivative symbols should be understood as partial derivatives-LF]; the semi-colon denotes covariant differentiation in the three-dimensional space with the metric $g_{\alpha\beta}$, and R_α^β is the three-dimensional Ricci curvature tensor based on $g_{\alpha\beta}$.

We shall only be considering the first and third equations. Near the singularity, the components of the stress-energy tensor can be neglected. If one further neglects R_α^β in the third equations then the solution to the first and third will be the aforementioned Kasner solution. Continuing this solution to the singularity lets us determine its region of applicability. In fact further analysis shows that it will break down as $t \to 0$.

Then there are four differential equations relating the metric coefficients a, b, and c.

These four equations coincide with the corresponding equations for the Bianchi type IX model for the period when one Kasner epoch changes into another. Locally, therefore, the solution describing the metric in the general case has an oscillatory character at each spatial point as the singularity is approached, as depicted in Figure 57. Expansion from the singularity follows from changing the direction of the flow of time.

This behavior was described in Part IV. It was investigated by Misner and collaborators as a hypothetical effort to explain the large-scale homogeneity and isotropy of the universe. It did not work.

As stated at the beginning of the section, the beginning of the expansion of the Universe is not necessarily described by the general solution.

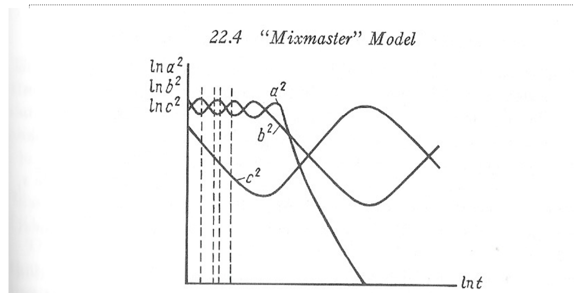


FIG. 57.—The evolution of the "mixmaster" model of the Universe as the singularity is approached.

FIG. 1. Figure 57 from text.

24. Physical Processes Near a Singularity and Developments in the Theory of Gravitation

24.1 Introduction

The existence of a singular state means that infinite values of ST curvature and matter density are achieved, at least formally, so a solution cannot be continued further into the past. More precisely, the world lines of massive and massless particles cannot be continued beyond a certain point.

But charge conservation laws in the express, in language of particles, the continuity of their world lines and impossibility that they can begin or end somewhere. *Physics therefore requires that singularities be avoided or that our conception of them be changes. On the other hand, when the curvature and density approach infinity, it is wholly probable that the GTR and the properties of matter change. Thus, behavior near a singularity suggest going beyond those ideas considered earlier.* [my emphasis-LF]

Whereas the previous discussions of physical processes and temperature were based on well-known physical laws verified in the laboratory, we will here speculate beyond those laws. First we shall consider the statistical properties of a plasma containing nucleons and anitnucleons under various assumptions related to the strong interactions and the excited states of the nucleon. Second we shall address specific quantum phenomena which occur in a gravitational field, but when the metric and and the geometry will be considered classical and predetermined.

Then we will study the back reaction of the quantum fields on the metric, which under definite conditions can be regarded as corrections to the classical GTR equations.

In a later section we shall examine quantization of the Universe as a whole by considering the possible evolutionary paths of the Universe as trajectories having various complex probability amplitudes; in this regard, Wheeler introduced the notion of a superspace of trajectories. Since, however, the Universe is so vast—even if it is closed, it contains $\sim 10^{80}$ baryons and $\sim 10^{90}$ photons—such a global approach is unphysical if local effects are ignored.

Additionally, can the physical laws change near a singularity? A 100% charge-aysmmetric cosmo hypothesis is possible. In this case, there is a cold initial state for the Universe and a unique spectrum of geometry and density perturbs. The long-wavelength part of the spectrum is then responsible for galaxy formation, while the short wavelength part gives acoustic waves, which upon decaying create entropy.

Consideration is also given to Mach's principle, whereby the entire Universe has a direct effect on local phenomena. Dirac perceives manifestations of this through numerical coincidences between local physical constants and quantities characteristic of the Universe as a whole.

Finally, we consider both generalizations of GTR and topological issues.

24.2 Cosmological Consequences of Hagedorn's Theory

Hagedorn's theory is characterized by two features: (1) a maximum temperature satisfying $kT_{max} \approx m_\pi c^2$, so that it is of order $T_{max} \approx (1-2)\times 10^{12}$ K; and (2) a "soft" equation of state, $P \ll \varepsilon$, for temperatures approaching the maximum. For lower temperatures, the theory's thermodynamic properties do not differ from those of ordinary theory, based on the existence of a finite and small number of kinds of particles.

The first feature is important regarding the number of quarks or patrons remaining after expansion of the Universe causes cooling. For a sufficiently large quark mass, $m_q > 7m_p$, the initial equilibrium number of quarks is less than that predicted for the period after the freezeout by ordinary particle theory.

If accelerator or cosmic-ray experiments prove that stable, heavy, free quarks exist, then Hagedorn's theory would save contemporary cosmo. However, they have not been observed, so the aforementioned prediction remains in the air. But the absence of cold quarks in such a situation is not a confirmation of Hagedorn's theory.[Confusing to me-LF]

Apropos the second feature, given the present values of the density and the Hubble parameter, all of the changes are confined to the period when $T \sim T_{max}$, so that there are no changes during the state of nucleosynthesis and recombination.

Regarding the development of perturbs, thermodynamic fluctuations in Hagedorn's theory are much greater (in the limit of $\rho > \infty$, by an infinite factor) that those in the theory of noninteracting particles. In cosmology, however, the conditions necessary for obtaining thermo equilibrium among fluctuations during the early stage do not exist.

A more natural formulation consists of a specifying the initial perturb of the metric. According to Lifschitz's solution, a metric perturb remains constant near the singularity, going neither to infinity or zero. Another way of saying this is that during this period, separate parts of the a perturbed Friedman solution develop independently, with the metric preserved.

Let us now find the characteristic wavelength at the instant when the Hagedorn equation of state turns into the ordinary equation of state $P = \varepsilon/3$. At this time, $T \approx 10^{12}$ K, $t \approx 10^{-4}$ s, $n_b \approx 10^{24}$ cm^{-3}, $N = n(ct)^3 \approx 3 \times 10^{33}$, and $M = 10^{-23} M_\odot$; the mass of the perturbed region is insignificant! But for any larger mass, the metric perturbs survive without changing until the period when all of the resonances annihilate and decay, when the Hagedorn effects disappear. [Note: (1) the wavelength is not given; (2) the values are those from the more recent Russian edition; and (3) the numbers seem inconsistent if nb = n.-LF]

Finally, the softness of the equation of the Hagedorn equation of state is important for the formation of primordial black holes and accretion thereto.

24.3 Quantum Phenomena that Take Place near Singular States of a Metric and in Strong Gravitational Fields

Under extreme conditions it is necessary to take simultaneous account of GTR and of quantum effects. Quantum effects will introduce important theoretical changes in the classical GTR deductions. Under what conditions?

GTR does not bring new physical quantities into the theory. The cosmological constant Λ is important only over large scales for which there are no quantum effects. Planck, introducing his eponymous constant h, with $\hbar = 1.05 \times 10^{-27}$ g cm^2 s^{-1}, understood quantization as applying to all of natural science. Considering c, G, and $\hbar$ as three fundamental quantities, he showed that quantities of any dimensionality can be expressed in terms of them alone. In particular:

$$l_{pl} = (G\hbar/c^3)^{1/2} \approx 1.6 \times 10^{-33}\text{cm};$$

$$t_{pl} = l_{pl}/c = (G\hbar/c^5)^{1/2} \approx 5.3 \times 10^{-44} \text{ s};$$

$$m_{pl} = (\hbar c/G)^{1/2} \approx 2.2 \times 10^{-5} \text{ g; and}$$

$$\rho_{pl} = m_{pl}/l_{pl}^3 = c^5/(G^2\hbar) \approx 5 \times 10^{93} \text{ g cm}^{-3}.$$

Consider the similarity between the Coulomb law and Newton's gravitational law, with e^2 and Gm^2 having the same dimensionality, so that $Gm^2/(\hbar c)$ is similar to the fine structure constant $e^2/(\hbar c) = 1/137$. For the electron and the proton respectively, we have

$$Gm_e^2/(\hbar c) \approx 2 \times 10^{-45} \text{ and } Gm_p^2/(\hbar c) \approx 6 \times 10^{-39}.$$

The condition $Gm^2/(\hbar c) = 1$ yields the mass m_{pl}. The length l_{pl} is the "Compton wavelength" for the mass m_{pl}, so $l_{pl} = \hbar/(m_{pl}c)$.

[In the system of units used in elementary-particle theory where $\hbar = c = 1$, length and time have the same dimensionality, inverse to mass. The product Gm^2 is dimensionless, so G corresponds to an area or cross section $\sim 2.5 \times 10^{-66}$ cm^2 or 2.5×10^{-42} barns.]

These quantities characterize the conditions where quantum gravitational effects play the principal role. For example, the ST curvature $R^{\alpha\gamma}_{\beta\delta}$ would be of order $\sim l_{pl}^{-2} \sim 10^{66}$ cm^{-2}. Such a curvature can arise in a vacuum but is not a necessary attribute of a vacuum. However, if a matter density achieves a value of order ρ_{pl}, the corresponding GTR equation curvature is order l_{pl}^{-2}; in this sense such a curvature is necessary.

Whereas in "ordinary" quantum physics, space and time are considered as determinate quantities, in a quantum gravitational regime, space and time themselves well might acquire probabilistic properties.

In cosmology, one method for dealing with this is to pose questions and calculate quantities for the period when the universe is no longer in the singular state, so both the curvature and the matter density are nowhere huge. This would be similar to S-matrix theory. But even in elementary particle physics, this approach is insufficient to address all questions, so the Lagrangian approach is used instead.

A quantum gravitational theory is especially necessary in cosmology because the Universe certainly passed through such a state. A quantum gravitational theory might also indicate how to choose from among the large number of classical cosmological solutions. [Is this still an issue?-LF]. A complete quantum-gravitational cosmological theory does not now exist. There are only partial results. One such result contemplates the formation of primeval black holes near the singularity and their subsequent quantum evaporation, with effective baryon nonconservation. Thus, quantum gravitational theory is doubtless enormously significant for theoretical cosmology and, through an inferential chain, for observational cosmology too.

Henceforth we shall set forth hypotheses and questions underlying the research as well as the firmly established facts.

QED serves as an analog for quantum gravitational theory. Via systematic calculations done in the realm of QED, physicists have obtained remarkable agreement with experimental results, e.g., for the Lamb shift in the H atom and the anomalous magnetic moment of the electron [and the muon?-LF]. These results did, however, require the introduction of such new concepts as mass and charge renormalization and vacuum polarization. [The next section addresses QED in more detail.-LF]

Quantum gravitational corrections to quantities observable in the laboratory can be computed in analogy with the schema for QED. The first step was the quantization of linear gravitational waves. Considered as small perturbs of flat space, the quanta of this field are understood today as gravitons: they have energy $\hbar\omega$, are spin 2 bosons, and have zero rest mass. But the nonlinearity of GTR means that gravitons are themselves sources of gravitational fields. Though the motion of particles in gravitational fields does not depend on the constant G but only on the ST metric, the converse effect of particles on the metric does depend on G.

In cosmology, the situation is quite different, since for $t \approx t_{pl}$, quantum gravitational effects are of order unity, so their character is of much interest. The most important effect is probably the creation of particles or pairs of particles in strong gravitational fields.

Geometrical optics provides some insight through two concepts. The first is the ray, which is to a wave packet what trajectory is to a particle. The second is adiabatic invariance, according to which the energy of a wave field changes in proportion to its frequency. Consequently the ratio of the energy to the frequency is an adiabatic invariant, remaining constant within the realm of geometrical optics.

But upon a rapid change in the metric, the adiabatic invariance is lost. The number of quanta changes as quanta are created or destroyed - not because of the presence of sources, but rather due to the interaction with the ST geometry.

Another important effect is that of zero-point oscillations of the field. which leads to the creation from the vacuum of real, observable particles.

Suppose that the metric has this form:

$$ds^2 = dt^2 - [1 + h(t)]dx^2 - [1 - h(t)]dy^2 - dz^2,$$

with $h(t) = \delta f(\omega t) = \delta e^{(-\omega^2 t^2)}$ or $h(t) = \delta(1+\omega^2 t^2)^{-1}$, so $h = 0$ for $t = \pm\infty$; $h \sim \delta$; and moreover $\delta \ll 1$. Then, to first order with respect to δ, the anisotropic stresses satisfy

$$T_{11} = -T_{22} = \delta\omega^4 \psi(\omega t), T_{33} = \varepsilon = 0.$$

The function ψ also goes to zero for $t = \pm\infty$. Then to second order in δ, there are terms for T_{ii}.

The first-order terms violate the energy dominance conditions when the metric changes are important! They describe vacuum polarization. The energy conservation condition is

$$d\varepsilon/dt = h(T_{11} - T_{22})$$

The first order in T_{11} and T_{22} gives a second order contribution in ε, which is not vanishing at $t = +\infty$ and which describes creation of real particles. For a typical power-law dependence of the metric on time, the characteristic time for a change in the metric equals the time since the singularity. Consequently, the waves are nonadiabatic, with $\omega < t^{-1}$ and $k < (ct)^{-1}$. Given that creation takes place at the average rate of one quantum per mode, this leads to an order of magnitude of the energy density of created quanta of

$$\begin{aligned}\varepsilon &\approx \hbar\omega k^3 \approx \hbar/(c^3t^4)\\ &\approx 1.05\times10^{-34}/[(3\times10^8)^3\times(5.3\times10^{-44})^4]\\ &\approx 4.9\times10^{113}\ \text{joules/m}^3\end{aligned}$$

[This compares to the currently observed value of 10...-JC] Note that G does not enter even though the we are considering particle creation in a gravitational field. What does the strong dependence of ε on t mean? Strictly speaking, we have found in order of magnitude the energy density of particles created during the time between t and 2t. Here there is a great difference between a collapse singularity in the future and the cosmological singularity in the past.

For collapse, quantum gravitational theory would apply only at the last moment. The singularity is at $t = 0$, so earlier events are for $t < 0$. Particles—pairwise as appropriate—are created, manifesting the destruction of adiabatic invariance. For $t \to 0$, the energy density of these newly created particles grows more rapidly than that of the preexisting matter. Upon the approach to t_{pl}, the effect of newly created particles becomes predominant, influencing further changes to the metric even if there was effectively a vacuum prior. Thus, for the period $t_{pl} < t < 0$, jointly solving the equations for the metric change and particle creation would be required.

For cosmology, the situation is totally different. Suppose that at an instant t_1 in a homogeneous vacuum model, the metric, expansion rate, and the type of space are specified.

Over the very short time from t_1 to $2t_1$, particles with an energy density $\sim \varepsilon \sim \varepsilon_1 = \hbar/(c^3t_1^4)$ arise. At the later instant $t_2 > t_1$, the energy density of the later created particles will be $\varepsilon_2 = \hbar/(c^3t_2^4)$. But the particles created earlier at $t \sim t_1$ do not disappear; rather the volume they occupy expands, so their energy density becomes $\varepsilon_1' = \hbar/(c^3t_1^4)\times(t_1/t_2)^{4/3}$. Since $\varepsilon_1' > \varepsilon_2$, in distinction to collapse, the energy density at time t depends on both the time when the particle creation begins and the manner in which it turns on.

To summarize, for a certain time before t_{pl} and perhaps later in collapse, analysis is possible without regard to the limitations of existing quantum gravitational theory. In cosmology however, the Universe "remembers" the initial conditions at every time.

But there is an important additional fact based on the principle of conformal invariance. The initial stage of the isotropic Friedmann model can be brought to the conformally flat form, while the anisotropic Kasner models cannot, and a GTR singularity is in general anisotropic. For collapse, a general solution must be realized, while for cosmology, a particular solution could be relied on near the singularity. [I am perplexed!-LF]

A metric change is conformal if lengths and times can differ among different ST points but must be the same for all spatial directions and time at a given point. For example, the flat Minkowski metric can transform into a conformally flat metric by the change

$$\begin{aligned}&ds^2 = dt^2 - dx^2 - dy^2 - dz^2\\ \to\ &ds^2 = a^2(x',y',z',t')[dt'^2 - dx'^2 - dy'^2 - dz'^2]\end{aligned}$$

Now since $ds = 0$ for photon propagation in the Minkowski metric, so also is this true in the conformally flat metric. But there is no particle creation in the flat Minkowski metric, so there would not be any in the conformally flat metric either.

On the other hand, the Kasner metric cannot be brought to the conformally flat form, so it would permit particle creation. The dimensional estimates above with $\varepsilon \sim t^{-4}$ in reality apply only to such an anisotropic singularity.

One can hypothesize that an isotropic emergence from the singularity occurs simply because particle creation would lead to theoretical contradictions otherwise. In the collapse case, general principles indicate that particle creation decreases the anisotropy of contraction. In cosmology, one can assume that the effect has the same sign, so that initiating particle creation at a time $t_1 \to 0$, the residual anisotropy also tends to zero.?????

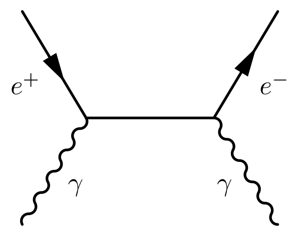
One ought to assume that the true solution is locally isotropic, so for the Universe as a whole, such a consideration leads to a quasi-isotropic solution described earlier, whose properties agree well with observations. What remains unknown is the scale and amplitude of metric deviations from homogeneity.

24.4 The Creation of Charged Particles in Electrodynamics

Particle creation in gravitational fields is an exotic, unobserved process from the viewpoint of laboratory physics. There is no hope for an experimental study of it. On the other hand, particle creation in EM fields is observed experimentally and QED predictions have been verified. We shall review some of the ideas and results of QED, some of which have analogs in gravitational theory.

Consider the well-studied annihilation direct process $e^+ + e^- \to 2\gamma$. According to general quantum theory principles, the possibility of the reverse process $2\gamma \to e^+ + e^-$ follows from the existence of the first. If the kinetic energy of the $e^+ + e^-$ in the center of momentum system is of the order m_ec^2 (not ultra relativistic), then the cross sections for the direct and reverse processes are of the same order, namely, $e^4/(m_e^2c^4) = r_0^2$, where is r_0 is the so-called classical electron radius satisfying $e^2/r_0 = m_ec^2$.

The reverse process, pair creation, is depicted in Figure 62, with time increasing upward. The positron e+ can also be thought of as an electron moving backward in time. Since two photons participate in the pair creation, there are two "vertices" in the diagram where two electron lines and one photon line come together. At each vertex the interaction is proportional to the electron charge e, so that the net effect for the two vertices is a matrix element proportional to to e^2 and a cross section or probability for the process proportional to e^4.



Feynman diagram of e^+e^- creation (reproduction of Figure 62 with time increasing upward).

A situation similar to that in Figure 62 but with only one free photon, and all three particles connecting at a single vertex, is incompatible with energy and momentum conservation laws. But such a situation can occur at a distance r from a radiating system, leading to a momentum indeterminancy of $\Delta p = \hbar/r$. The EM field at a distance of order $\hbar/m_e c$ from the radiating system is capable of creating pairs. Indeed, e^+e^- pairs can be so created even when there is no radiation of real photons at all.

This situation arises in "zero-zero" transitions, giving rise to pair creation and observed in the carbon and oxygen excited states $^{12}\mathbf{C}^*$ and $^{16}\mathbf{O}^*$, which transform into the ground state by emission of an e^+e^- pair.

In all of these cases, the energy of the photon or photons, free or not, must exceed the sum of the e^+ and e^- masses. Since a photon's energy is connected with frequency by the relation $E = \hbar\omega$, pair creation requires a field with a high frequency satisfying

$$\omega > 2m_e c^2/\hbar \text{ or } \omega_1 + \omega_2 > 2m_e c^2/\hbar$$

The quantum nature of the photons is not important here!

This reasoning leads to an expression for the probability of pair creation in a classical EM field $\mathcal{E}$ that depends on the energy density. For the situation of Figure 62, it is proportional to $e^4\mathcal{E}^4$ and for the single photon case it is proportional to $e^2\mathcal{E}^2$.

It is conceivable that pair creation could occur in a static field with $\omega = 0$, but by a different mechanism, one based on separation of the created pair by a distance $x_1 - x_2$ such that the energy characteristic of this separation is sufficiently large. With ΔE_0 the additional energy required,

$$\Delta E_0 = 2m_e c^2 - e\mathcal{E}(x_1 - x_2).$$

Note that ΔE_0 is less than zero if the if the separation is sufficiently large. But the required potential difference must be greater than 10^6 volts. The process is classically impossible because the different charges cannot emerge at different positions simultaneously. Quantum mechanics does allow this to happen by means of tunneling through the potential barrier. Another calculation yields the required field strength of 10^{16} volts cm^{-1}. The process has not been observed, but improvements in laser technology give hope.[but done at RHIC-MS]

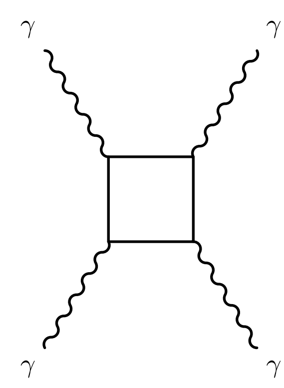
The process is also plausible near an atomic nucleus. The field for U reaches 10^{18} volts cm^{-1}, decreasing rapidly with distance. The spontaneous creation of pairs could only take place for hypothetical nuclei with an atomic number greater than 180.[137?-MS]

Another possibility is the the creation of virtual pairs in a low-frequency or weak field. Intuitively, one can imagine that pairs are created and then immediately disappear. No free charged particles are ever present, but there nevertheless appears a nonzero, time-averaged charge density and electric current. This is called the "polarization of the vacuum."

Intuitively, the vacuum polarization partially screens the charge.

One consequence of this is the following: Calculations show that this vacuum polarization leads to a renormalization of charge, but since its effect is the same for each elementary particle, the vacuum polarization per se is not observable. It is the renormalized value for charge that is so successful in QED calculations of experimental results.

Another consequence is observable, a shift in the atomic levels near the atomic nucleus. Higher order effects are also possible, such as that depicted in Figure 66. Though the scattering of light by light shown there has not been observed, a related process, the scattering of light by the electric field of a nucleus, has been.



Feynman diagram for a nonlinear process in electrodynamics; the scattering of light by light (reproduction of Figure 66 with tiome increasing upward).

Real pairs are not different from external charges. However, their number, motion, and the current to which they give rise all depend on the entire previous history of the field. One cannot express them in terms of instantaneous local quantities. For problems involving pair creation, therefore, quantum effects necessarily require changes on the right-hand side of Maxwell's equations.

A consistent description of pair creation includes vacuum polarization first. This is because only a joint consideration of vacuum polarization and pair creation removes the paradoxes involving causality violation and local nonconservation of charge. It is then appropriate to collect all of the effects - both virtual and real - on the right-hand side of the equation.

Putting quantum effects on the right side of Maxwell's equations will serve as a model for the examination of quantum phenomena in a gravitational field near the singularity.

Two theses arise as the moral from the history of electrodynamics. First, one considers those and only those effects and correction terms which are direct, necessary, and inevitable consequences of the unchanged initial assumption that the nature of the interaction of the EM field with charged particles is as originally described. Second, experiment in all its detail confirms the theoretical deductions.

24.5 The Mathematical Theory of Particles Creation

This section is an elaboration of Section 24.3 that follows the work of Zel'dovich and Starobinsky (1971). They examine a scalar wave equation in the background of a flat, spatially homogeneous metric. By positing solutions in the form of oppositely moving waves, they show that pairwise creation of waves is possible with conservation of total momentum.

Though this is a classical argument, its reformulation in terms of a quantum field of harmonic oscillators yields energy levels $E = \hbar\omega/2 + n\hbar\omega$. Renormalization is required to eliminate the infinite sum of zero-point energies $\hbar\omega/2$ for all possible oscillations..

A. Hawking's Mechanism of Black Hole Particle Creation

Akin to the Penrose mechanism for extracting rotational energy from a rotating black hole by particle decay in the ergosphere, where one product of the decay is captured by the black hole and the other escapes with more energy than the original particle. The same result should follow for a classical wave incident on a black hole, a phenomenon called "superradiance."

From the analogy between classical and quantum vacuum effects, a rotating black hole will radiate spontaneously — without incoming waves — losing mass and angular momentum. If a massless field is radiated, the characteristic wavelength is of order the gravitational radius $R_g = 2GM/c^2$ and the time scale for angular momentum loss of order $G^2M^3/(c^4h)$, which for the cosmological time $\sim 10^{10}$ years corresponds to $\sim 10^{15}$ g black holes.

Later, based on Bekenstein's introduction of a black hole temperature connected with the idea that black hole area behaves like entropy, Hawking computed the radiation from a nonrotating black hole. He showed it to be equilibrium black body radiation at $T(K) \sim 10^{27}M^{-1}(\mathrm{g})$.

For outgoing waves, adiabatic invariance is broken, corresponding to particle creation. This permits, and indeed requires, eventual black hole disappearance:

$$dM/dt = -\alpha/M^2; M(t) = (M_o^3 - 3\alpha t)^{1/3}$$

so the evaporation time is $t_0 = M_o^3/3\alpha$, and $t_0(s) \sim 10^{-28}M_o^3(\mathrm{g})$. The resulting time for solar mass black holes is $\sim 10^{44}$years; but the effect would be overwhelmed by accretion. The most important consequences relate to primordial black holes, with mass $\leq 10^{15}$g.

B. The Formation and Properties of Primordial Black Holes

. . .

24.6 Superspace and Mini-Superspace

The quantization of the evolution of the Universe as a whole is the ultimate goal of the theory of superspace developed by Wheeler and coworkers. The method used is based the "principle of least action" and the extremal value of the time integral of the Lagrangian between given initial and final states. Feynman proposed for quantum theory that the probability for a system evolution between states x_1 and x_2 be given by

$$[G(x_1, x_2)]^2 = [Se^{(i/\hbar \int_{x_1}^{x_2} \mathcal{L}dt)}]^2 \tag{1}$$

The trajectories that are close to the extremal or classical trajectory give the greatest contribution to the transition amplitude G.

Wheeler applied this principle to the Universe as a whole — the 4-dimensional manifold with the contained matter. The resulting space, whose points represent all possible states for the Universe is called "superspace." A calculation of the action integral for all possible trajectories followed by selection of the group of trajectories giving the main contribution to G leads to the trajectories corresponding to GTR. One can simultaneously also determine the degree of quantum deviations from the classical trajectory.

But carrying out this program would be enormously difficult. Misner proposed the more limited problem of considering the evolution of a closed, homogeneous Bianchi type IX universe described by the three parameters a, b, and c. In this approximation, superspace reduces to a space of three dimensions, “mini-superspace.” [Bianchi was an Italian mathematician who formulated a classification of homogeneous 3-D spaces by their geometrical properties based on Killing vectors; there are nine primary spaces. - LF]

Then the problem can be addressed via the Lagrangian formulation or the Schrodinger equation for a wave function $\Psi(a, b, c; t)$, the square of whose modulus gives the probability distribution. Terms that enter depend on the momenta or the three velocities $a' = da/dt$, b' and c'.

Other terms enter that depend upon the spatial curvature $R_{\alpha\beta}$, involving qualities of the type $a^2/(bc)^2$ and play the role of potential energy. But near the zeros of a, b, and c the potential energy becomes infinite. The peculiar feature of the quantum problem is thus that a transition from cosmological collapse to expansion is in principle possible, with some similarity to the scattering of a charged particle by a Coulomb center. [See in this connection Section 24.16 following. - LF] In the classical GTR problem, the collapse of a homogeneous Universe inevitably leads to a state of infinite density, and zero volume V.

How does the quantum theory of mini-superspace help? To see how, we shall give an estimate of the quantum number of that state to which the average motion of the Universe corresponds now. To begin, a rough estimate gives for the kinetic energy

$$\varepsilon \approx (c_\gamma^2 V/G)(a'/a)^2 \approx c_\gamma^2 (a')^2 a/G,$$

Here we take a = b = c ; and c_γ — not c — is the speed of light. The momentum corresponding to the coordinate a satisfies $p_a = d\varepsilon/da' = c_\gamma^2 a' a/G$, In order of magnitude,

$$ap_a = n\hbar, n = c_\gamma^2 a^2 a'/(G\hbar) = (t_0/t_{pl})^2 = 10^{120}.$$

Here $t_0 = 10^{17}$s is the age of the Universe, and $t_{pl} = (G\hbar/c^5)^{1/2} = 10^{-43}$s. The quantum corrections to the motion of the Universe now are of the order of $n^{-1} = 10^{-120}$, inconceivably small. The number n is an adiabatic invariant. For some solutions, $V \sim a^3 \sim t$. Therefore, $a^2 a' = const$ holds for these models roughly, where we assume again $a \approx b \approx c$.

Determining the characteristic parameters of greatest contraction in such a theory is much more difficult. The simplest estimate assumes $a \sim b \sim c \sim l_{pl} = (G\hbar/c^3)^{1/2} \sim 10^{-33}$cm, giving

$$t_1/t_0 = (l_{pl}/ct_0)^3 = (t_{pl}/t_0)^3, t_1 = t_{pl}(t_{pl}/t_0)^2 = 10^{-163}s$$

The time t_1 is when the contraction stops.

Even if the details of anisotropic contraction are considered, though, the contraction stopping time when contraction stops satisfies $t_1 \ll t_{pl}$.

This result is not surprising since the Universe is a huge — indeed the hugest — object. It is then natural that, for the Universe as a whole, quantum effects arise closer to the singularity than for its parts. Therefore Misner's mini-superspace theory, based on the GTR equations, can hardly be literally correct.

24.7 The Hypothesis of Baryon Nonconservation and the Charge Asymmetry of the Elementary Particles

A radical approach to the puzzle of charge asymmetry of the universe consists of a repudiation of the notion of baryon conservation. With this approach, however, one of the important motivations behind a theory of transitions through a singularity loses appeal, i.e., the continuity of baryon world lines. In one attempt, it is assumed that the process $p \rightleftharpoons 2\mu^+ + \mu^-$ is possible.

The ultimate goal is to obtain the current charge-asymmetric state given an initially charge-symmetric state for the plasma. To obtain this result, baryon nonconservation is necessary but not sufficient.

Before 1964, it was believed that all probabilities and cross sections are invariant under particle-antiparticle interchange. Now, however, we know that there is not an exact symmetry between the properties of particles and antiparticles. While many properties of particles and antiparticles are similar, there are several decay modes that exhibit a violation of charge symmetry, such as the decay probability for the long-lived neutral kaon, for which the probability of $K_L \to \pi^+ + e^- + \bar{\nu}$ is greater that that for the antiparticle to decay by $\bar{K}_L \to \pi^- + e^+ + \nu$. Suppose then that there are some hypothetical bosons W such that the probability for $W^+ \rightleftharpoons p + \nu$ is greater than the probability for $W^- \rightleftharpoons \bar{p} + \bar{\nu}$.

It is then guessed that after cooling and annihilation of $p + \bar{p} \to \pi, \gamma, \ldots$, the excess protons survive and give the observed $p : \gamma$ ratio. This primitive theory is contradictory to the known stability of protons and other “stable” nuclei. Obviously, if there are at least two channels, $W^+ \to p + \nu$ and $W^+ \to e^+ + \nu$, then there should also be a process through virtual W^+,

$$p \to W^+ + \bar{\nu} \to e^+ + \nu + \bar{\nu}$$

Another theory involves instability of quarks against leptonic decay. But since neutrons and protons are made of three quarks, they would be unstable only if three quarks simultaneously decay, with extremely low probability.

Baryon nonconservation could also arise based on the notion of BH evaporation. Can one count the baryons captured by a BH? Apparently not; a BH seems to be a collective "graveyard" of unknown baryons. [There continue to be papers addressing this notion.-LF] But BH evaporation means that the graveyard is lost. Perhaps the matter captured by a BH goes into a separate closed universe, disconnected from our space after the end of evaporation, when the mass seen by an outside observer is zero.

24.8 The Cold Universe and the Perturbation Spectrum

...

24.9 The Steady-State Universe Theory of the Universe

...

24.10 Mach's Principle and the Coincidences Among Large Numbers Occurring in Physics and Cosmology

Astronomers and physicists have borrowed ideas from one another, and the practice continues. Here we shall consider two connected ideas relevant to this theme - Mach's principle and the relation between the gravitational constant and the properties of the Universe as a whole.

Recall first that cosmology is that regime where the laws of gravitation, quantum mechanics, and elementary particle theory simultaneously play a role. We have seen that G, ℏ, c, e, and m_e or m_p characteristic of these theories yield two similar dimensionless qualities:

$$Gm^2/(\hbar c) = g \text{ and } e^2/(\hbar c) = \alpha = (137.03)^{-1}$$

The ratio of the electrostatic energy of the electron in the H atom to the rest energy of the electron is of order α^2. That $\alpha < 1$ allowed the creation of a nonrelativistic quantum theory of the atom, then a relativistic theory, and then QED and renormalization theory, with very precise results. In cases where a large dimensionless constant plays a role, i.e. strong interaction and nuclear physics, analogously precise calculations are not possible.[Has this been negated by lattice gauge theory?-LF]

The Coulomb interaction between two particles involves an energy e^2/r, while the gravitational interactions involves an energy Gm^2/r; hence Gm^2 plays the same role in gravity theory as e^2 plays in electrostatic theory so they have the same dimensionality. Substituting into g the proton and electron masses respectively gives

$$g_p = Gm_p^2/(\hbar c) = 6x10^{-39} \text{ and}$$
$$g_e = Gm_e^2/(\hbar c) = 2 \times 10^{-45};$$

Though both are extraordinarily small, their logarithms do not differ too much from each other, and we shall henceforth consider the rounded values $-lng = 92$ and $g = 10^{-40}$.

Physicists have the conviction that dimensionless quantities differing substantially from unity must be explained qualitatively. Furthermore, the closeness in values of large dimensionless numbers arising in theories of various natural phenomena indicates intrinsic connections among the phenomena, connections which illuminate the path for further scientific development.

For example, Landau noted that $\alpha \sim (lng)^{-1}$. [1/137 ∼ 1/92 ?? To me, this is a stretch! - LF] A possible conceptual basis for this coincidence arises from the cutoff momentum that enters into a QED expression for the observable charge of an elementary particle. In this context, Landau interprets the relation between α and g as showing that gravitational effects limit the momenta to be considered.

Another example is the maximum mass estimate for a white dwarf star, the Chandrasekhar limit. The maximum number of nucleons in a white dwarf N is $g_p^{-3/2}$. The approximate numerical comparison takes $M = 1.2M_\odot$, corresponding to 1.5×10^{57} baryons—while $g_p^{-3/2} = 2 \times 10^{57}$. The same dimensional argument could be applied to the neutron star mass limit of the order of 2 or 3 $M_\odot$, where we rely on the same order of magnitude for the masses of all strongly interacting particles.

One research direction consists of inferring the properties of the Universe from the theory of gravitation, i.e, under the assumption that g, measurable locally in laboratory physics, is basic to cosmology. Another direction is based on the hypothesis — called Mach's principle — that the Universe as a whole has a definite role in determining the local laws of physics.

In its embryonic form, Mach's principle addresses the Newtonian question of the centrifugal force in a rotating body. Is the rotation absolute, or is it relative to the entire remaining universe? More generally, are there centrifugal forces in a world without distant matter or stars? There can be no experimental answer, because there is no such Universe. There can only be a thought experiment. Mach's assumption is that no "absolute" rotation exists: no inertial (particularly no centrifugal) forces would exist in the empty Universe. But he never worked this assumption into quantitative theory.

In GTR inertia and gravity are closely connected, so with Mach's principle as motivation, the question arose of whether or not gravity depends on properties of the Universe as a whole. If it does, then g can depend on the large numbers characterizing the Universe as a whole.

So Mach's principle might have some connection with the role of large numbers in laboratory physics. By this reasoning a group of questions falls under the heading of the "many faces of Mach."

Consider now two questions concerning large numbers, the first a factual side. For the relation of $g^{-1} = 10^{40}$ with local physics, suppose that the total number of nucleons in the Universe N characterizes the Universe. Dirac obtains this number by taking the radius to be $c/H = 10^{28}$cm. Since the critical density of 10^{28} g cm^{-3} corresponds to a particle number density $n = 10^{-5}$ cm^{-3}, we have the remarkable coincidence

$$N = (4\pi/3)(c/H)^3 n = 4x10^{79} \approx g^{-2}.$$

Secondly, substitute into the relation $N = g^{-2}$ the expressions for the critical density $\rho_c = 3H^2/(8\pi G m_p)$ and the number density of nucleons, $n = 3H^2/(8\pi G m_p)$, from which $N \sim c^3/(HGm_p)$ follows. Omitting numerical factors and doing some algebra leads to another remarkable relation:

$$H = (m_p c^2/\hbar)(Gm_p^2/\hbar c) = g(m_p c^2)/\hbar,$$
$$\text{whence } t = H^{-1} = [\hbar/(m_p c^2)]g^{-1}.$$

Gamow called the unit of time characterizing the nucleon, $\hbar/(m_p c^2) = 10^{-24}$ s, a "tempon." In terms of that unit, the age of the universe is $\sim 3 \times 10^{41}$, a dimensionless number reminding us of $g^{-1} = 10^{40}$, characteristic of local gravitation.

Let us now make the physical ideas connected with these coincidences more precise. First, the idea of a total number of nucleons makes sense only for a closed Universe. To be more precise, one should take $\rho_c = \rho_c\Omega$ with $\Omega > 1$, the curvature radius $a = cH^{-1}f(\Omega)$ and the volume $V = 2\pi^2 a^3$.

The assumption of a connection between the total baryon number N in the closed Universe and the constant g is possible and does not lead to the variability of gravitational constants or masses of the elementary particles. There is a hidden assumption that even the baryons outside the particle horizon are still influencing the local physics. The connection between H or t and g is now only a temporary coincidence approximately valid for the part of the closed Universe evolution near the maximum expansion. For an open Universe, the best interpretation is that N is the number of baryons inside the particle horizon that influence the local physics in a universe of infinite overall size.

But in this case all the physical quantities cannot be left constant. In the simplest case of a flat, matter-filled Universe, preserving the equation above for N requires that G, for example, varies with t. But then all the beautiful edifice of theoretical physics is ruined. The possible variation of G and the consequences thereof have been widely discussed. One consequence is that the Earth would have been closer to the Sun by a factor of 1.5 to 2. But the variation of G also means that the flux of solar radiation incident on the Earth would vary as G^9. Geological and paleontological evidence decisively contradict this.

There is finally an attempt at reading the formula for G in a way that allows the specific traits of the contemporary epoch—life, civilization, writing this book!—as allowed only when H falls to a value of $\sim g(m_p c^2/\hbar)$. Galaxy and star formation in the hot Universe must come only after recombination, i.e., after the end of the RD stage. Now this time (given $\Omega = 1$) corresponds to the cosmological time $t_r = 5 \times 10^{12}$ s or to an age of 5×10^{36} tempons, which does not differ significantly from g^{-1} or g_p^{-1} or from $e^2/(Gm_p^2) = 1.25 \times 10^{36}$. This is wholly compatible with ordinary physics in which G, h, c and the masses of elementary particles are constant.

Return now to Mach's many-faced principle, whose essence is that the inertia of a body is determined by its gravitational-inertial interaction with other bodies in the Universe. Though this principle influenced Einstein in creating GTR, it became clear the Mach's principle is not contained in it!

What does it mean that inertia is determined by interaction with other bodies? A straightforward answer is that the inertial mass of a test body, a measure of its resistance to an applied force, is determined by its interaction with other bodies; if they did not exist, the test body would have no inertial mass. On the other hand, Einstein stated that "the inertia of a body must increase when ponderable masses are piled up in its neighborhood." (This error by him resulted from a misinterpretation of a formula he obtained.)

Neither of these assertions is fulfilled in the theory of relativity. First, SRT is valid in empty space, where a body possesses inertia and there are centrifugal forces in a rotating system. Second, a given force such as a compressed spring, always transmits the same acceleration to a given body independent of nearby heavy masses. Thus confirmation of the theory of relativity is a blow to Mach's principle.

But other ideas sometimes connected with Mach's principle survive. A body must experience an accelerating force in the same direction when nearby masses are accelerated. And a rotating body must generate inside of itself a Coriolis field which deflects moving bodies in the sense of the rotation and a radial centrifugal field as well. Both effects exist in GTR. They are not connected with a change in the inertial properties of a test body, but rather by a change in an inertial reference system due to the motion of neighboring bodies. These effects are theoretically of the same character as the change in an inertial system of reference in the presence of a motionless gravitating mass. Such a change is in no way connected with a change in the inertial properties of a body.

Having reclassified these two other effects as non-

Machian, we conclude that Mach's principle does not exist in GTR. But more radical views exist.

If one turns to our Universe, then the distant stars and the RR in it actually distinguish one system of reference at rest. Translational and rotational motion relative to the distant stars can be observed and measured. Thus translational motion should show a blueshift in the direction of motion — a higher temperature than 2.7 K — and a red shift in the opposite direction — a lower temperature. [Actually observed-LF]. Mach's Principle would return us not only to Newton, to teach how to distinguish rotation in the Universe, but also to Aristotle, to teach how to determine absolute rest in the Universe.

If experiment shows that the laws of nature are Lorentz invariant, then this directly demonstrates — independent of what the laws actually are — the absence of the influence of distant bodies. There is every reason to believe in Lorentz invariance, so the action of distant bodies is strongly undermined.

In summary, we do not regard as probable the radical revision of GTR in the direction of Mach's ideas. We moreover consider the problem of large numbers to be realistic only in a formulation that connects local properties and which therefore does not necessitate the variability of physical quantities such as the fundamental constants.

24.11 The General Theory of Relativity and the Topological Structure of the Universe

...

24.12 Local Topology, White Holes, Black Holes and Cosmology

...

24.13 Statistical Physics and Gravitation

...

24.14 The Brans-Dicke Theory of Gravitation and its Cosmological Consequences

...

24.15 Cosmology and New Hypotheses in Field Theory

...

24.16 An Oscillating Universe?

Certain ideas in the steady state model of the universe are attractive. Most importantly, baryon charge conservation favors the eternal existence of the universe, which possesses a nonzero baryon charge.

Cannot one consider the eternal existence of the universe as stationary on average by assuming the evolution is oscillatory? Thus contraction would follow expansion and change sign into anticollapse through the singular state. The question of of the possibility of a collapse-anticollapse transition at the singularity — with quantum and other phenomena accounted for — remains open today.

[But the Universe is now known to be expanding at an accelerated rate, so this scenario seems observationally excluded.-LF]

One could imagine cycles with the same maximum expansion or with growing maxima because of the increase of entropy.

For the latter, growing entropy would mean that the cycles would differ from one another and that there were only a finite number of cycles in the past because the specific entropy of the Universe, the entropy per baryon is finite. And also a finite number in the future to avoid having infinite specific entropy.

24.17 The Creation of Gravitons near the Singularity

...

24.18 The Singularity and Conformal Invariance

...

24.19 The Direction of Time

There have been proposals to define the future as that direction of time for which growth in entropy takes place.

In our opinion, attempts to connect the direction of time with some particular complicated phenomena are theoretically incorrect. Physics is local, so phenomena taking place that involve only a single pair of particles must not depend on the growth in the entropy of some other complex system not interacting with the particles in question, or on the recession of galaxies.

Because of the time reversibility of the equations of mechanics and electrodynamics, one can specify an EM field configuration consisting of converging waves in such a way that the waves' action on the charges would increase the energy of the system; but such a formulation of the problem is artificial. There is no reason to for the incident external waves to have just the frequency and phase necessary to drive the system. On the other hand, the waves emitted by the system are found in phase with the motion of the system by virtue of their origin, thereby slowing the motion of the charges and removing energy from them. Thus the radiation condition, or more generally, the principle of causality, allows the definition of the future and the past.

Concerning thermodynamics, it is convenient to use entropy growth as a practical means from determining the direction of increasing time. But this does not bear on the theoretical question of the direction of the arrow of time.

Finally, there is no foundation behind attempts to establish a connection between the arrow of time and cosmology, i.e., with the Universe's expansion. If this were the case, then the arrow of time would change direction in a closed Universe at the time of maximum expansion. The situation is in some sense similar to the trajectory of a projectile; if the projectile changes direction at its maximum altitude, then we would analogously expect light from stars to reverse direction at the instant of the Universe's maximum expansion. But this is obviously nonsense.

The connection between the arrow of time and the expansion is a property of the Universe now. But this connection cannot be used as a definition of the arrow of time — as a definition of the concept of future.

WHEW !!!!